Maria Isabel da Rocha Ferreira Caiado

CARACTERIZAÇÃO E ESTUDO DE ALGUMAS PROPRIEDADES DO MÉTODO DE NEWTON MODIFICADO

Universidade do Minho
Janeiro de 1999

Maria Isabel da Rocha Ferreira Caiado

CARACTERIZAÇÃO E ESTUDO DE ALGUMAS PROPRIEDADES DO MÉTODO DE NEWTON MODIFICADO

Dissertação do mestrado em Matemática Computacional apresentada para a obtenção do grau de Mestre.

Universidade do Minho
Janeiro de 1999

Ao meu Avô

Ao longo deste trabalho pude contar com o apoio de diversas pessoas.
Em primeiro lugar quero agradecer à Prof. Doutora Edite Fernandes pela sua disponibilidade, paciência demonstrada e motivação transmitida ao longo deste trabalho. Gostaria também de agradecer ao Eng. Paulo Ferreira pelos seus comentários e sugestões e a todos os meus colegas que contribuíram, de alguma forma, para que este trabalho se concretizasse.

Durante as provas de apresentação e defesa desta dissertação foram feitos alguns comentários e sugestões pelo Prof. Luís Nunes Vicente, dos quais resultaram as correcções agora introduzidas.

Resumo

O problema que consiste em minimizar uma função não linear sem restrições nas variáveis é há muito estudado. Existem vários métodos para o resolver, sendo o de Newton o método clássico e o mais conhecido. Devido às falhas deste método várias modificações têm sido sugeridas.

Neste trabalho estuda-se uma modificação específica ao método de Newton que se designa por método de Newton modificado.

Após a sua descrição prova-se a convergência global do método pelo menos r-linear. Em seguida prova-se que a matriz associada ao método é, sob certas condições, definida positiva.

No que diz respeito à direcção gerada pelo método, prova-se que é do tipo gradiente no sentido de Ortega e Rheinboldt (1970), e que satisfaz uma fracção do decréscimo óptimo, condição esta associada às técnicas de regiões de confiança. Por fim prova-se que a direcção modificada não é ortogonal à direcção de descida máxima.

Abstract

The nonlinear unconstrained minimization problem has been widely studied. There is a variety of methods to solve it. The best known classical method is the Newton's method. Due to its failures several modifications have been proposed. In this work a special modification, called modified Newton method, is considered.

We prove that the global convergence to a stationary point is at least r-linear and that the associated matrix is, under suitable conditions, positive definite.

As far as the modified search direction is concerned, it's proved that it is gradient-related according to Ortega and Rheinboldt (1970) and satisfies a fraction of the optimal decrease. This condition is usually associated with the trust region techniques. Finally we prove that the new direction is bounded away from orthogonality to the negative gradient.

Conteúdo

Capítulo 1

Introdução

Este trabalho foi realizado, sob orientação da Professora Doutora Edite Fernandes, no âmbito da dissertação do mestrado em Matemática Computacional, edição de 1996, da Universidade do Minho.

A Optimização Não Linear sem restrições nas variáveis, domínio no qual este trabalho se insere, pertence ao contexto mais amplo da Optimização Não Linear. A complexidade dos problemas desta área surge do facto de quer as funções quer as restrições envolvidas na formulação matemática dos problemas serem não lineares nas variáveis. Dada a dificuldade de resolver analiticamente este tipo de problemas é vulgar recorrer-se a métodos iterativos para obter a solução. Embora enquadrando-se nesta área, o trabalho aqui apresentado é de índole teórica, não sendo, por isso, incluídos quaisquer resultados de implementações computacionais.

A importância da programação sem restrições, designadamente dos métodos do tipo Newton, é mais abrangente do que a sua aplicação directa aos problemas sem restrições possa sugerir. Hoje em dia, cada vez mais, se recorre aos métodos de penalização e de pontos interiores para a resolução de problemas com restrições (de igualdade, desigualdade e com limites simples). Como os problemas daqui resultantes são, em geral, mais complexos do que os problemas sem restrições de base, o desenvolvimento de métodos eficientes e robustos torna-se cada vez mais necessário e premente.

O objectivo primordial deste trabalho é caracterizar, quanto à convergência global, uma modificação do método de Newton para o cálculo do mínimo de uma função. Pretende-se, ainda, estudar outras propriedades do método modificado.

Para uma melhor exposição, esta dissertação está estruturada da seguinte forma. No Capítulo dois é formulado o problema de optimização multidimensional sem restrições nas variáveis e são apresentadas as condições suficientes para que um ponto seja solução do problema. Em seguida apresenta-se o método de Newton para a sua resolução e a modificação a este método que nos propomos estudar neste trabalho. No Capítulo seguinte são enunciados alguns resultados relativos à convergência local quer do método clássico quer do método modificado, sendo ainda apresentadas duas estratégias de globalização do método de Newton.

A caracterização do método modificado acontece no terceiro Capítulo. Neste capítulo é estabelecido e provado o teorema da convergência global do método modificado. São ainda estabelecidas e provadas algumas propriedades do método e da direcção, nomeadamente que, sob certas condições, a matriz associada ao método é definida positiva, a direcção é do tipo gradiente, satisfaz a fracção do decréscimo óptimo, e não se torna ortogonal ao vector gradiente. O último Capítulo apresenta as conclusões e propostas de trabalho futuro.

Capítulo 2

O Método de Newton

Vamos iniciar este capítulo com a formulação de um problema de minimização sem restrições e com a apresentação das condições suficientes para que o problema tenha solução. Em seguida descrevemos o método de Newton básico. Este capítulo termina com a apresentação do método de Newton modificado.

Como referência, para este capítulo, tomamos os trabalhos de Dennis e Schnabel [1] e de Fernandes [3].

2.1 Caracterização do Problema

O problema da determinação do minimizante de uma função real de n variáveis reais, não linear e sem restrições nas variáveis pode ser formulado matematicamente como

$$\min_{x \in \mathbb{R}^n} f(x)$$

onde $f : \mathbb{R}^n \longrightarrow \mathbb{R}$.

Neste trabalho, denotamos o vector gradiente das primeiras derivadas da função f por ∇f, a matriz hessiana das segundas derivadas por H e o minimizante da função por x^*.

A noção de minimizante local recorre ao conceito de vizinhança. Dada uma norma vectorial $\|.\|$ em $\mathbb{R}^n$, definimos $V_\delta(x)$ como a vizinhança aberta de centro em x e raio $\delta > 0$, isto é

$$V_\delta(x) = \{\bar{x} \in \mathbb{R}^n \,|\, \|x - \bar{x}\| < \delta\}.$$

Um ponto $x^* \in \mathbb{R}^n$ é um minimizante local forte da função f se existir uma vizinhança $V_\delta(x^*)$ tal que, para todo $x \in V_\delta(x^*)$ com $x \neq x^*$

$$f(x^*) < f(x).$$

Se a desigualdade for em sentido lato e x^* não for um minimizante local forte dizemos que x^* é um minimizante local fraco. No caso de $f(x^*) < f(x)$, para todo x pertencente a algum subconjunto D de $\mathbb{R}^n$ que contenha x^*, então x^* é um minimizante global de f em D.

Definição 1 *Seja $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ duas vezes continuamente diferenciável no aberto D e $x^* \in D$.*

a) x^* é um ponto de estacionaridade da função f se $\|\nabla f(x^*)\| = 0$,

b) um ponto de estacionaridade x^* da função f é um minimizante se a hessiana de f, $H(x^*)$, for uma matriz definida positiva, isto é, se $s^T H(x^*) s > 0$, para todo o vector não nulo $s \in \mathbb{R}^n$.

2.2 O Método de Newton

Consideremos a aproximação quadrática da função f em torno de um ponto $\bar{x}$, numa vizinhança de raio d:

$$m(\bar{x} + d) = f(\bar{x}) + \nabla f(\bar{x})^T d + \frac{1}{2} d^T H(\bar{x}) d,$$

onde $d = x - \bar{x}$.

Se $H(x)$ for definida positiva, m tem um único minimizante global, $\bar{x}^*$, que é a solução do sistema

$$\nabla f(\bar{x})^T + d^T H(\bar{x}) = 0.$$

O minimizante do modelo quadrático é, então, tomado como uma aproximação a x^*:

$$x^* \simeq \bar{x}^* = \bar{x} - H(\bar{x})^{-1} \nabla f(\bar{x}).$$

Assim, se H for não singular, é possível definir um processo iterativo que, em cada iteração, determina uma direcção

$$d_k = -H(x_k)^{-1} \nabla f(x_k), \quad k = 0, 1, 2, ... \tag{2.1}$$

e uma aproximação ao minimizante da função, dada por

$$x_{k+1} = x_k + d_k.$$

À direcção definida pela equação (2.1) chamamos direcção Newton, que denotamos por d_k^N. O processo iterativo acima descrito é o *método de Newton básico*, cujo algoritmo apresentamos em seguida.

Algoritmo 1 (Método de Newton Básico) Dada $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ de classe C^2 em $V_\delta(x^*)$ e $x_0 \in V_\delta(x^*)$

1. fazer $k = 0$
2. calcular $f(x_k)$
3. calcular o vector gradiente $\nabla f(x_k)$
4. calcular a matriz hessiana $H(x_k)$
5. resolver o sistema
$$H(x_k)d_k^N = -\nabla f(x_k)$$
6. definir
$$x_{k+1} = x_k + d_k^N.$$
7. fazer $k = k + 1$ e ir para 2

Sempre que a matriz hessiana for definida positiva, a direcção Newton é de descida. Os métodos cujas direcções de procura satisfazem, em cada iteração, esta propriedade dizem-se métodos de descida.

Definição 2 *Seja $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ uma função cujas primeiras derivadas são contínuas no ponto x e tomemos $d \in \mathbb{R}^n$. Dizemos que a direcção d aponta no sentido descendente da função f (direcção de descida) em x se e só se $\nabla f(x)^T d < 0$.*

A direcção $d \in \mathbb{R}^n$ ser de descida significa que existe um escalar $\alpha > 0$ tal que

$$f(x + \alpha d) < f(x).$$

Se $H(x_k)$ for definida positiva, a direcção Newton calculada por (2.1) é realmente de descida relativamente a f, pois

$$\begin{aligned} \nabla f(x_k)^T d_k &= \nabla f(x_k)^T \left[-H(x_k)^{-1}\nabla f(x_k)\right] \\ &= -\underbrace{\nabla f(x_k)^T \left[H(x_k)^{-1}\nabla f(x_k)\right]}_{>0} < 0. \end{aligned}$$

2.3 O Método de Newton Modificado

O método de Newton básico pode falhar devido a dificuldades originadas quer pela matriz hessiana quer pela própria direcção d_k^N calculada por (2.1).

Quando a matriz $H(x)$ é singular a direcção Newton não está sequer definida. Existindo a inversa da hessiana, esta pode não ser definida positiva pelo que a direcção d_k^N pode não ser de descida. Se a matriz $H(x)^{-1}$ existe e é definida positiva mas f não é bem aproximada pelo modelo quadrático pode acontecer que $f(x_{k+1}) > f(x_k)$. Isto é, o novo iterando não é melhor que o anterior pois o valor da função não foi reduzido. Pode, ainda, ocorrer que d_k^N seja quase ortogonal ao vector $-\nabla f(x_k)$ não sendo, então, possível progredir ao longo desta direcção.

