

# Métodos de regiões de confiança

Marina Andretta

ICMC-USP

15 de setembro de 2014

Baseado no livro Numerical Optimization, de J. Nocedal e S. J. Wright.

Veremos agora métodos de **regiões de confiança** para resolução de problemas de minimização irrestrita, ou seja,

$$\text{Minimizar } f(x)$$

(1)

onde

- $x \in \mathbf{R}^n$ ;
- $f \in \mathbf{R}^n \rightarrow \mathbf{R}$  uma função suave.

# Métodos de regiões de confiança

Tanto **métodos de busca linear** como de **regiões de confiança** usam um modelo quadrático da função objetivo, mas de maneiras diferentes.

**Métodos de busca linear** usam o modelo quadrático no cálculo da direção de busca e depois se concentram em encontrar o tamanho de passo para andar nesta direção.

**Métodos de região de confiança** definem uma região dentro da qual acreditam que o modelo quadrático é uma boa aproximação da função objetivo e, então, calculam o minimizador do modelo nesta região. A direção e o tamanho de passo são calculados simultaneamente.

Se o passo não for aceito, o raio da região de confiança é reduzido e calcula-se uma nova direção. Assim, a direção muda sempre que o raio é modificado.

# Métodos de regiões de confiança

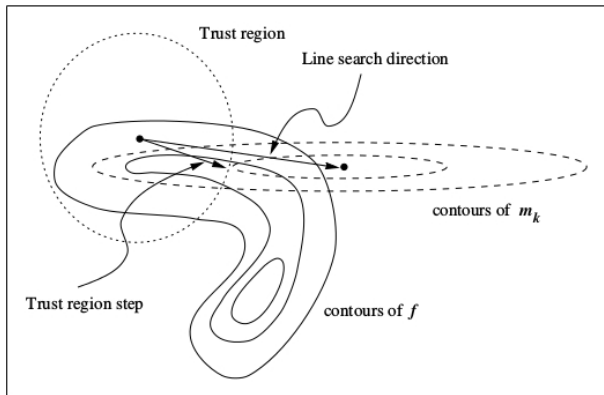


Figura : Regiões de confiança e busca linear (Figura 4.1 de Numerical Optimization, de J. Nocedal e S. J. Wright)

O raio da região de confiança é muito importante:

- Se ele for muito grande, o modelo quadrático pode estar muito longe da função objetivo e, então, o minimizador do modelo quadrático pode não produzir decréscimo suficiente, o raio deve ser diminuído e o minimizador do modelo recalculado.
- Por outro lado, se o raio da região de confiança for muito pequeno, podem ser calculados passos muito pequenos e, então, são necessárias muitas iterações de regiões de confiança para que a solução do problema (1) seja encontrada.

Na prática, escolhemos o raio da região de confiança de acordo com o resultado obtido em cada iteração:

- Se o modelo quadrático se mostra adequado para aproximar a função objetivo, o raio é mantido ou aumentado.
- Se o modelo quadrático se mostra uma aproximação ruim, o raio é reduzido.

# Métodos de regiões de confiança

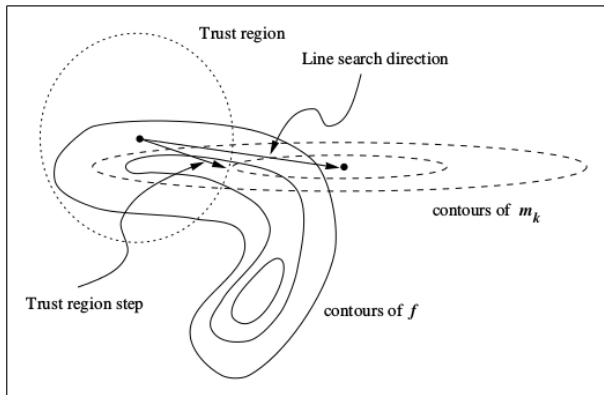


Figura : Solução do subproblema de regiões de confiança para diferentes raios (Figura 4.4 de Numerical Optimization, de J. Nocedal e S. J. Wright)

Supomos que os primeiros dois termos dos modelos quadráticos  $m_k$  em cada iterando  $x_k$  são idênticos aos primeiros dois termos da expansão de Taylor de  $f$  em torno de  $x_k$ . Ou seja,

$$\begin{aligned}m_k(p) &= f(x_k) + \nabla f(x_k)^T p + \frac{1}{2} p^T B_k p \\ &= f(x_k) + \nabla f(x_k)^T p + O(\|p\|^2),\end{aligned}$$

onde  $B_k$  é uma matriz simétrica.



