

Busca de Wolfe: Parâmetros Eficientes Obtidos Matematicamente

N. M. NASCIMENTO^{1*}, T. A. FERNANDES² e A. D. V. SANTIAGO³

Recebido em 20 de outubro de 2025 / Aceito em 21 de agosto de 2026

RESUMO. Encontrar zeros do gradiente de uma função real de n -variáveis reais é um problema matemático que pode ser usado para resolver aplicações práticas do nosso cotidiano, por exemplo, reconhecimento de fala e tradução automática. Para solucionar este problema, na área de Otimização Contínua, são desenvolvidos métodos iterativos que geram sequências que convergem para o zero do gradiente. Tais métodos devem ser computacionalmente eficientes. Uma das teorias matemáticas que contribuem para esta eficiência é a Busca Linear, responsável por encontrar os melhores tamanhos de passos a cada iterada da sequência gerada. De modo geral, estas buscas são definidas por parâmetros que interferem diretamente na eficiência numérica para solucionar o problema. Este aspecto, torna relevante a investigação de parâmetros ótimos. Apesar disso, a literatura atual não tem apresentado parâmetros ótimos obtidos matematicamente. Neste trabalho, com a finalidade de encontrarmos os zeros do gradiente de uma função acadêmica, determinamos matematicamente um intervalo de parâmetros da busca linear de Wolfe. Tal intervalo assegura tamanhos de passos mais eficientes quando comparados com outros tamanhos de passos fora do intervalo. Simulações numéricas mostram a validade do resultado teórico apresentado.

Palavras-chave: parâmetros ótimos, busca de Wolfe, otimização contínua.

1 INTRODUÇÃO

O problema de interesse deste trabalho está no contexto de encontrar zeros do gradiente de uma função real de n -variáveis reais. A obtenção desses zeros permite, por exemplo, minimizar ou maximizar funções [2, 4]. Consequentemente, a partir disso surgem várias aplicações nos mais diversos campos científicos e tecnológicos, como na resolução de problemas da engenharia civil [1] e Deep Learning [7]. Encontrar zeros do gradiente envolve, na Otimização Contínua, o desenvolvimento de métodos iterativos (ou métodos de otimização) que geram sequências que convergem para uma solução do problema. Ferramentas matemáticas para aprimorar computacionalmente esses métodos, tem sido estudadas na literatura atual, por exemplo, direção de descida e Busca Linear (foco deste trabalho).

*Corresponding author: Nalanda Moreira Nascimento – E-mail: nalandamoreira10@gmail.com

¹ Universidade Estadual do Sudoeste da Bahia (UESB), BA, Brasil <https://orcid.org/0009-0001-0288-2938>

² Universidade Estadual do Sudoeste da Bahia (UESB), BA, Brasil <https://orcid.org/0000-0002-1755-8532>

³ Universidade Estadual do Sudoeste da Bahia (UESB), BA, Brasil <https://orcid.org/0000-0002-0759-5807>

Na literatura matemática, podemos encontrar diversas buscas lineares, tal como as buscas de Armijo, Wolfe e Goldstein. Veja em [5, 6]. Elas são determinadas matematicamente com o uso de alguns parâmetros. Tais parâmetros tornam os métodos iterativos altamente sensíveis no seu desempenho numérico. Como exemplo, temos a busca de Armijo que em certos problemas pode obter, numericamente, tamanhos de passo muito pequenos. Para evitar isso, em geral, utiliza-se a busca de Wolfe. Essa busca é composta pela condição de Armijo e por uma segunda condição que tem o papel de evitar tamanhos de passos muito pequenos. Tais buscas fazem o uso de parâmetros numéricos η_1 e η_2 .

Teoricamente, a convergência global de um algoritmo que utiliza a busca de Wolfe para a escolha do tamanho do passo exige que $0 < \eta_1 < \eta_2 < 1$. Todavia, percebe-se que valores aleatórios escolhidos, ainda que satisfaçam esse intervalo, podem não fornecer bons desempenhos numéricos a esses algoritmos. Apesar disso, não têm sido investigados critérios matemáticos para obter valores de η_1 e η_2 que assegurem um desempenho numérico eficiente.

Motivados por esse problema, a fim de encontrar os zeros do gradiente de uma função acadêmica, estabelecemos matematicamente um intervalo ótimo de parâmetros η_2 que garante que a busca de Wolfe execute o seu papel de maneira eficiente. Fizemos simulações numéricas as quais evidenciaram maior eficiência do algoritmo de Wolfe quando consideramos parâmetros η_2 dentro do intervalo obtido.

O restante deste trabalho está estruturado da seguinte forma: na próxima seção mostramos os conceitos básicos e resultados preliminares. Na sequência, apresentamos a Seção 3 onde encontra-se o principal resultado obtido. Abordamos os experimentos numéricos na Seção 4. Os comentários finais compõem a última seção deste trabalho.

2 CONCEITOS PRELIMINARES

Nesta seção, apresentamos os principais resultados necessários ao desenvolvimento deste trabalho. Basicamente, nos referenciamos por [3, 5]. Iniciamos com a noção de direção de descida.