Algumas modificações a este método que têm sido introduzidas para resolver estes problemas podem ser encontradas, por exemplo, em [1], [5], [10] e [15]. Neste trabalho analisamos uma modificação específica do método de Newton que designamos por *método de Newton modificado* e que de seguida descrevemos.

A solução dos problemas anteriores, apresentada pelo método de Newton modificado, consiste em definir uma nova direcção de procura como uma combinação linear da direcção Newton e da direcção de descida máxima[1], $-\nabla f(x)$. A direcção modificada é,

$$d = (1 - w(x))d^N + w(x)\left[-\nabla f(x)\right], \tag{2.2}$$

em que d^N é calculada por (2.1).

A nova direcção é de descida em x se d^N não o for e $w(x)$ satisfizer a seguinte condição

$$w(x) > \frac{\nabla f(x)^T d^N}{\nabla f(x)^T \nabla f(x) + \nabla f(x)^T d^N} = a(x).$$

No caso em que $H(x)$ é invertível e definida positiva, tomamos $w(x) = 0$, ou seja $d = d^N$. Quando a matriz hessiana for singular tomamos $w(x) = 1$ o que equivale a considerar $d = -\nabla f(x)$, isto é, d é a direcção de descida máxima.

Nos restantes casos calculamos

$$w(x) = \min\left(1, \frac{\nabla f(x)^T d^N}{\|\nabla f(x)\|^2 + \nabla f(x)^T d^N} + \varepsilon(x)\right) \tag{2.3}$$

onde $\varepsilon(x)$ é uma função que tende para zero à medida que x tende para x^* e $0 \leq a(x) + \varepsilon(x) \leq 1$. Atendendo à definição, temos $0 \leq w(x) \leq 1$.

[1] Tradução do termo "steepest descent". Outros investigadores têm sugerido como tradução "descida abrupta".

Passamos a descrever o algoritmo do método de Newton modificado. Denotemos $w\left(x_k\right)$ por w_k.

Algoritmo 2 (Método de Newton Modificado) Dada $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ de classe C^2 em $V_\delta(x^*)$ e $x_0 \in V_\delta(x^*)$

1. fazer $k = 0$
2. calcular $f(x_k)$
3. calcular o vector gradiente $\nabla f(x_k)$ e a matriz hessiana $H(x_k)$
4. determinar a direcção
$$H(x_k)d_k^N = -\nabla f(x_k)$$
e se a hessiana for singular, fazer $w_k = 1$ e ir para 7
5. se $\nabla f(x_k)^T d_k^N < -\varepsilon \left\|\nabla f(x_k)\right\| \left\|d_k^N\right\|$, fazer $w_k = 0$ e ir para 7
6. calcular w_k usando (2.3)
7. calcular
$$d_k = (1 - w_k)d_k^N + w_k\left[-\nabla f(x_k)\right]$$
8. definir
$$x_{k+1} = x_k + d_k.$$
9. fazer $k = k + 1$ e ir para 2

Capítulo 3

Análise da Convergência

Neste capítulo apresentamos as condições de convergência local do método de Newton. Para isso introduzimos a definição de convergência e alguns conceitos a ela associados. De seguida descrevemos dois processos distintos para a globalização do método: o primeiro, com a introdução de procura unidimensional, o segundo com introdução de regiões de confiança. Terminamos com a descrição das condições de convergência local do método modificado.

3.1 Conceitos

As definições e lemas que apresentamos nesta secção têm como referência o trabalho de Ortega e Rheinboldt [10].

Comecemos por introduzir as noções de convergência $q-$linear e $r-$linear.

Definição 3 *Seja $\{x_k\} \subset I\!R^n$ uma sequência convergente para $x^* \in I\!R^n$. Dizemos que a convergência é $q-$linear se, para $k_0 > 0$,*

$$\|x_{k+1} - x^*\| \leq \gamma_q \|x_k - x^*\|, \quad \forall k \geq k_0$$

com $\gamma_q \in (0,1)$, para alguma norma $\|.\|$ de $I\!R^n$.

Caso tenhamos $\|x_{k+1} - x^*\| \leq c \|x_k - x^*\|^2$, para todo $k \geq k_0$, com $c > 0$, dizemos que a sequência $\{x_k\}$ converge $q-$quadraticamente para x^*.

Definição 4 *Seja $\{x_k\} \subset I\!R^n$ uma sequência que converge para $x^* \in I\!R^n$. Dizemos que a convergência é r-linear se existe $k_0 > 0$, tal que*

$$\|x_k - x^*\| \leq \gamma_r^k, \quad \forall k \geq k_0$$

com $\gamma_r \in (0,1)$.

Contrariamente ao que acontece com a $q-$convergência, a $r-$convergência é independente da norma.

Se uma sequência $\{x_k\}$ converge $q-$linearmente para x^* então também converge r-linearmente. No entanto o inverso não é verdadeiro.

Usaremos estas noções para estabelecer as condições de convergência global do método de Newton modificado. Para melhor entendermos as diferenças entre o método modificado e o método de Newton básico começamos por caracterizar o segundo quanto à convergência.

3.2 Convergência Local do Método de Newton

Relativamente ao método de Newton básico, descrito na secção 2.2, é do conhecimento generalizado que a sequência por ele gerada, sob certas condições, converge q-quadraticamente para o minimizante da função f. Essas condições são estabelecidas no teorema que se segue.

Teorema 1 ([1, pp90])*Seja* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *uma função duas vezes continuamente diferenciável no aberto* D. *Suponhamos que existem* $x^* \in D$ *e* $r, \beta > 0$ *tais que* $V_r(x^*) \subset D$, $\nabla f(x^*) = 0$, $H(x^*)^{-1}$ *existe,* $\left\|H(x^*)^{-1}\right\| \leq \beta$ *e* H *é Lipschitz contínua de constante* γ *em* $V_r(x^*)$.

Então, existe $\varepsilon > 0$ *tal que, para todo o* $x_0 \in V_\varepsilon(x^*)$, *a sequência* $\{x_k\}$ *gerada pelo método de Newton está bem definida, converge para o minimizante local* x^* *de* f *e satisfaz*

$$\|x_{k+1} - x^*\| \leq \beta\gamma \|x_k - x^*\|^2 , \quad k = 0, 1, 2, ...$$

Notemos que este teorema só garante a convergência local do método, pois a condição $x_0 \in V_\varepsilon(x^*)$, impõe que a aproximação inicial e o minimizante estejam suficientemente próximos. Muitas vezes esta exigência é muito forte pois não conseguimos obter x_0 nestas condições.

3.3 Globalização do Método de Newton

Como acabámos de apresentar, o método de Newton básico converge localmente. Nesta secção vamos estudar a introdução da técnica de procura unidimensional no método de Newton como uma estratégia de globalização. Passamos, em seguida, ao estudo da estratégia de regiões de confiança como outro processo de globalização.

Como referências tomamos, nesta secção, [1], [14] e [15].

3.3.1 Procura Unidimensional

No caso do passo Newton não ser o mais apropriado, a procura unidimensional introduz um escalar α que determina o comprimento do passo a dar ao longo desta direcção, por forma a que o valor da função objectivo diminua da forma desejada.

Podemos exigir que, em x_{k+1}, a função atinja o seu mínimo ao longo da direcção de procura. Neste caso, pretendemos calcular α_k^* que minimize $f(x_k + \alpha d_k)$ como função de α. Uma vez que $\alpha \in \mathbb{R}$, este é um problema de optimização unidimensional, conhecido por procura unidimensional exacta. Uma alternativa ao cálculo do α óptimo é procurar um valor deste parâmetro aceitável, isto é, um valor para o qual $f(x_k + \alpha d_k) < f(x_k)$. Esta condição, por si só, não garante a convergência do método. Outra alternativa, e a mais utilizada, é mais exigente que a anterior. Impõe que f, em cada iteração, sofra uma redução significativa.

Duas condições clássicas a exigir à função, para que esta sofra uma redução significativa, reúnem-se sob a *condição de Wolfe-Armijo*

$$\begin{array}{ll} f(x_k + \alpha d_k) \leq f(x_k) + \lambda \alpha \nabla f(x_k)^T d_k, & \lambda \in (0,1) \\ \nabla f(x_k + \alpha d_k)^T d_k \geq \beta \nabla f(x_k)^T d_k, & \beta \in (\lambda, 1)\,. \end{array} \tag{3.1}$$

O método de Newton com procura unidimensional, depois de calcular a direcção Newton, determina um escalar $\alpha > 0$, utilizando um dos vários métodos de optimização unidimensional, que satisfaça a condição de (3.1) e define o novo iterando como

$$x_{k+1} = x_k + \alpha_k d_k^N.$$

As condições de convergência do método de Newton com procura unidimensional podem ser encontradas no trabalho de Dennis e Schnabel [1, pp123] e podem ser resumidas no seguinte teorema.

Teorema 2 *Seja $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ uma função duas vezes continuamente diferenciável e consideremos o método de Newton com procura unidimensional satisfazendo (3.1). Suponhamos que existe $x_0 \in \mathbb{R}^n$ tal que f é limitada inferiormente em $\{x \in \mathbb{R}^n \,|\, f(x) \leq f(x_0)\}$. Se $H(x_k)$ for definida positiva e o seu número de condição for uniformemente limitado para todo o k, então o método converge globalmente. Além disso, a convergência é q-quadrática se a sequência $\{\alpha_k\} \subset \mathbb{R}$ convergir rapidamente para 1.*

3.3.2 Regiões de Confiança

A ideia do modelo de regiões de confiança é aceitar, para minimizante de f, o minimizante de um modelo quadrático desde que este reflicta adequadamente o comportamento da função objectivo.

Suponhamos que o passo Newton não produz, na função, uma redução satisfatória. No método de Newton com procura unidimensional calculávamos o melhor comprimento do passo a dar sobre a direcção de procura d^N. Na estratégia de regiões de confiança começamos por escolher o comprimento do passo e, em seguida, usamos o modelo quadrático para determinar a direcção de procura. A decisão de aceitar, ou não, o modelo é baseada no valor da norma desta direcção.

A formulação matemática do modelo de regiões de confiança é

$$\begin{array}{ll} \min & m(x+d) \\ s.a. & \|d\| \leq \delta. \end{array} \tag{3.2}$$

para algum $\delta > 0$ e onde $m(x+d) = f(x) + \nabla f(x)^T d + \frac{1}{2} d^T H(x) d$.

A solução deste problema é dada pelo seguinte lema.

Lema 1 ([1, pp131])*Seja $f : I\!R^n \longrightarrow I\!R$ duas vezes continuamente diferenciável, $H(x) \in I\!R^{n \times n}$ simétrica e definida positiva. O problema (3.2) tem como solução*

$$d(\mu) \triangleq -\left(H(x) + \mu I\right)^{-1} \nabla f(x)$$

para o único $\mu \geq 0$ tal que $\|d(\mu)\| = \delta$, a menos que $\|d(0)\| \leq \delta$, onde $\|.\|$ representa a norma l_2. Neste caso $d(0) = d^N$ é a solução. Para todo $\mu \geq 0$, $d(\mu)$ define uma direcção de descida de f em x.

