

MÉTODOS DE REGIÃO DE CONFIANÇA EM OTIMIZAÇÃO

Andrés Alfonso Carrillo López

TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS PROGRAMAS DE PÓS-GRADUAÇÃO EM ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

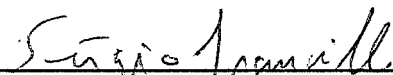
Aprovada por:



Nelson Maculan Filho, D.Sc.
(Presidente)



Paulo Roberto de Oliveira, Dr.Ing.



Sérgio Granville, Ph.D.

RIO DE JANEIRO, RJ - BRASIL

JUNHO DE 1990

CARRILLO, ANDRES ALFONSO

Métodos de Região de Confiança em Otimização (Rio de Janeiro) 1990.

VIII, 102 p. 29,7 cm (COPPE/UFRJ, M.Sc., Engenharia de Sistemas e Computação, 1990)

Tese-Universidade Federal do Rio de Janeiro, COPPE.

1. Métodos de Região de Confiança I. COPPE/UFRJ

II. Título (série).

A María Eugenia y a
nuestras hijas: Eugenia María y
Andrea Luisa.

AGRADECIMIENTOS

En los momentos de agradecer a todas las personas que de una u otra forma hicieron posible que llegara este momento de culminación de un trabajo difícil y largo; mi pensamiento se vuelca primeramente a mi querida esposa y a mis hijitas, por el cariño, comprensión y estímulo que siempre me brindaron. A mis colegas del Instituto Profesional de Santiago de Chile (IPS), que trabajaron sobrecargando sus horarios de trabajo para que yo me perfeccionara. A todos los profesores de la COPPE/UFRJ que me enseñaron y mostraron otros caminos de la Matemática Aplicada; destacando a mi profesor guía Nelson Maculan F., quién siempre me apoyó en la realización de este trabajo. A mi co-orientador Rafael Correa F., quién siempre tuvo fé en mi esfuerzo. A miscompañeros de curso ("turma del 88"), con quienes llegué a tener una amistad que ha crecido en el tiempo. A mi amigo Fernando Paredes, de quién siempre recibí solidaridad y estímulo en los buenos y malos momentos.

Rio de Janeiro, Junio de 1990.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

MÉTODOS DE REGIÃO DE CONFIANÇA EM OTIMIZAÇÃO

Andrés Alfonso Carrillo López

Junho de 1990

Orientador: *Nelson Maculan Filho.*

Co-Orientador: *Rafael Correa Fontecilla.*

Programa: *Engenharia de Sistemas e Computação.*

RESUMO

Neste trabalho, apresentamos os Métodos de Região de Confiança (para Otimização irrestrita); começando com uma evolução histórica do tema, continuamos com uma revisão das propriedades das Funções Quadráticas; para, nos restantes capítulos, desenvolver in-extenso os Métodos, dando ênfase a algumas demonstrações e exemplos ilustrativos que ajudarão a uma melhor compreensão do tema.

Resumen de la Tesis presentada a la COPPE/UFRJ como parte de los requisitos necesarios para la obtención del grado de Magister en Ciencias

METODOS DE REGION DE CONFIANZA EN OPTIMIZACION

Andrés Alfonso Carrillo López

Junio 1990

Orientador: *Nelson Maculan Filho*

Co-orientador: *Rafael Correa Fontecilla*

Programa: *Ingeniería de Sistemas y Computación*

RESUMEN

Se presentan en este trabajo Métodos de Región de Confianza (para Optimización sin restricciones); partiendo con una evolución histórica del tema, continuamos con una revisión de las propiedades de las Funciones Cuadráticas: para en los restantes Capítulos, desarrollar in-extenso los Métodos; dando énfasis a algunas demostraciones y ejemplos ilustrativos que servirán de ayuda para una mejor comprensión del tema.

Abstract of Thesis presented to COPPE/UFRJ as partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

THE TRUST REGION METHODS IN OPTIMIZATION

Andrés Alfonso Carrillo López

June 1990

Chairman: *Nelson Maculan Filho*

Co-Chairman: *Rafael Correa Fontecilla*

Department: *Systems Engineering and Computer Science*

ABSTRACT

This work is concerned with the Trust Region Methods in Unconstrained Optimization. We start with a review on the subject and a study of the properties of Quadratics Functions. Then, we extensively develop the method, with emphasis in some proofs and illustrating examples which should help in a better understanding of the area.

ÍNDICE

<u>CAPÍTULO I</u>	- INTRODUÇÃO	1
<u>CAPÍTULO II</u>	- ALGUNS ANTECEDENTES HISTÓRICOS.....	4
<u>CAPÍTULO III</u>	- PRELIMINARES MATEMÁTICAS.....	11
III.1	- Definições e teoremas prévios.....	11
III.2	- Propriedades das funções quadráticas..	17
<u>CAPÍTULO IV</u>	- MÉTODO DE REGIÃO DE CONFIANÇA.....	24
IV.1	- Descrição e idéias gerais.....	24
IV.2	- Algoritmo Geral.....	29
IV.3	- Resultados sobre convergência do Algoritmo.....	33
<u>CAPÍTULO V</u>	- ESTRATÉGIAS PARA O CÁLCULO DO PASSO...	41
V.1	- Passo Localmente Restrito ("HOOK")....	41
V.2	- O Passo Dogleg e Duplo Dogleg.....	52
V.3	- Atualização da Região de Confiança....	62
<u>CAPÍTULO VI</u>	- SOBRE UMA EXTENSÃO DO MÉTODO GERAL DE REGIÃO DE CONFIANÇA (REFERIDO A ESTRATÉGIAS DE BUSCA DO PASSO PARA MATRIZ HESSIANA DEFINIDA NEGATIVA).....	73
<u>CAPÍTULO VII</u>	- DESCRIÇÃO E EXEMPLO DO ALGORITMO DE REGIÃO DE CONFIANÇA.....	90
<u>REFERÊNCIAS BIBLIOGRÁFICAS.</u>	101

CAPITULO I

INTRODUÇÃO

Os Métodos chamados de "Região de Confiança" (TRUST-REGION METHODS) são de recente aparecimento em Análise Numérica. Na maioria dos Algoritmos de Otimização se determina primeiramente uma DIREÇÃO DE DESCIDA, ao longo da qual se faz uma busca unidimensional (chamada na literatura: "busca linear") com o objetivo de encontrar uma aproximação do mínimo nesta direção. A seguir, esta aproximação é considerada como o novo ponto de partida. Nesta categoria, aparecem os algoritmos do tipo Gradiente Conjugado e aqueles do tipo Quasi-Newton clássico.

A idéia da iteração dada pelo Método de Região de Confiança é a seguinte: Contrõe-se no ponto x_k um modelo (aqui considerado quadrático) da função objetivo e define-se, ao redor desse ponto x_k , uma bola, na qual "acredita-se" que este modelo seja uma aproximação adequada da função (ou funcional). Essa bola é chamada Região de Confiança. Logo, considera-se um passo s_k que garanta um decréscimo suficiente do modelo quadrático na interseção desta Região de Confiança e o conjunto viável. A função objetivo é avaliada neste novo ponto. Se seu valor decresceu o suficiente, o novo ponto é aceito como o seguinte iterado e se aumenta o raio da Região de Confiança. Caso contrário, o novo ponto é descartado e o raio da Região de Confiança é reduzido.

A diferença fundamental com os chamados Métodos de Busca linear, está no fato de como eles utilizam o modelo quadrático para a eleição do comprimento do passo. Assim, enquanto na "busca linear", o modelo quadrático é usado para obter a direção de busca minimizando o modelo quadrático; no caso do Método de Região de Confiança, escolhe-se primeiro um passo de prova de comprimento Δ_c e a seguir se usa o modelo quadrático para selecionar o melhor passo com esse comprimento. Para isso se resolve o problema:

$$P: \quad \text{Mín } m_c(x_c + s) = f(x_c) + g_c^t s + (1/2)s^t H_c s$$

$$s \in \mathbb{R}^n$$

$$\text{s. a. } \|s_c\| \leq \Delta_c$$

onde g_c é o gradiente da função f avaliada no ponto x_c e H_c é a matriz Hessiana definida no ponto x_c .

A idéia aqui é que o passo de prova Δ_c seja considerado um estimador de até onde podemos "confiar" no modelo quadrático.

Assim, depois que s_c foi encontrado (sendo s_c a solução do problema de minimização (P)), o Método de Região de Confiança efetua uma avaliação de $f(x_c + s_c)$ para ver se $(x_c + s_c)$ é satisfatório ou não. O que foi expresso anteriormente de forma esquemática é tratado com certo detalhe no Capítulo IV e parte do Capítulo V deste trabalho.

Como resolver numericamente o problema (P)? Para facilitar sua resolução, efetuam-se duas relaxações, a saber:

i) Fazendo que a restrição respeito da Região de Confiança só possa ser cumprida de forma aproximada, isso dá origem aos chamados Métodos de passo localmente restringido ("HOOK"), e

ii) O modelo quadrático pode ser minimizado só aproximadamente, o que dá origem aos chamados Métodos "dogleg" e "double dogleg".

Ambas relaxações são tratadas no Capítulo V (especificamente nas seções V.1 e V.2).

Os Métodos de Região de Confiança são atrativos, já que trata-se de um grupo de métodos que asseguram convergência global e na literatura matemática existem numerosos teoremas

deste tipo. Isso faz com que se possa esperar uma excelente robustez dos algoritmos existentes para estes métodos (no sentido que o algoritmo continue trabalhando bem; ainda que os dados sejam "ruins", assim por exemplo: com uma matriz mal condicionada, o algoritmo continue fornecendo bons resultados).

As definições e teoremas requeridos para justificar convergência nos algoritmos que se apresentam são expostos no Capítulo III.

No Capítulo VI, como indica seu nome, se estuda uma extensão do Método de Região de Confiança, especialmente referido a mostrar estratégias de busca do passo quando a matriz Hessiana é definida negativa. Tudo isso além de alguns exemplos desenvolvidos com certo detalhe. No último Capítulo se apresenta, a modo de resumo, uma descrição com um exemplo do Algoritmo de Região de Confiança.

CAPITULO II

ALGUNS ANTECEDENTES HISTÓRICOS.

O desenvolvimento dos Métodos de Região de Confiança começa a tomar forma com os trabalhos de LEVENBERG(1944) y MARQUARDT (1963); no que diz respeito a problemas de Mínimos Quadráticos não-lineares.

