Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, v. 12, n. 1, 2026.

Trabalho apresentado no XLIV CNMAC, Fundação Getulio Vargas - Rio de Janeiro - RJ, 2025.

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Uma Introdução ao Método do Gradiente em Otimização Contínua

Eduarda F. Zanatta[1] Roger Behling[2]
UFSC, Blumenau, SC

Nesta pesquisa de Iniciação Científica foi estudado o *método do gradiente* aplicado à minimização de funções continuamente diferenciáveis, com foco na sua convergência e eficácia em problemas de programação quadrática convexa. Além da análise teórica, foram realizadas implementações e experimentos que verificaram as propriedades de convergência do algoritmo. O desenvolvimento da pesquisa abrangeu em detalhes o processo iterativo do método, exemplos práticos em que o algoritmo foi aplicado, e a velocidade de convergência do método. O estudo evidenciou que, sob condições específicas, o método de gradiente é eficaz para encontrar mínimos de funções convexas, com uma taxa de convergência dependente de propriedades da função a ser otimizada. As fontes principais da literatura que forneceram uma base teórica sólida e clássica sobre o tema foram: Convex Optimization [1], Elementos de Programação Não-Linear [2], Numerical Optimization [3], Aspectos Teóricos e Computacionais [4] e Otimização: Condições de Otimalidade, Elementos de Analise Convexa e Dualidade [5].

Método do Gradiente com busca exata para minimizar uma função $f : \mathbb{R}^n \to \mathbb{R}$ continuamente diferenciável:

1. Dado $x_0 \in \mathbb{R}^n$
2. Para $k = 0, 1, 2, \ldots$ até convergência:
   (a) Calcule $d^k = -\nabla f(x^k)$;
   (b) Determine $t_k \in \mathbb{R}$ minimizador de $\varphi(t) := f(x^k + td^k)$;
   (c) Atualize $x^{k+1} = x^k + t_k d^k$.

A noção de convergência na prática leva em consideração algum critério de parada. Um critério bastante comum é pedir que $\|\nabla f(x^k)\| < \epsilon$, para alguma tolerância $\epsilon > 0$, o que vai ao encontro da condição necessária de otimalidade de primeira ordem $\nabla f(x) = 0$. Outro aspecto com relevância prática é considerar algum tipo de minimização unidimensional inexata no item (b), por exemplo, algo como a famosa busca de Armijo. Tal estratégia visa garantir uma redução suficiente da função objetivo a cada iteração, sem exigir a determinação exata do ponto minimizador ao longo da direção de descida. A ideia principal consiste em evitar passos longos, que podem afastar o iterado da região de interesse, bem como passos demasiadamente curtos, que resultam em progresso ineficiente. Assim, busca-se um equilíbrio entre avanço significativo na direção de descida e economia computacional.

Um caso simplificado, mas ao mesmo tempo relevante, de convergência do Método do Gradiente é quando a função objetivo f é quadrática e estritamente convexa, ou seja, quando f pode ser descrita como

[1]eduardazanatta6@gmail.com
[2]rogerbehling@gmail.com

$$f(x) = \frac{1}{2}x^T Ax - b^T x, \tag{1}$$

em que $A \in \mathbb{R}^{n\times n}$ é simétrica definida positiva e b é um vetor dado do $\mathbb{R}^n$. Estas hipóteses sobre (1) garantem que existe um único $x^* \in \mathbb{R}^n$ tal que $\nabla f(x^*) = 0$. Em outras palavras, $Ax^* - b = 0$, ou seja, x^* é solução do sistema linear $Ax = b$. Além disso, x^* é minimizador global único de f.

Quanto ao Método do Gradiente para minimizar f quadrática como (1), tem-se os seguintes resultados:

(i) O tamanho de passo $t_k \geq 0$ do item (b) tem fórmula fechada apresentada na equação (2).

$$t_k = \frac{(d^k)^T d^k}{(d^k)^T Ad^k}. \tag{2}$$

(ii) Dado qualquer ponto inicial $x^0 \in \mathbb{R}^n$, a sequência $\{x^k\}$ gerada pelo método do gradiente converge a x^*.

(iii) Mais que isso, a convergência é linear com taxa $\gamma \in [0, 1)$ dada por $\gamma = \sqrt{1 - \frac{\lambda_n}{\lambda_1}}$ sendo $\lambda_n > 0$ e $\lambda_1 > 0$ o menor e maior autovalor da matriz A, respectivamente. Tem-se inclusive que $\forall k$ vale (3).

$$\|x^{k+1} - x^*\| \leq \gamma\|x^k - x^*\|. \tag{3}$$

Os resultados (i), (ii) e (iii) são válidos para funções quadráticas estritamente convexas. Notamos, no entanto, que resultados similares podem ser obtidos sob hipóteses mais gerais. Estudos nesse sentido também foram conduzidos nessa Iniciação Científica.

Em aplicações práticas, é comum que a matriz A seja mal condicionada ou quase-singular, o que compromete tanto a estabilidade numérica quanto a velocidade de convergência do método. Nestes casos, algumas estratégias podem ser adotadas para contornar essas dificuldades. Uma possibilidade consiste na utilização de pré-condicionadores, cujo objetivo é transformar o problema original em um equivalente, mas com propriedades numéricas mais favoráveis. Outras estratégias incorporando algum tipo de regularização no método também poderiam ser consideradas.

Referências

[1] L. Boyd S. ; Vanderberghe. **Convex Optimization**. United States of America by Cambridge University Press, 2004. ISBN: 0521833787.

[2] A. Friedlander. **Elementos de Programação Não-Linear**. Campinas: UNICAMP, 1994. ISBN: 9788526803046.

[3] S. J. Noceda J.; Wright. **Numerical Optimization**. New York: Springer, 2006. ISBN: 9780387227429.

[4] E. W. Ribeiro A. A.; Karas. **Otimização Contínua: Aspectos Teóricos e Computacionais**. São Paul: LTC, 2013. ISBN: 9788522115013.

[5] A. Solodov M.; Izmailov. **Otimização: Condições de Otimalidade, Elementos de Analise Convexa e Dualidade**. Rio de Janeiro: Associação Instituto Nacional de Matemática Pura e Aplicada -IMPA, 2014. ISBN: 9788524402388.