A uma direcção de procura d pode-se exigir que satisfaça uma de duas condições: a fracção do decréscimo de Cauchy ou a fracção do decréscimo óptimo.

Denotemos $m(x_k + d)$ por $m_k(d)$ e a solução óptima do problema (3.2) por d^*. Dizemos que d satisfaz uma fracção do decréscimo óptimo se

$$m_k(0) - m_k(d) \geq \beta_2 \left(m_k(0) - m_k(d^*)\right) \tag{3.3}$$

com $\beta_2 \in (0, 1]$.

Dizemos que d_k satisfaz uma fracção do decréscimo de Cauchy se

$$m_k(0) - m_k(d) \geq \beta_1 \left(m_k(0) - m_k(c_k)\right) \tag{3.4}$$

onde $\beta_1 \in (0, 1]$ e c_k é o passo de Cauchy. Este passo é definido como a solução do problema

$$\begin{array}{ll} \min & m_k(d) \\ s.a. & \|d\| \leq \delta_k \\ & d \in span\{-\nabla f(x_k)\}^1 \end{array}$$

e é dado por

$$c_k = \begin{cases} -\dfrac{\|\nabla f(x_k)\|^2}{\nabla f(x_k)^T H\left(x_k\right) \nabla f(x_k)} \nabla f(x_k) & \text{se } \dfrac{\|\nabla f(x_k)\|^3}{\nabla f(x_k)^T H\left(x_k\right) \nabla f(x_k)} \leq \delta_k \\ -\dfrac{\delta_k}{\|\nabla f(x_k)\|} \nabla f(x_k), & \text{caso contrário.} \end{cases}$$

Enquanto a condição (3.4) relaciona o decréscimo verificado no modelo com a direcção d, na iteração corrente, com o obtido caso fosse dado o passo de Cauchy, a fracção do decréscimo óptimo (3.3) relaciona essa redução como o decréscimo resultante do passo óptimo do problema (3.2).

Facilmente se prova que a condição (3.3) implica a condição (3.4). Para isso mostra-se a desigualdade

$$\beta_2 \left(m_k(0) - m_k(d^*)\right) \geq \beta_1 \left(m_k(0) - m_k(c_k)\right).$$

Atendendo à definição de d^* temos $m_k(c_k) \geq m_k(d^*)$ pelo que

$$m_k(0) - m_k(d^*) \geq \beta_1 \left(m_k(0) - m_k(d^*)\right) \geq \beta_1 \left(m_k(0) - m_k(c_k)\right),$$

para $\beta_2 = \beta_1$.

A convergência global superlinear do algoritmo de regiões de confiança para um ponto de estacionaridade de f foi provada por Powell em [11]. Posteriormente em [12], o mesmo autor, provou um resultado semelhante em que o limite sobre a matriz que aproxima $H\left(x_k\right)$ depende linearmente de k. As condições gerais para a convergência global destes algoritmos são referidas por Vicente em [14]:

[1] $span\{v\}$, com $v \in \mathbb{R}^n$, denota $\{\alpha v | \alpha \in \mathbb{R}\}$

Teorema 3 *Seja* $f : I\!R^n \longrightarrow I\!R$ *uma função limitada inferiormente e* ∇f *uniformemente contínua. Se* d_k *satisfaz uma fracção do decréscimo de Cauchy e* $\left\|\bar{H}(x_k)\right\|$ *é uniformemente limitada, onde* $\bar{H}(x_k)$ *é uma aproximação à matriz hessiana, então* $\{x_k\}$ *satisfaz*

$$\lim_{k\to\infty} \|\nabla f(x_k)\| = 0.$$

Se, além disso, f *for duas vezes continuamente diferenciável e* d_k *satisfizer uma fracção do decréscimo óptimo, então* $\{x_k\}$ *tem um ponto de acumulação em* x^* *onde* $H(x^*)$ *é semi-definida positiva.*

3.4 Convergência Local do Método de Newton Modificado

Tal como o método de Newton básico, também o método modificado converge localmente, sob certas condições, q-quadraticamente para um ponto de estacionaridade de f. É possível estabelecer e resumir as condições de convergência no seguinte teorema e cuja demonstração pode ser encontrada em [8].

Teorema 4 *Seja* $f : I\!R^n \longrightarrow I\!R$ *uma função duas vezes continuamente diferenciável no aberto* $D \subset I\!R^n$. *Suponhamos que* $x^* \in D$ *é tal que* $\nabla f(x^*) = 0$, *existe uma constante* $\beta > 0$ *tal que* $H(x^*)$ *é não singular e* $\left\|H(x^*)^{-1}\right\| \leq \beta$ e H *é Lipschitz contínua de constante* γ *numa vizinhança de* x^* *contida em* D.

Então, existe $\delta > 0$ *tal que, para todo* $x_0 \in V_\delta(x^*)$, *a sequência* $\{x_k\}$ *gerada pelo método de Newton modificado está bem definida e converge pelo menos* $q-$*linearmente.*

Além disso, se

$$h^T H(x_k)^{-1} h \leq \mu_k h^T h, \quad \forall x_k \in V_\delta(x^*),\ h \in I\!R^n$$

e $\mu_k < 1$, *para* $w(x_k)$ *definido por* (2.3) *a sequência* $\{x_k\}$ *gerada pelo método de Newton modificado converge q-quadraticamente para* x^*, *sempre que* $0 \leq w(x_k) < 1$.

O teorema anterior considera $0 \leq w(x_k) < 1$ pois, para $w(x_k) = 1$, temos direcções de descida máxima que induzem a convergência linear, [15, pp86].

Capítulo 4

Convergência Global do Método de Newton Modificado

Neste capítulo estabelecemos um resultado sobre a convergência global do método de Newton modificado.

Começamos por incluir um conjunto de definições, lemas e teoremas essenciais à demonstração que nos propomos apresentar. Em seguida estabelecemos e provamos o resultado pretendido. Provamos também algumas propriedades quer do método quer da direcção de procura.

Todas as definições, lemas e teoremas seguem, salvo outra referência, [10], pelo que, neste capítulo, usamos o vector $p \in \mathbb{R}^n$ como direcção de procura, sendo $p = -d$.

4.1 Teorema da Convergência Global

Notemos que, de acordo com a definição de d^N, a direcção modificada apresentada em (2.2) pode ser reescrita como

$$\begin{aligned} d &= -\left[(1 - w\,(x))H(x)^{-1} + w\,(x)\,I\right]\nabla f(x) \\ &= -\left[B(x) + w\,(x)\,I\right]\nabla f(x), \end{aligned} \tag{4.1}$$

com $B(x) = (1 - w\,(x))H(x)^{-1}$.

O teorema seguinte estabelece as condições de convergência global do método de Newton modificado.

Teorema 5 *Seja $f : D \subset I\!R^n \longrightarrow I\!R$ uma função duas vezes continuamente diferenciável no aberto D. Suponhamos que existe $x_0 \in D$ tal que $L^0 = \{x \in D \,|\, f(x) \leq f(x_0)\}$ é compacto e que f tem um número finito de pontos de estacionaridade em L^0. Suponhamos, ainda, que $B : L^0 \longrightarrow I\!R^{n\times n}$ onde $B(x) = (1 - w(x))H(x)^{-1}$, com $w(x) \in [0, 1)$, é contínua para todo $x \in L^0$ e que*

$$\mu_0 h^T h \leq h^T B(x) h \leq \mu_1 h^T h, \quad \forall x \in L^0,\, h \in I\!R^n \tag{4.2}$$

$$h^T H(x) h \leq \gamma_1 h^T h, \quad \forall x \in L^0,\, h \in I\!R^n \tag{4.3}$$

onde μ_0, μ_1 e γ_1 não são necessariamente todos positivos, mas ou $\gamma_1 \leq 0$ e $\mu_1 > 0$ ou $\gamma_1 > 0$ e $-\frac{\gamma_1}{2+\gamma_1} \leq \mu_1 < \frac{2}{\gamma_1} - 1$.

Consideremos

$$x_{k+1} = x_k - \left[(1 - w(x_k))H(x_k)^{-1} + w(x_k) I\right] \nabla f(x_k), \quad k = 0, 1, 2, ...$$

com

$$\frac{1}{2}\hat{\gamma}_1 - \mu_0 < \eta_0 \leq w(x_k) \leq \frac{1}{2}\hat{\gamma}_1 + \frac{\mu_1}{\mu_1 + 1} < 1, \quad \hat{\gamma}_1 = \max(0, \gamma_1) \tag{4.4}$$

onde $\eta_0 < 1$.

Então

$$\{x_k\} \subset L^0,$$

$$\lim_{k\to\infty} x_k = x^* \qquad e \qquad \nabla f(x^*) = 0$$

e se $H(x^*)$ é não singular, a convergência é pelo menos r-linear.

Para demonstrar este teorema começamos por apresentar os conceitos de G e F derivada de uma função e a relação entre elas.

Definição 5 *Uma função $f : D \subset I\!R^n \longrightarrow I\!R$, D aberto, é Gâteaux (ou G) diferenciável se para algum $x \in D$ e todoh $\in I\!R^n$ o limite*

$$f'(x)(h) := \lim_{t\to 0} \frac{1}{t}\left(f(x + th) - f(x)\right) \tag{4.5}$$

existir e $f'(x)$ for contínua.

A aplicação $f'(x)$ é a G-derivada de f em x.

No caso de existir a aplicação linear $f'(x) : \mathbb{R}^n \longrightarrow \mathbb{R}$ contínua que satisfaz a condição

$$\lim_{h\to 0} \frac{1}{\|h\|} \left\| f(x + h) - f(x) - f'(x)(h) \right\| = 0 \tag{4.6}$$

então $f'(x)$ é a derivada de Fréchet ou F-derivada de f no ponto x.

A definição de F-derivada é mais geral do que a de G-derivada. Isto é, uma função que seja F-diferenciável num ponto é, necessariamente, G-diferenciável nesse ponto. O inverso não é verdadeiro. Daqui temos que qualquer propriedade válida para G-derivadas é válida para F-derivadas.

No entanto, se $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ admite G-derivada f' em cada ponto duma vizinhança (aberta) de x, e f' é contínua em x então f é F-derivável em x, [10, pp71].

É também importante referir que se f admite G-derivada, f', em todo o ponto de $D_0 \subset D$, e esta admite G-derivada, que denotamos por f'', então f'' é a segunda G-derivada da função. Por sua vez, se f' admitir F-derivada então f'' é a segunda F-derivada de f, [10, pp75].

Outro conceito do qual necessitamos, já ulitlizado no enunciado do teorema 5, é o de conjunto de nível de uma função.

Definição 6 *Se $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$, então qualquer conjunto não vazio da forma*

$$L(\gamma) = \{x \in D \,|\, f(x) \leq \gamma\}, \quad \gamma \in \mathbb{R} \tag{4.7}$$

é um conjunto de nível de f.