A idéia é que dada uma função continuamente diferenciada $F: \mathbb{R}^n \longrightarrow \mathbb{R}^m$; o problema de Mínimos Quadráticos não linear é encontrar o mínimo da função:

$$\| F(x) \|_2 = \left[\sum_{i=1}^m f_i(x)^2 \right]^{1/2}$$

O Algoritmo de Levenberg-Marquardt pode ser visto como um método para gerar uma sucessão de iterados $\{x_k\}$ onde o passo s_k entre os iterados é uma solução para o problema:

$$\min \left\{ \|F(x_k) + F'(x_k)w\|_2 : \|D_k w\| \leq \Delta_k \right\} \quad (\text{II.1})$$

para uma variação Δ_k e uma matriz de escala D_k . A norma $\|\cdot\|$ que aparece neste problema é arbitrária, mas geralmente se escolhe como norma l_2 , já que com isto MARQUARDT (1963) provou que se o passo s_k se determina resolvendo o sistema linear:

$$\left[F'(x_k)^t F'(x_k) + \lambda_k D_k^t D_k \right] s_k = -F'(x_k)^t F(x_k)$$

para um parâmetro $\lambda_k > 0$ então s_k resolve o problema (II.1) com $\Delta_k = \|D_k s_k\|$.

A idéia fundamental do Método de Região de Confiança, é que o passo s_k seja a solução de um subproblema, com uma variação sobre o passo. O subproblema é escolhido de forma que a solução s_k melhore a aproximação do problema de

Otimização (em questão). -

Na versão do Método de Newton feita por GOLDFELD, QUANDT e TROTTER (1966), para a minimização sem restrições de uma função $f: \mathbb{R}^n \longrightarrow \mathbb{R}$, o passo s_k resolve um subproblema da forma:

$$\min \left\{ \nabla f(x_k)^t w + (1/2) w^t \nabla^2 f(x_k) w ; \| D_k w \| \leq \Delta_k \right\} \quad (\text{II.2})$$

para alguma cota Δ_k e alguma matriz de escala D_k . Os autores anteriormente mencionados, escolheram como norma $\| \cdot \|$, a norma l_2 e provaram que: se s_k satisfaz:

$$\left[\nabla^2 f(x_k) + \lambda_k D_k^t D_k \right] s_k = - \nabla f(x_k)$$

para um parâmetro $\lambda_k \geq 0$, com $\left[\nabla^2 f(x_k) + \lambda_k D_k^t D_k \right]$, matriz semi-definida positiva, então s_k , resolve o problema de minimização (II.2), com $\Delta_k = \| D_k s_k \|$.

Os algoritmos de Goldfeld-Quandt-Trotter e de Levenberg-Marquardt são bastante parecidos. Em ambos, o subproblema tem a forma:

$$\min \left\{ \psi_k(w) : \| D_k w \| \leq \Delta_k \right\} \quad (\text{II.3})$$

onde a função $\psi_k(w)$ é uma aproximação quadrática da real redução na função objetivo. No algoritmo de Levenberg-Marquardt, $\psi_k(w) - \psi_k(0)$ é uma aproximação de primeiro ordem, embora no Algoritmo de Goldfeld-Quandt-Trotter, $\psi_k(w) - \psi_k(0)$ seja uma aproximação de segunda ordem. A diferença é que no Algoritmo de Levenberg-Marquardt, a função aproximação quadrática ψ_k é convexa e, portanto, é relativamente fácil se obter a solução global do problema (II.3).

As primeiras versões, tanto do Algoritmo de Levenberg-Marquardt, quanto do Algoritmo de Goldfeld-

Quandt-Trotter, controlavam indiretamente s_k ; alterando o parâmetro $\lambda_k (\geq 0)$. Isto não parece ser compatível com a filosofia dos Métodos de Região de Confiança, que requer um controle direto da cota Δ_k e determina s_k como a solução do subproblema (II.3). O controle direto de λ_k tem vários inconvenientes. Um deles é que não parece razoável eleger automaticamente um λ_0 . Por outro lado um valor razoável de Δ_0 é freqüentemente uma pequena fração do tamanho $\|D_0 x_0\|$ do ponto de partida. Outro problema ocorre quando $x_k + s_k$ conduz a um aumento da função objetivo.

Neste caso se pode usar a informação obtida da função e de sua derivada em x_k e $(x_k + s_k)$ para estimar o decréscimo requerido em Δ_k , mas não está claro que esta informação possa ser usada para estimar um valor razoável para $\lambda_{k+1} > \lambda_k$.

Assim se temos, por exemplo:

$$\phi(\tau) = \|F(x_k + \tau s_k)\|_2$$

então; no Algoritmo de Levenberg-Marquardt já conhecemos: $\phi(0)$, $\phi(1)$ y $\phi'(0) < 0$. Se ocorre :

$$\phi(1) > \phi(0)$$

então a quadrática que interpola estes dados tem um mínimo $\gamma < 1$, e é razoável deixar $\Delta_{k+1} = \gamma \|D_k s_k\|$ na maioria dos casos.

Seguindo a idéia do Algoritmo de Goldfeld-Quandt-Trotter, HEBDEN (1973) foi o primeiro a propor um algoritmo razoável para a solução aproximada do problema (II.3). Seu algoritmo é basicamente bom, porém apresenta dificuldades quando ψ_k não é convexa.

MORE (1978) apresentou resultados teóricos e numéricos que demonstraram que as idéias de Hebden podiam ser usadas para realizar uma implementação confiável e eficiente do Algoritmo de Levenberg-Marquardt. Esta implementação, em

particular, não tem muitas das decisões ad-hoc das implementações prévias.

Os Algoritmos de Hebden e Moré não buscavam uma solução exata do problema (II.3), ao invés disto, eles se satisfaziam com uma solução ótima aproximada do problema (II.3). Resultados numéricos e teóricos seguros podem ser obtidos se o passo s_k , satisfaz a desigualdade:

$$\psi_k(s_k) \leq \beta_1 \min \left\{ \psi_k(w) : \|D_k w\| \leq \Delta_k \right\} ; \|D_k s_k\| \leq \beta_2 \Delta_k \quad (\text{II.4})$$

sendo β_1 e β_2 constantes positivas.

Para muitos problemas, o custo com a obtenção do passo s_k cumprindo as condições anteriores é bastante razoável, porém em outro tipo de problemas (como por exemplo, problemas de grandes dimensões), o custo pode ser proibitivo. Uma alternativa, neste último caso, está em determinar o passo s_k que resolve o problema com restrições da forma:

$$\min \left\{ \psi_k(w) : w \in S_k ; \|D_k w\| \leq \Delta_k \right\} \quad (\text{II.5})$$

sendo S_k um subespaço. A vantagem em resolver este subproblema, é que a solução de (II.5) pode ser consideravelmente menos custosa que a do subproblema (II.3). E ainda mais, se S_k é escolhido adequadamente, podem ser obtidos bons resultados, tanto teóricos como práticos.

Resultados de convergência podem ser obtidos se:

$$\psi_k(s_k) \leq \beta_1 \min \left\{ \psi_k(w) : w = v s_k^G ; \|D_k w\| \leq \Delta_k \right\} ; \|D_k s_k\| \leq \beta_2 \Delta_k \quad (\text{II.6})$$

com $\beta_1 > 0$ y $\beta_2 > 0$ e onde s_k^G é a direção de descida mais

profunda para ψ_k no que respeito a norma $\|D_k(\cdot)\|$ (vcte. >0)

POWELL (1970) foi o primeiro a propor um Método de Região de Confiança "acompanhado" de um subproblema do tipo (II.5). No Algoritmo de Powell para resolver um sistema de equações algébricas não-lineares, $F: \mathbb{R}^n \longrightarrow \mathbb{R}^n$; o passo s_k resolve um problema da forma:

$$\min \left\{ \|F(x_k) + J_k w\|_2 : w = \alpha s_k^c + (1 - \alpha) s_k^N ; \|D_k w\| \leq \Delta_k \right\} \quad (\text{II.7})$$

onde s_k^c é o passo de Cauchy, s_k^N é o passo Newton, e a matriz J_k é ou a aproximação por diferenças da matriz jacobiana $F'(x_k)$ ou uma versão escalada (ou a escala) da aproximação quase-Newton obtida usando-se a atualização de BROYDEN (1965).

Os Métodos chamados de Região de Confiança foram também desenvolvidos para resolver problemas de Otimização com restrições lineares do tipo:

$$\min \left\{ f(x) : c^t x \leq d \right\} \quad (\text{II.8})$$

no qual o subproblema associado tem a forma:

$$\min \left\{ \nabla f(x_k)^t w + \frac{1}{2} w^t B_k w : c^t (x_k + w) \leq d ; \|D_k w\| \leq \Delta_k \right\} \quad (\text{II.9})$$

sendo B_k uma matriz simétrica. Fletcher sugeriu que a norma $\|\cdot\|$ fosse escolhida como a norma l_∞ já que assim o subproblema (II.9) é um problema de programação quadrática. O Algoritmo de Fletcher é uma generalização do Algoritmo de GRIFFITH y STEWART (1961) na qual $B_k=0$ e a norma $\|\cdot\|$ é a norma l_∞ . A vantagem de escolher $B_k=0$ está em que o subproblema (II.9) se transforma em um problema de programação linear. Deve-se, contudo, notar que a escolha de $B_k=0$ pode conduzir a uma razão de convergência

inaceitavelmente baixa.

Até aqui sempre dissemos que a função ψ_k é uma função do tipo quadrática, muito embora nem sempre ocorra assim. MASDEN (1975) desenvolveu um Algoritmo para a solução minimax de um sistema de equações não-lineares $F: \mathbb{R}^n \longrightarrow \mathbb{R}^m$. O problema minimax requer averiguar o mínimo da função:

$$\| F(x) \|_{\infty} = \text{Máx} \left\{ | f_i(x) | : 1 \leq i \leq m \right\}$$

Em seu Algoritmo, o passo s_k , resolve o subproblema:

$$\text{mín} \left\{ \| F(x_k) + F'(x_k)w \|_{\infty} : \| D_k w \| \leq \Delta_k \right\} \quad (\text{II.10})$$

A norma $\| \cdot \|$ é arbitrária, entretanto Madsen usou a norma l_{∞} já que assim o problema (II.10) pode ser formulado como um problema de programação linear. Note que o subproblema (II.10) é da forma do subproblema (II.3), com ψ_k : função poliedral convexa.

