



Desempenho de Métodos Numéricos Clássicos em Problemas de Otimização Irrestrita*

Performance of Classic Numerical Methods in Unrestricted Optimization Problems

Jonatan Ismael Eisermann¹
Maritza Camilli Almeida Brito²

Resumo

Minimizar uma função real, com as mais diversas particularidades, tem configurado um dos principais desafios matemáticos no campo da Otimização. Dentre tais problemas, a minimização de funções sem restrições, conhecida como otimização irrestrita, tornou-se um ramo de estudo específico sob o qual foram desenvolvidas estratégias matemáticas e computacionais que, ante certas condições e limitações, garantem a identificação de pontos críticos – possíveis candidatos a minimizadores globais. Pensando em explorar tais técnicas, o presente estudo visa analisar as potencialidades e fragilidades de cinco desses métodos iterativos: método do Gradiente, método de Newton, método Quase-Newton BFGS, método de Região de Confiança e método do Gradiente Conjugado Não Linear. Para efetivar os objetivos da pesquisa, além da análise teórica e implementação computacional de cada método, buscou-se examinar seu desempenho em um conjunto de dez funções-teste propostas por Moré, Garbow e Hillstom (1981). Os resultados evidenciam um comparativo que aponta vantagens e desvantagens de cada método estudado, de acordo com o problema, a convergência e as possíveis dificuldades que podem ser encontradas ao longo do processo de otimização.

Palavras-chave: Otimização irrestrita. Minimização de funções. Métodos iterativos.

*Submetido em 08/03/2020 - Aceito em 28/08/2021

¹Licenciado em Matemática pelo Instituto Federal de Educação, Ciência e Tecnologia Farroupilha (2019), Mestre em Matemática pela Universidade Federal de Santa Catarina (2021) e doutorando em Matemática Aplicada pela Universidade Federal do Rio Grande do Sul, Brasil – jonatan.eisermann@hotmail.com

²Licenciada em Matemática pela Universidade Estadual do Sudoeste da Bahia (2019), Mestre em Matemática pela Universidade Federal de Santa Catarina (2021) e doutoranda em Matemática pela Universidade Federal de Santa Catarina, Brasil – maritzacamilli@gmail.com

Abstract

Minimizing a real function, with many different characteristics, has been one of the main mathematical challenges in the Optimization field. Among the problems, the minimization of unrestricted functions, known as unrestricted optimization, has become a specific field of study under which mathematical and computational strategies have been developed that, under certain conditions, ensure the identification of critical points – possible global minimizer candidates. The present study aims to analyze the potentialities and weaknesses of five of these iterative methods: Gradient method, Newton's method, Quasi-Newton BFGS method, Trust Region method and Nonlinear Conjugate Gradient method. For this, besides the theoretical analysis and computational implementation of each method, we sought to examine its performance in a set of ten test-functions proposed by Moré, Garbow and Hillstom (1981). The results, obtained through comparison, indicate the advantages and disadvantages of each studied method, according to the problem, the convergence, and possible difficulties that can be found during the optimization process.

Keywords: Unrestricted optimization. Function minimization. Iterative methods.

1 INTRODUÇÃO

O interesse em procurar a melhor solução para determinado problema é o fator determinante que vem impulsionando o desenvolvimento de estudos no campo matemático conhecido como Otimização. Para além da análise teórica, o desenvolvimento e aperfeiçoamento de técnicas específicas que identifiquem tais soluções, bem como sua posterior implementação computacional, tornam-se ferramentas propícias para resolver uma classe de problemas que atendem determinadas especificidades.

Dada uma função $f : \Omega \rightarrow \mathbb{R}$, na qual $\Omega \subseteq \mathbb{R}^n$, $n \in \mathbb{N}$, a classe de problemas em que o domínio não possui restrições é o foco de estudo da otimização irrestrita. Nesse contexto, otimizar f significa minimizar $f(x)$ ou, de forma equivalente, maximizar $-f(x)$, com $x \in \Omega$, quando isto é possível. Dessa forma, faz-se necessário que o contradomínio das funções estudadas seja o conjunto dos números reais, tendo em vista que esse se trata de um corpo ordenado e, portanto, torna factível a busca de minimizadores e maximizadores.

Com a necessidade de identificar pontos ótimos das mais diversas funções e suas respectivas especificidades, muitos pesquisadores deram início ao desenvolvimento de métodos matemáticos iterativos, que, ante certas condições e limitações, garantem a convergência para um ponto crítico. Pensando em explorar tais técnicas, o presente estudo visa analisar as potencialidades e fragilidades de cinco destes métodos: método do Gradiente, método de Newton, método Quase-Newton BFGS, método de Região de Confiança e método do Gradiente Conjugado Não Linear.

Para isto, os métodos supracitados foram estudados, implementados computacionalmente no *software* Matlab, e postos a resolver dez funções-teste propostas por Moré, Garbow e Hillstom (1981) em seu artigo intitulado de *Testing Unconstrained Optimization Software*. Tal produção apresenta um conjunto de funções com características e dificuldades específicas, que possibilitam a avaliação da robustez e eficiência de métodos de otimização irrestrita – o que justifica os recursos empregados neste estudo. Convém salientar que a escolha de uma quantidade relativamente pequena de funções em comparação com outras pesquisas de cunho similar, a exemplo dos estudos de Borges (1985), Urrea (2008) e de Ribas (2017), se deve ao foco deste trabalho em discutir, com maior minuciosidade, as particularidades de cada função e do respectivo desempenho dos métodos, prezando pela didática, clareza e concisão das ideias.

A fim de almejar seus objetivos, esta produção está estruturada em três partes. Na primeira seção são apresentados os principais aportes teóricos que condicionam a existência de minimizadores e dão suporte para os métodos estudados. Na segunda ocorre a apresentação dos métodos e sua estruturação geral. Por fim, na terceira é realizada a análise do desempenho dos métodos nas funções-teste escolhidas para o estudo, destacando suas principais potencialidades e fragilidades na resolução dos referidos problemas.

2 PRESSUPOSTOS TEÓRICOS

Tendo em vista o interesse em resolver problemas de otimização irrestrita e de padronizar as notações a serem utilizadas neste estudo, utilizar-se-á como objetivo a seguinte formulação matemática dos problemas:

$$\min_{x \in \Omega} f(x) \quad (1)$$

na qual $\Omega \subseteq \mathbb{R}^n$ e tem-se a hipótese inicial de f ser de classe C^2 , ou seja, ter todas as derivadas de segunda ordem contínuas. Para esse problema, a função f é denominada *função objetivo* e o conjunto Ω é denominado *conjunto viável*. Como no presente contexto o domínio de f não é definido por um conjunto de igualdades e desigualdades (restrições), considerar-se-á $\Omega = \mathbb{R}^n$, ou seja, todo $x \in \mathbb{R}^n$ será considerado ponto viável.