Refira-se que se a função f for contínua e o seu conjunto de nível compacto, existe $\bar{x}^* \in D$ tal que $f(\bar{x}^*) \leq f(x)$, para todo $x \in D$, [10, pp98].

Após a apresentação dos resultados que se seguem e cujas provas podem ser encontradas no trabalho de Ortega e Rheinboldt, estamos aptos a iniciar a prova do teorema 5.

Teorema 6 (Teorema do valor médio) ([10, pp79])*Seja $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ e suponhamos que f admite segunda G-derivada em cada ponto do convexo $D_0 \subset D$. Então, para quaisquer x, $y \in D_0$, existe $t \in (0,1)$ tal que*

$$f(x) - f(y) = -\nabla f(x)^T (y - x) - \frac{1}{2}(y - x)^T H(x + t(y - x))(y - x).$$

Lema 2 ([10, pp480])*Seja $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ uma função contínua no aberto D e G-diferenciável no compacto $L^0 = L(f(x_0))$ para algum $x_0 \in D$. Para quaisquer $x \in L^0$ e $p \in \mathbb{R}^n$ com $\nabla f(x)^T p > 0$ existe $\alpha^* > 0$ tal que*

$$f(x - \alpha^* p) = f(x) \quad e \quad [x, x - \alpha^* p] \subset L^0.$$

Em particular, se $\eta > 0$ for tal que

$$f(x-\alpha p) < f(x), \quad \forall x-\alpha p \in (x, x-\eta p] \cap L^0$$

então $[x, x-\eta p] \subset L^0$.

Lema 3 ([10, pp476])*Seja* $f : D \subset I\!R^n \longrightarrow I\!R$ *uma função continuamente diferenciável no compacto* $D_0 \subset D$ *e* Ω *o conjunto, finito, dos pontos de estacionaridade de* f *em* D_0*. Consideremos uma qualquer sequência* $\{x_k\} \subset D_0$ *para a qual*

$$\lim_{k\to\infty} (x_k - x_{k+1}) = 0 \quad e \quad \lim_{k\to\infty} \nabla f(x_k) = 0.$$

Então

$$\lim_{k\to\infty} x_k = x^* \quad e \quad \nabla f(x^*) = 0$$

com $x^* \in \Omega$.

Lema 4 ([10, pp477])*Seja* $f : D \subset I\!R^n \longrightarrow I\!R$ *uma função G-diferenciável no aberto* $D_0 \subset D$ *e suponhamos que* $\{x_k\} \subset D_0$ *converge para* $x^* \in D_0$.

Suponhamos, ainda, que $\nabla f(x^*) = 0$*,* f *admite segunda F-derivada em* x^**, que a matriz hessiana* $H(x^*)$ *é invertível e que existe* $\eta > 0$ *e* k_0 *para os quais*

$$f(x_k) - f(x_{k+1}) \geq \eta \left\|\nabla f(x_k)\right\|^2, \quad \forall k \geq k_0.$$

Então a convergência da sequência $\{x_k\}$ *é pelo menos* r*-linear.*

Demonstração do teorema 5 Das desigualdades (4.2) e (4.4) vem, para qualquer $x_k \in L^0$ e $h \in \mathbb{R}^n$

$$(\mu_0 + \eta_0) h^T h \leq h^T \left[B(x_k) + w_k I\right] h < (\mu_1 + 1) h^T h \tag{4.8}$$

para $w_k \in [\eta_0, 1)$, pelo que a direcção

$$p_k = \left[B(x_k) + w_k I\right] \nabla f(x_k)$$

está bem definida em L^0.

Além disso, como de (4.4), $0 \leq \frac{1}{2}\hat{\gamma}_1 < \mu_0 + \eta_0$, temos

$$\begin{aligned}
\nabla f(x_k)^T p_k &= \nabla f(x_k)^T \left[B(x_k) + w_k I\right] \nabla f(x_k) \\
&\geq (\mu_0 + \eta_0) \nabla f(x_k)^T \nabla f(x_k) \\
&\geq 0.
\end{aligned}$$

Assim, $\nabla f(x_k)^T p_k > 0$, a menos que $\|\nabla f(x_k)\| = 0$, isto é, a menos que o processo termine em x_k.

Consideremos, agora, $\alpha \in (0,1]$ tal que $[x_k, x_k - \alpha p_k] \subset L^0$. Pelo teorema do valor médio existe $\hat{\alpha} \in (0, \alpha)$ para o qual

$$f(x_k) - f(x_k - \alpha p_k) = \alpha \nabla f(x_k)^T p_k - \frac{1}{2}\alpha^2 p_k^T H(x_k - \hat{\alpha} p_k) p_k.$$

Ora

$$\begin{aligned}
\nabla f(x_k)^T p_k &= \nabla f(x_k)^T \left[B(x_k) + w_k I\right] \nabla f(x_k) \\
&= \nabla f(x_k)^T \left[B(x_k) + w_k I\right] \left[B(x_k) + w_k I\right]^{-1} \left[B(x_k) + w_k I\right] \nabla f(x_k) \\
&= p_k^T \left[B(x_k) + w_k I\right]^{-1} p_k \\
&> \frac{1}{\mu_1 + 1} p_k^T p_k,
\end{aligned}$$

por (4.8), pelo que, atendendo a (4.3),

$$\begin{aligned}
f(x_k) - f(x_k - \alpha p_k) &> \alpha \frac{1}{\mu_1 + 1} p_k^T p_k - \frac{1}{2}\alpha^2 p_k^T H(x_k - \hat{\alpha} p_k) p_k \\
&> \alpha \frac{1}{\mu_1 + 1} p_k^T p_k - \frac{1}{2}\alpha^2 \gamma_1 p_k^T p_k \\
&> \alpha \left[\frac{1}{\mu_1 + 1} - \frac{1}{2}\alpha \hat{\gamma}_1\right] p_k^T p_k.
\end{aligned}$$

Assim,

$$f(x_k) - f(x_k - \alpha p_k) > \alpha \left[1 - \left(\frac{\mu_1}{\mu_1 + 1} + \frac{1}{2}\hat{\gamma}_1\right)\right] p_k^T p_k > 0 \tag{4.9}$$

novamente por (4.4). Do lema 2 vem $x_{k+1} = x_k - p_k \in L^0$ donde, por indução sobre k, $\{x_k\} \subset L^0$.

Atendendo à desigualdade (4.8) e fazendo $h = \left[B(x_k) + w_k I\right]^{1/2} \nabla f(x_k)$ temos

$$\begin{aligned}
& (\mu_0 + \eta_0) \nabla f(x_k)^T \left[B(x_k) + w_k I\right]^{1/2} \left[B(x_k) + w_k I\right]^{1/2} \nabla f(x_k) \\
\leq\ & \nabla f(x_k)^T \left[B(x_k) + w_k I\right]^{1/2} \left[B(x_k) + w_k I\right] \left[B(x_k) + w_k I\right]^{1/2} \nabla f(x_k) \\
<\ & (\mu_1 + 1) \nabla f(x_k)^T \left[B(x_k) + w_k I\right]^{1/2} \left[B(x_k) + w_k I\right]^{1/2} \nabla f(x_k)
\end{aligned}$$

donde

$$\begin{aligned}
(\mu_0 + \eta_0) \nabla f(x_k)^T \left[B(x_k) + w_k I\right] \nabla f(x_k) &\leq \nabla f(x_k)^T \left[B(x_k) + w_k I\right] \left[B(x_k) + w_k I\right] \nabla f(x_k) \\
&< (\mu_1 + 1) \nabla f(x_k)^T \left[B(x_k) + w_k I\right] \nabla f(x_k).
\end{aligned}$$

Assim,

$$(\mu_0+\eta_0)^2\nabla f(x_k)^T\nabla f(x_k) \le {p_k}^T p_k < (\mu_1+1)^2\nabla f(x_k)^T\nabla f(x_k)$$

pelo que

$$(\mu_0+\eta_0)^2\left\|\nabla f(x_k)\right\|^2 \le \|p_k\|^2 < (\mu_1+1)^2\left\|\nabla f(x_k)\right\|^2. \tag{4.11}$$

Desta forma podemos reescrever a desigualdade (4.9) como

$$f(x_k)-f(x_k-\alpha p_k) > \alpha\left[1-\left(\frac{\mu_1}{\mu_1+1}+\frac{1}{2}\hat{\gamma}_1\right)\right](\mu_0+\eta_0)^2\left\|\nabla f(x_k)\right\|^2.$$

Uma vez que L^0 é compacto, o lado esquerdo da desigualdade anterior converge para zero e então $\lim\limits_{k\to\infty}\|\nabla f(x_k)\|=0$.

Por outro lado, como $p_k = x_{k+1}-x_k$, de (4.11) temos $\lim\limits_{k\to\infty}(x_{k+1}-x_k)=0$. Por aplicação do lema 3 temos

$$\lim_{k\to\infty} x_k = x^* \quad \text{e} \quad \nabla f(x^*)=0.$$

Nestas condições, o lema 4 garante a convergência, pelo menos r-linear da sequência $\{x_k\}$. ■

Os escalares μ_0 e μ_1 usados no teorema 5 podem ser entendidos, respectivamente, como um minorante e um majorante dos valores próprios de $B(x)$ em L^0.

4.2 Propriedades do Método de Newton Modificado

No teorema 5 imposemos algumas condições aos escalares γ_1 e μ_1. A justificação de tais res-trições é dada pelo lema 5.

Lema 5 *Seja $f : D \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ uma função duas vezes continuamente diferenciável no aberto D e L^0 o conjunto de nível da função f associado a $x_0 \in D$. Suponhamos que $B : L^0 \longrightarrow \mathbb{R}^{n\times n}$ onde $B(x) = (1-w(x))H(x)^{-1}$, com $w(x)\in[0,1)$, é contínua para todo $x \in L^0$ e que*

$$\mu_0 h^T h \le h^T B(x) h \le \mu_1 h^T h, \quad \forall x \in L^0,\ h\in\mathbb{R}^n, \tag{4.12}$$

$$h^T H(x) h \le \gamma_1 h^T h, \quad \forall x \in L^0,\ h\in\mathbb{R}^n, \tag{4.13}$$

onde μ_0, μ_1 e γ_1 não são necessariamente todos positivos, mas ou $\gamma_1 \le 0$ e $\mu_1 > 0$ ou $\gamma_1 > 0$ e $-\frac{\gamma_1}{2+\gamma_1} \le \mu_1 < \dfrac{2}{\gamma_1} - 1$, então

$$0 \le w(x) \le \frac{1}{2}\hat{\gamma}_1 + \frac{\mu_1}{\mu_1+1} < 1, \quad \hat{\gamma}_1 = \max(0,\gamma_1). \tag{4.14}$$

Demonstração Em toda a demonstração w denota $w(x)$.