Definição 2.1 (Direção de Descida). *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$, d é direção de descida para f a partir de x se $\exists \delta > 0$ tal que $f(x + \alpha d) < f(x)$, $\forall \alpha \in (0, \delta]$.*

O teorema a seguir estabelece uma condição necessária para que determinada direção seja de descida.

Teorema 2.1. *Se $\langle \nabla f(x), d \rangle < 0$, d é direção de descida.* Uma demonstração desse teorema pode ser encontrada em [3, Cap. 4, Teorema 4.2].

Observação 1. *Neste trabalho, fixamos $d = -\nabla f(x) \neq 0$, a qual é uma direção de descida. De fato, $\langle \nabla f(x), d \rangle = \langle \nabla f(x), -\nabla f(x) \rangle = -\|\nabla f(x)\|^2 < 0$.*

Agora, apresentamos formalmente as buscas lineares de Armijo e Wolfe.

Teorema 2.2 (Busca de Armijo). *Sejam $f : \mathbb{R}^n \rightarrow \mathbb{R}$ diferenciável, $\hat{p}, d \in \mathbb{R}^n$, onde d é direção de descida e $\eta_1 \in (0, 1)$. Então $\exists \delta > 0$ tal que*

$$f(\hat{p} + \alpha d) \leq f(\hat{p}) + \alpha \eta_1 \langle \nabla f(\hat{p}), d \rangle, \quad \forall \alpha \in (0, \delta). \quad (2.1)$$

Proof. Veja [3, Cap. 4, pag. 44]. □

Teorema 2.3 (Busca de Wolfe). *Sejam $f : \mathbb{R}^n \rightarrow \mathbb{R}$ continuamente diferenciável, $\hat{p}, d \in \mathbb{R}^n$, com d direção de descida e suponha que f é limitada inferiormente no conjunto $\Omega = \{\hat{p} + \alpha d; \alpha > 0\}$. Então se $0 < \eta_1 < \eta_2 < 1$ existem pelo menos δ_1 e δ_2 tais que*

$$f(\hat{p} + \alpha d) \leq f(\hat{p}) + \alpha \eta_1 \langle \nabla f(\hat{p}), d \rangle, \quad \forall \alpha \in (\delta_1, \delta_2) \quad (2.2)$$

$$\langle \nabla f(\hat{p} + \alpha d), d \rangle \geq \eta_2 \langle \nabla f(\hat{p}), d \rangle, \quad \forall \alpha \in (\delta_1, \delta_2). \quad (2.3)$$

Proof. Primeiro, vamos provar (2.2). Suponha que $\langle \nabla f(\hat{p}), d \rangle = 0$. Desde que d é uma direção de descida, $\exists \delta > 0$ tal que $f(\hat{p} + \alpha d) < f(\hat{p}) \quad \forall \alpha \in (0, \delta)$. Dessa última desigualdade, concluímos $f(\hat{p} + \alpha d) < f(\hat{p}) + \alpha \eta_1 \langle \nabla f(\hat{p}), d \rangle \quad \forall \alpha \in (0, \delta)$. Considere δ_1 e δ_2 tais que $0 < \delta_1 < \delta_2 < \delta$. Dessa forma, podemos escrever

$$f(\hat{p} + \alpha d) < f(\hat{p}) + \alpha \eta_1 \langle \nabla f(\hat{p}), d \rangle \quad \text{com} \quad \delta_1 < \alpha < \delta_2. \quad (2.4)$$

No caso que $\langle \nabla f(\hat{p}), d \rangle < 0$, da definição de derivada direcional podemos concluir $\lim_{\alpha \rightarrow 0} (f(\hat{p} + \alpha d) - f(\hat{p})) / \alpha = \langle \nabla f(\hat{p}), d \rangle$. Portanto, $\forall \varepsilon > 0, \exists \delta_\varepsilon > 0$ tal que

$$\left| \frac{f(\hat{p} + \alpha d) - f(\hat{p})}{\alpha} - \langle \nabla f(\hat{p}), d \rangle \right| < \varepsilon, \quad \text{com} \quad 0 < \alpha < \delta_\varepsilon. \quad (2.5)$$

Assim, como $0 < \eta_1 < 1$ e $\langle \nabla f(\hat{p}), d \rangle < 0$, então $\langle \nabla f(\hat{p}), d \rangle < \eta_1 \langle \nabla f(\hat{p}), d \rangle$. Consequentemente, $\eta_1 \langle \nabla f(\hat{p}), d \rangle - \langle \nabla f(\hat{p}), d \rangle = (\eta_1 - 1) \langle \nabla f(\hat{p}), d \rangle > 0$. Logo, se tomarmos $\hat{\varepsilon} = (\eta_1 - 1) \langle \nabla f(\hat{p}), d \rangle$, $\exists \delta_{\hat{\varepsilon}} > 0$ tal que de (2.5) obtemos

$$\frac{f(\hat{p} + \alpha d) - f(\hat{p})}{\alpha} - \langle \nabla f(\hat{p}), d \rangle < (\eta_1 - 1) \langle \nabla f(\hat{p}), d \rangle \quad \text{com} \quad 0 < \alpha < \delta_{\hat{\varepsilon}}.$$