O principal objetivo da resolução de um problema de minimização consiste em identificar um ou mais minimizadores globais da função f , isto é, pontos em que $f(x)$ atinge seu menor valor. Porém, nem todo problema torna possível tal identificação, e, em alguns casos, consegue-se apenas encontrar pontos em que a função atinge seu menor valor localmente. Nesse contexto, torna-se necessário distinguir as duas definições a seguir:

Definição 2.1 (*Minimizador global*). Um ponto x^* é um minimizador global se $f(x^*) \leq f(x)$, para todo $x \in \mathbb{R}^n$.

Definição 2.2 (*Minimizador local*). Um ponto x^* é um minimizador local se houver uma vizinhança U de x^* , tal que $f(x^*) \leq f(x)$, para todo $x \in U$.

Nessa perspectiva, é importante questionar-se de que maneira é possível reconhecer um minimizador local, pois a estratégia imediata de avaliar o valor da função em todos os pontos de sua vizinhança, para garantir que nenhum deles tenha valor de função menor, mostra-se um processo impossível. Ao assumir que a função objetivo é continuamente diferenciável permite-se o desenvolvimento de maneiras mais práticas e eficientes de identificar quando um ponto é minimizador local. Para além disto, se a função objetivo for duas vezes continuamente diferenciável, isto é, de classe C^2 , é possível verificar se x^* é um minimizador local examinando apenas seu gradiente e sua hessiana nesse ponto, representados, respectivamente, por $\nabla f(x^*)$ e $\nabla^2 f(x^*)$. A partir destes dois operadores matemáticos fundamentam-se as condições de otimalidade para os problemas de minimização irrestrita.

Antes de enunciar tais condições, é de fundamental importância apresentar dois resultados matemáticos que permitem a construção básica de grande parte da teoria de Otimização: o Teorema de Taylor e o Teorema de Weierstrass. Aquele possibilita a realização de aproximações para a função objetivo, bem como seu gradiente e hessiana; enquanto este, sob determinadas hipóteses, garante a existência de pontos extremos da função avaliada, e, por consequência, de um minimizador global.

Teorema 2.1 (Teorema de Taylor). Seja $f : V \rightarrow \mathbb{R}$ uma função definida no aberto $V \subset \mathbb{R}^n$, q vezes diferenciável no ponto $x \in V$. Então para todo $p \in \mathbb{R}^n$ tal que $x + p \in V$, vale que

$$f(x + p) = f(x) + df(x) \cdot p + \frac{1}{2}d^2f(x) \cdot p^2 + \cdots + \frac{1}{q!}d^qf(x) \cdot p^q + r_q(p), \quad (2)$$

onde $\lim_{p \rightarrow 0} \frac{r_q(p)}{|p|^q} = 0$. Além disso, se $[x, x + p] \subset V$, f é de classe C^q , $q + 1$ vezes diferenciável no segmento aberto $(x, x + p)$ então existe $\theta \in (0, 1)$ tal que

$$r_q(p) = \frac{1}{(q + 1)!} + d^{q+1}f(x + \theta p) \cdot p^{q+1}. \quad (3)$$

Corolário 2.1 Suponha que $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é continuamente diferenciável e que $p \in \mathbb{R}^n$. Então

$$f(x + p) = f(x) + \nabla f(x + tp)^T p, \quad (4)$$

para algum $t \in (0, 1)$. Além disso, se f é duas vezes continuamente diferenciável, então

$$f(x + p) = f(x) + \nabla f(x)^T p + \frac{1}{2}p^T \nabla^2 f(x + tp)p, \quad (5)$$

para algum $t \in (0, 1)$.⁴

Teorema 2.2 (Teorema de Weierstrass). Se $D \subset \mathbb{R}^n$ é um conjunto compacto e $f : D \rightarrow \mathbb{R}$ é uma função contínua, então existe $x^* \in D$ minimizador global de f em D .⁵

Utilizando de tais resultados, é possível enunciar e provar as condições necessárias para identificar mínimos de uma função continuamente diferenciável e sem restrições, a serem apresentadas a seguir.

Teorema 2.3 (Condições necessárias de primeira ordem). Se x^* é um minimizador local e f é uma função continuamente diferenciável em uma vizinhança aberta de x^* , então $\nabla f(x^*) = 0$.⁶

Teorema 2.4 (Condições necessárias de segunda ordem). Se x^* é um minimizador local e $\nabla^2 f$ existe e é contínua em uma vizinhança aberta de x^* , então $\nabla f(x^*) = 0$ e $\nabla^2 f(x^*)$ é positiva semidefinida.⁷

Já as condições suficientes são resultados considerados mais fortes do que as condições necessárias, uma vez que oferecem garantias de um determinado ponto ser um minimizador local.

Teorema 2.5 (Condições suficientes de segunda ordem). Suponha que $\nabla^2 f$ é contínua em uma vizinhança aberta de x^* , que $\nabla f(x^*) = 0$ e que $\nabla^2 f(x^*)$ é definida positiva. Então x^* é um minimizador local estrito de f .⁸

³Demonstração: Disponível em Lima (2015).

⁴Demonstração: Disponível em Nocedal e Wright (1998).

⁵Demonstração: Disponível em Lima (2015).

⁶Demonstração: Disponível em Nocedal e Wright (1998).

⁷Demonstração: Disponível em Nocedal e Wright (1998).

⁸Demonstração: Disponível em Nocedal e Wright (1998).

A partir desses resultados, é possível apresentar os fundamentos que sustentam os principais métodos de otimização irrestrita. Este é o objetivo da seção seguinte.

3 MÉTODOS CLÁSSICOS DE OTIMIZAÇÃO IRRESTRITA

As condições de otimalidade estabelecidas na seção anterior configuram a essência dos métodos matemáticos desenvolvidos para o problema de minimização de funções irrestritas.

Para que um ponto do domínio da função f seja um candidato a minimizador global é preciso que se cumpra a condição necessária de primeira ordem, isto é, o gradiente de f deve anular-se nesse ponto. Porém, calcular as derivadas parciais para identificar o gradiente de uma função, e resolver o consequente sistema para verificar qual(is) sua(s) respectiva(s) raiz(es) (encontrar o(s) ponto(s) que gera(m) o gradiente nulo) nem sempre é uma tarefa simples e possível de ser realizada manualmente, sendo necessária a utilização de métodos computacionais que resolvam tal problema.

Esses métodos utilizam, em geral, a hipótese da função objetivo ser duas vezes continuamente diferenciável, e do conjunto $\Omega \subset \mathbb{R}^n$ ser fechado e limitado (compacto). A partir de um ponto inicial $x_0 \in \Omega$, é construída, então, uma sequência $\{x_k\}_{k \in \mathbb{N}}$ de forma que $x_k \rightarrow x^*$, quando $k \rightarrow \infty$, em que x^* é ponto crítico de f em Ω . Nesse contexto, quando a convergência a uma solução não depende do ponto inicial viável escolhido, o método é classificado como global.