Para $\gamma_1 \leq 0$, por definição de $\hat{\gamma}_1$ temos , $\hat{\gamma}_1 = 0$ pelo que a desigualdade (4.14) se reduz a

$$w \leq \frac{\mu_1}{\mu_1 + 1} < 1. \tag{4.15}$$

Além disso, como $\mu_1 > 0$ temos, naturalmente, $0 < \frac{\mu_1}{\mu_1+1} < 1$ donde

$$0 \leq w \leq \frac{1}{2}\hat{\gamma}_1 + \frac{\mu_1}{\mu_1 + 1} < 1.$$

Para $\gamma_1 > 0$, novamente por definição de $\hat{\gamma}_1$, temos $\hat{\gamma}_1 = \gamma_1$. Assim, a condição (4.14) vem

$$w \leq \frac{1}{2}\gamma_1 + \frac{\mu_1}{\mu_1 + 1} < 1. \tag{4.16}$$

Uma vez que $-\frac{\gamma_1}{2+\gamma_1} \leq \mu_1$ temos

$$\begin{aligned}
1 - \frac{\gamma_1}{2 + \gamma_1} &\leq \mu_1 + 1 \\
\frac{2 + \gamma_1 - \gamma_1}{2 + \gamma_1} &\leq \mu_1 + 1 \\
1 + \frac{1}{2}\gamma_1 &\geq \frac{1}{\mu_1 + 1} \\
-\frac{\mu_1}{\mu_1 + 1} &\leq \frac{1}{2}\gamma_1.
\end{aligned}$$

Por outro lado como $\mu_1 < \frac{2}{\gamma_1} - 1$ vem

$$\mu_1 + 1 < \frac{2}{\gamma_1} \Rightarrow \frac{1}{2}\gamma_1 < \frac{1}{\mu_1 + 1}.$$

Notemos que

$$\frac{1}{2}\gamma_1 + \frac{\mu_1}{\mu_1 + 1} < 1 \Leftrightarrow \frac{1}{2}\gamma_1 - \frac{1}{\mu_1 + 1} < 0.$$

Temos assim, para $\gamma_1 > 0$ e $-\frac{\gamma_1}{2+\gamma_1} \leq \mu_1 < \frac{2}{\gamma_1} - 1$,

$$0 \leq w \leq \frac{1}{2}\gamma_1 + \frac{\mu_1}{\mu_1 + 1} < 1. \blacksquare$$

Na secção 2.2, quando considerámos o método de Newton básico, referimos que se a matriz hessiana fosse definida positiva teríamos um método de descida. Uma questão interessante será saber se, ou quando, a matriz utilizada pelo método modificado, que na prática funciona como uma aproximação à matriz H, é definida positiva.

Lema 6 *Sendo λ_k o menor valor próprio de $H(x)^{-1}$, com $x \in D$, a matriz $(1-w(x))H(x)^{-1} + w(x)I$ é definida positiva se*

$$0 < w(x) < \frac{\lambda_k}{\lambda_k - 1} < 1.$$

Demonstração Denotemos novamente $w(x)$ por w.

Sejam $\lambda_1, \lambda_2, ..., \lambda_n$ os valores próprios da matriz $H(x)^{-1}$, com $x \in D$. Como esta matriz é simétrica todos os seus valores próprios são reais e admite uma decomposição espectral. Isto é existe uma matriz real U, ortogonal tal que

$$H(x)^{-1} = U\Omega U^T$$

onde $\Omega = diag(\lambda_1, \lambda_2, ..., \lambda_n)$. Então, atendendo a que $UU^T = I$

$$\begin{aligned}
(1-w)H(x)^{-1} + wI &= (1-w)U\Omega U^T + wI \\
&= (1-w)U\Omega U^T + wUU^T \\
&= U\left[(1-w)\Omega + wI\right]U^T \\
&= U diag\left((1-w)\lambda_1 + w, ..., (1-w)\lambda_n + w\right)U^T \\
&= U\bar{\Omega}U^T.
\end{aligned}$$

Para que a matriz $(1-w)H(x)^{-1} + wI$ seja definida positiva é necessário que

$$\min_{1 \leq i \leq n}\left([1-w]\lambda_i + w\right) > 0.$$

Seja λ_k (< 0) o menor valor próprio de $H(x)^{-1}$. Então

$$(1-w)\lambda_k + w > 0 \Rightarrow w < \frac{\lambda_k}{\lambda_k - 1}, \qquad \lambda_k \neq 1.$$

Como λ_k é o menor valor próprio de $H(x)^{-1}$, que não é definida positiva, a condição $\lambda_k \neq 1$ é verificada automaticamente. Além disso, $0 < \frac{\lambda_k}{\lambda_k - 1} < 1$ pois $\lambda_k < 0$. ■

Nesta prova consideramos $w(x) \in (0,1)$ pois os extremos $w(x) = 0$ e $w(x) = 1$ são casos que, no algoritmo modificado, representam situações especiais. A primeira corresponde a $H(x)$ ser definida positiva e $d_k = d_k^N$, a segunda verifica-se quando $H(x)$ não é invertível e, neste caso, $d_k = -\nabla f(x_k)$.

Poderíamos usar, para provar a convergência global no método de Newton modificado, o seguinte teorema enunciado por Ortega e Rheinboldt em [10, pp495] e que os autores utilizam para estabelecer a convergência global de vários métodos.

Teorema 7 *Suponhamos que $f : D \subset I\!R^n \longrightarrow I\!R$ é continuamente diferenciável no aberto D e que existe $x_0 \in D$ tal que $L^0 = L(f(x_0))$ é compacto. Suponhamos, ainda, que f tem um único ponto de estacionaridade x^* em D e consideremos $x_{k+1} = x_k - \xi_k \alpha_k p_k$, onde ξ_k e α_k são calculados por forma a que $x_k \in L^0$ para todo o k e que $\lim_{k\to+\infty} \frac{\nabla f(x_k)^T p_k}{\|p_k\|} = 0$. Suponhamos, finalmente, que $\{p_k\} \subset I\!R^n$ é do tipo gradiente para $\{x_k\}$.*

Então

$$\lim_{k\to+\infty} x_k = x^*.$$

Sendo $f : D \subset \mathrm{I\!R}^n \longrightarrow \mathrm{I\!R}$ uma função G-diferenciável em D e $\{x_k\} \subset D$ uma dada sequência, dizemos que a sequência $\{p_k\} \subset \mathrm{I\!R}^n$ de vectores não nulos é do tipo gradiente para a sequência $\{x_k\}$ se existe uma F-função σ tal que

$$\frac{\left|\nabla f(x_k)^T p_k\right|}{\|p_k\|} \geq \sigma\left(\|\nabla f(x_k)\|\right).$$

Entendemos por F-função uma qualquer aplicação $\sigma : [0, +\infty) \longrightarrow [0, +\infty)$ tal que

$$\lim_{k\to+\infty} \sigma\left(t_k\right) = 0 \quad \text{implica} \quad \lim_{k\to+\infty} t_k = 0$$

para qualquer sequência $\{t_k\} \subset \mathrm{I\!R}^n$.

Para estabelecer o resultado final de convergência global do método de Newton modificado, usando o teorema 7, teríamos que estudar o factor de amortecimento ξ_k e o escalar resultante da procura unidimensional α_k e que provar que $\lim_{k\to+\infty} \frac{\left|\nabla f(x_k)^T p_k\right|}{\|p_k\|} = 0$. As condições para que a direcção modificada seja do tipo gradiente são estabelecidas no lema que se segue.

Lema 7 *A direcção d_k, gerada pelo método de Newton modificado, é uma direcção do tipo gradiente (no sentido de Ortega e Rheinboldt) para a sequência $\{x_k\}$, definida por $x_{k+1} = x_k + d_k$.*

Demonstração Em toda a demonstração w_k denota $w\left(x_k\right)$.

Seja $f : D \subset \mathrm{I\!R}^n \longrightarrow \mathrm{I\!R}$ uma função duas vezes continuamente diferenciável em D. Consideremos a direcção d_k, de descida, escrita na forma (4.1)

$$d_k = -\left[B(x_k) + w_k I\right] \nabla f(x_k)$$

onde $B(x_k) = (1 - w_k) H(x_k)^{-1}$.

Consideremos as constantes μ_0, μ_1 e γ_1, não necessariamente todas positivas, tais que

$$\mu_0 h^T h \leq h^T B(x) h \leq \mu_1 h^T h, \quad \forall x \in D,\, h \in \mathbb{R}^n \tag{4.17}$$

e

$$h^T H(x) h \leq \gamma_1 h^T h, \quad \forall x \in D,\, h \in \mathbb{R}^n.$$

Tomemos $\eta_0 < 1$ e $\mu_2 \in \mathbb{R}^+$ tais que

$$\frac{1}{2}\hat{\gamma}_1 - \mu_0 < \eta_0 \leq w \leq \frac{1}{2}\hat{\gamma}_1 + \frac{\mu_1}{\mu_1 + 1} < 1, \quad \hat{\gamma}_1 = \max(0, \gamma_1) \tag{4.18}$$

e

$$\|B(x) + wI\| \leq \mu_2, \quad \forall x \in D. \tag{4.19}$$

De (4.17) e de (4.18) vem, para $x \in D$, $h \in \mathbb{R}^n$ e $w_k \in [\eta_0, 1)$

$$(\mu_0 + \eta_0) h^T h \leq h^T \left[B(x) + wI\right] h < (\mu_1 + 1) h^T h. \tag{4.20}$$

Dados uma matriz $C(x)$, $x \in D$ e $h \in \mathbb{R}^n$ temos

$$\begin{aligned} \|h\| &= \left\|C(x)C(x)^{-1}h\right\| \\ &\leq \|C(x)\| \left\|C(x)^{-1}h\right\|. \end{aligned}$$

Em particular para $C(x) = B(x) + wI$, por (4.19)

$$\|h\| \leq \mu_2 \left\|C(x)^{-1}h\right\|$$

donde $\left\|C(x)^{-1}h\right\| \geq \mu_2^{-1}\|h\|$. Daqui vem, considerando $C(x)h = \bar{h}$

$$\|h\| = \left\|C(x)^{-1}C(x)h\right\| = \left\|C(x)^{-1}\bar{h}\right\| \geq \mu_2^{-1}\|C(x)h\|. \tag{4.21}$$

Notemos que, por definição de $\hat{\gamma}_1$, $\mu_0 + \eta_0 > 0$ e que, sendo d_k de descida

$$\nabla f(x_k)^T d_k < 0 \Leftrightarrow \nabla f(x_k)^T \left[-\left(B(x_k) + w_k I\right)\right] \nabla f(x_k) < 0.$$

Assim, para $x_k \in D$, atendendo a (4.20) e (4.21)

$$\begin{aligned} \left|\nabla f(x_k)^T d_k\right| &= \left|-\nabla f(x_k)^T \left[B(x_k) + w_k I\right] \nabla f(x_k)\right| \\ &= \nabla f(x_k)^T \left[B(x_k) + w_k I\right] \nabla f(x_k) \\ &\geq (\mu_0 + \eta_0) \|\nabla f(x_k)\|^2 \end{aligned}$$

$$
\begin{aligned}
&= (\mu_0+\eta_0)\,\|\nabla f(x_k)\|\,\underbrace{\|\nabla f(x_k)\|} \\
&\geq (\mu_0+\eta_0)\,\|\nabla f(x_k)\|\,\mu_2^{-1}\,\|(B(x_k)+w_kI)\,\nabla f(x_k)\|\,, \\
&= \gamma\,\|\nabla f(x_k)\|\,\|d_k\|\,, \qquad (4.22)
\end{aligned}
$$

onde $\gamma = \frac{\mu_0+\eta_0}{\mu_2} > 0$, isto é

$$\left|\nabla f(x_k)^T d_k\right| \geq \sigma\left(\|\nabla f(x_k)\|\right)\|d_k\|\,,$$

onde $\sigma(t) = \gamma t$, é uma F-função. ■

4.3 Propriedades da Direcção Newton Modificada

Como vimos na secção 3.3.2, nas técnicas de regiões de confiança é usual exigir-se que a direcção de procura satisfaça uma fracção do decréscimo óptimo dada por (3.3). De seguida estabelecemos as condições necessárias para que a direcção gerada pelo método modificado satisfaça tal condição.