FLETCHER (1981) estendeu o Algoritmo de Madsen ao Problema de Otimização Não-Diferenciável:

$$\text{Mín} \left\{ f(x) + \phi(F(x)) : x \in \mathbb{R}^n \right\} \quad (\text{II.11})$$

onde as funções $f: \mathbb{R}^n \longrightarrow \mathbb{R}$ e $F: \mathbb{R}^n \longrightarrow \mathbb{R}^m$ são diferenciáveis; porém a função $\phi: \mathbb{R}^m \longrightarrow \mathbb{R}$ se supõe somente convexa. Assim os subproblemas no Algoritmo de Fletcher são da forma (II.3), com:

$$\psi_k(w) = \nabla f(x_k)^t w + \frac{1}{2} w^t B_k w + \phi \left[F(x_k) + F'(x_k)w \right]$$

sendo B_k uma matriz simétrica. Nestes subproblemas, a função ψ_k é a soma de uma função quadrática, com uma função geral convexa.

O problema de Otimização não-diferenciável (II.11) inclui problemas de minimização diferenciável quando $\phi = 0$ e problemas de minimização convexa quando $f = 0$.

Recentemente ZOWE (1988) estuda a minimização de uma função convexa f , não necessariamente diferenciável, sujeita a restrições lineares. Ele utiliza planos de corte para modelar f e a minimização deste modelo dá uma direção para a qual se efetua uma busca linear. Esta aproximação prova ser muito boa, entretanto seu êxito depende de uma cuidadosa maneira de escolher certos parâmetros ξ_k em cada iteração. Os mencionados ξ_k podem ser interpretados como os raios ρ_k de uma bola na qual se acredita que f está modelada de maneira suficientemente boa. Infelizmente, ainda é um mistério como escolher estes ξ_k e ρ_k inteligentemente. Para suprir esta dificuldade, ZOWE utiliza as idéias do Método de Região de Confiança. Para isso, em vez de fixar algum ρ no ponto de iteração x_k , se resolve o modelo para vários ρ , tratando de ajustar sistematicamente ρ (e simultaneamente o modelo) até encontrar o raio ρ da bola, no qual o modelo se aproxima bem de f . A minimização do modelo sobre esta bola, origina um novo iterado, com um decréscimo substancial em f .

CAPITULO III
PRELIMINARES MATEMÁTICAS

III.1. DEFINIÇÕES E TEOREMAS PRÉVIOS

DEFINIÇÃO 1: Uma sucessão $\{x_k\}$ converge linearmente a x^* se:

$$\|x_{k+1} - x^*\| < \alpha \|x_k - x^*\|; k \geq 0 \quad (\text{III.1})$$

para uma constante $\alpha \in] 0, 1 [$

Observação 1: Se α é pequena, então (III.1) é adequada para um algoritmo de Otimização razoável.

Se α se aproxima da unidade ($\alpha \geq 0.9$) então (III.1) é inadequado (para um algoritmo de otimização razoável).

DEFINIÇÃO 2: Uma sucessão $\{x_k\}$ converge quadraticamente a x^* se:

$$\|x_{k+1} - x^*\| \leq \beta \|x_k - x^*\|^2; k \geq 0 \quad (\text{III.2})$$

para alguma constante $\beta > 0$.

Observação 2: Como

$$\frac{\|x_{k+1} - x^*\|}{\|x^*\|} \leq \beta \|x^*\| \left(\frac{\|x_k - x^*\|}{\|x^*\|} \right)^2$$

a convergência quadrática implica que o número de dígitos significativos de x_k , como uma aproximação a x^* , é o dobro em cada iteração.

Existe um meio termo entre as definições anteriores (e suas correspondentes desigualdades). Assim tem-se, a seguinte definição:

DEFINIÇÃO 3: Uma sucessão $\{x_k\}$ converge superlinearmente a x^* se:

$$\|x_{k+1} - x^*\| \leq \beta_k \|x_k - x^*\| ; k \geq 0$$

para alguma sucessão $\{\beta_k\}$ para qual converge a 0.

Observação 3: Uma sucessão convergente superlinearmente é linearmente convergente. Além disto uma sucessão que converge quadraticamente é superlinearmente convergente.

Observação 4: Como:

$$| \|x_{k+1} - x_k\| - \|x_k - x^*\| | \leq \|x_k - x^*\|$$

tem-se que:

$$\lim_{k \rightarrow \infty} \frac{\|x_{k+1} - x_k\|}{\|x_k - x^*\|} = 1$$

quando $\{x_k\}$ converge superlinearmente a x^* . Esta é uma propriedade importante já que implica que $\|x_{k+1} - x_k\|$ pode ser usado para estimar a distância entre x_k e x^* .

Para entender melhor as propriedades dos mínimos locais, nos referiremos aqui especificamente a classe de funções f que sejam duas vezes continuamente diferenciáveis. Desta maneira as propriedades dos mínimos locais podem ser escritas em termos da função quadrática:

$$\psi(w) = \nabla f(x)^t w + (1/2) w^t \nabla^2 f(x) w \quad (\text{III.3})$$

onde como sabemos, $\nabla f(x)$ representa o gradiente da função f no ponto x , e $\nabla^2 f(x)$ representa a matriz Hessiana da função f no ponto x .

Notas: i) A i -ésima componente do gradiente se denota por

$$\partial_i f(x).$$

ii) O elemento (i,j) da matriz Hessiana se denota

$$\text{por } \partial_{i,j} f(x).$$

Como:

$$f(x+w) = f(x) + \psi(w) + o(\|w\|^2) \quad (\text{III.4})$$

a função quadrática ψ é o modelo quadrático local em x da possível redução em f .

TEOREMA 1: Seja $f: \mathbb{R}^n \rightarrow \mathbb{R}$ duas vezes continuamente diferenciável em um conjunto aberto D . Se $x^* \in D$ é um mínimo local de f , então $\nabla f(x^*) = 0$ e $\nabla^2 f(x^*)$ é semi-definida positiva. Se $\nabla f(x^*) = 0$ e $\nabla^2 f(x^*)$ é definida positiva para algum $x^* \in D$ então x^* é um mínimo local separado de f .

Demonstração: (\Rightarrow)

i) Seja ψ o modelo quadrático local em x^* da possível redução de f :

Se x^* é um mínimo de f então tem-se (usando (III.4)) que: $\forall p \in \mathbb{R}^n$ e α suficientemente pequeno:

$$0 \leq \psi(\alpha p) + o(\|\alpha\|^2) = \alpha \nabla f(x^*)^t p + \alpha^2 / 2 p^t \nabla^2 f(x^*) p + o(\alpha^2)$$

ii) A expressão anterior implica que $\nabla f(x^*)^t p = 0$ e além disso que:

$$p^t \nabla^2 f(x^*) p \geq 0.$$

Como p é arbitrário, conclue-se necessariamente que $\nabla f(x^*)$ é igual a 0 e que $\nabla^2 f(x^*)$ é semidefinida positiva.

(\Leftarrow)

i) Se $\nabla f(x^*) = 0$ e $\nabla^2 f(x^*)$ é definida positiva, então, tem-se que :

$$\psi(w) = (1/2) w^t \nabla^2 f(x^*) w \quad (\text{III.5})$$

Seja $\lambda > 0$ o menor valor próprio da matriz Hessiana $\nabla^2 f(x^*)$ então em (III.5) tem-se que:

$$\psi(w) = (1/2) w^t \nabla^2 f(x^*) w \geq \lambda \|w\|^2 > 0$$

ii) Logo, tem-se que x^* deve ser um mínimo local separado para a função f (ver em (III.4)). -

DEFINIÇÃO 4: Um ponto $x^* \in \mathbb{R}^n$ tal que $\nabla f(x^*) = 0$ é chamado PONTO CRÍTICO de f .

DEFINIÇÃO 5: Os pontos críticos podem ser: mínimos locais, máximos locais ou pontos sela.

OBSERVAÇÃO: Os algoritmos para minimização irrestrita de uma função $f: \mathbb{R}^n \longrightarrow \mathbb{R}$, são geralmente Métodos de descida que levam a cabo a seguinte idéia básica:

Partindo de um ponto inicial x_0 , o método gera uma sucessão de aproximações $\{x_k\}$ ao mínimo local com a propriedade que:

$$f(x_{k+1}) < f(x_k) ; \forall k \geq 0$$

Somente a condição de descida não garante suficientemente para que a sucessão de iterações $\{x_k\}$ convirja ao mínimo local. São requeridas condições mais fortes (por exemplo, sobre o gradiente e o Hessiano de f) para obrigar a sucessão $\{x_k\}$ a estar vizinha ao mínimo local, e uma vez que as iterações x_k estão na referida vizinhança, os métodos de descida admitem uma rápida convergência local.

Neste sentido vejamos o seguinte teorema:

TEOREMA 2: Seja $F: \mathbb{R}^n \longrightarrow \mathbb{R}^m$ continuamente diferenciável, definida em um conjunto aberto D e suponhamos que $F(x^*) = 0$ para algum $x \in D$, e $F'(x^*)$ não é singular. Existe então um conjunto S tal que para qualquer $x_0 \in S$, as iterações de Newton:

$$x_{k+1} = x_k - F'(x_k)^{-1} F(x_k) \quad (\text{III.6})$$

$k \geq 0$, estão bem definidas, seguem em S e converge a x^* .

Demostração: -Seja α ; uma constante $\in]0, 1[$.

-Como F' é contínua em x^* e $F'(x^*)$ é não singular; existe uma bola aberta S tal que:

$$S \equiv \left\{ x : \|x - x^*\| < \varepsilon \right\}$$

e uma constante $\mu > 0$, tal que $\forall x, y \in S$

$$\|F'(x)^{-1}\| \leq \mu ; \|F'(y) - F'(x)\| \leq \frac{\alpha}{\mu}$$

-Suponhamos que $x_k \in S$. Como x_{k+1} satisfaz (II.6) e $F(x^*)=0$ tem-se que:

$$\begin{aligned} x_{k+1} - x^* &= x_k - F'(x_k)^{-1}F(x_k) - x^* \\ &= -F'(x_k)^{-1} \left[F(x_k) - F(x^*) - F'(x_k)x_k + F'(x_k)x^* \right] \\ &= -F'(x_k)^{-1} \left[F(x_k) - F(x^*) - F'(x_k)(x_k - x^*) \right] / \| \| \end{aligned}$$

$$\|x_{k+1} - x^*\| \leq \|F'(x_k)^{-1}\| \|F(x_k) - F(x^*) - F'(x_k)(x_k - x^*)\| \quad (\text{III.7})$$

$$\leq \mu$$

-Porém, do Teorema Fundamental do Cálculo tem-se que:

$$F(x_k) - F(x^*) - F'(x_k)(x_k - x^*) = \int_0^1 \left[F'(x^* + \xi(x_k - x^*)) - F'(x_k) \right] * (x_k - x^*) d\xi.$$

Logo em (III.7):

$$\|x_{k+1} - x^*\| \leq \mu \left\{ \text{Máx} \|F'(x^* + \xi(x_k - x^*)) - F'(x_k)\| \right\} \|x_k - x^*\|$$

$$\leq \alpha / \mu \quad (\text{III.8})$$

Logo:

$$\|x_{k+1} - x^*\| \leq \alpha \|x_k - x^*\| \quad (\text{III.9})$$

Como $\alpha < 1$, a desigualdade (III.9) implica que se $x_0 \in S$, então $x_k \in S$; para $k > 0$ e tem-se além disso que $\{x_k\}$ converge a x^* . ■■

Observação: O teorema 2 estabelece que o Método de Newton é localmente convergente, no sentido que se o ponto de partida

(ou prefixação inicial) está suficientemente próximo da solução x^* , então a sucessão $\{x_k\}$ gerada pelo Método de Newton, converge a x^* .