Como

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

para um escalar  $t \in (0, 1)$ , a diferença entre  $m_k$  e  $f(x_k + p)$  é  $O(\|p\|^2)$ .

Assim, o erro da aproximação é muito pequeno quando  $p$  é pequeno.

# Métodos de regiões de confiança

Quando  $B_k$  é a Hessiana verdadeira  $\nabla^2 f(x_k)$ , o modelo quadrático é igual a expansão de Taylor em 3 termos, o que dá um erro de  $O(\|p\|^3)$ .

O método baseado em  $B_k = \nabla^2 f(x_k)$  é chamado de método de Newton de regiões de confiança.

Daqui em diante, não nos preocuparemos com a definição da matriz  $B_k$ .

Para a obtenção de cada passo, procuramos a solução do subproblema

$$\begin{array}{ll} \text{Minimizar} & m_k(p) = f(x_k) + \nabla f(x_k)^T p + \frac{1}{2} p^T B_k p \\ \text{Sujeita a} & \|p\| \leq \Delta_k, \end{array} \quad (2)$$

onde  $\Delta_k$  é o raio da região de confiança e  $\|\cdot\|$  é a norma euclidiana.

A solução  $p_k^*$  de (2) é o minimizador de  $m_k$  na bola de raio  $\Delta_k$ .

Assim, o método de regiões de confiança precisa que seja resolvida uma seqüência de subproblemas (2) nos quais tanto a função objetivo como a restrição são quadráticas.

Quando  $B_k$  é definida positiva e  $\|B_k^{-1}\nabla f(x_k)\| \leq \Delta_k$ , a solução de (2) é fácil de ser identificada. É, simplesmente, o minimizador irrestrito  $p_k^B = -B_k^{-1}\nabla f(x_k)$  do modelo quadrático  $m_k(p)$ .

Neste caso, chamamos  $p_k^B$  de passo completo.

Em outros casos, a solução de (2) não é óbvia, mas, geralmente, pode ser encontrada sem muito custo.

De qualquer forma, para garantir convergência e bons resultados práticos, é necessário apenas uma aproximação da solução de (2).

# Métodos de regiões de confiança

Para definir o método de regiões de confiança, é preciso escolher como atualizar o raio da região de confiança  $\Delta_k$  a cada iteração.

Baseamos essa atualização na concordância entre o decréscimo obtido na função objetivo e o decréscimo do modelo quadrático na iteração anterior.

Dado um passo  $p_k$ , definimos

$$\rho_k = \frac{f(x_k) - f(x_k + p_k)}{m_k(0) - m_k(p_k)}.$$

Chamamos o numerador de redução real (*actual reduction*) e o denominador de redução predita (*predicted reduction*).

# Métodos de regiões de confiança

Note que, como o passo  $p_k$  é calculado minimizando o modelo  $m_k$  numa região que inclui  $p = 0$ , a **redução predita sempre é não-negativa**.

Assim, se  $\rho_k$  é **negativo**, significa que o valor de  $f(x_k + p_k)$  é maior que  $f(x_k)$ , e o **passo deve ser rejeitado**.

Por outro lado, se  $\rho_k$  é **perto de 1**, o modelo quadrático  $m_k$  é uma boa aproximação para a função objetivo  $f$  neste ponto e, assim, é seguro **aumentar o raio da região de confiança** para a próxima iteração.

Se  $\rho_k$  é positivo, mas não está próximo de 1, não alteramos o raio de região de confiança.

# Métodos de regiões de confiança

**Método de regiões de confiança:** Dados  $\bar{\Delta} > 0$ ,  $\Delta_0 \in (0, \bar{\Delta})$ ,  $\eta \in [0, \frac{1}{4})$ , um ponto inicial  $x^{(0)}$  e um escalar  $\epsilon > 0$ .

**Passo 1:** Faça  $k \leftarrow 0$ .

**Passo 2:** Se  $\|\nabla f(x_k)\| \leq \epsilon$ , então pare e devolva  $x_k$  como solução.

**Passo 3:** Calcule  $p_k$  uma solução aproximada de (2).

**Passo 4:** Calcule  $\rho_k = \frac{f(x_k) - f(x_k + p_k)}{m_k(0) - m_k(p_k)}$ .

**Passo 5:** Se  $\rho_k < \eta$ , então faça  $\Delta_k = \frac{1}{4}\|p_k\|$  e volte para o Passo 3.

**Passo 6:** Se  $\rho_k < \frac{1}{4}$ , então faça  $\Delta_{k+1} = \frac{1}{4}\|p_k\|$ .

Senão

Se  $\rho_k > \frac{3}{4}$  e  $\|p_k\| = \Delta_k$ , então faça  $\Delta_{k+1} = \min(2\Delta_k, \bar{\Delta})$ .