Teorema 8 *Seja $f : D \subset I\!R^n \longrightarrow I\!R$ uma função duas vezes continuamente diferenciável no aberto D. Suponhamos que $B : D \longrightarrow I\!R^{n\times n}$, onde $B(x) = (1-w(x))H(x)^{-1}$ e $w(x) \in [0,1)$, é contínua para todo $x \in D$ e que*

$$\mu_0 h^T h \leq h^T B(x) h \leq \mu_1 h^T h, \quad \forall x \in D, h \in I\!R^n, \qquad (4.23)$$

$$\|H(x)\| \leq \gamma_2, \quad \forall x \in D \qquad (4.24)$$

com

$$\frac{1}{2}\gamma_2 - \mu_0 < \eta_0 \leq w(x) \leq \frac{1}{2}\gamma_2 + \frac{\mu_1}{\mu_1+1} < 1, \qquad (4.25)$$

onde μ_0 e μ_1, não são necessariamente positivos, mas $-\frac{\gamma_2}{2+\gamma_2} \leq \mu_1 < \frac{2}{\gamma_2} - 1$ e $0 < \gamma_2 \leq \frac{3}{2}\frac{1}{\mu_0+\eta_0}$ e $\eta_0 < 1$.

Então a direcção

$$d_k = -\left[B(x_k) + w(x_k)\,I\right]\nabla f(x_k)$$

satisfaz uma fracção do decréscimo óptimo, isto é

$$m_k(0) - m_k(d_k) > \beta\left(m_k(0) - m_k(d_k^*)\right)$$

onde $\beta \in (0,1]$, $m_k(d) = f(x_k) + \nabla f(x_k)^T d + \frac{1}{2} d^T H(x_k) d$ e d_k^* é a solução óptima do problema *de regiões de confiança*

$$\begin{array}{ll} \min & m_k(d) \\ s.a. & \|d\| \leq \delta_k. \end{array}$$

Demonstração Denotemos $w(x_k)$ por w_k.

Uma vez que $w_k \in [0,1)$, comecemos por tomar $w_k = 0$. Neste caso

$$d_k = d_k^N = -H(x_k)^{-1}\nabla f(x_k),$$

que é a solução óptima de $\min m_k(d)$. Assim, $m_k(d_k^N) \leq m_k(d_k^*)$ pelo que

$$m_k(0) - m_k(d_k^N) \geq m_k(0) - m_k(d_k^*) \geq \beta\left(m_k(0) - m_k(d_k^*)\right),$$

onde $\beta \in (0,1]$.

Consideremos, agora, $w_k \in (0,1)$.

Das condições (4.23) e (4.25) vem, para qualquer $x_k \in D$ e $h \in \mathbb{R}^n$

$$0 < (\mu_0 + \eta_0)\, h^T h \leq h^T \left[B(x_k) + w_k I\right] h. \tag{4.26}$$

Suponhamos que $\left\|d_k^N\right\| = \left\|-H(x_k)^{-1}\nabla f(x_k)\right\| > \delta_k$. Então por (4.24) existe $c \geq 1$ tal que

$$\left\|H(x_k)^{-1}\right\| \leq \frac{c}{\gamma_2} \tag{4.27}$$

donde

$$\begin{aligned} \|d_k^*\| \leq \delta_k &< \left\|-H(x_k)^{-1}\nabla f(x_k)\right\| \\ &\leq \left\|H(x_k)^{-1}\right\| \|\nabla f(x_k)\| \\ &\leq \frac{c}{\gamma_2}\|\nabla f(x_k)\|. \end{aligned} \tag{4.28}$$

Da definição de m_k e pelas propriedades da norma, temos

$$\begin{aligned} m_k(0) - m_k(d_k^*) &= -\left[\nabla f(x_k)^T d_k^* + \frac{1}{2}(d_k^*)^T H(x_k) d_k^*\right] \\ &\leq \left\|\nabla f(x_k)^T d_k^*\right\| + \frac{1}{2}\left\|(d_k^*)^T H(x_k) d_k^*\right\| \\ &\leq \|\nabla f(x_k)\| \|d_k^*\| + \frac{1}{2}\|d_k^*\|^2 \|H(x_k)\|. \end{aligned}$$

mas por aplicação, a esta desigualdade, das condições (4.24) e (4.28) vem

$$\begin{aligned} m_k(0) - m_k(d_k^*) &\leq \|\nabla f(x_k)\| \delta_k + \frac{1}{2}\delta_k^2 \|H(x_k)\| \\ &\leq \delta_k \|\nabla f(x_k)\| + \frac{1}{2}\gamma_2 \delta_k^2 \\ &< \delta_k \|\nabla f(x_k)\| + \frac{1}{2}\gamma_2 \delta_k \frac{c}{\gamma_2} \|\nabla f(x_k)\| = \frac{2+c}{2}\delta_k \|\nabla f(x_k)\| . \end{aligned}$$

Desta forma,

$$\frac{2}{2+c}\left[m_k(0) - m_k(d_k^*)\right] < \delta_k \|\nabla f(x_k)\| . \tag{4.29}$$

Novamente por definição de m_k

$$m_k(0) - m_k(d_k) = -\nabla f(x_k)^T d_k - \frac{1}{2} d_k^T H(x_k) d_k,$$

mas atendendo a (4.24) e a (4.26) temos

$$\begin{aligned} -\nabla f(x_k)^T d_k &= \nabla f(x_k)^T \left[B(x_k) + w_k I\right] \nabla f(x_k) \\ &\geq (\mu_0 + \eta_0) \|\nabla f(x_k)\|^2 \end{aligned}$$

e

$$\begin{aligned} d_k^T H(x_k) d_k &\leq \|H(x_k)\| \|d_k\|^2 \\ &\leq \gamma_2 \|d_k\|^2 \\ &= \gamma_2 \left\|- \left[B(x_k) + w_k I\right] \nabla f(x_k)\right\|^2 \\ &\leq \gamma_2 \|B(x_k) + w_k I\|^2 \|\nabla f(x_k)\|^2 \\ &\leq \gamma_2 \frac{1}{\gamma_2^2} \left[c\,(1-w_k) + \gamma_2 w_k\right]^2 \|\nabla f(x_k)\|^2 \end{aligned}$$

pois da definição de $B(x)$ e por (4.27)

$$\begin{aligned} \|B(x_k) + w_k I\| &\leq \|B(x_k)\| + \|w_k I\| \\ &\leq \left\|(1-w_k)\, H(x_k)^{-1}\right\| + w_k \\ &\leq (1-w_k) \left\|H(x_k)^{-1}\right\| + w_k \\ &\leq (1-w_k)\frac{c}{\gamma_2} + w_k \\ &= \frac{1}{\gamma_2}\left[c\,(1-w_k) + \gamma_2 w_k\right]. \end{aligned}$$

Assim, denotando $c\,(1-w_k) + \gamma_2 w_k$ por σ,

$$
\begin{aligned}
m_k(0) - m_k(d_k) &\geq (\mu_0 + \eta_0) \|\nabla f(x_k)\|^2 - \frac{1}{2}\frac{\sigma^2}{\gamma_2} \|\nabla f(x_k)\|^2 \\
&= \frac{1}{2}\left[2(\mu_0 + \eta_0) - \frac{\sigma^2}{\gamma_2}\right] \|\nabla f(x_k)\|^2 \\
&> \frac{1}{2}\left[2(\mu_0 + \eta_0) - \frac{\sigma^2}{\gamma_2}\right] \frac{\gamma_2}{c}\delta_k \|\nabla f(x_k)\|,
\end{aligned}
$$

pois $\frac{\gamma_2}{c}\delta_k < \|\nabla f(x_k)\|$, de (4.28).

Recorrendo a (4.29), a desigualdade anterior pode ser escrita como

$$
m_k(0) - m_k(d_k) > \frac{\gamma_2}{2c}\left[2(\mu_0 + \eta_0) - \frac{\sigma^2}{\gamma_2}\right] \frac{2}{2+c} (m_k(0) - m_k(d_k^*))
$$

ou ainda

$$
m_k(0) - m_k(d_k) > \beta (m_k(0) - m_k(d_k^*))
$$

com $\beta = \frac{\gamma_2}{c(2+c)}\left[2(\mu_0 + \eta_0) - \frac{\sigma^2}{\gamma_2}\right]$.

Mostremos agora que $0 < \beta \leq 1$.

Comecemos por notar que $\sigma = w_k(\gamma_2 - c) + c$, onde $w_k \in (0, 1)$. Assim, se $\gamma_2 > c$ temos

$$
\begin{gathered}
0 < w_k(\gamma_2 - c) < \gamma_2 - c \\
c < \sigma < \gamma_2 \\
c^2 < \sigma^2 < \gamma_2^2 \\
-\gamma_2 < -\frac{\sigma^2}{\gamma_2} < -\frac{c^2}{\gamma_2}
\end{gathered}
$$

donde

$$
\frac{\gamma_2}{c(2+c)}[2(\mu_0 + \eta_0) - \gamma_2] < \beta.
$$

Como $0 < \gamma_2$ e $\gamma_2 < 2(\mu_0 + \eta_0)$ de (4.25) vem

$$
\frac{\gamma_2}{c(2+c)}[2(\mu_0 + \eta_0) - \gamma_2] > 0.
$$

Por outro lado se $\gamma_2 \leq c$ temos

$$
\begin{gathered}
\gamma_2 - c \leq w_k(\gamma_2 - c) \leq 0 \\
\gamma_2 \leq \sigma \leq c \\
\gamma_2^2 \leq \sigma^2 \leq c^2 \\
-\frac{c^2}{\gamma_2} \leq -\frac{\sigma^2}{\gamma_2} \leq -\gamma_2
\end{gathered}
$$

donde

$$\frac{\gamma_2}{c\,(2+c)}\left[2\,(\mu_0+\eta_0)-\frac{c^2}{\gamma_2}\right]\leq\beta.$$

Uma vez que $\gamma_2>0$,

$$\frac{\gamma_2}{c\,(2+c)}\left[2\,(\mu_0+\eta_0)-\frac{c^2}{\gamma_2}\right]>0\Rightarrow\gamma_2>\frac{c^2}{2\,(\mu_0+\eta_0)}.$$

Desta forma, como $\gamma_2\leq c$,

$$\begin{aligned}\frac{\gamma_2^2}{2\,(\mu_0+\eta_0)}\leq\frac{c^2}{2\,(\mu_0+\eta_0)}<\gamma_2 &\Rightarrow \frac{\gamma_2^2}{2\,(\mu_0+\eta_0)}<\gamma_2\\ &\Rightarrow \gamma_2<2\,(\mu_0+\eta_0)\end{aligned}$$

que é válido atendendo, novamente, a (4.25).