Dessa forma, considerando δ_1, δ_2 dados em (2.4) e ajustando $\delta_{\hat{\varepsilon}}$, se necessário, temos

$$f(\hat{p} + \alpha d) < f(\hat{p}) + \alpha \eta_1 \langle \nabla f(\hat{p}), d \rangle \quad \text{com} \quad \delta_1 < \alpha < \delta_2.$$

Isso conclui a prova de (2.2). Agora, vamos provar (2.3). Desde que f é limitada inferiormente em $\Omega = \{\hat{p} + \alpha d; \alpha > 0\}$, então existe uma constante $c \in \mathbb{R}$ tal que

$$c < \phi(\alpha) := f(\hat{p} + \alpha d), \quad \forall \alpha > 0. \quad (2.6)$$

Além disso,

$$l(\alpha) := f(\hat{p}) + \alpha \eta_1 \langle \nabla f(\hat{p}), d \rangle, \quad \forall \alpha > 0 \quad (2.7)$$

é função afim decrescente. Como d é uma direção de descida, existe $\delta > 0$ tal que

$$f(\hat{p} + \alpha d) < f(\hat{p}), \forall \alpha \in (0, \delta). \quad (2.8)$$

De (2.6), (2.7) e (2.8), podemos concluir que existe $\alpha' > 0$ tal que

$$f(\hat{p} + \alpha' d) = f(\hat{p}) + \alpha' \eta_1 \langle \nabla f(\hat{p}), d \rangle. \quad (2.9)$$

Tomando o intervalo $(0, \alpha')$, pelo Teorema do Valor Médio $\exists \alpha'' \in (0, \alpha')$ tal que

$$f(\hat{p} + \alpha' d) - f(\hat{p}) = \alpha' \langle \nabla f(\hat{p} + \alpha'' d), d \rangle. \quad (2.10)$$

De (2.9) e (2.10) obtemos que

$$\langle \nabla f(\hat{p} + \alpha'' d), d \rangle = \eta_1 \langle \nabla f(\hat{p}), d \rangle. \quad (2.11)$$

Desde que $\eta_1 < \eta_2$ e $\langle \nabla f(\hat{p}), d \rangle < 0$, concluímos

$$\eta_1 \langle \nabla f(\hat{p}), d \rangle > \eta_2 \langle \nabla f(\hat{p}), d \rangle. \quad (2.12)$$

De (2.11) e (2.12) segue que $\langle \nabla f(\hat{p} + \alpha'' d), d \rangle > \eta_2 \langle \nabla f(\hat{p}), d \rangle$ com $0 < \alpha'' < \alpha'$. Logo, como f é continuamente diferenciável existe $\delta > 0$ tal que $\langle \nabla f(\hat{p} + \alpha d), d \rangle > \eta_2 \langle \nabla f(\hat{p}), d \rangle$ com $\alpha \in (0, \delta)$. Considerando δ_1, δ_2 dados em (2.4) e ajustando δ , se necessário, temos

$$\langle \nabla f(\hat{p} + \alpha d), d \rangle > \eta_2 \langle \nabla f(\hat{p}), d \rangle \quad \text{com} \quad \delta_1 < \alpha < \delta_2.$$

□

Com o intuito de facilitar a aplicação do Teorema 2.3 no objetivo principal desse trabalho, denotamos

$$\phi(\alpha) := f(\hat{p} + \alpha d), \quad (2.13)$$

$$l(\alpha) := f(\hat{p}) + \alpha \eta_1 \langle \nabla f(\hat{p}), d \rangle, \quad (2.14)$$

$$g(\alpha) := \langle \nabla f(\hat{p} + \alpha d), d \rangle, \quad (2.15)$$

$$C(\eta_2) := \eta_2 \langle \nabla f(\hat{p}), d \rangle. \quad (2.16)$$

Dessa forma, de (2.2) e (2.3) segue que

$$\phi(\alpha) \leq l(\alpha), \quad \forall \alpha \in (\delta_1, \delta_2) \quad (2.17)$$

$$g(\alpha) \geq C(\eta_2), \quad \forall \alpha \in (\delta_1, \delta_2). \quad (2.18)$$

Observação 2. As buscas de Armijo e Wolfe garantem a existência de pelo menos um conjunto de valores possíveis para os tamanhos de passos, o qual chamamos intervalos admissíveis. No caso da busca de Armijo os intervalos admissíveis são constituídos pelos passos que satisfazem a condição (2.1), já na condição de Wolfe devemos admitir apenas aqueles que tornam válidas as desigualdades (2.2) e (2.3). Em nosso estudo, enfatizamos a influência do parâmetro η_2 da desigualdade (2.3) na obtenção de melhores tamanhos de passos, ou seja, melhores intervalos admissíveis.

Utilizando a busca de Wolfe [5, Cap. 3, (3.6a)–(3.6b)] no Algoritmo de Cauchy [3, Cap. 5, Alg. 5.1], apresentamos o método a seguir.