Senão, faça  $\Delta_{k+1} = \Delta_k$

**Passo 7:** Faça  $x_{k+1} \leftarrow x_k + p_k$ .

**Passo 8:** Faça  $k \leftarrow k + 1$  e volte para o Passo 2.



# Métodos de regiões de confiança

Neste método,  $\bar{\Delta}$  é um limitante para o tamanho do passo.

Note que o raio da região de confiança só é aumentado quando o passo  $p_k$  atinge, de fato, a borda da região de confiança.

Quando a borda não é atingida, pode-se inferir que o tamanho da região de confiança não está atrapalhando o cálculo do passo e, por isso, o raio  $\Delta_k$  é mantido para a próxima iteração.

Para completar a definição do método é necessária uma maneira de executar o Passo 3. Veremos como isto pode ser feito de duas maneiras que garantem a convergência do método: usando **ponto de Cauchy** e o **método dogleg**.

Lembrando que, para executar o Passo 3 do método de regiões de confiança, precisamos calcular a solução para o subproblema

$$\begin{array}{ll} \text{Minimizar} & m_k(p) = f(x_k) + \nabla f(x_k)^T p + \frac{1}{2} p^T B_k p \\ \text{Sujeita a} & \|p\| \leq \Delta_k, \end{array} \quad (3)$$

onde  $\Delta_k$  é o raio da região de confiança e  $\|\cdot\|$  é a norma euclidiana.

Foi visto que para métodos de busca linear serem globalmente convergente não é necessário ser calculado tamanho de passo ótimo. O mesmo se aplica a métodos de regiões de confiança.

Apesar de, em princípio, querermos calcular a solução ótima de (3), é suficiente para convergência global do método que seja calculada uma aproximação de  $p_k$  que esteja dentro da região de confiança e que forneça decréscimo suficiente do modelo quadrático.

O decréscimo suficiente pode ser quantificado pelo **ponto de Cauchy**, denotado  $p_k^C$ , calculado da seguinte maneira:

**Cálculo do ponto de Cauchy:** Dado  $\Delta_k > 0$ .

**Passo 1:** Calcule  $p_k^S$  solução da versão linear de (3)

$$\begin{array}{ll} \text{Minimizar} & f(x_k) + \nabla f(x_k)^T p \\ \text{Sujeita a} & \|p\| \leq \Delta_k. \end{array}$$

**Passo 2:** Calcule o escalar  $\tau_k > 0$  solução de

$$\begin{array}{ll} \text{Minimizar} & m_k(\tau p_k^S) \\ \text{Sujeita a} & \|\tau p_k^S\| \leq \Delta_k. \end{array}$$

**Passo 3:** Faça  $p_k^C = \tau_k p_k^S$ .

Há uma fórmula fechada para o cálculo de  $p_k^C$ .

No Passo 1,  $p_k^S = -\frac{\Delta_k}{\|\nabla f(x_k)\|} \nabla f(x_k)$ .

Para calcular  $\tau_k$  no Passo 2 é necessário considerar 2 casos:

- Se  $\nabla f(x_k)^T B_k \nabla f(x_k) \leq 0$ , a função  $m_k(\tau p_k^S)$  decresce conforme  $\tau$  cresce, para  $\nabla f(x_k) \neq 0$ . Assim,  $\tau_k$  é o maior escalar que faz  $\tau_k p_k^S$  permanecer na região de confiança. Ou seja,  $\tau_k = 1$ .
- Se  $\nabla f(x_k)^T B_k \nabla f(x_k) > 0$ , a função  $m_k(\tau p_k^S)$  é uma quadrática convexa, então ou  $\tau_k$  é o minimizador irrestrito da quadrática  $\|\nabla f(x_k)\|^3 / (\Delta_k \nabla f(x_k)^T B_k \nabla f(x_k))$  ou  $p_k^C$  está na borda (ou seja,  $\tau_k = 1$ ), o que vier antes.

