GLOBALIZAÇÃO DO MÉTODO DE NEWTON: ANÁLISE COMPARATIVA ENTRE ESTRATÉGIAS DE BUSCAS LINEARES EM UM MÉTODO HÍBRIDO

Maria Clara Brito dos Reis[1], **Marcio Antônio de Andrade Bortoloti**[2]

Universidade Estadual do Sudoeste da Bahia, BA, Brasil

`maria.cbritoreis@gmail.com`[1], `mbortoloti@uesb.edu.br`[2]

Resumo

Neste estudo, analisamos o desempenho de um método híbrido que usa as direções de descida de Newton e do gradiente para encontrar o minimizador de uma função. O desempenho deste método foi estudado numericamente empregando as buscas lineares de Armijo, Goldstein e Wolfe, em diferentes combinações com respeito às direções de descida. O experimentos numéricos evidenciaram que o emprego das buscas de Goldstein-Wolfe e Armijo-Armijo para as direções do gradiente e Newton, respectivamente, obtiveram melhor desempenho.

Palavras-Chave: Otimização Contínua; Globalização do método de Newton; Buscas lineares.

Introdução

O Método de Newton destaca-se por apresentar taxa de convergência superlinear, podendo chegar a quadrática, sob algumas hipóteses. Contudo, uma dificuldade deste método reside na exigência de um chute inicial próximo à solução. Para contornar essa dificuldade, pode-se empregar estratégias para o fornecimento de chutes iniciais mais adequados. Uma delas consiste em utilizar a direção de descida do gradiente da função objetivo, característica do Método do gradiente, que garante a convergência global da sequência gerada, mas com taxa linear [1]. Buscando reunir a característica global do Método do gradiente com as boas taxas de convergência do Método de Newton, consideramos um método híbrido que gera uma sequência com o emprego da direção de descida do gradiente. Este fornece um ponto apropriado para as direções de descida dadas pelo Método de Newton. Essas direções são obtidas pelas soluções d^k da equação linear

$$H(x^k)d^k = -\nabla f(x^k), \tag{1}$$

onde $H(x^k)$ é a matriz Hessiana de f no ponto x^k. Neste trabalho, consideraremos as buscas monotônicas de Armijo, Goldstein e Wolfe, equipadas no método híbrido. Estas serão testadas tanto na fase inicial conduzida pelo método do gradiente quanto na fase de aceleração promovida pelo método de Newton.

Fundamentação Teórica

Para cada direção de descida, empregamos uma busca linear que tem como objetivo garantir o decrescimento da função objetivo. As propriedades matemáticas das buscas empregadas neste trabalho podem ser encontradas em [1]. A primeira busca que apresentaremos é a Regra de Armijo, que consiste em encontrar um comprimento de passo $\alpha_k > 0$ que assegure uma redução suficiente em f, isto é, satisfazer, para $\sigma \in (0, 1)$ a condição

$$f(x^k + \alpha_k d^k) \leq f(x^k) + \sigma\alpha_k\langle\nabla f(x^k), d^k\rangle. \tag{2}$$

A segunda busca que apresentaremos é a Regra de Goldstein. Esta consiste na adição de uma segunda desigualdade à Regra de Armijo, com o intuito de descartar comprimentos de passo muito pequenos. Essa regra busca determinar, para $0 < \sigma_1 < \sigma_2 < 1$, um $\alpha_k > 0$ tal que

$$f(x^k) + \sigma_1\alpha_k\langle\nabla f(x^k), d^k\rangle \leq f(x^k + \alpha_k d^k) \leq f(x^k) + \sigma_2\alpha_k\langle\nabla f(x^k), d^k\rangle. \tag{3}$$

Por fim, apresentamos a Regra de Wolfe, que adiciona uma condição de curvatura à Regra de Armijo. Esta regra consiste em encontrar $\alpha_k > 0$, tal que

$$\begin{aligned} f(x^k + \alpha_k d^k) &\leq f(x^k) + \sigma_1\alpha_k\langle\nabla f(x^k), d^k\rangle \\ \langle\nabla f(x^k + \alpha_k d^k), d^k\rangle &\geq \sigma_2\langle\nabla f(x^k), d^k\rangle, \end{aligned} \tag{4}$$

para $0 < \sigma_1 < \sigma_2 < 1$.

Desenvolvimento e Metodologia

Para avaliar a eficiência das diferentes combinações das buscas lineares definidas por (2), (3) e (4), analisamos problemas de Otimização irrestritos de dimensão menor que, ou igual a 100, classificadas na categoria de soma de quadrados, fornecidos pela biblioteca de problemas teste CUTEst, [3]. Desses problemas, selecionamos os 75 que foram solucionados por pelo menos uma das estratégias de busca linear. O algoritmo a seguir detalha o método híbrido.

Algoritmo 1: MÉTODO HÍBRIDO

1 Tome um ponto inicial $x^0 \in \mathbb{R}^n$ e $\delta > 0$.
2 Defina $k = 0$.
3 Repita enquanto $\nabla f(x^k) \neq 0$.
4   Se $||\nabla f(x^k)|| > \delta$.
5   Tome $d^k = -\nabla f(x^k)$.
6   Senão
7   Calcule d^k por (1).
8 Calcule o comprimento de passo α_k usando (2), (3) ou (4).
9 Defina $x^k = x^k + \alpha_k d^k$.
10 Defina $k = k + 1$ e retorne para o passo 3.