Algorithm 1 Método do Gradiente com Busca de Wolfe

1. Definir $0 < \eta_1 < \eta_2 < 1$, p_k e $d_k \in \mathbb{R}^n$, $f : \mathbb{R}^n \rightarrow \mathbb{R}$;
2. Pare somente quando $\|\nabla f(p_k)\| \leq 10^{-6}$;
3. Definir $d_k = -\nabla f(p_k)$;
4. Escolha α_k tal que

$$f(p_k + \alpha_k d_k) \leq f(p_k) + \eta_1 \alpha_k \langle \nabla f(p_k), d_k \rangle,$$

$$\langle \nabla f(p_k + \alpha_k d_k), d_k \rangle \geq \eta_2 \langle \nabla f(p_k), d_k \rangle;$$

5. Faça $p_{k+1} = p_k + \alpha_k d_k$;
 6. $k = k + 1$.
-

Nosso interesse está em obter tamanhos de passos suficientemente grandes no **item 4** do algoritmo acima. Em [3, Cap. 4, Teorema 4.18] é provado que esse algoritmo, quando equipado com a busca de Armijo, converge globalmente. Utilizando ideias dessa prova, obtivemos uma demonstração do Teorema 2.4 apresentado a seguir. Observamos que esse teorema exige que $0 < \eta_1 < \eta_2 < 1$. Contudo, podem ser tomados η_1 e η_2 nessa condição de modo que o desempenho numérico do algoritmo não seja satisfatório. Na Seção 4, determinamos um intervalo para η_2 de modo a assegurar um bom desempenho numérico do Algoritmo 1. Destacamos que a prova do Teorema 2.3 garante que o passo 4 do Algoritmo 1 está bem definido.

Teorema 2.4 (Convergência Global do Algoritmo 1). *Sejam $f : \mathbb{R}^n \rightarrow \mathbb{R}$ de classe C^1 , $\hat{p}, d \in \mathbb{R}^n$, com $d = -\nabla f(\hat{p})$ e $0 < \eta_1 < \eta_2 < 1$. O Algoritmo 1 é globalmente convergente.*

Proof. Seja (p_k) uma sequência gerada pelo Algoritmo 1 e seja $\bar{p}$ um ponto de acumulação de (p_k) . Então existe uma subsequência $(p_{k'})$ tal que $p_{k'} \rightarrow \bar{p}$. Suponha que $\nabla f(\bar{p}) \neq 0$, isto é, $\bar{p}$ não é ponto crítico de f .

Como f é contínua, segue que $f(p_{k'}) \rightarrow f(\bar{p})$. Além disso, considerando a Observação 1, como $d_k = -\nabla f(p_k)$ é direção de descida, temos $f(p_{k+1}) < f(p_k)$, logo $(f(p_k))$ é monótona decrescente e possui subsequência convergente. Logo, temos $f(p_k) \rightarrow f(\bar{p})$. Por outro lado, do Teorema 2.3, existe α_k tal que

$$f(p_{k+1}) = f(p_k - \alpha_k \nabla f(p_k)) \leq f(p_k) - \eta_1 \alpha_k \langle \nabla f(p_k), \nabla f(p_k) \rangle.$$

Afirmamos que $\alpha_{k'} \rightarrow 0$. De fato, desta última desigualdade temos que

$$f(p_{k+1}) - f(p_k) \leq -\eta_1 \alpha_k \langle \nabla f(p_k), \nabla f(p_k) \rangle = -\eta_1 \alpha_k \|\nabla f(p_k)\|^2. \quad (2.19)$$

Desde que $\eta_1 > 0$ e $\alpha_k > 0$ concluímos que

$$-\eta_1 \alpha_k \|\nabla f(p_k)\|^2 \leq 0. \quad (2.20)$$

De (2.19) e (2.20) temos $f(p_{k+1}) - f(p_k) \leq -\eta_1 \alpha_k \|\nabla f(p_k)\|^2 \leq 0$. Já que $p_{k'} \rightarrow \bar{p}$, do Teorema do Confronto, provamos nossa afirmação, isto é

$$\alpha_{k'} \rightarrow 0. \quad (2.21)$$

Pela condição (2.3) do Teorema 2.3,

$$-\langle \nabla f(p_k - \alpha_k \nabla f(p_k)), \nabla f(p_k) \rangle \geq -\eta_2 \|\nabla f(p_k)\|^2. \quad (2.22)$$

Sendo f continuamente diferenciável, $p_{k'} \rightarrow \bar{p}$, levando em conta (2.21) e tomando o limite em (2.22) quando $k' \rightarrow \infty$, obtemos

$$-(1 - \eta_2) \|\nabla f(\bar{p})\|^2 \geq 0.$$

Como $0 < \eta_2 < 1$, concluímos desta última desigualdade que $\nabla f(\bar{p}) = 0$, o que contradiz a hipótese inicial. Portanto, todo ponto de acumulação da sequência gerada pelo Algoritmo 1 é ponto crítico de f . $\square$