Resumindo,

$$p_k^C = -\tau_k \frac{\Delta_k}{\|\nabla f(x_k)\|} \nabla f(x_k),$$

com

$$\tau_k = \begin{cases} 1 & , \text{ se } \nabla f(x_k)^T B_k \nabla f(x_k) \leq 0 \\ \min \left( 1, \frac{\|\nabla f(x_k)\|^3}{(\Delta_k \nabla f(x_k)^T B_k \nabla f(x_k))} \right) & , \text{ caso contrário.} \end{cases}$$

# Ponto de Cauchy

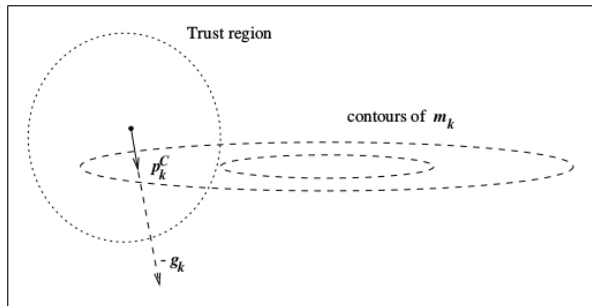


Figura : Ponto de Cauchy (Figura 4.2 de Numerical Optimization, de J. Nocedal e S. J. Wright)

# Ponto de Cauchy

O ponto de Cauchy é fácil de ser calculado e tem muita importância para decidir se a aproximação da solução de (3) é aceitável.

Na verdade, um método de regiões de confiança é globalmente convergente se os passos  $p_k$  obtêm, em cada iteração, decréscimo do modelo quadrático  $m_k$  de, pelo menos, um múltiplo do decréscimo obtido pelo ponto de Cauchy.

Apesar de fácil de calcular e de produzir decréscimo do modelo quadrático, **usar sempre o ponto de Cauchy como passo produz um método de máxima descida com um tamanho de passo particular.** Como visto, mesmo quando os tamanhos de passo usados são ótimos, métodos de máxima descida têm convergência lenta.



O ponto de Cauchy não depende fortemente de  $B_k$ , que é usada apenas no cálculo do tamanho do passo. Para obter convergência rápida (superlinear, por exemplo), é necessário usar  $B_k$  também no cálculo do passo  $p_k$ .

Há vários métodos que começam a aproximação da solução de (3) com o ponto de Cauchy e tentam melhorar esta aproximação.

Em geral, tenta-se usar o passo completo  $p_k^B = -B_k^{-1}\nabla f(x_k)$  sempre que  $B_k$  é definida positiva e  $\|p_k^B\| \leq \Delta_k$ . Em particular, quando  $B_k$  é a Hessiana verdadeira ou uma aproximação quase-Newton, espera-se um método deste tipo tenha convergência superlinear.

Lembrando que estamos interessados em resolver o seguinte subproblema:

$$\begin{array}{ll} \text{Minimizar} & m_k(p) = f(x_k) + \nabla f(x_k)^T p + \frac{1}{2} p^T B_k p \\ \text{Sujeita a} & \|p\| \leq \Delta_k, \end{array} \quad (4)$$

onde  $\Delta_k$  é o raio da região de confiança e  $\|\cdot\|$  é a norma euclidiana.

No caso do **método dogleg**, estamos interessados apenas no caso em que  $B_k$  é definida positiva.

# Método dogleg

A solução de (4) será denotada por  $p^*(\Delta_k)$ . Vejamos agora como o valor de  $\Delta_k$  afeta a solução de (4).

Quando  $B_k$  é definida positiva, o minimizador irrestrito de  $m_k$  é o passo completo  $p_k^B = -B_k^{-1}\nabla f(x_k)$ . Quando este ponto está na região de confiança, a solução é dada por

$$p^*(\Delta_k) = p_k^B, \text{ quando } \|p_k^B\| \leq \Delta_k.$$

# Método dogleg

Quando  $\Delta_k$  é pequeno, a restrição  $\|p\| \leq \Delta_k$  garante que a parte quadrática do modelo  $m_k$  não tem muita influência em seu valor. Neste caso, a solução verdadeira  $p^*(\Delta_k)$  pode ser aproximada pela solução de

$$\begin{array}{ll} \text{Minimizar} & f(x_k) + \nabla f(x_k)^T p \\ \text{Sujeita a} & \|p\| \leq \Delta_k, \end{array}$$

Ou seja,

$$p^*(\Delta_k) \approx -\Delta_k \frac{\nabla f(x_k)}{\|\nabla f(x_k)\|}, \text{ quando } \Delta_k \text{ é pequeno.}$$

# Método dogleg

O método dogleg calcula uma solução aproximada de (4), trocando a curva correta por uma aproximação composta por dois segmentos de reta.

