

Convergência global de um método para otimização irrestrita sem derivadas

Priscila S. Ferreira* Elizabeth W. Karas Mael Sachine

Universidade Federal do Paraná

Departamento de Matemática

CP 19081, 81531-980, Curitiba, PR

E-mail: pri.savulski@hotmail.com, ewkaras@ufpr.br, mael@ufpr.com

Resumo: *Apresenta-se um método para minimização irrestrita de uma função $F : \mathbb{R}^n \rightarrow \mathbb{R}$ duas vezes diferenciável cujas derivadas estão indisponíveis. Considera-se para tal, um algoritmo iterativo de região de confiança sem derivadas. Durante as iterações a função objetivo é aproximada por modelos quadráticos através de interpolações polinomiais. De uma iteração para outra, o conjunto interpolador pode sofrer alteração em no máximo um elemento. Além disso, a cada iteração a função objetivo é avaliada uma única vez. O método proposto possui dois tipos de iterações, região de confiança e alternativa. As do tipo de região de confiança visam otimalidade. Já as alternativas visam melhorar a geometria dos pontos interpoladores. Prova-se que todo ponto de acumulação da sequência gerada por este método é um ponto estacionário da função F .*

Palavras-chave: *Método de região de confiança, teoria de convergência, otimização sem derivadas, minimização irrestrita.*

1 O método

Em muitas aplicações industriais e de engenharia necessita-se resolver problemas de otimização em que as derivadas da função objetivo não estão disponíveis. A ausência de derivadas computáveis naturalmente proíbem o uso de modelos de Taylor, tão comuns em otimização com derivadas. Além disso, acredita-se que o contexto sem derivadas seja mais difícil, uma vez que tenta-se obter um ponto de mínimo com menos informação. A otimização sem derivadas recebe muita atenção na obra de Conn, Scheinberg e Vicente [2].

O objetivo deste trabalho é apresentar um algoritmo baseado no método proposto por Powell em [3]. Neste, resolve-se um problema irrestrito

$$\min_{x \in \mathbb{R}^n} F(x), \quad (1)$$

onde as derivadas da função objetivo não estão disponíveis. Consequentemente, avaliações desta função devem ser evitadas.

A cada iteração a função objetivo é aproximada por um modelo através de interpolação polinomial com $n + 1$ pontos afim-independentes. A Hessiana do modelo é pré estabelecida como uma matriz simétrica cuja sua norma é limitada superiormente, a qual pode particularmente ser nula.

O conjunto interpolador afim-independente de uma iteração de índice $k \in \{1, 2, 3, \dots\}$ é denotado por

$$P^k = \{y_0, \dots, y_n\}, \quad (2)$$

*Bolsista do Programa de Pós-Graduação em Matemática Aplicada PPGMA/CAPES-REUNI

onde os pontos interpoladores $y_i \in \mathbb{R}^n$ são indexados de tal forma que

$$F(y_0) \leq F(y_i), \quad (3)$$

para $i = 1, \dots, n$. O primeiro conjunto interpolador é um dado de entrada. Durante uma iteração k o conjunto P^{k+1} difere do conjunto P^k em no máximo um elemento, permanecendo com $n + 1$ elementos afim-independentes.

A sequência gerada pelo método é inicializada com $x^1 = y_0$, para $y_0 \in P^1$. No decorrer da k -ésima iteração, determina-se um passo d^k a partir do ponto x^k . Se este novo passo satisfizer certas exigências, a sequência gerada é atualizada de modo que

$$x^{k+1} = \begin{cases} x^k + d^k, & \text{se } F(x^k + d^k) < F(x^k) \\ x^k, & \text{caso contrário,} \end{cases} \quad (4)$$

e o novo ponto $x^k + d^k$ substitui um dos pontos interpoladores do conjunto P^k com exceção do ponto y_0 , obtendo assim o novo conjunto P^{k+1} . Desta maneira, por (3) e (4) conclui-se que

$$x^k = y_0, \quad (5)$$

onde $y_0 \in P^k$.

A cada iteração k , o modelo quadrático $Q^k : \mathbb{R}^n \rightarrow \mathbb{R}$ que interpola F em torno de x^k é definido por

$$Q^k(x) = F(x^k) + (x - x^k)^T g^k + \frac{1}{2}(x - x^k)^T G^k (x - x^k), \quad (6)$$

onde $G^k \in \mathbb{R}^{n \times n}$ é uma matriz dada pelo método e $g^k \in \mathbb{R}^n$ é unicamente determinado pela condição de interpolação

$$Q^k(y_i) = F(y_i), \quad (7)$$

para $y_i \in P^k$ e $i = 1, \dots, n$. Para convergência do modelo exige-se que as matrizes G^k sejam uniformemente limitadas.

A cada iteração realizam-se tentativas de região de confiança e alternativas. Dentre as alternativas tem-se dois tipos: alternativa alpha e alternativa beta. Se durante uma iteração alguma dessas tentativas avalia a função objetivo em um ponto qualquer, então essa iteração recebe o nome desta tentativa e o contador de iterações é incrementado. Caso contrário, um outro tipo de tentativa é realizado sem incrementar a iteração.