As duas estratégias de globalização comumente exploradas nos métodos de otimização são a busca linear e as regiões de confiança. Em ambas, ao longo do processo iterativo, um iterado somente é aceito se ele for melhor do que o anterior, isto é, se ele fornecer um decréscimo suficientemente grande para a função a ser minimizada. Mais especificamente, nos métodos de busca linear, a cada iteração o algoritmo define uma direção que minimiza a função objetivo e, em seguida, encontra um tamanho de passo adequado às exigências de decréscimo da função. Já nos métodos de região de confiança, a cada iteração 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 em pontos distantes de x_k , restringimos a busca por um minimizador de m_k para alguma região em torno de x_k — a região de confiança. Esse minimizador (restrito) será tomado como passo da k -ésima iteração se fornecer um decréscimo suficientemente grande para a função f .

Os métodos de otimização, em geral, constroem um processo iterativo a partir de um ponto inicial $x_0 \in \mathbb{R}^n$, no qual cada novo iterado é gerado a partir da soma entre o iterado atual e um vetor que fornece a direção a ser percorrida pelo método na iteração corrente. Tal vetor, denominado neste estudo de p_k , é multiplicado por um escalar α , cuja funcionalidade consiste em adequar o tamanho do passo a ser dado na referida direção, a fim de minimizar f . Assim, o método realiza o processo $x_{k+1} = x_k + \alpha p_k$ até atingir os critérios de parada utilizados. Dentre tais critérios, destacam-se um pré-fixado número máximo de iterações atingido, e a

norma euclidiana do gradiente da função no ponto ser suficientemente pequena (indicando a identificação de um ponto crítico). Nessa perspectiva, os algoritmos para otimização irrestrita distinguem-se pela forma como é obtida a direção p_k e o tamanho do passo α a cada iteração.

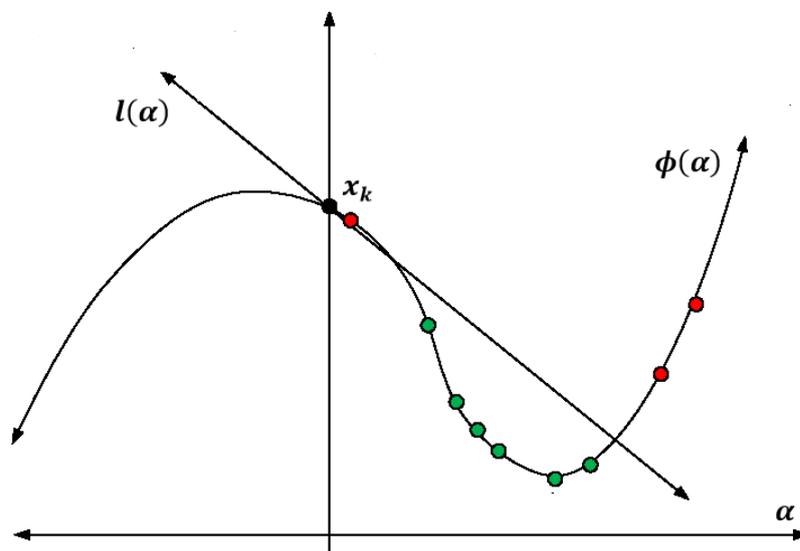
Conforme já mencionado, para os métodos de busca linear é fundamental que $f(x_{k+1}) < f(x_k)$, enquanto $\nabla f(x_k) \neq 0$. Nesse sentido, a implementação de condições de decréscimo suficiente e de curvatura, conhecidas como condições de Wolfe, na referida classe de métodos, torna-se uma etapa imprescindível para a garantia de convergência a um minimizador da função.

A primeira condição de Wolfe, conhecida como condição de Armijo, estipula que α deve, antes de mais nada, fornecer decréscimo suficiente da função f , como medido pela desigualdade

$$f(x_k + \alpha p_k) \leq f(x_k) + c_1 \alpha \nabla f(x_k)^T p_k, \quad (6)$$

para alguma constante $c_1 \in (0, 1)$. Como, nesses casos, $\alpha > 0$ e $c_1 > 0$, a efetividade da condição de Armijo exige que $\nabla f(x_k)^T p_k \leq 0$, isto é, de que p_k seja uma direção de descida. Note que na k -ésima iteração, o lado direito da equação (6) é uma função linear dependente de α , que será denotada por $l(\alpha)$; já o lado esquerdo, é uma função não linear dependente de α , e será denotada por $\phi(\alpha)$. Assim, deseja-se que $\phi(\alpha) \leq l(\alpha)$, relação que, pelo viés geométrico, pode ser interpretada como a exigência de que o valor de α a ser utilizado deixe a função ϕ abaixo da função l . Esta interpretação da condição de Armijo pode ser visualizada na Figura 1, onde os pontos de cor verde são considerados valores aceitáveis de α , enquanto os pontos de cor vermelha não são.

Figura 1 – Primeira condição de Wolfe



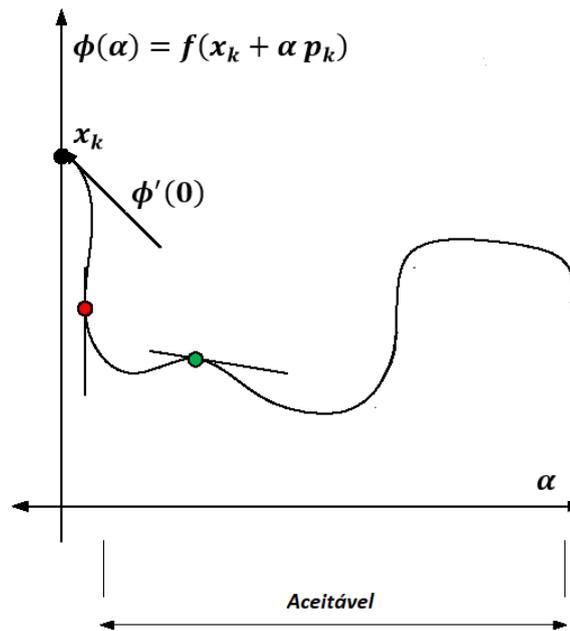
Fonte: Os autores.

A condição de Armijo não é suficiente para garantir que o algoritmo faça o progresso desejado, pois em uma direção de descida essa condição pode ser satisfeita para valores pequenos de α . Com o propósito de eliminar tamanhos de passo inaceitavelmente curtos, é introduzida uma segunda condição, chamada de condição de curvatura, e expressa por

$$\nabla f(x_k + \alpha p_k)^T p_k \geq c_2 \nabla f(x_k)^T p_k, \quad (7)$$

para alguma constante $c_2 \in (c_1, 1)$, em que c_1 é a constante da condição de Armijo. O lado esquerdo da equação (7) é simplesmente a derivada $\phi'(\alpha)$, o que faz com que a condição de curvatura exija que a inclinação de ϕ em α seja suficientemente grande com relação à inclinação inicial $\phi'(0)$. Isso faz sentido, pois se a inclinação $\phi'(\alpha)$ é fortemente negativa, então é possível reduzir f significativamente movendo-se mais ao longo da direção escolhida. Por outro lado, se $\phi'(\alpha)$ é fracamente negativa ou mesmo positiva, então não se pode esperar muito mais diminuição de f nessa direção – fator que faz com que seja concluído o processo de busca linear. Na Figura 2, o ponto verde, diferentemente do vermelho, satisfaz a condição de curvatura.