TEOREMA 3: Seja $F: \mathbb{R}^n \longrightarrow \mathbb{R}^n$ que satisfaça as hipóteses do Teorema 2. Então a sucessão $\{x_k\}$ definida pela iteração:

$$x_{k+1} = x_k - [F'(x_k)]^{-1} F(x_k); \text{ para } k \geq 0$$

$$\text{(ou seja: } x_{k+1} = x_k - [\nabla f(x_k)]^{-1} f(x_k), \text{ com } k \geq 0)$$

converge superlinearmente a x^* .

Além disto se:

$$\|F'(x) - F'(x^*)\| \leq \gamma \|x - x^*\|; \text{ com } x \in D \quad (\text{III.10})$$

para alguma constante $\gamma > 0$, então a sucessão $\{x_k\}$ converge quadraticamente a x^* .

Demonstração: i) A convergência da sucessão $\{x_k\}$ já está demonstrada no Teorema 2.

ii) Vejamos agora a razão de convergência.

Para isto se define:

$$\beta_k = \mu \left\{ \text{Máx}_{0 \leq \xi \leq 1} \|F'(x^* + \xi(x_k - x^*)) - F'(x_k)\| \right\}$$

com μ ; constante > 0 , que aparece definida no Teorema 2 e onde $x_0 \in S$ com:

$$S \equiv \left\{ x: \|x^* - x\| \leq \varepsilon; \varepsilon > 0 \right\}$$

iii) A hipótese (III.10) e a convergência da sucessão $\{x_k\}$ a x^* asseguram que a sucessão $\{\beta_k\}$, converge a 0.

iv) Além disso, usando a desigualdade (III.8) que aparece no Teorema 2, tem-se que:

$$\|x_{k+1} - x^*\| \leq \beta_k \|x_k - x^*\| \quad (\text{III.11})$$

o que demonstra que a sucessão $\{x_k\}$ converge superlinearmente a x^* .

v) Por outro lado, utilizando (III.10) e como:

$$\beta_k = \mu \left\{ \max_{0 \leq \xi \leq 1} \|F'(x^* + \xi(x_k - x^*)) - F'(x_k)\| \right\} \leq \mu \max_{0 \leq \xi \leq 1} \gamma (1 + \xi) \|x^* - x_k\|$$

$$\leq \gamma \mu \max_{0 \leq \xi \leq 1} (1 + \xi) \|x^* - x_k\|$$

$$\geq 2$$

$$\leq 2 \mu \gamma \|x_k - x^*\|$$

$$> 0$$

Logo, substituindo em (III.11) tem-se:

$$\|x_{k+1} - x^*\| \leq 2 \mu \gamma \|x_k - x^*\|^2$$

o que significa que a sucessão $\{x_k\}$,
converge quadraticamente a x^* . ■■

III.2. - PROPRIEDADES DAS FUNÇÕES QUADRÁTICAS. -

As funções quadráticas têm um papel importante no desenvolvimento de Algoritmos aplicados a problemas de Otimização. Assim por exemplo, sabemos que em uma vizinhança de um mínimo local de uma função f , onde $f: \mathbb{R}^n \longrightarrow \mathbb{R}$, o Método de Newton pode ser obtido requerendo que o passo seja o mínimo do modelo quadrático local:

$$\psi_k(w) = \nabla f(x_k)^t w + (1/2) w^t \nabla^2 f(x_k) w \quad (\text{III.12})$$

da redução esperada em f . Logo é importante entender as propriedades das funções quadráticas e fornecer algoritmos numericamente estáveis para a minimização deles.

O primeiro resultado, descreve completamente a minimização irrestrita de funções quadráticas:

LEMA 1: Seja $\psi: \mathbb{R}^n \longrightarrow \mathbb{R}$; a função quadrática:

$$\psi(w) = g^t w + (1/2) w^t B w$$

onde $g \in \mathbb{R}^n$ e $B \in \mathbb{R}^{n \times n}$ é uma matriz simétrica.

- a) A função quadrática ψ tem um mínimo se e só se B é semidefinida positiva e g está no nível de B .
- b) A função quadrática ψ tem um único mínimo se e só se B é definido positivo.
- c) Se B é semidefinida positiva então cada solução da equação $Bp = -g$ é um mínimo global.

Demonstração: (Ver MORE, J. J. e SORENSEN D.C. (1981))

Dada uma função quadrática ψ , existe um excelente procedimento numérico para achar seu mínimo. Primeiro se tenta uma fatorização de Cholesky da matriz B . Esta fatorização existe se e só se B é semidefinida positiva e daí se obtém uma matriz triangular superior R , tal que:

$$B = R^t R$$

Observação: Se se encontra durante o processo de factorização, uma diagonal negativa, então B não é semidefinida positiva e o Lema 1 mostra que a função quadrática ψ não tem mínimo (negação da parte a) no Lema 1).

Se a factorização dá certo e R é não-singular, então o mínimo se calcula resolvendo o sistema $Bp = -g$ ou de forma equivalente:

$$R^t v = -g ; R p = v$$

Se a factorização dá certo, mas R é singular então B é semi-definida positiva e singular. E ainda é possível calcular uma solução p ; mas do ponto-de-vista numérico, este cálculo é instável, já que uma perturbação arbitrariamente pequena, pode transformar B em uma matriz definida positiva ou em uma matriz indefinida.

O Teorema 1 deste Capítulo, mostra que em uma vizinhança de um mínimo local de f , podemos esperar que a matriz Hessiana ($\nabla^2 f(x) = H$) seja definida positiva, e então

o Lema 1 mostra que o modelo quadrático local (III.12) tem um único mínimo. Assim, neste caso, o mínimo do modelo quadrático local é um caminho razoável para um algoritmo de minimização. Entretanto, longe de um mínimo local, a matriz Hessiana $\nabla^2 f(x)$, pode ter um valor próprio negativo e então o Lema 1 nos diz que este modelo quadrático local não tem um mínimo. De fato o modelo não está limitado inferiormente.

Existem alguns paleativos para esta dificuldade. Uma possibilidade é modificar o modelo quadrático, acrescentando uma matriz semi-definida positiva $E(x)$, tal que:

$$\nabla^2 f(x) + E(x)$$

seja definida positiva. (Quando B é substituída por esta matriz, o cálculo do passo pode efetuar-se como foi dito anteriormente).

Outro paleativo possível, e é aqui que centraremos este trabalho, é restringir a região na qual suporemos que o modelo quadrático local é apropriado. Localmente o modelo ainda provê uma excelente aproximação a redução que se espera em f , assim é razoável restringir ψ à bola $\{w: \|w\| \leq \Delta\}$, para algum $\Delta > 0$ e calcular o passo como o mínimo de ψ nesta bola.

O seguinte Lema caracteriza as soluções do problema de minimização de uma função quadrática na região que foi restringida:

LEMA 2: Seja $\psi: \mathbb{R}^n \rightarrow \mathbb{R}$ a função quadrática:

$$\psi(w) = g^t w + (1/2) w^t B w$$

e seja $\Delta > 0$ dado.

Um ponto $p \in \mathbb{R}^n$ resolve o problema:

$$\min \left\{ \psi(w) : \|w\| \leq \Delta \right\} \quad (\text{III.13})$$

se e somente se, existe $\lambda > 0$ tal que:

$$(B + \lambda I) p = -g \quad ; \quad \lambda \left[\Delta - \|p\| \right] = 0 \quad (\text{III.14})$$

com $B + \lambda I$ semidefinida positiva.

Demonstração: i) (\leftarrow)

Suponhamos que λ e p satisfazem (III.14) com $B + \lambda I$, semidefinida positiva. Então, pelo Lema 1 (parte (a)) tem-se que p minimiza a função quadrática:

$$\hat{\psi}(w) = g^t w + (1/2) w^t (B + \lambda I) w$$

Assim, $\hat{\psi}(p) \leq \hat{\psi}(w)$, quer dizer:

$$g^t p + (1/2) p^t (B + \lambda I) p \leq g^t w + (1/2) w^t (B + \lambda I) w$$

donde tem-se que:

$$g^t w + (1/2) w^t B w \geq g^t p + (1/2) p^t B p + (\lambda/2) (p^t p - w^t w) \quad (\text{III.15})$$

para todo $w \in \mathbb{R}^n$.

Como além, de (III.14) tem-se que:

$$\lambda p^t p = \lambda \Delta^2 \quad \text{e} \quad \lambda \geq 0,$$

segue de (III.15) que $\psi(w) \geq \psi(p)$, sempre que $\|w\| \leq \Delta$; assim p resolve o problema (III.13).

ii) (\rightarrow)

Suponhamos agora que p resolve o problema (III.13). Se $\|p\| < \Delta$, então p é um mínimo irrestrito de ψ , logo o Lema 1 implica que (III.14) se cumpre com $\lambda = 0$ e que B é semi-definida positiva. Se $\|p\| = \Delta$, então p também resolve o problema de minimização com restrições de igualdade:

$$\min \left\{ \psi(w) : \|w\| = \Delta \right\}$$

Portanto, o Método de Lagrange assegura a existência de λ tal que:

$$\nabla L(p) = 0, \quad \text{onde} \quad L(w) = \psi(w) + (\lambda/2) (w^t w - \Delta^2)$$

Isto implica que (III.14) se cumpre para λ e p .