3 OBTENÇÃO DE PARÂMETROS η_2 QUE ASSEGURAM PASSOS EFICIENTES

Nesta seção, apresentamos valores de η_2 que asseguram matematicamente que o passo obtido por Wolfe é mais eficiente do que aquele obtido por Armijo. Para tanto, consideramos $p_0 = (0; 0)$, $d_0 = -\nabla f(0; 0) = (85/6; 0)$, onde $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ é definida por

$$f(x, y) = \frac{1}{6} \left[\left(6x - 6 \right)^2 \left(\frac{x}{6} - 1 \right)^2 - x + y^2 - 8 \right]. \quad (3.1)$$

Com isso, primeiramente, vamos obter tamanhos de passo exclusivamente pela busca de Armijo. Deste modo, consideramos o seguinte problema

$$\min f(p_0 + \alpha d_0), \quad \alpha > 0.$$

Levando em conta que $p_0 + \alpha d_0 = (85\alpha/6; 0)$ e utilizando a igualdade (3.1) temos

$$\phi(\alpha) := f(p_0 + \alpha d_0) = \frac{1}{6} \left[\left(85\alpha - 6 \right)^2 \left(\frac{85\alpha}{36} - 1 \right)^2 - \frac{85\alpha}{6} - 8 \right].$$

Usando $\eta_1 = 10^{-2}$, p_0 , d_0 definidos acima e (3.1), temos

$$l(\alpha) := f(p_0) + \alpha \eta_1 \langle \nabla f(p_0), d_0 \rangle = \frac{14}{3} - \frac{289}{144} \alpha.$$

Pela condição (2.1), existe $\delta > 0$ tal que $\phi(\alpha) \leq l(\alpha), \forall \alpha \in (0, \delta)$. Para analisarmos quais os valores de α satisfazem esta última desigualdade, a partir da interpretação geométrica da Figura 1, observamos que l é uma função que limita superiormente ϕ . Logo,

$$\alpha \in]0, 0.22] \cup [0.28, 0.49] \quad \text{satisfaz a condição de Armijo.} \quad (3.2)$$

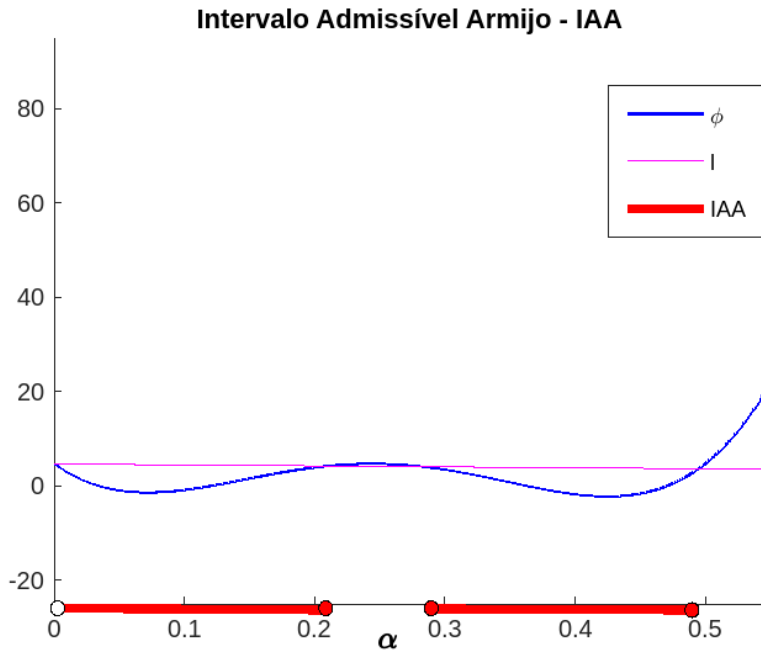


Figura 1: Intervalos admissíveis da busca de Armijo.

Agora, nossa proposta é apresentar um intervalo de parâmetros para η_2 que irá assegurar tamanhos de passo obtidos por Wolfe mais eficientes que os obtidos em (3.2). Antes disso, considerando $g(\alpha) := \langle \nabla f(p_0 + \alpha d_0), d_0 \rangle$, $C(\eta_2) := \eta_2 \langle \nabla f(p_0), d_0 \rangle$, $p_0 = (0; 0)$ e $d_0 = -\nabla f(0; 0) = (85/6; 0)$ obtemos

$$g(\alpha) = -\frac{7225}{36} + \frac{440725}{108} \alpha - \frac{4298875}{216} \alpha^2 + \frac{52200625}{1944} \alpha^3 \quad (3.3)$$

$$C(\eta_2) := -\frac{7225}{36}\eta_2, \quad \eta_1 < \eta_2 < 1. \quad (3.4)$$

$$r_{\eta_2}(\alpha) = (0, C(\eta_2)) + \alpha(1, 0), \quad \alpha \in \mathbb{R}^+, \quad (3.5)$$

$$s_{\bar{\alpha}}(\alpha) = (0, g(\bar{\alpha})) + \alpha(1, 0), \quad \alpha \in \mathbb{R}^+ \quad (3.6)$$

onde $\bar{\alpha}$ é um minimizador local de g .