Figura 2 – Segunda condição de Wolfe



Fonte: Os autores.

A segunda condição de Wolfe é, comumente, substituída nos métodos de otimização irrestrita por uma estratégia denominada *backtracking*. Partindo de $\alpha = 1$, reduz-se seu valor, multiplicando-o por um escalar $\gamma \in (0, 1)$, toda vez que não for satisfeita a condição de Armijo.

O lema a seguir garante a aplicabilidade das condições de Wolfe.

Lema 3.1 *Suponha que $f : \mathbb{R}^n \rightarrow \mathbb{R}$ seja continuamente diferenciável. Seja p_k uma direção de descida em x_k e suponha que f seja limitada inferiormente ao longo da semirreta*

$$\{x_k + \alpha p_k : \alpha > 0\}. \quad (8)$$

*Então, se $0 < c_1 < c_2 < 1$, existem intervalos de comprimentos de passo satisfazendo as condições de Wolfe.*⁹

Nas subseções a seguir serão apresentadas as ideias básicas que caracterizam cada um dos métodos, destacando-se a maneira como cada um define a direção e o tamanho do passo a ser utilizado em cada iteração. As referidas abordagens teóricas estão baseadas nos estudos de

⁹Demonstração: Disponível em Nocedal e Wright (1998).

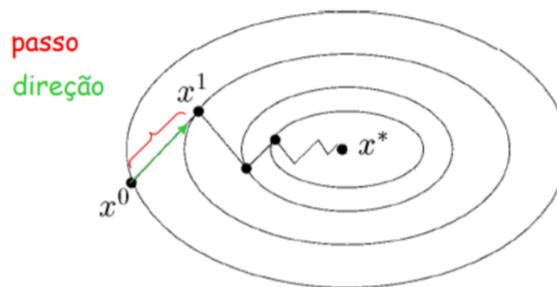
Nocedal e Wright (1998), o que torna recomendável consultar tal bibliografia caso haja interesse em obter maiores informações acerca dos métodos abordados.

3.1 Método do Gradiente

Um dos mais clássicos métodos de otimização irrestrita, o método do Gradiente possui sua essência em explorar o fato de que o gradiente de uma função continuamente diferenciável sempre será um vetor que direcionará para o(s) ponto(s) de máximo da função; consequentemente, o oposto do vetor gradiente apontará para o(s) ponto(s) em que a função atinge seu mínimo.

Partindo de um ponto inicial $x_0 \in \mathbb{R}^n$, a cada iteração o método do Gradiente calcula a direção $p_k = -\nabla f(x_k)$, e toma $x_{k+1} = x_k + \alpha p_k$. Nesse contexto, o tamanho do passo, representado por α , pode ser obtido por *backtracking*. Os passos realizados pelo método tornam o processo iterativo um *zigue-zague* entre x_0 e x^* , como mostra a Figura 3.

Figura 3 – Direções perpendiculares que caracterizam o método do Gradiente



Fonte: Andretta (2008).

O método do Gradiente é considerado lento por ter convergência linear (NOCEDAL; WRIGHT, 1998). De certa forma, essa velocidade se mostra condizente para o referido método, uma vez que a direção $-\nabla f(x_k)$ minimiza o modelo linear da função f ao redor de x_k baseado na equação (4) do Corolário 2.1. Esse modelo, denotado por l_k , é dado por

$$l_k(x_k + p) = f(x_k) + \nabla f(x_k)^T p. \quad (9)$$

3.2 Método de Newton

Buscando agilizar a convergência ao minimizador da função, diferentemente do método do Gradiente, o método de Newton oferece, sob condições adequadas, convergência quadrática (NOCEDAL; WRIGHT, 1998). Dessa vez, porém, a direção a ser utilizada será baseada no minimizador do modelo quadrático da função f ao redor de x_k baseado na equação (5) do

Corolário 2.1, e expresso por

$$m_k(x_k + p) = f(x_k) + \nabla f(x_k)^T p + \frac{1}{2} p^T \nabla^2 f(x_k) p. \quad (10)$$

Assim, assumindo que $\nabla^2 f(x)$ é definida positiva, obtemos a direção de Newton

$$p_k = -(\nabla^2 f(x_k))^{-1} \nabla f(x_k), \quad (11)$$

que será, no método, submetida a um processo de *backtracking* a fim de se obter um tamanho de passo adequado. Evidentemente, quando $\nabla^2 f(x_k)$ é definida positiva, essa direção é de descida, já que satisfaz $\nabla f(x_k)^T p < 0$. Portanto, o método se torna eficiente quando o ponto inicial está em uma vizinhança suficientemente próxima do minimizador, no qual a hessiana preserva tal característica devido a sua continuidade ao supormos que f é de classe C^2 . Tal fato implica na necessidade de uma boa aproximação do modelo quadrático em relação à função objetivo nessa região de trabalho.

Apesar de apresentar velocidade de convergência maior que a do método do Gradiente, o método de Newton é considerado altamente custoso devido à necessidade do cálculo e armazenamento da matriz hessiana a cada iteração, que, apesar de ser simétrica, tem ordem igual ao número de variáveis do vetor x . Aliado a esse fator, a necessidade de obtenção da inversa de uma matriz aumenta ainda mais o custo computacional do método e o torna numericamente instável.

3.3 Método Quase-Newton BFGS

Calcular a matriz hessiana de uma função pode configurar uma tarefa trabalhosa e extremamente difícil, já que nem sempre as suas derivadas de segunda ordem estarão disponíveis ou serão de fácil obtenção. Pensando em sanar essa e outras dificuldades do método de Newton, sem perder a agilidade proporcionada por ele, foram desenvolvidos os métodos Quase-Newton, cujo princípio consiste em construir um processo iterativo similar ao método de Newton, porém substituindo $\nabla^2 f(x_k)$ por uma matriz de aproximação B_k , preferencialmente mais fácil de ser obtida e invertida.

Dentre as aproximações Quase-Newton, as atualizações secantes ganham destaque por proporcionar aproximações boas e pouco custosas para as matrizes $\nabla^2 f(x_k)$. Nelas, o processo iterativo é iniciado com uma matriz não singular B_0 , que é atualizada nas iterações seguintes a partir de matrizes que satisfazem a equação $B_{k+1} s_k = y_k$, em que $s_k = x_{k+1} - x_k$ e $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$. Em outros termos, essas atualizações são construídas, a cada iteração, a partir de informações de curvatura obtidas na iteração anterior. Para isso, usa-se o fato de que a diferença entre valores sucessivos que o gradiente assume descreve, de certa forma, o comportamento local da hessiana. A maneira como cada método propõe o cálculo de B_{k+1} é o que diferencia os métodos inerentes a essa classe estratégica da otimização irrestrita.