O primeiro segmento vai da origem até o minimizador irrestrito na direção de máxima descida, definida por

$$p_k^U = -\frac{\nabla f(x_k)^T \nabla f(x_k)}{\nabla f(x_k)^T B_k \nabla f(x_k)} \nabla f(x_k).$$

O segundo segmento vai de  $p_k^U$  ao passo completo  $p_k^B$ .

Definimos então a **trajetória de dogleg**  $\tilde{p}_k(\tau)$ , para  $\tau \in [0, 2]$ , como

$$\tilde{p}_k(\tau) = \begin{cases} \tau p_k^U & , 0 \leq \tau \leq 1 \\ p_k^U + (\tau - 1)(p_k^B - p_k^U) & , 1 \leq \tau \leq 2 \end{cases}$$

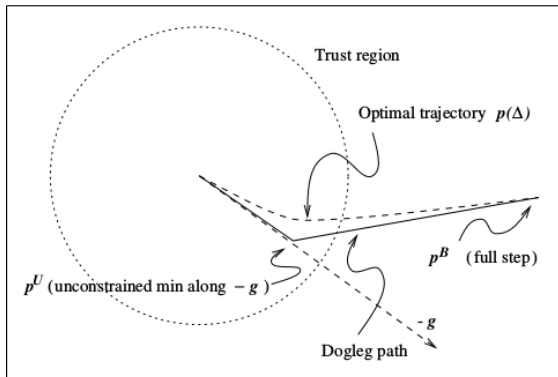


Figura : Trajetória da solução  $p^*(\Delta_k)$  para valores intermediários de  $\Delta_k$  (Figura 4.3 de Numerical Optimization, de J. Nocedal e S. J. Wright)

O método dogleg escolhe  $p_k$  que minimiza o modelo  $m_k$  nesta trajetória, sujeito a  $p_k$  pertencer à região de confiança.

Não é necessário fazer uma busca para calcular  $p_k$ , pois a trajetória  $\tilde{p}_k$  intersecta a borda da região de confiança no máximo uma vez e este ponto de intersecção pode ser **calculado analiticamente**.

**Lema:** *Seja  $B_k$  definida positiva. Então*

- 1  $\|\tilde{p}_k(\tau)\|$  é uma função crescente de  $\tau$ , e
- 2  $m_k(\tilde{p}_k(\tau))$  é uma função decrescente de  $\tau$ .



Pelo lema anterior podemos ver que  $\tilde{p}_k(\tau)$  intersecta a borda da região de confiança  $\|p_k\| = \Delta_k$  em exatamente um ponto se  $\|p_k^B\| \geq \Delta_k$ .

Como  $m_k$  decresce ao longo da trajetória,  $p_k$  será igual a  $p_k^B$  quando  $\|p_k^B\| \leq \Delta_k$  e o ponto de intersecção com a borda, caso contrário.

Para calcular o ponto de intersecção com a borda, precisamos analisar dois casos.

Se  $\|p_k^U\| \leq \Delta_k$ , a intersecção é dada por  $\frac{\Delta_k}{\|p_k^U\|} p_k^U$ .

Caso contrário, basta calcular  $\tau$  solução de

$$\|p_k^U + (\tau - 1)(p_k^B - p_k^U)\|^2 = \Delta_k^2.$$

# Método dogleg

Portanto, o passo calculado pelo método dogleg é dado por

$$p_k = \begin{cases} p_k^B, & \text{se } \|p_k^B\| \leq \Delta_k, \\ \frac{\Delta_k}{\|p_k^U\|} p_k^U, & \text{se } \|p_k^B\| > \Delta_k \text{ e } \|p_k^U\| \leq \Delta_k, \\ p_k^U + (\tau - 1)(p_k^B - p_k^U), & \text{caso contrário,} \end{cases}$$

com

- $p_k^B = -B_k^{-1} \nabla f(x_k)$ ,
- $p_k^U = -\frac{\nabla f(x_k)^T \nabla f(x_k)}{\nabla f(x_k)^T B_k \nabla f(x_k)} \nabla f(x_k)$ ,
- $\tau$  solução de  $\|p_k^U + (\tau - 1)(p_k^B - p_k^U)\|^2 = \Delta_k^2$ .