1.1 Tentativa de região de confiança

As tentativas de região de confiança têm como objetivo minimizar o modelo Q^k dentro de uma determinada região na esperança de que pelo menos uma porcentagem dessa redução seja herdada pela função objetivo. Durante as iterações, o raio da região de confiança ρ pode diminuir ou permanecer o mesmo, seu índice é omitido para não sobrecarregar a notação. Dado um raio ρ , representa-se por ν o número da primeira iteração que considerou este raio.

Ao longo dessas tentativas resolve-se o problema

$$\begin{aligned} \min \quad & Q^k(x^k + d) \\ \text{s. a.} \quad & \|d\| \leq \rho. \end{aligned} \quad (8)$$

Defina d_c^k como um múltiplo do gradiente g^k tal que dentro da região de confiança minimize o modelo Q^k . O passo de região de confiança é qualquer vetor d^k que satisfaz

$$Q^k(x^k + d^k) \leq Q^k(x^k + d_c^k) \quad \text{e} \quad \|d^k\| \leq \rho, \quad (9)$$

exceto no caso em que $k \geq \nu + 5$ e

$$\eta^k = \begin{cases} 0, & \text{se } k = \nu \\ \max \{|Q^j(x^j + d^j) - F(x^j + d^j)| : j = \nu, \nu + 1, \dots, k - 1\}, & \text{if } k > \nu. \end{cases} \quad (10)$$

seja nulo. Nesse caso d^k deve ser uma solução exata do problema (8). A solução exata é exigida a fins teóricos para a prova de convergência.

Para evitar excesso na quantidade de avaliações da função objetivo, o novo valor da função $F(x^k + d^k)$ só é calculado quando são satisfeitas as condições

$$Q^k(x^k) - Q^k(x^k + d^k) > \gamma\eta^k \quad \text{e} \quad \|d^k\| \geq \frac{1}{2}\rho, \quad (11)$$

para um $\gamma > 0$ dado. Caso d^k não satisfaça (11), a tentativa é definida como insucesso e um outro tipo de tentativa é realizado sem incrementar k . Caso contrário, a função objetivo é avaliada em $x^k + d^k$ e compara-se a redução real e a predita

$$F(x^k) - F(x^k + d^k) \geq 0.1 \{Q^k(x^k) - Q^k(x^k + d^k)\}. \quad (12)$$

Caso d^k não satisfaça a condição acima esta tentativa também é insucesso, porém como $F(x^k + d^k)$ foi calculado, incrementa-se o valor de k . Dizemos que uma tentativa de região de confiança é sucesso quando as condições (11) e (12) são satisfeitas. No caso de sucesso algum ponto $y_t \in P^k$ não estará no conjunto P^{k+1} para que o novo ponto aceito $x^k + d^k$ entre em seu lugar. Para tal, são encontrados os valores θ_i , para $i = 0, \dots, n$, tal que

$$x^k + d^k = \sum_{i=0}^n \theta_i y_i \quad (13)$$

e

$$\sum_{i=0}^n \theta_i = 1. \quad (14)$$

Como $x^k = y_0$, para $y_0 \in P^k$, podemos reescrever estas condições como

$$d^k = \sum_{i=1}^n \theta_i (y_i - y_0). \quad (15)$$

Observe que os valores de θ_i não são todos nulos, para $i = 1, \dots, n$. Pois do contrário $d^k = 0$, o que é impossível para um passo d^k de região de confiança de sucesso. O índice $t \in \{1, \dots, n\}$ que indica o ponto $y_t \in P^k$ que não entrará no conjunto interpolador P^{k+1} é tal que

$$|\theta_t| \geq |\theta_i|, \quad (16)$$

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

1.2 Tentativa alternativa

Essas tentativas visam melhorar a disposição dos pontos interpoladores. Nestas, primeiro calcula-se o candidato a deixar o conjunto interpolador para depois obter o passo alternativo d^k . A escolha do passo é a mesma para as alternativas alphas e betas, a diferença está na escolha do ponto a deixar o conjunto interpolador.

- **Tentativa alpha** Nesta tentativa visa-se maximizar o volume da envoltória convexa formada pelo conjunto interpolador. Fixado $i \in \{1, \dots, n\}$, defina \mathcal{H}_i o hiperplano que contém os pontos interpoladores y_j para $j \in \{0, \dots, n\} \setminus \{i\}$. Considere σ_i a distância de y_i ao hiperplano \mathcal{H}_i , para $i \in \{1, \dots, n\}$. O ponto y_t candidato a deixar o conjunto interpolador é o ponto talque $\sigma_t \leq \sigma_i$, para $i = 1, \dots, n$. A fim de minimizar o número de avaliações da função objetivo, este ponto só deixará o conjunto interpolador quando esta distância for pequena o suficiente, ou seja, quando

$$\sigma_t < \alpha\rho, \quad (17)$$

para $\alpha \in (0, 1)$. Caso esta condição seja satisfeita, esta tentativa será sucesso.