A partir da geometria (obtida usando os softwares Geogebra e MatLab) apresentada pela função g dada em (3.3) e pelas retas r e s dadas em (3.5) e (3.6), notamos que, quando a reta r está abaixo da reta s , “ α ’s pequenos” admitidos por (2.17) (condição de Armijo) não seriam descartados pela desigualdade (2.18) (condição de Wolfe). Por outro lado, quando a reta r está acima da reta s , ou seja,

$$C(\eta_2) \geq g(\bar{\alpha}), \quad (3.7)$$

a condição de Wolfe elimina significativamente “ α ’s pequenos” admitidos pela condição de Armijo.

Desde que $l(\alpha) = f(p_0) + \alpha\eta_1 \langle \nabla f(p_0), d_0 \rangle = \frac{14}{3} - \frac{289}{144}\alpha$ e impondo

$$l(\alpha) = C(\eta_2)$$

obtemos,

$$\alpha_{\eta_2} = 100\eta_2 + \frac{672}{289}. \quad (3.8)$$

Por outro lado, levando em conta (3.3), (3.7) e assumindo $l(\alpha) = g(\bar{\alpha})$ concluímos

$$\hat{\alpha} = -\frac{144}{289}g(\bar{\alpha}) + \frac{672}{289}. \quad (3.9)$$

Portanto, por (3.7), (3.8) e (3.9) queremos $\alpha_{\eta_2} \leq \hat{\alpha}$. Já que $g(\bar{\alpha}) \approx -59.16$, temos $\eta_2 < 0.29$. Assim, dessa última desigualdade bons parâmetros satisfazem

$$\eta_2 \in (\eta_1, 0.29). \quad (3.10)$$

Para exemplificar que (3.10) fornece passos eficientes, vamos apresentar a geometria do problema para $\eta_2 = 0.02$. Dessa forma, de (3.3) e (3.4) temos

$$C(0, 02) = -\frac{289}{72},$$

$$g(\alpha) = -\frac{7225}{36} + \frac{440725}{108}\alpha - \frac{4298875}{216}\alpha^2 + \frac{52200625}{1944}\alpha^3.$$

Pelas condições (2.17), (2.18) e analisando a Figura 2 nota-se que $\alpha \in [0.07, 0.22] \cup [0.43, 0.49]$ satisfaz o critério de Wolfe.

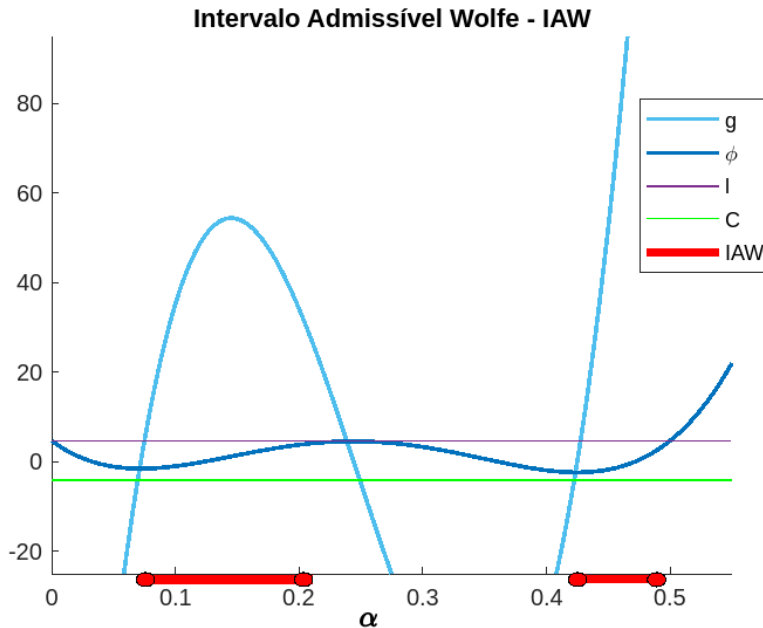


Figura 2: Eficiência da busca de Wolfe quando $\eta_2 = 0,02$.

Comparando esse intervalo com o intervalo admissível de Armijo obtido em (3.2), é possível perceber que

$$\alpha \in (0, 0.07) \cup [0.28, 0.43] \quad \text{não são aceitos.} \quad (3.11)$$

Por outro lado, se $\eta_2 = 0,85 \notin (\eta_1, 0.29)$, com cálculos análogos notamos que $\alpha \in [0.007, 0.22] \cup [0.28, 0.49]$ satisfaz o critério de Wolfe. Comparando (veja a Figura 3) esse intervalo com o intervalo admissível de Armijo obtido em (3.2), é possível perceber

$$\alpha \in (0, 0.007) \quad \text{não são aceitos.} \quad (3.12)$$

Logo, a partir de (3.11) e (3.12), concluímos que $\eta_2 = 0,02 \in (\eta_1, 0.29)$ eliminou tamanhos de passos pequenos de maneira mais significativa do que quando tomamos $\eta_2 = 0,85 \notin (\eta_1, 0.29)$.

4 RESULTADOS NUMÉRICOS

Nesta seção, apresentamos o desempenho numérico do Algoritmo 1 para encontrar os pontos críticos da função dada em (3.1). Analisamos seu desempenho quando utilizamos os parâmetros $\eta_2 \in (\eta_1, 0.29)$ e $\eta_2 \notin (\eta_1, 0.29)$. Pode ser visto nas Figuras 4a e 4b que o melhor desempenho do algoritmo ocorre para $\eta_2 \in (\eta_1, 0.29)$ obtido em (3.10).