Assim, para todo $\gamma_2>0$ temos $0<\beta$.

Mostremos então que $\beta\leq 1$. Dado que $\sigma\geq 0$ temos

$$\begin{aligned}-\frac{\sigma^2}{\gamma_2} &\leq 0\\ 2\,(\mu_0+\eta_0)-\frac{\sigma^2}{\gamma_2} &\leq 2\,(\mu_0+\eta_0)\\ \beta &\leq 2\frac{\gamma_2}{c\,(c+2)}\,(\mu_0+\eta_0)\,.\end{aligned}$$

Além disso, uma vez que $c\geq 1$ temos ainda $\frac{1}{c(c+2)}\leq\frac{1}{3}$ donde

$$\beta\leq\frac{2}{3}\gamma_2\,(\mu_0+\eta_0)\leq 1$$

pois $\gamma_2\leq\frac{3}{2}\frac{1}{\mu_0+\eta_0}$. ■

Na demonstração anterior considerámos $\left\|d_k^N\right\|>\delta_k$. No caso de $\left\|d_k^N\right\|\leq\delta_k$, a direcção de Newton modificada é a de Newton, que se prova satisfazer a condição do decréscimo óptimo.

O caso de $\|d_k\|>\delta_k$, onde d_k é a direcção modificada, não é tomado em consideração. Este facto deve-se a "supormos" que, de alguma forma, a escolha de $w\,(x_k)$ limita o tamanho da norma da direcção, pelo que o passo associado é sempre inferior a δ_k.

Reunindo as condições do teorema 8 que envolvem o escalar γ_2 podemos concluir que

$$\gamma_2\leq\min\left(2\,(\mu_0+\eta_0)\,,\frac{3}{2\,(\mu_0+\eta_0)}\right).$$

Sob certas condições, uma direcção que satisfaça uma fracção do decréscimo óptimo, origina uma redução significativa do modelo quadrático, $m(d)$, em todas as iterações. Podemos estabelecer essas condições no lema que se segue e para o qual desenvolvemos a prova apresentada.

Lema 8 *Se $d \in I\!R^n$ satisfaz uma fracção do decréscimo óptimo e $\nabla f(x_k)^T d \leq -c_1 \|d\| \|\nabla f(x_k)\|$, com $c_1 \in (0,1]$, então existem $\bar{c}_1, \sigma_1 > 0$ tais que, para todo $\delta > 0$,*

$$m_k(0) - m_k(d) \geq \bar{c}_1 \|\nabla f(x_k)\| \min\left(\delta, \sigma_1 \frac{\|\nabla f(x_k)\|}{\|H(x_k)\|}\right).$$

Demonstração Seja, para $a > 0$,

$$\begin{aligned} h(a) &= -\left[m_k(0) - m_k(ad)\right] \\ &= a\nabla f(x_k)^T d + \frac{a^2}{2} d^T H(x_k) d, \end{aligned}$$

então

$$\begin{aligned} h'(a) &= \nabla f(x_k)^T d + a d^T H(x_k) d \\ h''(a) &= d^T H(x_k) d. \end{aligned}$$

Considere-se o problema

$$\begin{array}{ll} \min & h(a) \\ s.a. & a\|d\| \leq \delta \end{array}$$

e seja $s^* = a^* d$, onde a^* é solução do problema.

Se $d^T H(x_k) d > 0$ então

$$a^* = \begin{cases} -\frac{\nabla f(x_k)^T d}{d^T H(x_k) d} & \text{se } -\frac{\nabla f(x_k)^T d}{d^T H(x_k) d} \leq \frac{\delta}{\|d\|} \\ \frac{\delta}{\|d\|} & \text{caso contrário.} \end{cases}$$

No primeiro caso

$$\begin{aligned} m_k(0) - m_k(s^*) &= m_k(0) - m_k(a^* d) \\ &= \frac{\nabla f(x_k)^T d}{d^T H(x_k) d} \nabla f(x_k)^T d - \frac{1}{2}\left(\frac{\nabla f(x_k)^T d}{d^T H(x_k) d}\right)^2 d^T H(x_k) d \\ &= \frac{1}{2} \frac{\left(\nabla f(x_k)^T d\right)^2}{d^T H(x_k) d} \geq \frac{1}{2} c_1^2 \frac{\|d\|^2 \|\nabla f(x_k)\|^2}{d^T H(x_k) d}. \end{aligned}$$

Como

$$d^T H(x_k) d \leq \left\|d^T H(x_k) d\right\| \leq \|d\|^2 \|H(x_k)\|$$

temos

$$\frac{\|d\|^2 \|\nabla f(x_k)\|^2}{d^T H(x_k) d} \geq \frac{\|d\|^2 \|\nabla f(x_k)\|^2}{\|d\|^2 \|H(x_k)\|}$$

pelo que

$$m_k(0) - m_k(s^*) \geq \frac{1}{2} c_1^2 \frac{\|\nabla f(x_k)\|^2}{\|H(x_k)\|}. \tag{4.30}$$

No segundo caso, isto é se $-\frac{\delta}{\|d\|} > \frac{\nabla f(x_k)^T d}{d^T H(x_k) d}$,

$$\begin{aligned}
m_k(0) - m_k(s^*) &= -\frac{\delta}{\|d\|} \nabla f(x_k)^T d - \frac{1}{2} \frac{\delta^2}{\|d\|^2} d^T H(x_k) d \\
&> -\frac{\delta}{\|d\|} \nabla f(x_k)^T d + \frac{1}{2} \frac{\delta}{\|d\|} \nabla f(x_k)^T d \\
&= -\frac{1}{2} \frac{\delta}{\|d\|} \nabla f(x_k)^T d \\
&\geq \frac{c_1}{2} \frac{\delta}{\|d\|} \|d\| \, \|\nabla f(x_k)\| \\
&= \frac{c_1}{2} \delta \, \|\nabla f(x_k)\| . \qquad (4.31)
\end{aligned}$$

Se $d^T H(x_k) d \leq 0$, $a^* = \delta / \|d\|$, pelo que

$$\begin{aligned}
m_k(0) - m_k(s^*) &= -\frac{\delta}{\|d\|} \nabla f(x_k)^T d - \frac{1}{2} \frac{\delta^2}{\|d\|^2} d^T H(x_k) d \\
&\geq -\frac{\delta}{\|d\|} \nabla f(x_k)^T d \\
&\geq \frac{\delta}{\|d\|} c_1 \|d\| \, \|\nabla f(x_k)\| \\
&\geq c_1 \delta \, \|\nabla f(x_k)\| . \qquad (4.32)
\end{aligned}$$

Atendendo a (4.30), (4.31), e (4.32) temos

$$\begin{aligned}
m_k(0) - m_k(s^*) &\geq \min \left(\frac{1}{2} c_1^2 \frac{\|\nabla f(x_k)\|^2}{\|H(x_k)\|}, \frac{c_1}{2} \delta \, \|\nabla f(x_k)\| , c_1 \delta \, \|\nabla f(x_k)\| \right) \\
&= \min \left(\frac{1}{2} c_1^2 \frac{\|\nabla f(x_k)\|^2}{\|H(x_k)\|}, \frac{c_1}{2} \delta \, \|\nabla f(x_k)\| \right) \\
&= \frac{c_1}{2} \|\nabla f(x_k)\| \min \left(c_1 \frac{\|\nabla f(x_k)\|}{\|H(x_k)\|}, \delta \right).
\end{aligned}$$

Daqui, para $\beta \in (0, 1]$, vem

$$\begin{aligned}
\beta \left(m_k(0) - m_k(s^*) \right) &\geq \beta \frac{c_1}{2} \|\nabla f(x_k)\| \min \left(c_1 \frac{\|\nabla f(x_k)\|}{\|H(x_k)\|}, \delta \right) \\
&= \bar{c}_1 \|\nabla f(x_k)\| \min \left(\sigma_1 \frac{\|\nabla f(x_k)\|}{\|H(x_k)\|}, \delta \right)
\end{aligned}$$

com $\bar{c}_1 = \beta \frac{c_1}{2} > 0$ e $\sigma_1 = c_1 > 0$.

Atendendo a que por hipótese, para $\beta \in (0,1]$,

$$m_k(0) - m_k(d) \geq \beta\,(m_k(0) - m_k(s^*))$$

temos

$$m_k(0) - m_k(d) \geq \bar{c}_1 \left\| \nabla f(x_k) \right\| \min\left(\sigma_1 \frac{\|\nabla f(x_k)\|}{\|H(x_k)\|}, \delta\right). \blacksquare$$

Uma vez que já provámos que a direcção gerada pelo método de Newton modificado satisfaz uma fracção do decréscimo óptimo, vamos estabelecer condições idênticas para provar que a direcção (4.1) origina uma redução significativa no modelo.

O termo "significativa" diz respeito à relação entre a quantidade de redução do comprimento do passo e o gradiente. O facto de se verificar, de iteração para iteração, uma redução no valor da função não significa convergência para um ponto de estacionaridade. A não convergência pode ocorrer de duas formas.

Na primeira, os passos calculados são demasiado grandes em relação ao decréscimo verificado no valor de f. Embora a sequência $\{x_k\}$ conduza a valores decrescentes da função, a sequência $\{f(x_k)\}$ converge para dois pontos de não estacionaridade.

Na segunda, os passos tornam-se demasiado pequenos e o decréscimo da função é reduzido relativamente à norma do gradiente. Neste caso, a sequência converge novamente para um ponto de não estacionaridade.

O lema seguinte particulariza o lema 8 para o caso da direcção modificada calculada por (2.2).

Lema 9 *Nas condições do teorema* 8 *existem* $\bar{c}_1, \sigma_1 > 0$ *tais que, para todo* $\delta > 0$,

$$m_k(0) - m_k(d_k) \geq \bar{c}_1 \left\| \nabla f(x_k) \right\| \min\left(\delta, \sigma_1 \frac{\|\nabla f(x_k)\|}{\|H(x_k)\|}\right)$$

onde d_k *é a direcção gerada pelo método de Newton modificado.*

Demonstração Atendendo ao teorema 8 e ao lema 8 basta mostrar que, com $c_1 \in (0,1]$,

$$\nabla f(x_k)^T d_k \leq -c_1 \left\| d_k \right\| \left\| \nabla f(x_k) \right\|.$$

Ora, da demonstração do lema 7 sabemos que de (4.23) e de (4.25) vem

$$\left|\nabla f(x_k)^T d_k\right| > \gamma \left\|\nabla f(x_k)\right\| \left\|d_k\right\|, \qquad \text{(cf.(4.22))}$$

onde $\gamma = \frac{\mu_0+\eta_0}{\mu_2} > 0$, sendo $\mu_2 \geq \|B(x_k) + w_k I\|$.