- **Tentativa beta** O objetivo desta tentativa é impedir a dispersão dos pontos. Para isto as distâncias dos pontos interpoladores ao centro da região x^k da região são calculadas. Nesta tentativa faz-se uso de um conjunto auxiliar $\mathcal{B} \subset \{1, \dots, n\}$. O candidato a deixar o conjunto interpolador nesta tentativa é o ponto y_t , para $t \in \mathcal{B}$, talque

$$\|y_t - x^k\| \geq \|y_i - x^k\|, \quad (18)$$

para todo $i \in \mathcal{B}$. Para evitar avaliações desnecessárias de F , este ponto sairá do conjunto interpolador quando

$$\mathcal{B} \neq \emptyset \quad \text{e} \quad \|y_t - x^k\| > \beta\rho, \quad (19)$$

para um dado $\beta > 1$. Neste caso esta tentativa será de sucesso.

Deste modo, a tentativa alpha ou beta é de sucesso quando a condição (17) ou (19) vale, respectivamente. Neste caso, o passo alternativo d^k é tal que o ponto $x^k + d^k$ maximiza o volume da envoltória convexa formada por $P^k \setminus \{y_t\}$. Prova-se que d^k é o passo perpendicular ao hiperplano formado pelos pontos $P^k \setminus \{y_t\}$ e de tamanho ρ . Dentre os dois sentidos possíveis, escolhe-se o de menor valor do modelo.

Algumas condições para o método são exigidas, as reduções nos raios ocorrem sempre que, para um determinado raio, tenha-se uma tentativa de região de confiança de insucesso seguida de uma beta de insucesso antes da próxima tentativa de região de confiança. Após uma redução no raio, as primeiras tentativas devem ser do tipo alpha e região de confiança, respectivamente. Outra exigência é que o conjunto \mathcal{B} seja completo, $\mathcal{B} = 1, \dots, n$, sempre que uma tentativa de região de confiança seja sucesso, ou sempre que o raio reduza. Além disso, sempre que uma tentativa alpha ou beta seja sucesso e o ponto y_t saia do conjunto interpolador, o conjunto \mathcal{B} perde o elemento t , a menos que este não tenha sido removido anteriormente. Para forçar tentativas de otimalidade exige-se que entre duas tentativas de região de confiança com um determinado raio exista no máximo uma tentativa alpha e uma tentativa beta. Após cada tentativa de região de confiança de insucesso, deve-se realizar uma tentativa beta antes da próxima tentativa de região de confiança. As Hessianas dos modelos Q^k podem ser arbitrárias, exceto no caso em que $k \geq \nu + 5$ e $\eta^{k+1} = 0$, neste caso, toma-se $G^{k+1} = G^k$. Conseqüentemente, $Q^{k+1} = Q^k$. Permite-se que as Hessianas dos modelos sejam, em particular, nulas. Para garantir uma boa aproximação da função objetivo, exige-se que com um mesmo raio tenham no máximo $\tau_\alpha \geq 1$ ($\tau_\beta \geq 1$) tentativas de região de confiança seguidas desde a última tentativa alpha (beta).

2 Convergência

A partir do método proposto na seção anterior obtém-se um algoritmo bem definido. Além disso, propriedades importante sobre os conjuntos interpoladores são obtidas. Por fim, apresenta-se que todo ponto de acumulação da sequência gerada é um ponto estacionário da função objetivo.

Teorema *Se a sequência gerada (x^k) é infinita, as Hessianas G^k dos modelos Q^k são uniformemente limitadas e a função objetivo F é limitada inferiormente, duas vezes diferenciável e suas Hessianas são uniformemente limitadas. Então*

$$\lim_{k \rightarrow \infty} \nabla F(x^k) = \vec{0}.$$

Dem.: A demonstração pode ser obtida em [3].

3 Conclusão e trabalhos futuros

Apresentou-se neste trabalho um método de minimização irrestrita baseado em região de confiança sem derivadas em busca da convergência global. Deseja-se implementar o algoritmo obtido. Posteriormente, melhorar o algoritmo visando aplicabilidade na prática tomando o cuidado de manter a convergência global do método.

Referências

- [1] A. R. Conn, N. I. M. Gould e P. L. Toint. “Trust-Region Methods”. MPS-SIAM Series on Optimization, SIAM, Filadelfia, (2000).
- [2] A. R. Conn, K. Scheinberg e L. N. Vicente. “Introduction to Derivative-Free Optimization”. MPS-SIAM Series on Optimization, Filadelfia, (2009).
- [3] M. J. D. Powell, “On the Convergence of Trust Region Algorithms for Unconstrained Minimization without Derivatives”. Technical report, Dep. of Applied Mathematics and Theoretical Physics, Inglaterra, (2011). Presented at the Workshop on Nonlinear Optimization, Variational Inequalities and Equilibrium Problems, Erice, Itália, (2010).