Além disso, como por hipótese p resolve (III.13), tem-se que (III.15) é válida para λ e p , sempre que $\|w\| = \|p\|$. Usando (III.14), para substituir g e ordenando termos em (III.15), tem-se que:

$$(1/2) (w - p)^t (B + \lambda I) (w - p) \geq 0$$

para cada w , com norma $\|p\|$.

Desta última desigualdade tem-se que $(B + \lambda I)$ é semi-definida positiva.

Para demonstrar que $\lambda \geq 0$, notar que o Lema 1 implica que (III.15) é válido $\forall w \in \mathbb{R}^n$. Agora se $\lambda < 0$, então (III.15) implica que $\psi(w) \geq \psi(p)$ sempre que $\|w\| \geq \|p\|$. Como p resolve o problema (III.13), devemos saber que p é um mínimo irrestrito de ψ e do Lema 1 implica que $\lambda = 0$.

Logo, tem-se que $\lambda \geq 0$. - ■■

Observação: Para encontrar uma solução numérica aproximada do problema (III.13), é necessário ter um pouco de cuidado: Uma complicação imediata é que devido a restrição não-linear ($\|w\| < \Delta$) não existe um método geral direto para resolver o problema mencionado (III.13). De fato, quando $g=0$ é uma solução p de (III.13) deve ser um vetor próprio de norma Δ , que corresponde ao menor valor próprio de B . Logo, um método geral para resolver o problema (III.13) deve resolver um problema de valor próprio simétrico neste caso especial.

-A solução do problema (III.13) é direta se não existem soluções na fronteira da bola: $\{ w: \|w\| \leq \Delta \}$.

(ver MORE, J. J. e SORENSEN, D. C. (1981)).

-Se existe uma solução na fronteira de $\{ w: \|w\| < \Delta \}$ para o problema (III.13), então o Lema 2 mostra que é razoável esperar que a equação não-linear:

$$\| \underbrace{-(B + \alpha I)^{-1}g}_{p_\alpha} \| = \Delta \quad (\text{III.16})$$

tenha uma solução $\lambda \geq 0$ em $(-\lambda_1, \infty)$, onde λ_1 é o menor valor próprio da matriz B. -

Observação: Resolver a equação (III.16) é equivalente a achar um zero de um problema unidimensional em α , que pode ser resolvido (entre outros) pelo Método de Newton. Entretanto, notar que cada avaliação de p_α necessita da solução de um sistema de equações lineares, o que é importante para resolver (III.16) com muito poucas avaliações de p_α .

Para resolver (III.16), REINSCH (1967,1971) e HEBDEN (1973), sugeriram que era muito mais eficiente aplicar à função o Método de Newton:

$$\phi_1(\alpha) = \frac{1}{\Delta} - \frac{1}{\|p_\alpha\|}$$

Esta função $\phi_1(\alpha)$ não tem extremos e é quase linear perto da solução de (III.16).

A iteração de Newton aplicada para achar um 0 de $\phi_1(\alpha)$ é a seguinte:

ALGORITMO 1:

- 1) Sejam $\lambda_0 > 0$ e $\Delta > 0$ dados.
- 2) Para $k = 0, 1, 2, \dots$, até "que haja convergência":

a) Fatorar $B + \lambda_k I = R_k^t R_k$.

b) Resolver: $R_k^t R_k p_k = -g$.

c) Resolver: $R_k^t q_k = p_k$.

d) $\lambda_{k+1} = \lambda_k + \left[\frac{\|p_k\|}{\|q_k\|} \right]^2 \left[\frac{\|p_k\| - \Delta}{\Delta} \right]$

Observações: i) Este algoritmo pode ser usado com algumas precauções (por exemplo, a respeito da matriz B) para resolver em muitos casos o problema (III.13)

$$\left\{ \min \{ \psi(w) : \|w\| \leq \Delta \} \right\}$$

ii) Quando B é indefinida, existem casos nos quais a equação (III.16) não tem solução em $(-\lambda_1, \infty)$ (logo o algoritmo anterior não funciona). Isto acontece, por exemplo: quando $g = 0$ [aí $\phi_1(\alpha) \longrightarrow -\infty$]. Pode ocorrer também, quando $g \neq 0$. Para isto, vejamos o seguinte exemplo:

EJEMPLO:

$$\text{Se } B = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \quad ; \quad g = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Aí, $\det(B - \lambda I) = 0 \Rightarrow \lambda_1 = -1$ e $\lambda_2 = 1$
e se $\alpha > 1$, então sabemos que:

$$\|p_\alpha\|^2 = \|(B + \alpha I)^{-1} g\|^2$$

$$\Rightarrow (B + \alpha I)^{-1} g = \begin{bmatrix} 1/(-1+\alpha) & 0 \\ 0 & 1/1+\alpha \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1/1+\alpha \end{bmatrix}$$

$$\Rightarrow \|p_\alpha\|^2 = 1 / |1 + \alpha|$$

$$\text{logo: } \|p_\alpha\|^2 < 1/2$$

de tal forma que a equação (III.16) não tem solução no intervalo $(-\lambda_1, \infty)$. -

CAPITULO IVMÉTODO DE REGIÃO DE CONFIANÇAIV.1. - DESCRIÇÃO E IDÉIAS GERAIS.

Uma das dificuldades que apresenta o Método de Newton, é dar um ponto de partida x_0 , que esteja suficientemente perto do mínimo local. Para se afastar esta dificuldade se investigou consideravelmente, com um custo bastante grande em Otimização Numérica. Aqui se discutirá um tipo de solução para o problema: os chamados "Métodos de Região de Confiança" ("TRUST REGION METHODS").

No Método de Newton, com busca linear, a Hessiana é modificada quando não é suficientemente definida positiva. Esta modificação ao modelo quadrático garante convergência, mas parece ignorar o papel do modelo quadrático como uma aproximação local a função objetivo. Vamos aqui considerar uma aproximação alternativa, na qual, o modelo quadrático não é modificado, mas em seu lugar (ou em vez disto), o modelo quadrático é considerado somente em uma "região de confiança" restrito. O uso desta técnica levou a algoritmos com fortes propriedades de convergência.

Seja $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, uma função duas vezes diferenciável continuamente. No Método de Newton, com uma estratégia de região de confiança, cada iterado x_k , tem uma cota Δ_k , tal que:

$$f(x_k + w) \approx f(x_k) + \psi_k(w); \quad \|w\| \leq \Delta_k$$

onde:

$$\psi_k(w) = \nabla f(x_k)^t w + (1/2)w^t \nabla^2 f(x_k) w$$

é o modelo quadrático da possível redução em f , dentro de uma vizinhança do iterado x_k .

Isto sugere que seria desejável calcular um passo s_k , o

qual resolve aproximadamente o problema:

$$\min \left\{ \psi_k(w) : \|w\| \leq \Delta_k \right\} \quad (\text{IV.1})$$

Se o passo é satisfatório, no sentido que $x_k + s_k$ produza uma redução suficiente em f , então Δ_k pode ser aumentado. Se ao contrário, o passo não é satisfatório, então a cota Δ_k deve ser diminuída.

Observação 1: O problema (IV.1) é equivalente ao problema:

$$\min m_c(x_c + s) = f(x_c) + \nabla f(x_c)^t s + (1/2) s^t \nabla^2 f(x_c) s$$

$$\text{s.a.: } \|s\|_2 \leq \delta_c \quad (\text{IV.2})$$

(Nesta equivalência tem-se que: $x_k = x_c$)

$$s = w; \delta_c = \Delta_k$$

Observação 2: O nome dado de "Região de Confiança", surge do fato da observação de δ_c como uma região na qual podemos confiar m_c como um modelo adequado de f (que em nosso caso será um modelo de tipo quadrático).

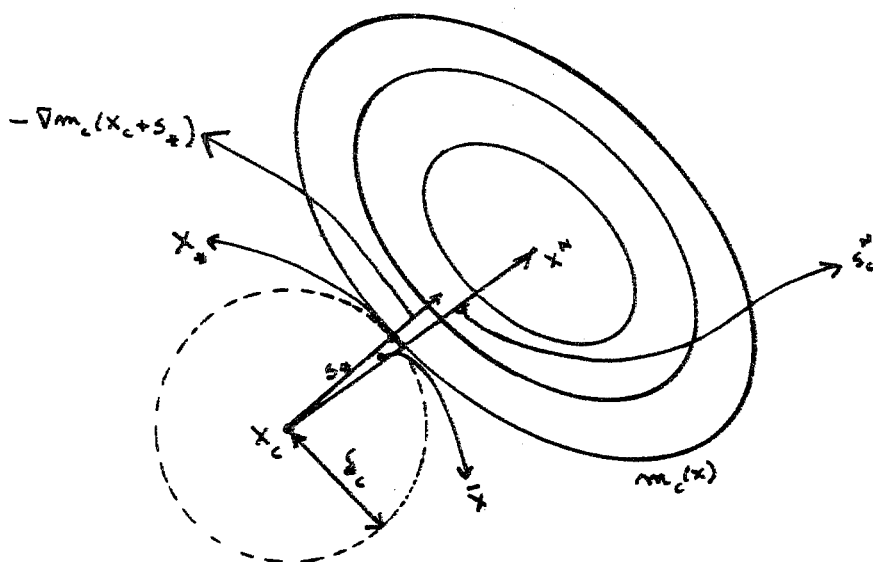
LEMA 1: Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$, duas vezes continuamente diferenciável; $H_c \in \mathbb{R}^{n \times n}$; simétrica e definida positiva e seja $\|\cdot\|$ a norma l_2 . Então o problema (IV.2) é resolvido por:

$$s(\mu) = - (H + \mu I)^{-1} \nabla f(x_c) \quad (\text{IV.3})$$

para o único $\mu > 0$, tal que $\|s(\mu)\| = \delta_c$, a menos que $\|s(0)\| \leq \delta_c$; em cujo caso $s(0) = s_c^N$ é a solução.

Para qualquer $\mu \leq 0$, $s(\mu)$ define uma direção de descida para f , a partir do ponto x_c .

Demonstração:



Seja s_* ; a solução de (IV.2) e seja $x_* = x_c + s_*$. Como o ponto (de Newton) $x_c + s_c^N$ é o mínimo global de m_c , é claro que se $\|s_c^N\| \leq \delta_c$; então $s_c^N = s_*$.

Consideremos agora o caso quando $x_c + s_c^N$ está fora da região limitada por δ_c (como se mostra na figura acima).