O parâmetro δ define o momento em que o algoritmo deixa de usar a direção do gradiente e passa a utilizar a direção de Newton. Consideramos $\delta = 1$ e $\delta = 10^4$. Quando $||\nabla f(x^k)|| < 10^{-6}$, é declarado convergência do método. Quando a quantidade de iterações é maior que 1000, ou o comprimento de passo se torna menor que 10^{-12}, o algoritmo é interrompido. Os códigos estão disponíveis em: `https://github.com/petimatematica/HGNM`.

Resultados Numéricos

O estudo numérico realizado considerou o percentual de problemas resolvidos por cada estratégia de busca linear, representado por cada coluna nas figuras abaixo. Deste percentual, analisamos a quantidade de problemas resolvidos por cada estratégia, comparando quantos desses problemas foram solucionados com um número menor de iterações, em relação às demais (coluna verde). Denotamos cada estatégia usando a primeira letra de cada nome das buscas. A primeira letra diz respeito a busca equipada no método do gradiente enquanto a segunda refere-se à busca equipada no método de Newton. A seguir, as figuras detalham os resultados obtidos.

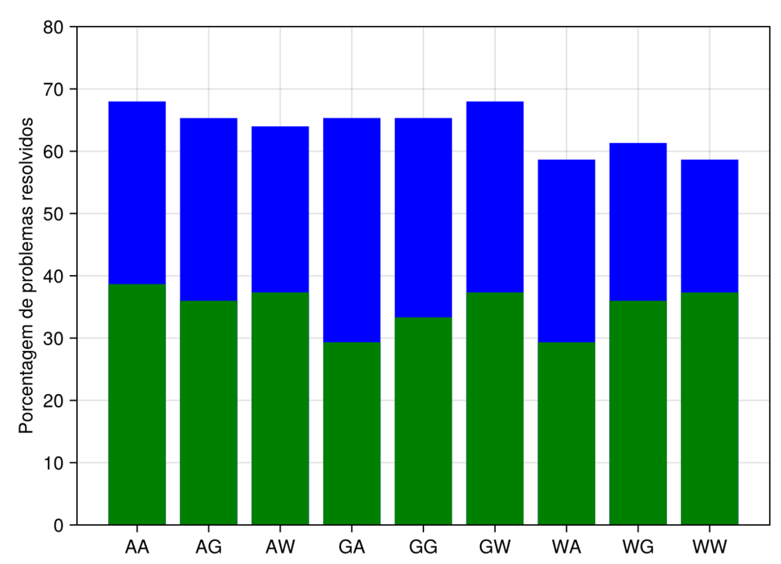

Figura 1: Buscas lineares empregadas no método híbrido ($\delta = 1$).

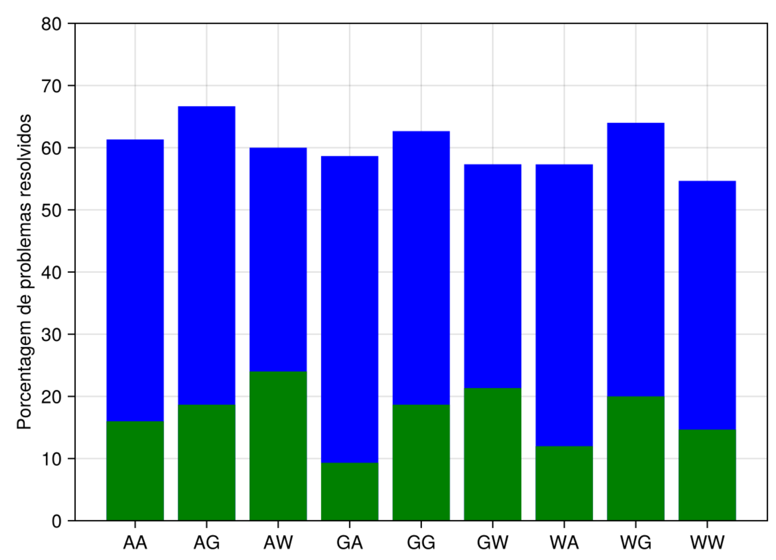

Figura 2: Buscas lineares empregadas no método híbrido ($\delta = 10^4$).

Observamos que houve problemas que foram resolvidos exclusivamente por apenas um dos parâmetros. Na figura 1, verificamos que as estratégias *GW* e *AA* se destacaram em relação às demais, considerando a porcentagem de problemas resolvidos e número de iterações. Na figura 2, observa-se que, para $\delta = 10^4$, a estratégia *AG* apresentou o melhor desempenho. Comparando as figuras 1 e 2, observamos que o parâmetro $\delta = 10^4$ resultou, de maneira geral, em um maior número de iterações. Isto pode ser atribuído ao emprego das direções de Newton em regiões distantes da solução, o que ocasiona na perda de eficiência, [2].

Referências

[1] IZMAILOV, Alexey; SOLODOV, Mikhail. *Otimização, volume 2: métodos computacionais*. 2018. Rio de Janeiro: IMPA.

[2] NOCEDAL, Jorge; WRIGHT, Stephen J. *Numerical optimization*. New York: Springer, 1999.

[3] GOULD, Nicholas IM; ORBAN, Dominique; TOINT, Philippe L. CUTEst: a constrained and unconstrained testing environment with safe threads for mathematical optimization. *Computational optimization and applications*, v. 60, p. 545-557, 2015.

Agradecimentos

A autora Maria Clara Brito dos Reis agradece ao Programa de Educação Tutorial Institucional da Universidade Estadual do Sudoeste da Bahia (PETI/UESB) pela bolsa de estudo.