Dentre as atualizações secantes, a atualização Broyden, Fletcher, Goldfarb e Shanno (BFGS) destaca-se pela construção de matrizes definidas positivas e, por isso, é uma das mais populares nesse meio. Nela, inicia-se o processo com uma matriz B_0 definida positiva qualquer (geralmente a matriz Identidade de ordem n) e, posteriormente, calcula-se a matriz B_{k+1} através da fórmula

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}, \quad (12)$$

toda vez que $y_k^T s_k > 0$. Caso contrário, toma-se $B_{k+1} = B_k$.

Assim, partindo de um ponto inicial viável o método toma a direção $p_k = -B_k^{-1} \nabla f(x_k)$ a cada iteração, tendo o tamanho do passo definido por *backtracking*.

3.4 Método do Gradiente Conjugado Não Linear

O método do Gradiente Conjugado é considerado um processo intermediário entre o método do Gradiente e o método de Newton, tendo em vista que sua estruturação exige, em geral, uma quantidade inferior de operações por iteração em comparação ao método de Newton, além de oferecer uma taxa de convergência maior que a do método do Gradiente.

Uma de suas principais propriedades refere-se à capacidade de trabalhar com um conjunto de vetores p_0, p_1, \dots, p_l conjugados, isto é, vetores linearmente independentes tais que, dado uma matriz simétrica definida positiva A , tem-se $p_i^T A p_j = 0$, para todo $i \neq j$. Nesse contexto, não é necessário conhecer previamente o conjunto de vetores conjugados, pois cada direção p_k pode ser calculada a partir de p_{k-1} . Tal propriedade aumenta a eficácia do método, já que possibilita o armazenamento computacional de uma quantidade pequena de informações.

Os diferentes métodos que utilizam direções conjugadas são muito empregados na resolução de problemas de grande dimensão. Neste estudo, o método a ser utilizado e analisado nas funções é o gradiente conjugado para problemas não lineares, a partir da abordagem de Fletcher e Reeves (1964). O tratamento dado pelos referidos pesquisadores parte do passo inicial $p_0 = -\nabla f(x_0)$, e a partir de então calcula-se

$$p_{k+1} = -\nabla f(x_{k+1}) + \beta_{k+1}^{FR} p_k, \quad (13)$$

em que

$$\beta_{k+1}^{FR} = \frac{\nabla f(x_{k+1})^T \nabla f(x_{k+1})}{\nabla f(x_k)^T \nabla f(x_k)}. \quad (14)$$

Assim como nos métodos abordados anteriormente, tendo α calculado via *backtracking*, o método computa $x_{k+1} = x_k + \alpha p_k$ até atingir os critérios de parada estabelecidos.

3.5 Método de Região de Confiança

O método de Região de Confiança constitui um processo iterativo focado em trabalhar sobre uma região ao redor de cada iterado x_k , na qual é definido um modelo m_k representativo da função objetivo f . Nesse contexto, o passo a ser considerado na k -ésima iteração para identificar x_{k+1} é baseado na minimização do modelo na referida região – denominada *região de confiança*. Porém, ele somente será aceito se fornecer uma determinada redução no valor da função objetivo; se isso não ocorrer, o passo é recusado e o tamanho da região de confiança é reduzido para encontrar um novo minimizador a partir de uma região na qual o modelo represente melhor a função objetivo. Diferentemente dos métodos de busca linear, a direção p_k pode mudar quando o tamanho da região é alterado, pois nesse método primeiro define-se o tamanho do passo, e, em seguida, a direção a ser percorrida.

Um dos modelos representativos muito utilizados nesses métodos é o modelo quadrático, definido pela expressão

$$m_k(p) = f(x_k) + \nabla f(x_k)^T p + \frac{1}{2} p^T B_k p, \quad (15)$$

no qual B_k é uma matriz simétrica de ordem n , podendo, portanto, ser a hessiana da função f em x_k . Assim, após a identificação do passo p através da minimização de $m_k(p)$ numa região de confiança ao redor do iterado x_k , verifica-se o valor da razão

$$r_k(p) = \frac{f(x_k) - f(x_k + p)}{m_k(0) - m_k(p)}, \quad (16)$$

sendo o numerador denominado a redução real da função e o denominador a redução prevista através da utilização do modelo.

Quando $r_k(p)$ é negativo, tem-se que $f(x_k + p)$ é maior que $f(x_k)$, implicando na rejeição do passo p e na redução da região de confiança. Esse procedimento também é realizado para os casos em que $r_k(p)$ é positivo, porém muito próximo de zero. Por outro lado, quando $r_k(p)$ é próximo de 1, o passo p é aceito, já que o modelo constituiu uma boa aproximação local para a função. Nesse caso, o tamanho da região de confiança para a próxima iteração pode ser mantido ou, até mesmo, expandido.

Neste estudo o modelo quadrático a ser usado em cada iteração será o modelo (15), com $B_k = \nabla^2 f(x_k)$. Já a minimização desse modelo na região de confiança, necessária para a obtenção do passo a ser dado na k -ésima iteração, será feita através da estratégia desenvolvida por Moré e Sorensen (1983).

4 DESEMPENHO DOS MÉTODOS NAS FUNÇÕES-TESTE

As funções-teste constituem os principais meios de análise e validação dos métodos matemáticos inerentes à Otimização. Baseados nessa premissa, Moré, Garbow e Hillstom (1981) desenvolveram um conjunto de trinta e cinco funções com as mais diversas características, projetadas para testes de confiabilidade e robustez da otimização sem restrições. A fim de potencializar tais recursos, foram definidos pontos iniciais estratégicos, as dimensões de cada problema e os respectivos minimizadores que devem ser identificados pelo método empregado.

A escolha das dez funções utilizadas neste estudo se deu baseada na diversidade das características, comportamento, dimensões e estruturas matemáticas que as escolhidas apresentam. A Tabela 1 mostra os problemas explorados neste trabalho e sua respectiva nomenclatura.

A implementação computacional dos métodos e das funções deu-se no programa Matlab, estando os arquivos disponíveis para acesso e *download*¹⁰. Já o problema de otimização irrestrita a ser considerado para as funções-teste apresentadas na Tabela 1 será o de minimizar a função

$$f(x) = \sum_{i=1}^n f_i^2. \quad (17)$$

Tendo em vista a utilização de funções de pequena dimensão neste estudo e das oscilações no tempo de processamento da resolução dos problemas pelo Matlab, limitou-se a análise do desempenho dos métodos abordados à quantidade de iterações e ao comportamento do processo iterativo em cada problema. O ponto inicial definido para cada problema foi o mesmo utilizado nos estudos de Moré, Garbow e Hillstom (1981). A partir dele, cada método construirá um processo iterativo que terminará quando a norma euclidiana do gradiente de f for menor que a tolerância 10^{-4} (indicando a identificação de um ponto estacionário), ou quando for atingido o limite previamente estabelecido de $2 \cdot 10^5$ iterações (indicando falha ao método).