Seja \bar{x} um ponto qualquer no interior da região restrita, ou seja tem-se que: $\|\bar{x} - x_c\| \leq \delta_c$. Como $\nabla m_c(\bar{x}) \neq 0$ é possível então decrescer m_c a partir de \bar{x} embora permaneça dentro da região restrita, por considerar pontos da forma $\bar{x} - \lambda \nabla m_c(\bar{x})$. Isto implica que \bar{s} não pode ser solução de (IV.2), logo \bar{s} deve satisfazer $\|\bar{s}\| = \delta_c$, exceto quando $\|s(0)\| \leq \delta_c$. Além disso como a região restrita no problema (IV.2) é fechada e compacta, então algum s para o qual $\|s\| = \delta_c$; deve ser a solução.

Consideremos agora um s tal que $\|s\| = \delta_c$. Por ser s a solução de (IV.2), é necessário que nos desloquemos a uma

distância arbitrariamente pequena a partir de $x_c + s$, em qualquer direção de descida, aumentamos a distância a partir de x_c . Uma direção de descida para m_c a partir de $x_c + s$ é qualquer vetor p para o qual:

$$p^t \nabla m_c (x_c + s) = p^t \left[\nabla f(x_c) + H_c s \right] < 0 \quad (\text{IV.4})$$

Do mesmo modo, uma direção \bar{p} a partir de $x_c + s$ aumenta a distância a partir de x_c ; sse:

$$\bar{p}^t s > 0 \quad (\text{IV.5})$$

O que se diz é que para que s resolva o problema (IV.2) qualquer p que satisfaça (IV.4) deve também satisfazer (IV.5). Como se sabe que $\nabla m_c(x_c + s) \neq 0$, isto só pode ocorrer se $\nabla m_c(x_c + s)$ e $-s$, apontam na mesma direção. Em outras palavras, se para algum $\mu > 0$,

$$\mu s = -\nabla m_c(x_c + s) = -\left[\nabla f(x_c) + H_c s \right]$$

donde se obtém que:

$$\left[H_c + \mu I \right] s = -\nabla f(x_c)$$

Assim, $s_* = s(\mu)$, para algum $\mu > 0$, quando $\|s_c^N\| > \delta_c$. Como $\mu \geq 0$ e H_c é simétrica e definida positiva, $(H_c + \mu I)$ é também simétrica e definida positiva, logo $s(\mu)$ é uma direção de descida para f , a partir de x_c .

Somente falta para terminar a demonstração, demonstrar que s_* é único, o qual é imediato (ou está diretamente implicado) pelo resultado (forte), que nos diz que se:

$$\mu_1 > \mu_2 \geq 0 \text{ então: } \|s(\mu_1)\| < \|s(\mu_2)\| :$$

Isto mostra que $\|s(\mu)\| = \delta_c$ tem uma solução única, que deve ser a solução do problema (IV.2). A demonstração é direta, já que se $\mu > 0$, e se define:

$$\eta(\mu) = \|s(\mu)\|_2 = \left\| (H_c + \mu I)^{-1} \nabla f(x_c) \right\|$$

então:

$$\eta'(\mu) = \frac{-\nabla f(x_1)^t (H_c + \mu I)^{-1} \nabla f(x_1)}{\|s(\mu)\|} = \frac{-s(\mu)^t (H + \mu I)^{-1} s(\mu)}{\|s(\mu)\|};$$

logo $\eta'(\mu) < 0$, sempre que $\nabla f(x_c) \neq 0$. Portanto η é uma função de μ , monotonamente decrescente. ■ ■

Observações: i) O problema de usar o Lema 1 como base de um passo em um algoritmo de minimização, é que não existe um método finito para determinar $\mu_c > 0$ tal que: $\|s(\mu_c)\| = \delta_c$, quando $\delta_c < \|H_c^{-1} \nabla f(x_c)\|_2$.

ii) Neste trabalho, descreveremos dois métodos que se usam para resolver aproximadamente o problema (IV.2):

a) O primeiro deles, chamado "passo otimal localmente restringido", (chamado também: "HOOK"), encontra μ_c tal que $\|s(\mu_c)\| \approx \delta_c$ e logo faz: $x_+ = x_c + s(\mu_c)$.

b) O segundo, chamado "PASSO DOGLEG", faz uma aproximação linear de pedaços da curva $s(\mu)$ e toma x_+ como o ponto desta aproximação para a qual $\|x_+ - x_c\| = \delta_c$.

iii) Não existe garantia de que x_+ seja o ponto seguinte, o qual resolve exata ou aproximadamente o problema (IV.2), ainda que se tenha a esperança que assim seja, sempre que δ_c seja uma boa cota.

Daremos agora um Algoritmo do tipo geral para o Modelo de Aproximação que chamamos: Região de Confiança.

IV.2. -ALGORITMO GERAL (ALGORITMO 1)

Seja $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, $\delta_c > 0$, $x_c \in \mathbb{R}^n$, $H_c \in \mathbb{R}^{n \times n}$ simétrica e definida positiva.

Repita:

(1) s_c = solução aproximada do problema (IV.2)

$$x_+ : = x_c + s_c$$

(2) Decidir se x_+ é aceitável e calcular um novo valor de δ_c .

Quando x_+ é o seguinte ponto aceitável:

$$\delta_+ = \delta_c$$

De forma mais explícita este Algoritmo também se escreve assim:

PASSO 0: Se definem as constantes: $\mu \in (0,1)$, $\eta \in (\mu,1)$;

$\gamma_0 ; \gamma_1 ; \gamma_2$, as quais satisfazem:

$$0 < \gamma_0 \leq \gamma_1 < 1 \leq \gamma_2$$

O ponto de partida x_0 , o valor da função no referido ponto, (ou seja $f(x_0)$) e o gradiente da função em x_0 : g_0 são dados. Também são dados o raio inicial $\delta_0 (\equiv \Delta_0)$, da região de confiança e B_0 : a aproximação inicial da Hessiano no ponto de partida.

Seja $k = 0$.

PASSO 1: Obtém-se o passo s_k , como:

$$s_k = - B_k^{-1} g_k$$

sempre que B_k seja definida positiva (aproximação da Hessiano $\nabla^2 f(x_k)$); $g_k = \nabla f(x_k)$).

PASSO 2: Se calcula $f(x_k + s_k)$, e

$$\rho_k = \frac{f(x_k) - f(x_k + s_k)}{f(x_k) - m_c(x_k + s_k)}$$

onde:

$$m_c(x_k + s_k) = f(x_k) + g_k^t s_k + (1/2) s_k^t B_k s_k$$

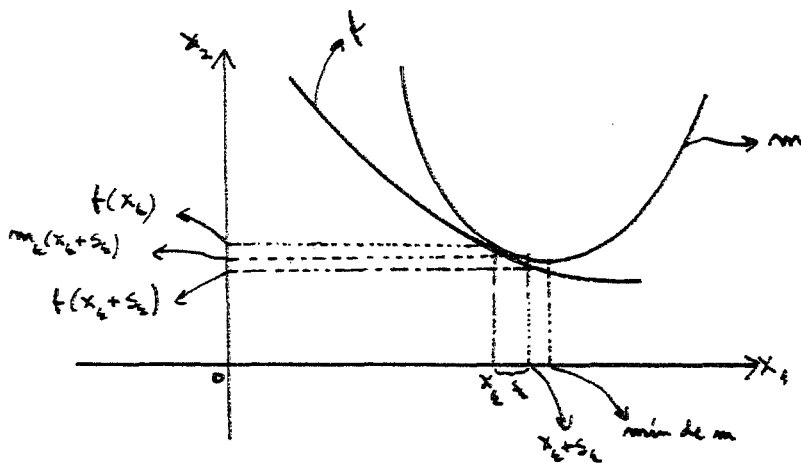
Notar que: ρ_k pode ser escrito também como:

$$\rho_k = \frac{f(x_k + s_k) - f(x_k)}{\psi_k(s_k)}$$

(ver a primeira página deste Capítulo IV)

PASSO 3: i) No caso em que $\rho_k > \mu$; ou seja, quando:

$$f(x_k) - f(x_k + s_k) > f(x_k) - m_k(x_k + s_k) \quad (\text{IV.6})$$



fazer:

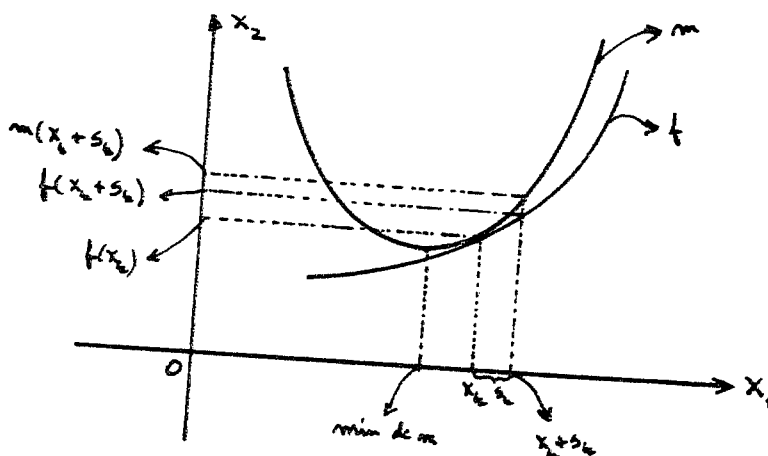
$$x_{k+1} = x_k + s_k \quad ; \quad g_{k+1} = \nabla f(x_{k+1})$$

e $\Delta_{k+1} \in [\Delta_k, \gamma_2 \Delta_k]$; se $\rho_k \geq \eta$

ou $\Delta_{k+1} \in [\gamma_1 \Delta_k, \Delta_k]$; se $\rho_k < \eta$

ii) No caso em que $\rho_k \leq \mu$, ou seja quando:

$$f(x_k) - f(x_k + s_k) \leq f(x_k) - m_k(x_k + s_k)$$



fazer:

$$x_{k+1} = x_k \quad ; \quad g_{k+1} = g_k$$

$$e \quad \Delta_{k+1} \in [\gamma_0 \Delta_k, \gamma_1 \Delta_k]$$

PASO 4: Atualizar a matriz B_k .

Incrementar k em 1 (fazer $k=k+1$) e voltar ao passo 1.