Assim,

$$\begin{aligned} & \left|\nabla f(x_k)^T d_k\right| > \gamma \left\|\nabla f(x_k)\right\| \left\|d_k\right\| \\ \Rightarrow & \nabla f(x_k)^T d_k < -\gamma \left\|\nabla f(x_k)\right\| \left\|d_k\right\| \quad \vee \quad \nabla f(x_k)^T d_k > \gamma \left\|\nabla f(x_k)\right\| \left\|d_k\right\| \\ \Rightarrow & \nabla f(x_k)^T d_k < -\gamma \left\|\nabla f(x_k)\right\| \left\|d_k\right\| \end{aligned}$$

pois d_k é de descida. ■

Tendo em consideração o trabalho de Shultz, Schnabel e Byrd [13, Lema 4.1], o lema 8 surge de forma imediata. Estes autores provam que qualquer direcção que satisfaça uma fracção do decréscimo de Cauchy induz uma redução significativa no modelo. Ora qualquer direcção que satisfaça uma fracção do decréscimo óptimo satisfaz uma fracção do decréscimo de Cauchy, logo o resultado é imediato. Consequentemente o lema 9 é também um corolário do lema apresentado no trabalho destes autores.

Um dos problemas que pode ocorrer com o método de Newton é o de a direcção por ele gerada ser quase ortogonal à direcção de descida máxima. Uma vez que o método modificado pretende ser uma alternativa a este método, temos interesse em saber se a direcção modificada se torna, ou não, ortogonal à direcção de descida máxima.

As condições para a não ortogonalidade são estabelecidas no lema seguinte. Este lema exige que o número de condição[1] da matriz do método de Newton modificado $B(x) + w(x) I$ seja limitado para que o co-seno do ângulo formado pela direcção modificada e o simétrico do vector gradiente, em todas iterações, não esteja numa vizinhança de zero.

Lema 10 *Nas condições do teorema 8 se o número de condição de* $B(x_k) + w_k I$ *satisfizer* $cn\,(B(x_k) + w_k I) \leq \nu$, $\nu > 0$ *finito, então*

$$\cos(\theta_k) = \frac{\nabla f(x_k)^T \left[B(x_k) + w_k I\right] \nabla f(x_k)}{\left\|\nabla f(x_k)\right\| \left\|\left[B(x_k) + w_k I\right] \nabla f(x_k)\right\|} \geq \frac{1}{\nu},$$

onde $\|.\|$ *denota a norma* l_2 *e* θ_k *é o ângulo formado por* $-\nabla f(x_k)$ *e* d_k.

[1] Usamos $cn(A)$ para denotar o número de condição da matriz A.

Demonstração Pelas propriedades das normas

$$\begin{aligned}\nabla f(x_k)^T\left[B(x_k)+w_kI\right]\nabla f(x_k) &\leq \|\nabla f(x_k)\|\,\|\left[B(x_k)+w_kI\right]\nabla f(x_k)\| \\ &\leq \|\nabla f(x_k)\|^2\,\|B(x_k)+w_kI\|\end{aligned}$$

donde

$$\frac{1}{\|\nabla f(x_k)\|\,\|\left[B(x_k)+w_kI\right]\nabla f(x_k)\|} \geq \frac{1}{\|\nabla f(x_k)\|^2\,\|B(x_k)+w_kI\|}$$

$$\Rightarrow \frac{\nabla f(x_k)^T\left[B(x_k)+w_kI\right]\nabla f(x_k)}{\|\nabla f(x_k)\|\,\|\left[B(x_k)+w_kI\right]\nabla f(x_k)\|} \geq \frac{\nabla f(x_k)^T\left[B(x_k)+w_kI\right]\nabla f(x_k)}{\|\nabla f(x_k)\|^2\,\|B(x_k)+w_kI\|}$$

$$\Rightarrow \frac{\nabla f(x_k)^T\left[B(x_k)+w_kI\right]\nabla f(x_k)}{\|\nabla f(x_k)\|\,\|\left[B(x_k)+w_kI\right]\nabla f(x_k)\|} \geq \frac{\xi_1}{\|B(x_k)+w_kI\|}$$

onde $\xi_1(>0)$ é o menor valor próprio de $B(x_k)+w_kI$.

Assim,

$$\begin{aligned}\frac{\nabla f(x_k)^T\left[B(x_k)+w_kI\right]\nabla f(x_k)}{\|\nabla f(x_k)\|\,\|\left[B(x_k)+w_kI\right]\nabla f(x_k)\|} &\geq \frac{\xi_1}{\|B(x_k)+w_kI\|} \\ &= \frac{1}{\frac{1}{\xi_1}\|B(x_k)+w_kI\|} \\ &= \frac{1}{\left\|(B(x_k)+w_kI)^{-1}\right\|\,\|B(x_k)+w_kI\|} \geq \frac{1}{\nu}. \blacksquare\end{aligned}$$

Se ξ_1 é o menor valor próprio de $B(x_k)+w_kI$, $\frac{1}{\xi_1}$ é o maior valor próprio de $\left[B(x_k)+w_kI\right]^{-1}$, pelo que

$$\left\|\left[B(x_k)+w_kI\right]^{-1}\right\| = \frac{1}{\xi_1}.$$

Refira-se que é necessário que as condições (4.23) e (4.25) se verifiquem para haver garantia de ξ_1 ser positivo.

Capítulo 5

Reflexões Finais

5.1 Conclusões

Após a realização deste trabalho, várias conclusões se podem tirar relativamente ao método de Newton modificado. Naturalmente a mais importante diz respeito à sua convergência global.

O facto do método de Newton modificado convergir globalmente permite-nos concluir que, sob certas condições, é uma boa alternativa ao método de Newton básico. Tal como este, também o novo método converge q-quadraticamente numa vizinhança de um ponto estacionário de f, no entanto, sendo globalmente convergente, pode funcionar como método de "arranque" para o processo iterativo.

Uma vez que o método é globalmente convergente e a direcção por ele gerada satisfaz uma fracção do decréscimo óptimo, condição esta associada aos algoritmos de regiões de confiança, somos induzidos a pensar que este método tem algumas semelhanças com a estratégia de regiões de confiança. Surgem, no entanto, algumas questões relacionadas com o controlo do passo. A estratégia de regiões de confiança define o comprimento máximo do passo (região) a tomar na iteração seguinte de acordo com o que terá acontecido na iteração corrente. A direcção Newton modificada, de certa forma é controlada pela necessidade, ou não, de calcular a função escalar $w(x)$. Não é, no entanto, até à data, óbvio que esta necessidade seja controlada pela direcção da iteração anterior. Tem sim a ver com a região onde se encontra a aproximação da iteração corrente.

Muitas modificações ao método básico exigem que a matriz hessiana seja definida positiva para que a direcção calculada seja de descida. Embora a direcção modificada tenha sido construída por forma a ser de descida, dependendo este facto da função escalar $w(x)$, não foi exigido, à matriz modificada, que gozasse desta propriedade. Mostrou-se, no entanto, que a propriedade de ser definida positiva depende da forma como $w(x)$ se relaciona com o menor valor próprio da inversa da hessiana da função.

Outra propriedade de que a direcção modificada goza é a de não se tornar ortogonal, sob certas condições, ao simétrico do vector gradiente, que é uma situação frequente no método de Newton. Notemos que, atendendo à definição de direcção do tipo gradiente, a prova deste facto é feita de duas formas distintas. A primeira usando a noção de direcção do tipo gradiente e a segunda mostrando que o co-seno do ângulo formado pelos dois vectores nunca está numa vizinhança de zero.

5.2 Trabalho Futuro

Após a realização deste trabalho há alguns pontos que merecem atenção futura.

O mais interessante seria a caracterização do método de Newton modificado como uma estratégia de regiões de confiança. Para tal há que estudar a relação entre as direcções geradas pelo método em iterações sucessivas e estabelecer o algoritmo modificado como um algoritmo de regiões de confiança.

Um outro ponto a abordar seria o da convergência global do método de Newton modificado recorrendo ao teorema de Ortega e Rheinboldt [10, pp495]. Para a aplicação deste resultado há que analisar o factor amortecimento, o escalar α e provar que $\lim_{k\to+\infty} \frac{\nabla f(x_k)^T p_k}{\|p_k\|} = 0$. Este teorema é mais fraco que o teorema da convergência global provado neste trabalho no sentido em que exige que f tenha um único ponto de estacionaridade em D. No entanto, o teorema de Ortega e Rheinboldt [*ibidem*] não impõe restrições à matriz dos coeficientes $B(x) + w(x) I$.

Ainda dentro da convergência global do método, uma reflexão e uma análise mais detalhada poderiam conduzir à definição de condições que garantissem uma ordem de convergência (global) superior à verificada neste trabalho.

Bibliografia

[1] J. E. Dennis e R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Inc., New Jersey, 1983.

[2] E. M. G. P. Fernandes, *Computação Numérica*, 2^a edição, Universidade do Minho, Braga, 1998.

[3] E. M. G. P. Fernandes, *Newton Based Exact Penalty Techniques for Nonlinear Optimization with Constraints* em U. Derigs, A. Bachem e A. Dreul (editores), Operations Research Proceedings 1994, Springer-Verlag, Berlim, 1995, (pp 39-44).

[4] R. Fletcher, *Practical Methods of Optimization*, 2^a edição, John Wiley & Sons Ltd, Chichester, 1991.

[5] P. E. Gill, W. Murray e M. H. Wright, *Pratical Optimization*, Academic Press, Inc., London, 1984.

[6] J. Jahn, *Introduction to the Theory of Nonlinear Optimization*, Springer-Verlag, Berlim, 1994.

[7] H. Lutkepohl, *Handbook of Matrices*, John Wiley & Sons Ltd, Chichester, 1996.

[8] M. T. Monteiro e E. M. G. P. Fernandes, *Local Convergence Analysis of a Modified Newton's Method*, pré-publicação interna, Braga, 1998.

[9] J. Nocedal, *Theory of Algorithms for Unconstrained Optimization*, em Acta Numerica 1992, University Press, Cambridge, 1992, (pp199-242).

[10] J.M. Ortega e W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, Inc., New York, 1970.

[11] M. J. D. Powell, *Convergence Properties of a Class of Minimization Algorithms* em O. L. Mangasarian, R. R. Meyer e S. M. Robinson (editores), Nonlinear Programming 2, Academic Press, New York, 1975, (pp 1-27).

[12] M. J. D. Powell, *On the Global Convergence of Trust Region Algorithms for Unconstrained Minimization*, Mathematical Programming, Vol. 29, 1984, (pp 297-303).

[13] G. A. Shultz, R. B. Schnabel e R. H. Byrd, *A Family of Trust-Region-Based Algorithms for Unconstrained Minimization With Strong Global Convergende Propreties*, SIAM Journal of Numerical Analysis, Vol 22, 1985, (pp 47-67).

[14] L. N. Vicente, *A Comparison Between Line Searches and Trust Regions for Nonlinear Optimization*, Investigação Operacional, Vol. 16, N^o 2, Dezembro 1996, (pp 173-179).

[15] M. A. Wolfe, *Numerical Methods for Unconstrained Optimization*, Van Nostrand Reinholds Company, 1978.