Para os métodos de busca linear (Gradiente, Newton, Quase-Newton BFGS e Gradiente Conjugado Não Linear), será utilizada a condição de Armijo com $c_1 = 10^{-4}$, e o processo de *backtracking* a partir do multiplicador $\gamma = 0.9$. Já para o método de Região de Confiança, o minimizador restrito p do modelo (15) na k -ésima iteração será aceito se a razão $r_k(p)$, descrita pela equação (16), for maior ou igual a 10^{-2} . Mais ainda, se $r_k(p) > 0.9$ dobrar-se-á o tamanho do raio da região de confiança para a próxima iteração. Por outro lado, se $r_k(p)$ for menor que 10^{-2} o raio de região de confiança será reduzido ao meio, e buscar-se-á novamente um p que satisfaça as condições de decréscimo da função esperadas.

Nas subseções a seguir serão apresentados os principais resultados do desempenho dos métodos nos mais diversos tipos de funções.

¹⁰Endereço: <<https://bit.ly/jonataneisermann>>

Tabela 1 – Funções-teste de Moré, Garbow e Hillstrom (1981)

<i>Função</i>	f_i
<i>Rosenbrock</i>	$f_1(x) = 10 \cdot (x_2 - x_1^2)$ $f_2(x) = 1 - x_1$
<i>Freudensteine e Roth</i>	$f_1(x) = -13 + x_1 + ((5 - x_2) \cdot x_2 - 2) \cdot x_2$ $f_2(x) = -29 + x_1 + ((x_2 + 1) \cdot x_2 - 14) \cdot x_2$
<i>Powell Badly Scaled</i>	$f_1(x) = 10^4 \cdot x_1 \cdot x_2 - 1$ $f_2(x) = e^{-x_1} + e^{-x_2} - 1.0001$
<i>Brown Badly Scaled</i>	$f_1(x) = x_1 - 10^6$ $f_2(x) = x_2 - 2 \cdot 10^{-6}$ $f_3(x) = x_1 \cdot x_2 - 2$
<i>Beale</i>	$f_1(x) = 1.5 - x_1 \cdot (1 - x_2)$ $f_2(x) = 2.25 - x_1 \cdot (1 - x_2^2)$ $f_3(x) = 2.625 - x_1 \cdot (1 - x_2^3)$
<i>Powell Singular</i>	$f_1(x) = x_1 + 10 \cdot x_2$ $f_2(x) = \sqrt{5} \cdot (x_3 - x_4)$ $f_3(x) = (x_2 - 2 \cdot x_3)^2$ $f_4(x) = \sqrt{10} \cdot (x_1 - x_4)^2$
<i>Wood</i>	$f_1(x) = 10 \cdot (x_2 - x_1^2)$ $f_2(x) = 1 - x_1$ $f_3(x) = \sqrt{90} \cdot (x_4 - x_3^2)$ $f_4(x) = 1 - x_3$ $f_5(x) = \sqrt{10} \cdot (x_2 + x_4 - 2)$ $f_6(x) = \frac{1}{\sqrt{10}} \cdot (x_2 - x_4)$
<i>Trigonométrica</i>	$f_i(x) = n - \sum_{j=1}^n \cos(x_j) + i \cdot (1 - \cos(x_i)) - \sin(x_i)$
<i>Linear – Posto Completo</i>	$f_i(x) = x_i - \frac{2}{m} \cdot (\sum_{j=1}^n x_j) - 1, \quad 1 < i \leq n$ $f_i(x) = \frac{-2}{m} \cdot (\sum_{j=1}^n x_j) - 1, \quad n < i \leq m$
<i>Linear – Posto Um</i>	$f_i(x) = i \cdot (\sum_{j=1}^n j \cdot x_j) - 1$

Fonte: Adaptado de Moré, Garbow e Hillstrom (1981).

4.1 Função Rosenbrock

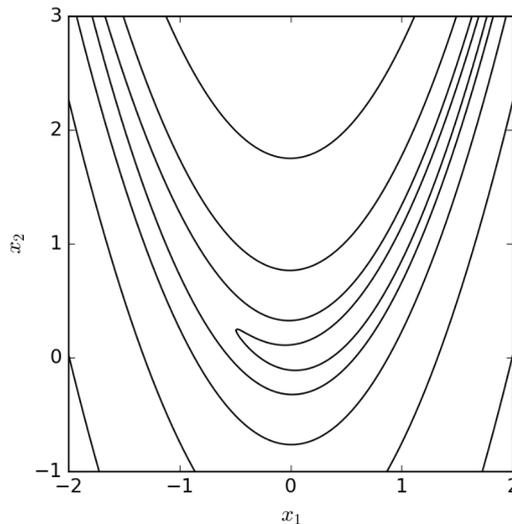
A primeira função analisada foi a função Rosenbrock. Ela é caracterizada como não convexa e implicou no desempenho dos métodos mostrado na Tabela 2.

Tabela 2 – Desempenho dos métodos na função Rosenbrock

<i>Método</i>	<i>Quantidade de Iterações</i>
Gradiente	55624
Newton	19
BFGS	20
Região de Confiança	30
Gradiente Conjugado Não Linear	379

Fonte: Os autores.

De imediato, percebe-se uma quantidade elevada de iterações do método do Gradiente em relação aos demais. Ao analisar a estrutura de f para a função Rosenbrock, percebe-se a existência de vales muito estreitos e longos, como pode ser visto na Figura 4, o que tornou o processo de *zigue-zague* do método do Gradiente expansivo e implicou na sua ineficiência quando aplicado ao problema.

Figura 4 – Curvas de nível da Função Rosenbrock

Fonte: Os autores.

O mínimo global da função ocorre em $x^* = (1, 1)$. Observa-se que os contornos não são espaçados uniformemente nessa figura; eles são espaçados logaritmicamente; portanto, a solução fica dentro de um vale muito profundo, estreito e em forma de banana – fator que torna a função Rosenbrock conhecida, também, como função Banana.

Diferentemente, os métodos Newton, BFGS e Região de Confiança mostram-se ideais para o problema, evidenciando que o modelo quadrático no qual cada um deles se baseia representou bem a função f . Considerando que as quantidades de iterações dos métodos de Newton e BFGS foram muito próximas, podemos, ainda, concluir que o acúmulo de informações de curvatura em cada iteração proporcionou ao método BFGS boas aproximações para a hessiana da função.

4.2 Função Freudenstein e Roth

A função Freudenstein e Roth é caracterizada por ser não convexa, e possui, além de um minimizador global, dois minimizadores locais. O ponto inicial $(0.5, -2)$ é suficientemente distante da solução para causar problemas no método de Newton, pois a hessiana deixa de ser definida positiva nesse ponto e, portanto, o método deixa de trabalhar sobre regiões convexas da função. Aliado a isso, o modelo quadrático nos quais se baseiam os métodos de Newton e de Região de Confiança não representa bem a função f em uma vizinhança do ponto inicial, fazendo com que os passos dados por esses métodos sejam bem pequenos nas iterações iniciais. A Tabela 3, a seguir, confirma tal informação.