O experimento numérico foi realizado por meio de um algoritmo desenvolvido na linguagem Java. A sua implementação completa pode ser acessada no Github <https://github.com/DavidVSantiago/CriterioDeWolfe>. Com relação às especificações de hardware, o experimento foi executado em notebook Dell modelo Inspiron 153511, com processador Intel Core

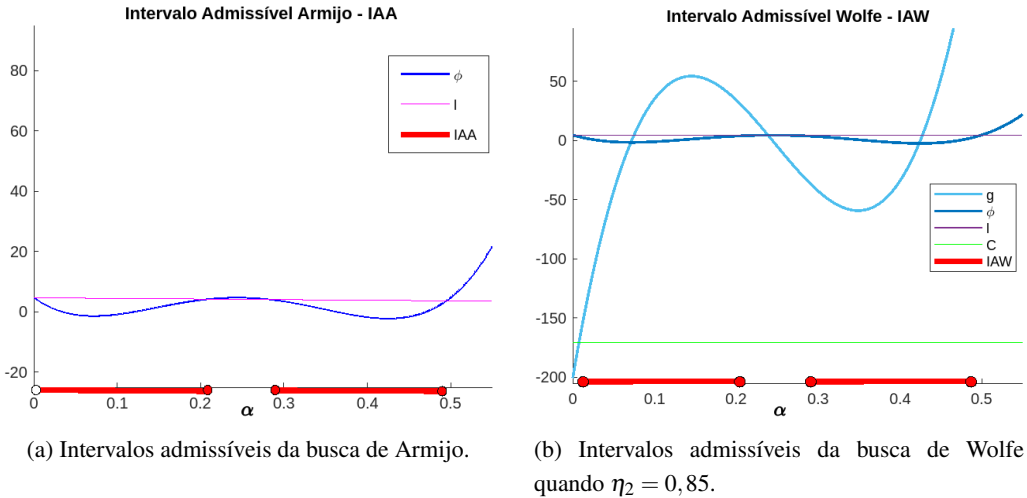


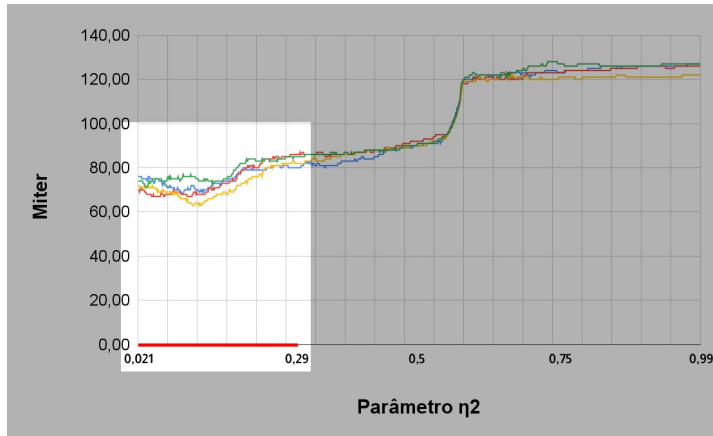
Figura 3: Ineficiência da busca de Wolfe.

$i5 - 1135G7$. Foram utilizados 400 pontos iniciais, critério de parada para a convergência $\|\nabla f(x_k)\| \leq 10^{-6}$, $\eta_1 = 10^{-2}$ e $\eta_2 \in (\eta_1, 1)$.

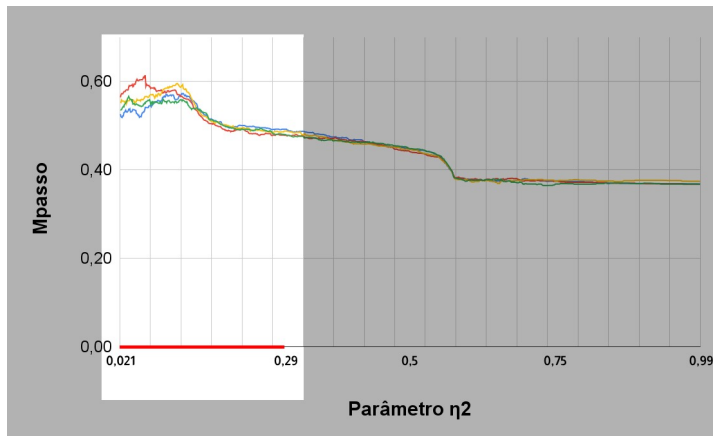
A Figura 4a apresenta o desempenho do Algoritmo 1 relacionado ao número de iterações em função do parâmetro η_2 . Cada curva foi construída a partir de um conjunto de 100 pontos iniciais gerados randomicamente. No eixo horizontal temos os valores de η_2 e no eixo vertical a variável $M_{iter} = 1/100 \sum_{i=1}^{100} iter_{x_{0i}}$, onde $iter_{x_{0i}}$ é o número de iterações para convergência da sequência gerada, por cada ponto inicial. Observe que, dentro do intervalo de parâmetros obtido em (3.10), isto é, para $\eta_2 \in (\eta_1, 0.29)$, o número de iterações é inferior ao número de iterações obtidas quando $\eta_2 \notin (\eta_1, 0.29)$.