Observações: i) Diremos que a iteração é vitoriosa se (IV.6) se cumpre, ou seja, quando executada a função redução:

$$f(x_k) - f(x_k + s_k);$$

ela é, bem grande comparada com a função redução anteriormente falada:

$$f(x_k) - m_k(x_k + s_k)$$

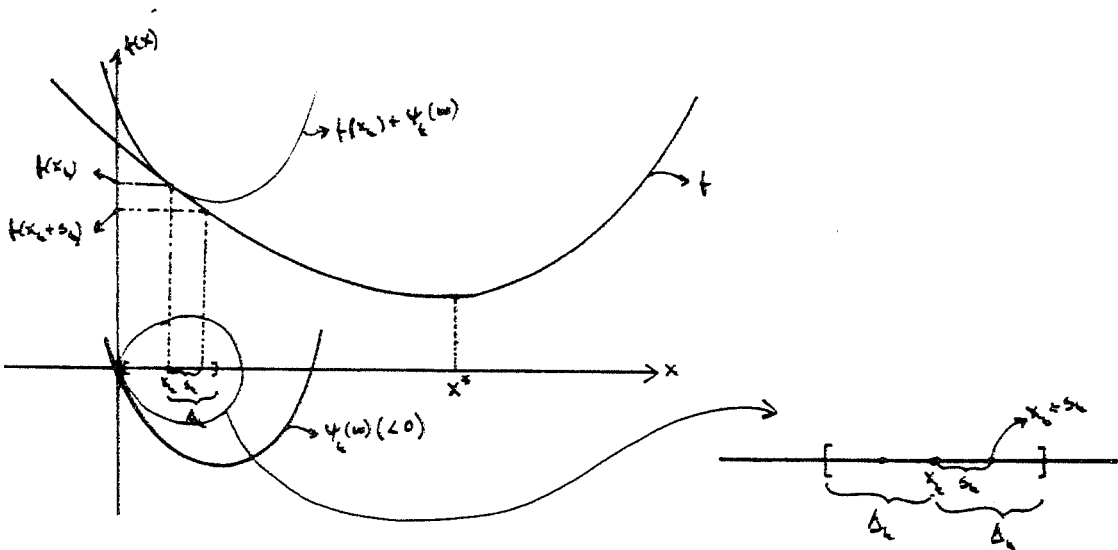
ii) Uma variante deste Algoritmo inclui uma matriz de escala para as variáveis. Com a referida variante, o problema (IV.2) toma a forma:

$$\min \left\{ \psi_k(w) : \|D_k w\| \leq \Delta_k \right\}$$

onde: $\psi_k(w) = f(x_k) + \nabla f(x_k)^t w + (\alpha/2)w^t H_c w$
e D_k é uma matriz não-singular.

(Em Resumo)

GRÁFICO:



f : função à minimizar

$$\psi_k(w) = \nabla f(x_k)^t w + (\alpha/2)w^t \nabla^2 f(x_k) w$$

$f(x_k + w) \approx f(x_k) + \nabla f(x_k)^t w + (\alpha/2)w^t \nabla^2 f(x_k) w = f(x_k) + \psi_k(w)$:
aproximação quadrática da função f no ponto x_k

$$\left. \begin{array}{l} \min \psi_k(w) \\ \text{s. a. } \|w\| \leq \Delta_k \end{array} \right\} \text{Problema (IV.1)}$$

-Se a solução do problema (IV.1), chamada s_k é tal que: $|s_k| < \Delta_k$, então s_k é a solução do sistema: $\nabla^2 f(x_k) w = -\nabla f(x_k)$, (o qual equivale a minimizar sem restrições a função $\psi_k(w)$).

-Se $\|s_k\| = \Delta_k$, então s_k é a solução do problema:

$$(H_k + \mu I) w = -\nabla f(x_k)$$

onde H_k é a matriz Hessiana definida em x_k e μ é o multiplicador de Lagrange associado a restrição equivalente:

$$\|s_k\|^2 = \Delta_k^2 .$$

Observação: Como se está encontrando um mínimo da função f , deve ter-se que:

$$f(x_k + w) < f(x_k) \quad (x, w \in \mathbb{R}^n)$$

o que é equivalente a dizer que:

$$\nabla f(x_k)^t w + (1/2)w^t \nabla^2 f(x_k) w \equiv \psi_k(w) < 0. \quad //$$

IV.3. - RESULTADOS SOBRE CONVERGÊNCIA DO ALGORITMO.

Não estamos interessados em resolver o problema modelo (IV.2) com grande precisão. Em vez disto, estamos interessados em dar condições mais relaxadas (ou mais suaves), para aceitar uma solução aproximada s_k do problema (IV.2), as quais garantam que a sucesão $\{x_k\}$ gerada pelo Algoritmo 1 é convergente ao ponto x^* , com $\nabla f(x^*)=0$ e $\nabla^2 f(x^*)$ semi-definida positiva.

Um requisito mínimo para s_k , é que existam constantes especificadas: $\beta_1 > 0$ e $\beta_2 > 0$ tais que:

$$-\psi_k(s_k) \geq \beta_1 |\psi_k^*| \quad , \quad \text{com } \|s_k\| \leq \beta_2 \Delta_k \quad (\text{IV.7})$$

onde ψ_k^* é o valor ótimo do problema (IV.2).

A condição (IV.7) pode também expressar-se de forma mais conveniente, para provas de convergência, da seguinte maneira:

Se $p_k \in \mathbb{R}^n$ é uma solução para o problema (IV.2), então, o Lema 2, do Capítulo III, implica que existe um parâmetro λ_k , tal que:

$$\begin{cases} \left[\nabla^2 f(x_k) + \lambda_k I \right] p_k = -\nabla f(x_k) \\ \lambda_k (\Delta_k - \|p_k\|) = 0 \end{cases}$$

Seja $R_k^t R_k$ a fatorização de Cholesky de $(\nabla^2 f(x_k) + \lambda_k I)$, então:

$$\| \psi_k^*(p_k) \| = 1/2 \left(\| R_k p_k \|^2 + \lambda_k \Delta_k^2 \right) \quad (\text{IV.8})$$

De fato: i) Sabemos que: $(\nabla^2 f(x_k) + \lambda_k I) p_k = -\nabla f(x_k)$

$$(R_k^t R_k) p_k = -\nabla f(x_k) / {}^t$$

$$[(R_k^t R_k) p_k] {}^t = -\nabla f(x_k) / p_k$$

$$-[(R_k^t R_k) p_k] {}^t p_k = \nabla f(x_k) p_k$$

ii) Como $(\nabla^2 f(x_k) + \lambda_k I) p_k = -\nabla f(x_k)$

$$\Rightarrow \nabla^2 f(x_k) p_k = -\nabla f(x_k) - \lambda_k I p_k / p_k {}^t$$

$$p_k {}^t \nabla^2 f(x_k) p_k = -p_k {}^t \nabla f(x_k) - \lambda_k p_k p_k {}^t$$

$$= -\lambda_k \|p_k\|^2 - p_k {}^t (-R_k^t R_k) p_k$$

$$\Rightarrow 1/2 p_k {}^t \nabla^2 f(x_k) p_k = -\lambda_k / 2 \|p_k\|^2 + 1/2 p_k {}^t R_k^t R_k p_k$$

$$\text{iii) } \psi_k^*(p_k) = -[(R_k^t R_k) p_k] {}^t p_k - \lambda_k / 2 \|p_k\|^2 + 1/2 p_k {}^t R_k^t R_k p_k$$

$$= -\|R_k p_k\|^2 - \lambda_k / 2 \|p_k\|^2 + 1/2 \|R_k p_k\|^2$$

$$= -1/2 (\|R_k p_k\|^2 + \lambda_k \|p_k\|^2)$$

e como por hipótese sabemos que $\Delta_k = \|p_k\|$

$$\Rightarrow \psi_k^*(p_k) = 1/2 (\|R_k p_k\|^2 + \lambda_k \Delta_k^2) \quad \blacksquare \square$$

Substituindo esta expressão para $|\psi_k^*|$ em (IV.7), tem-se que:

$$-\psi_k(s_k) \geq 1/2 \beta_1 (\|R_k p_k\|^2 + \lambda_k \Delta_k^2) \quad (\text{IV.9})$$

e assim, das expressões anteriores (IV.7) e (IV.9), os iterados $\{x_k\}$ gerados pelo Algoritmo dado, satisfazem:

$$-\psi_k(s_k) = -\rho_k (f(x_k + s_k) - f(x_k)) \geq 1/2 \beta_1 (\|R_k p_k\|^2 + \lambda_k \Delta_k^2)$$

$$\Rightarrow f(x_k) - f(x_k + s_k) \geq (1/2) (\beta_1 / \rho_k) (\|R_k p_k\|^2 + \lambda_k \Delta_k^2)$$

a qual pode escrever-se como:

$$f(x_k) - f(x_k + s_k) \geq (1/2) \mu \beta_1 (\|R_k p_k\|^2 + \lambda_k \Delta_k^2) \quad (\text{IV.10})$$

O desenvolvimento até aqui, servirá para demonstrar o seguinte resultado:

TEOREMA 1 :

Seja $f: \mathbb{R}^n \longrightarrow \mathbb{R}$, duas vezes continuamente diferenciável, sobre um conjunto aberto D e suponhamos que o ponto de partida x_0 é tal que, o conjunto de nível:

$$\Omega = \left\{ x \in D : f(x) \leq f(x_0) \right\}$$

é compacto. Se a sucessão $\{x_k\}$ produzida pelo Algoritmo, onde s_k satisfaz (IV.7), então: o Algoritmo termina em $x_l \in \Omega$ já que $\nabla f(x_l) = 0$ e $\nabla^2 f(x_l)$ é semi-definida positiva, ou $\{x_k\}$ tem um ponto limite x^* em Ω com $\nabla f(x^*) = 0$ e $\nabla^2 f(x^*)$ semi-definida positiva.

Demonstração:

i) Se $\nabla f(x_l) = 0$ e $\nabla^2 f(x_l)$ é semi-definida positiva para algum ponto (ou iterado) $x_l \in \Omega$, então o Algoritmo termina.

ii) Do contrário; como:

$$-\psi_k(s_k) \geq \beta_1 |\psi_k^*| \quad ; \text{ com } \|s_k\| \leq \beta_2 \Delta_k$$

$$\Rightarrow \psi_k(s_k) \leq -\beta_1 |\psi_k^*|$$

$$\Rightarrow \psi_k(s_k) < 0; \text{ para } k \geq 0$$

logo $\{x_k\}$ está bem definida e está em Ω .

iii) Provemos agora o resultado (ou a tese) supondo que a sucessão $\{\lambda_k\}$ é não limitada, longe de 0. Se alguma subsucessão de $\{\lambda_k\}$ converge para 0, então como Ω é compacto, sem perda de generalidade, podemos supor que a mesma subsucessão de $\{x_k\}$ converge para algum x^* no conjunto de nível Ω .