Tabela 3 – Desempenho dos métodos na função Freudenstein e Roth

<i>Método</i>	<i>Quantidade de Iterações</i>
Gradiente	8550
Newton	9077
BFGS	7
Região de Confiança	16153
Gradiente Conjugado Não Linear	236

Fonte: Os autores.

Os resultados evidenciam uma diferença considerável de iterações entre os métodos BFGS e Newton, destacando, para esse caso não convexo, a superioridade da utilização de matrizes de atualização definidas positivas em relação à hessiana da função f . Apesar dessa divergência, ambos identificaram o mínimo global $x^* = (5, 4)$, no qual $f(x^*) = 0$. O único método que não identificou tal minimizador foi o de Região de Confiança, que encontrou o mínimo local $x \cong (11.41, -0, 90)$ em que $f(x) \cong 48.98$.

Nesse contexto, ressalta-se a importância da conferência dos autovalores da hessiana no ponto encontrado, a fim de verificar se ela realmente é definida positiva e satisfaz as condições suficientes de segunda ordem estabelecidas no Teorema 2.5. Em $x^* = (5, 4)$, tem-se todos os autovalores da matriz hessiana positivos, diferentemente de quando $x \cong (11.41, -0.90)$.

4.3 Função Powell Badly Scaled

A função Powell Badly Scaled tem como característica marcante seu mau condicionamento, proporcionando a tendência de propagação de erros ao se trabalhar com sua matriz hessiana. Esse fenômeno é consequência do número de condição da matriz hessiana ser grande, ou seja, dos seus autovalores terem magnitudes muito distintas.

A Tabela 4 mostra que o método BFGS e o Gradiente Conjugado Não Linear conseguiram identificar minimizadores dentro dos critérios de parada estabelecidos, enquanto os demais não obtiveram sucesso (fenômeno representado nessa e em outras tabelas pelo símbolo *).

Tabela 4 – Desempenho dos métodos na função Powell Badly Scaled

<i>Método</i>	<i>Quantidade de Iterações</i>
Gradiente	*
Newton	*
BFGS	238
Região de Confiança	*
Gradiente Conjugado Não Linear	13355

Fonte: Os autores.

Nessa função, o método do Gradiente Conjugado Não Linear encontra um minimizador local, enquanto o método BFGS identifica o minimizador global. Já o método de Newton falha ao longo do processo iterativo por encontrar um ponto em que a matriz hessiana fica muito próxima de ser singular, dificultando a identificação de uma direção adequada. Tal problema novamente é superado no método BFGS, pois este assume a aproximação da hessiana como definida positiva e, portanto, impede-a de tornar-se singular ao longo do processo iterativo na busca por um minimizador.

Os métodos do Gradiente e Região de Confiança apresentaram convergência lenta e, assim, extrapolaram o limite de iterações estabelecido.

4.4 Função Brown Badly Scaled

Assim como a função anterior, a função Brown Badly Scaled também se caracteriza pelo seu mau condicionamento. Porém, nesta, os métodos, em geral, tiveram melhor desempenho e a maioria deles conseguiu identificar o minimizador global, como mostra a Tabela 5.

Tabela 5 – Desempenho dos métodos na função Brown Badly Scaled

<i>Método</i>	<i>Quantidade de Iterações</i>
Gradiente	*
Newton	11
BFGS	18
Região de Confiança	33
Gradiente Conjugado Não Linear	121905

Fonte: Os autores.

Como a solução está relativamente distante do ponto inicial, métodos com convergência mais lenta acabaram apresentando um número elevado de iterações. Ressalta-se, nesse contexto, o método do Gradiente Conjugado Não Linear, que atingiu mais de 10^5 iterações, e o método do Gradiente, que extrapolou o limite máximo de iterações sem ter identificado o ponto ótimo.

4.5 Função Beale

A função Beale é considerada uma função multimodal, isto é, que apresenta em sua curvatura diversos mínimos e máximos locais ocasionados por vales e picos inerentes a sua geometria. Apesar de tal fator dificultar, em geral, os algoritmos de otimização a encontrar a solução ótima, a maioria dos métodos utilizados conseguiu identificar o minimizador global como mostrado na Tabela 6.

Tabela 6 – Desempenho dos métodos na função Beale

<i>Método</i>	<i>Quantidade de Iterações</i>
Gradiente	665
Newton	11
BFGS	16
Região de Confiança	8
Gradiente Conjugado Não Linear	7704

Fonte: Os autores.

O único método que não superou as referidas adversidades de curvatura da função e não encontrou o minimizador global $x^* = (3, 0.5)$ em que $f(x^*) = 0$, foi o Gradiente Conjugado Não Linear. Partindo do ponto inicial $x_0 = (1, 1)$, o método direcionou-se a um vale da função e encontrou um mínimo local no qual $f(x) \cong 0.4537$.

Nesse problema, o método de Newton e o de Região de Confiança destacaram-se por uma quantidade pequena de iterações, evidenciando que o modelo quadrático utilizado aproximou bem a função ao longo do processo iterativo.

4.6 Função Powell Singular

A função Powell Singular é classificada, quanto à sua curvatura, como convexa, e, portanto, o único minimizador que possui é global. Uma de suas características marcantes é possuir matriz hessiana com dupla singularidade na solução, tornando-se um possível problema na identificação da solução ótima para os métodos que utilizam tal operador ao longo do processo iterativo. Apesar desse aspecto negativo, os métodos demonstraram, em geral, serem robustos na identificação do minimizador, como pode ser visto na Tabela 7.

Tabela 7 – Desempenho dos métodos na função Powell Singular

<i>Método</i>	<i>Quantidade de Iterações</i>
Gradiente	34482
Newton	13
BFGS	39
Região de Confiança	13
Gradiente Conjugado Não Linear	617

Fonte: Os autores.

Nota-se que o gradiente da função f é relativamente próximo do vetor nulo no ponto inicial $x_0 = (3, -1, 0, 1)$, o que fez o método do Gradiente tomar passos muito pequenos e, conseqüentemente, necessitar de um número elevado de iterações para identificar a solução ótima $x^* = (0, 0, 0, 0)$. Além disto, a convergência quadrática relativa ao método de Newton não é observada para esta função, assim como evidenciado teórica e numericamente nos estudos de Steihaug e Suleiman (2012).

4.7 Função Wood

A sétima função analisada foi a função Wood. Suas características de curvatura são similares às da função Rosenbrock, tendo em vista a presença de longos vales em sua geometria.

Tabela 8 – Desempenho dos métodos na função Wood

<i>Método</i>	<i>Quantidade de Iterações</i>
Gradiente	9049
Newton	10599
BFGS	36
Região de Confiança	60
Gradiente Conjugado Não Linear	2331

Fonte: Os autores.