# Convergência de métodos de regiões de confiança

Como já mencionado anteriormente, para obter convergência global do método de regiões de confiança, é necessário que a redução do modelo quadrático obtida pela solução aproximada de (2) seja pelo menos uma fração da redução obtida pelo ponto de Cauchy.

Agora estimaremos o decréscimo do modelo quadrático obtido pelo ponto de Cauchy e veremos que, se uma solução aproximada de (2) obtém uma fração do decréscimo do modelo quadrático obtido pelo ponto de Cauchy, então a seqüência de gradientes  $\{\nabla f_k\}$  gerada pelo método de regiões de confiança tem um ponto de acumulação em 0 ou converge para 0, dependendo da escolha de  $\eta = 0$  ou  $\eta$  estritamente positivo.

Primeiramente, vejamos que tanto o ponto de Cauchy como o passo calculado pelo método dogleg satisfazem a equação

$$m_k(0) - m_k(p_k) \geq c_1 \|\nabla f_k\| \min \left( \Delta_k, \frac{\|\nabla f_k\|}{\|B_k\|} \right), \quad (5)$$

para  $c_1 \in (0, 1]$ .

**Teorema 1:** *O ponto de Cauchy  $p_k^C$  satisfaz*

$$m_k(0) - m_k(p_k^C) \geq \frac{1}{2} \|\nabla f_k\| \min \left( \Delta_k, \frac{\|\nabla f_k\|}{\|B_k\|} \right).$$

# Convergência de métodos de regiões de confiança

Para **satisfazer a equação (5)** é necessário apenas que a solução aproximada  $p_k$  obtenha pelo menos uma fração  $c_2$  da redução obtida pelo ponto de Cauchy.

**Teorema 2:** *Seja  $p_k$  um vetor tal que  $\|p_k\| \leq \Delta_k$  e  $m_k(0) - m_k(p_k) \geq c_2(m_k(0) - m_k(p_k^C))$ . Então  $p_k$  satisfaz (5) com  $c_1 = c_2/2$ . Em particular, se  $p_k$  é a solução exata  $p_k^*$  de (2), então  $p_k$  satisfaz (5) com  $c_1 = \frac{1}{2}$ .*

Note que o **método dogleg** satisfaz (5) com  $c_1 = 1/2$ , já que o passo  $p_k$  é tal que  $m_k(p_k) \leq m_k(p_k^C)$ .

# Convergência de métodos de regiões de confiança

Resultados de convergência global de métodos de regiões de confiança apresentam dois resultados: um para o caso de  $\eta = 0$  e outro para o caso em que  $\eta$  é um escalar pequeno estritamente positivo.

Quando  $\eta = 0$  pede-se que o passo produza apenas decréscimo simples da função objetivo para que seja aceito. Neste caso, vamos mostrar que a seqüência de gradientes  $\{\nabla f_k\}$  tem um ponto limite em 0.

Quando  $\eta$  é estritamente positivo, o passo só é aceito se obtém um decréscimo da função objetivo de pelo menos uma fração do decréscimo obtido no modelo quadrático. Neste caso, veremos que a seqüência de gradientes converge para 0.



**Teorema 3:** Tome  $\eta = 0$  no método de regiões de confiança. Suponha que  $\|B_k\| \leq \beta$  para alguma constante  $\beta$ ,  $f$  possui primeira derivada contínua e limitada inferiormente no conjunto  $\{x | f(x) \leq f(x_0)\}$  e todas as soluções aproximadas de (2) satisfazem a desigualdade (5) e  $\|p_k\| \leq \gamma \Delta_k$ ,  $\gamma \geq 1$ , para constantes positivas  $c_1$  e  $\gamma$ . Vale que

$$\liminf_{k \rightarrow \infty} \|\nabla f_k\| = 0.$$

**Teorema 4:** Tome  $\eta \in (0, 1/4)$  no método de regiões de confiança. Suponha que  $\|B_k\| \leq \beta$  para alguma constante  $\beta$ ,  $f$  possui primeira derivada Lipschitz contínua e limitada inferiormente no conjunto  $\{x | f(x) \leq f(x_0)\}$  e todas as soluções aproximadas de (2) satisfazem a desigualdade (5) e  $\|p_k\| \leq \gamma \Delta_k$ ,  $\gamma \geq 1$ , para constantes positivas  $c_1$  e  $\gamma$ . Vale que

$$\lim_{k \rightarrow \infty} \nabla f_k = 0.$$