A Figura 4b apresenta o desempenho do Algoritmo 1 relacionado ao tamanho do passo em função do parâmetro η_2 . Cada respectiva curva em ambos os gráficos, foi construída a partir do mesmo conjunto de 100 pontos iniciais utilizados para gerar a Figura 4a. No eixo horizontal, temos os valores de η_2 e no eixo vertical a variável $M_{passo} := 1/100 \sum_{i=1}^{100} \left(1/iter_{x_{0i}} \sum_{k=1}^{iter_{x_{0i}}} \alpha_k \right)$, onde para cada ponto inicial, $iter_{x_{0i}}$ é o número de iterações para convergência da sequência gerada e α_k o tamanho do passo obtido. Observe que, dentro do intervalo de parâmetros obtido em (3.10), isto é, para $\eta_2 \in (\eta_1, 0.29)$, os tamanhos de passos são maiores em comparação aos tamanhos de passos obtidos quando $\eta_2 \notin (\eta_1, 0.29)$.

É importante ressaltar que as Figuras 4a e 4b apresentam o comportamento do Algoritmo 1 para um mesmo conjunto de pontos iniciais. A Figura 4b apresenta os maiores tamanhos de passos quando η_2 pertence ao intervalo obtido em (3.10) e a Figura 4a apresenta menor número de iterações também para este mesmo intervalo de η_2 . Consequentemente, observa-se a inversão da concavidade dessas curvas. Com isso, podemos concluir que os tamanhos de passos gerados



(a) Número de iterações para convergência do algoritmo 1 com η_2 dado em (3.10).



(b) Tamanhos de passos obtidos pelo algoritmo 1 com η_2 dado em (3.10).

Figura 4: Desempenho numérico do Algoritmo 1.

pelo Algoritmo 1, quando usado os parâmetros η_2 obtidos em (3.10), asseguraram uma menor quantidade de iterações para resolver o problema.

5 CONSIDERAÇÕES FINAIS

Neste trabalho, determinamos matematicamente um intervalo de parâmetros η_2 dado no passo 4 do Algoritmo 1 que garante boa eficiência deste algoritmo para encontrar os pontos críticos da função acadêmica dada em (3.1). Este resultado foi obtido por meio de manipulações algébricas nas desigualdades da busca linear de Wolfe e da geometria do problema. Seria interessante investigar este resultado para uma função qualquer de classe C^1 .

Disponibilidade de dados

Não é aplicável.

Editor associado: Luiz Rafael Santos

ABSTRACT. Finding zeros of gradient a n -variables real function is a task which may be applied to solve problems of the speech recognition and machine translation. On Continuous Optimization, this task can be solved by using iterative methods which generates sequences that converge to a solution. Such methods must be computationally efficient. Line Search generally are used in iterative method in order to ensured a good computational efficiency. This Line Search finding step sizes for each iteration of the generated sequence for iterative method. Generally, these line search are defined by parameters that directly affect the numerical efficiency of the method used. Therefore, this is relevant for a mathematical investigation. The current literature has being not presented optimal parameters obtained mathematically. In this work, in order to find zeros of the gradient of an academic function, we mathematically determine a range of parameters for Wolfe's linear search. This range ensures more efficient step sizes when compared to other step sizes outside of the range. Our numerical study demonstrate the validity of the theoretical result obtained.

Keywords: optimal parameters, Wolfe line search, continuous optimization.

REFERÊNCIAS

- [1] S. Handibag & P. Sutkar. Optimization algorithms and their applications. *Malaya Journal of Matematik*, **9**(1) (2021), 1006–1014.
- [2] H. Hassani, A. Karbasi, A. Mokhtari & Z. Shen. Stochastic Conditional Gradient++: (Non)Convex Minimization and Continuous Submodular Maximization. *SIAM Journal on Optimization*, **30**(4) (2020), 3315–3344.
- [3] E.W. Karas & A.A. Ribeiro. Um curso de Otimização. *Notas de aula, UFPR, Curitiba, Brasil*, (2011).
- [4] D. Kim & J.A. Fessler. Generalizing the optimized gradient method for smooth convex minimization. *SIAM Journal on Optimization*, **28**(2) (2018), 1920–1950.
- [5] J. Nocedal & S.J. Wright. Numerical optimization (1999).
- [6] D.M. Panonceli. Um estudo de buscas unidirecionais aplicadas ao método BFGS (2015).
- [7] D. Shulman. Optimization Methods in Deep Learning: A Comprehensive Overview (2023). doi:10.48550/arXiv.2302.09566.

How to cite

N.M. Nascimento, T.A. Fernandes & A.D.V. Santiago. Busca de Wolfe: parâmetros eficientes obtidos matematicamente. *Trends in Computational and Applied Mathematics*, **27**(2026), e01891. doi:10.5540/tcam.2026.027.e01891.