Nessa função, partindo do ponto inicial $x_0 = (-3, -1, -3, -1)$ e da utilização do método de Newton, a matriz hessiana permanece definida positiva até a sétima iteração. A partir de então, ela perde tal característica e o método passa a trabalhar sobre regiões não convexas da função, o que fez com que o método passasse a tomar direções de subida em algum momento do processo iterativo. Como alternativa para superar esse empecilho, adotou-se a direção do método do Gradiente a partir de quando $-\nabla f(x_k)^T (\nabla^2 f(x_k))^{-1} \nabla f(x_k) > 0$. Tal fator fez com que os passos passassem a ser menores e, por conseqüência, o número de iterações do método fosse alto, como pode ser visto na Tabela 8.

4.8 Função Trigonométrica

A função Trigonométrica foi uma das funções em que o método do Gradiente convergiu em menor quantidade de iterações, ocasionado, principalmente, pelo fato dela não ter um alto grau de não linearidade (Tabela 9).

Tabela 9 – Desempenho dos métodos na função Trigonométrica

<i>Método</i>	<i>Quantidade de Iterações</i>
Gradiente	90
Newton	24
BFGS	16
Região de Confiança	7
Gradiente Conjugado Não Linear	22

Fonte: Os autores.

Apesar de todos os métodos partirem do mesmo ponto inicial, o método de Região de Confiança encontrou um minimizador diferente e melhor que os demais, no qual a função assume um valor mais próximo de zero. Porém, diferentemente do ocorrido na Função Freudenstein e Roth, ambos os minimizadores identificados estão muito próximos da solução ótima, fazendo com que os autovalores da hessiana sejam todos positivos nos dois casos.

4.9 Função Linear de Posto Completo

A função Linear de Posto Completo é uma função que preserva hessiana definida positiva, para todo $x \in \mathbb{R}^n$. Para seu processo de minimização neste estudo utilizou-se $n = 3$, o que implicou na obtenção dos resultados mostrados na Tabela 10.

Tabela 10 – Desempenho dos métodos na função Linear de Posto Completo

<i>Método</i>	<i>Quantidade de Iterações</i>
Gradiente	*
Newton	1
BFGS	2
Região de Confiança	3
Gradiente Conjugado Não Linear	118

Fonte: Os autores.

De imediato, percebe-se a eficiência de alguns métodos para resolver problemas lineares de pequenas dimensões. A exceção deu-se no método do Gradiente que, em determinada parte do processo iterativo, começou a rebater dois valores interminavelmente e acabou deixando de convergir.

4.10 Função Linear de Posto 1

A função Linear de Posto 1, como o próprio nome já insinua, possui hessiana de posto incompleto, portanto, não inversível. Tal fato torna a função imprópria para a utilização do método de Newton devido à impossibilidade de calcular p_k . Diferentemente, os demais métodos resolveram com facilidade o problema, como mostra a Tabela 11.

Tabela 11 – Desempenho dos métodos na função Linear de Posto 1

<i>Método</i>	<i>Quantidade de Iterações</i>
Gradiente	78
Newton	*
BFGS	2
Região de Confiança	2
Gradiente Conjugado Não Linear	145

Fonte: Os autores.

Assim como ocorrido na função Linear de Posto Completo, a utilização das direções linearmente independentes p_k no método do Gradiente Conjugado Não Linear não se mostrou vantajosa sobre os demais métodos. Diferentemente da grande maioria dos problemas analisados, nesse caso, o referido método teve maior quantidade de iterações do que o método do Gradiente.

A análise dessa e das demais funções mostra-se, assim, uma ferramenta primordial na verificação do desempenho dos métodos computacionais a serem utilizados no processo de minimização. Cabe, portanto, a cada pesquisador usufruir do senso crítico e dos conhecimentos matemáticos, a fim de realizar as melhores escolhas ao lidar com problemas de Otimização.

5 CONSIDERAÇÕES FINAIS

Ao analisar exclusivamente a teoria que embasa a otimização irrestrita, muitos dos empecilhos que podem vir a ocorrer ao utilizá-la na prática passam despercebidos. Nesse sentido, a implementação computacional dos métodos estudados e a análise de seu desempenho frente às adversidades encaminha o estudante/pesquisador a desenvolver uma aprendizagem sólida e significativa em relação aos conceitos em foco.

A partir da prática desenvolvida neste estudo, fica nítida a fragilidade que os métodos de Otimização possuem para encontrar um minimizador global. Toda teoria que os sustenta não é suficiente para garantir que o resultado esperado será identificado; é necessário, por vezes, examinar o ponto inicial, analisar a curvatura da função, avaliar o gradiente e a hessiana, verificar autovalores, entre outros fatores que podem interferir no processo de convergência.

Dentre os maiores problemas constatados no desempenho dos métodos empregados neste estudo está a sua fragilidade em identificar minimizadores globais de funções não conve-

xas. Nessa perspectiva, a alteração do ponto inicial e a conferência dos autovalores da matriz hessiana no ponto encontrado mostram-se estratégias para classificar os diferentes minimizadores. Para além disto, a verificação da taxa de convergência de cada método, do seu custo computacional e da distância do ponto inicial ao minimizador, constituem fatores importantes a serem analisados no processo de minimização.

Por fim, ressalta-se que não existe um método de otimização irrestrita ideal para todos os problemas. A chave para a escolha do método que irá realizar o processo de minimização com maior eficiência e robustez está na análise da estrutura, comportamento e geometria da função na qual os métodos serão empregados. Otimizar 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.

REFERÊNCIAS

- ANDRETTA, M. **Sistemas lineares: Método dos gradientes conjugados**. Disponível em: <<http://sites.icmc.usp.br/andretta/ensino/aulas/sme0100-2-12/aula6-gradconj.pdf>>. Acesso em 20 nov. 2019.
- BORGES, P. R. T. **Aspectos sobre a aplicação de Métodos de Gradientes Conjugados em Otimização Irrestrita**. 1985. Dissertação (Mestrado em Engenharia) — Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1985.
- FLETCHER, R.; REEVES, C. M. Function minimization by conjugate gradients. **Computer journal**, v. 7, p. 149–154, 1964.
- LIMA, E. L. **Curso de análise**. Rio de Janeiro: IMPA, 2015. v. 2.
- MORÉ, J. J.; GARBOW, B. S.; HILLSTROM, K. E. Testing unconstrained optimization software. **ACM Transactions on Mathematical Software**, ACM New York, NY, USA, v. 7, p. 17–41, 1981.
- MORÉ, J. J.; SORENSEN, D. C. Computing a trust region step. **SIAM Journal on scientific and statistical computing**, v. 4, p. 553–572, 1983.
- NOCEDAL, J.; WRIGHT, S. **Numerical optimization**. New York: Springer, 1998.
- RIBAS, F. L. M. **Métodos de Região de Confiança para Minimização Irrestrita**. 2017. Monografia (Graduação em Matemática Industrial) – Universidade Federal do Paraná, Curitiba, 2017.
- STEIHAUG, T.; SULEIMAN, S. Global convergence and the powell singular function. **Journal of Global Optimization**, v. 56, n. 3, p. 845–853, 2012.
- URREA, L. J. **Um metodo de região de confiança para minimização irrestrita sem derivadas**. 2008. Dissertação (Mestrado) — Universidade Estadual de Campinas, Campinas, 2008.