-Como $\nabla^2 f(x_k) + \lambda_k I$ é semi-definida positiva, então: $\nabla^2 f(x^*)$ é também semi-definida positiva.

-Para provar que $\nabla f(x^*) = 0$, sabemos que:

$$\begin{aligned} -\nabla f(x_k) &= (\nabla^2 f(x_k) + \lambda_k I) p_k \\ &= \nabla^2 f(x_k) p_k + \lambda_k p_k / \|p_k\| \\ \|\nabla f(x_k)\| &\leq \|\nabla^2 f(x_k)\| \|p_k\| + \lambda_k \|p_k\| / \|\nabla f(x_k)\| \\ \|\nabla f(x_k)\|^2 &\leq \|\nabla^2 f(x_k)\| \|p_k\| (\|\nabla^2 f(x_k) + \lambda_k I\|) \quad (\text{IV.11}) \end{aligned}$$

Porém como também sabemos que:

$$\begin{aligned} \|\nabla f(x_k)\| &= \|R_k^t R_k p_k\| / \|p_k\| \\ \Rightarrow \|\nabla f(x_k)\| \|p_k\| &\leq \|R_k\|^2 \|p_k\|^2 = \|R_k p_k\|^2 \end{aligned}$$

Logo em (IV.11) tem-se que:

$$\|\nabla f(x_k)\|^2 \leq \|R_k p_k\|^2 (\|\nabla^2 f(x_k)\| + \lambda_k) \quad (\text{IV.12})$$

que é equivalente a escrever:

$$\|R_k p_k\|^2 \geq \frac{\|\nabla f(x_k)\|^2}{\|\nabla^2 f(x_k)\| + \lambda_k}$$

De (IV.10) tem-se que $\langle \|R_k p_k\| \rangle$ converge a 0, logo:

$$\nabla f(x^*) = 0.$$

** Pode-se demonstrar que $\langle \lambda_k \rangle$ não é limitada longe de 0, por contradição:

Se $\lambda_k \geq \varepsilon > 0$, então (IV.7) e (IV.9) implicam que:

$$-\psi_k(s_k) \geq (1/2) \beta_1 \lambda_k \Delta_k^2 \geq 1/2 \left[\frac{\beta_1}{\beta_2} \right] \varepsilon \|s_k\|^2 \quad (\text{IV.13})$$

De fato : De (IV.7) $\Rightarrow -\psi_k(s_k) \geq \beta_1 |\psi_k^*|$

De (IV.9) $\Rightarrow -\psi_k(s_k) \geq 1/2 \beta_1 (\|R_k p_k\|^2 + \lambda_k \Delta_k^2)$

$$\Leftrightarrow -\psi_k(s_k) \geq 1/2 \beta_1 \lambda_k \Delta_k^2 \quad (\text{IV.14})$$

Como sabemos: $\|s_k\| \leq \beta_2 \Delta_k$

$$\Rightarrow \Delta_k \geq \frac{\|s_k\|}{\beta_2} / 2$$

$$\Rightarrow \Delta_k^2 \geq \frac{\|s_k\|^2}{\beta_2^2}$$

Logo em (IV.14) tem-se que:

$$-\psi_k(s_k) \geq 1/2 \beta_1 \lambda_k \frac{\|s_k\|^2}{\beta_2^2} \text{ e como } \lambda_k \geq \varepsilon \geq 0$$

$$\Rightarrow -\psi_k(s_k) \geq 1/2 \left[\frac{\beta_1}{\beta_2^2} \right] \varepsilon \|s_k\|^2 \quad (\text{IV.15})$$

-Sabemos que $f(x_k + s_k) \approx f(x_k) + \psi_k(s_k)$, onde:

$$\psi_k(s_k) = \nabla f(x_k)^t s_k + 1/2 s_k^t \nabla^2 f(x_k) s_k$$

então uma estimativa padrão é:

$$|f(x_k + s_k) - f(x_k) - \psi_k(s_k)| \leq 1/2 \|s_k\|^2 \max_{0 \leq \xi \leq 1} \|\nabla^2 f(x_k + \xi s_k) - \nabla^2 f(x_k)\| \quad (\text{IV.16})$$

-Além de (IV.15) tem-se que:

$$\|s_k\|^2 \leq \frac{-\psi_k(s_k) \beta_2^2}{\beta_1 \varepsilon} \leq \frac{2 \psi_k(s_k) \beta_2^2}{\beta_1 \varepsilon}$$

Logo, recordando que $\rho_k = \frac{f(x_k + s_k) - f(x_k)}{\psi_k(s_k)}$, subs-

tituindo em (IV.16), tem-se que:

$$|\rho_k - 1| \leq (\beta_2^2 / \beta_1 \varepsilon) \text{Máx}_{0 \leq \xi \leq 1} \|\nabla^2 f(x_k + \xi s_k) - \nabla^2 f(x_k)\| \quad (\text{IV.17})$$

A desigualdade (IV.10) implica que $\langle \Delta_k \rangle$ converge para 0 e de (IV.13), temos que $\langle \|s_k\| \rangle$ também converge para 0. Deste modo, a continuidade uniforme de $\nabla^2 f$ em Ω conjuntamente com (IV.17) implicam que $\rho_k > \eta$; $\forall k$ suficientemente grande e portanto, as regras para atualizar Δ_k implicam que $\langle \Delta_k \rangle$ está limitado longe de 0. Isto é uma contradição com o fato de que $\langle \Delta_k \rangle$ converge para 0.

Observação: O teorema recém demonstrado é somente uma mostra dos resultados válidos de convergência para o Algoritmo 1; segundo a hipótese que:

$$-\psi_k(s_k) \geq \beta_1 |\psi_k^*| ; \text{con } \|s_k\| \leq \beta_2 \Delta_k \\ (\text{con } \beta_1 > 0 \text{ e } \beta_2 > 0)$$

Este resultado estende resultados de SORENSEN, (1982) e FLETCHER (1980), que admitem solução inexatas ao problema modelo (IV.2).

Os seguintes resultados são conhecidos: (ver: MORE, J. J. e SORENSEN, D. C. (1983))

- i) A sucessão $\{\nabla f(x_k)\}$ converge para 0.
- ii) Se x^* é um ponto limite separado de $\langle x_k \rangle$, então $\nabla^2 f(x^*)$ é semi-definida positiva.
- iii) Se $\nabla^2 f(x^*)$ é não-singular para algum ponto limite x^* de $\langle x_k \rangle$; então $\langle x_k \rangle$ converge para x^* .

Um resultado adicional que serve para estabelecer alguns resultados sobre razão de convergência; é o seguinte:

Se a sucessão $\{x_k\}$ converge para x^* e $\nabla^2 f(x^*)$ é definida positiva, então a sucessão das cotas $\{\Delta_k\}$ está limitada longe de 0. Para demonstrar esta proposição note que se $\varepsilon > 0$ é uma cota inferior dos valores próprios da matriz dos $\nabla^2 f(x_k)$, então de (IV.8) tem-se que:

$$|\psi_k^*| \geq 1/2 \varepsilon_0 \min \langle \Delta_k^2, \|s_k^N\|^2 \rangle \quad (\text{IV.18})$$

$$\text{onde: } s_k^N = -[\nabla^2 f(x_k)]^{-1} \nabla f(x_k) \quad (\text{passo Newton})$$

Além disso; como $\psi_k(s_k) < 0$ (já que estamos minimizando) tem-se que:

$$1/2 \|s_k\| \leq \|s_k\| = \|\nabla^2 f(x_k)^{-1}\| \|\nabla f(x_k)\|$$

Logo:

$$1/2 \|s_k\| \leq K \|s_k^N\|$$

onde K é uma cota superior do número de condição da matriz $\nabla^2 f(x_k)$.

Por outro lado, da desigualdade (IV.7) que corresponde a:

$$-\psi_k(s_k) \geq \beta_1 |\psi_k^*| \quad (\text{com } \|s_k\| \leq \beta_2 \Delta_k)$$

e de (IV.18); tem-se:

$$-\psi_k(s_k) \geq \beta_1 1/2 \varepsilon_0 \min \langle \Delta_k^2, \|s_k^N\|^2 \rangle \geq 1/2 \beta_1 \varepsilon_0 \|s_k\|^2 = 1/2 \varepsilon_1 \|s_k\|^2$$

(> 0)
ε₁

Assim, em nossa estimação padrão, temos que:

$$\left| \frac{f(x_k + s_k) - f(x_k)}{\psi_k(s_k)} - 1 \right| \leq |\psi_k(s_k)| \leq 1/2 \|s_k\|^2 \max_{0 \leq \xi \leq 1} \|\nabla^2 f(x_k + \xi s_k) - \nabla^2 f(x_k)\|$$

$$\begin{aligned}
 (\Leftrightarrow) \quad | \rho_k - 1 | &\leq [1 / (2 | \psi_k(s_k) |)] \| s_k \|^2 * \\
 &* \text{Máx}_{0 \leq \xi \leq 1} \| \nabla^2 f(x_k + \xi s_k) - \nabla^2 f(x_k) \|
 \end{aligned}$$

e usando (IV.16) tem-se que :

$$| \rho_k - 1 | \leq (1/\varepsilon_1) \text{máx}_{0 \leq \xi \leq 1} \| \nabla^2 f(x_k + \xi s_k) - \nabla^2 f(x_k) \|$$

e como $\rho_k > \eta$ (recordar que $0 < \mu < \eta < 1$); $\forall k$; suficientemente grande, tem-se que a sucessão $\{ \Delta_k \}$ está limitada longe de 0 ■ □ //

Observação: Os resultados sobre razão de convergência, podem ser obtidos, também, com a suposição ou hipótese adicional, que existe uma constante $\beta_3 > 0$, tal que se $\|s_k\| \leq \beta_3 \Delta_k$ então $\nabla^2 f(x_k)$ é definida positiva e $s_k = s_k^N$; onde s_k^N é o passo Newton [quer dizer: $s_k^N = -\nabla^2 f(x_k)^{-1} \nabla f(x_k)$]. Tendo presente estas idéias, suponhamos que $\{x_k\}$ converge para x^* e que $\nabla^2 f(x^*)$ é definida positiva. Então $\{s_k\}$ converge para 0 e portanto $\|s_k\| \leq \beta_3 \Delta_k$; $\forall k$; suficientemente grande. Assim, \exists um $k_0 \geq 0$; tal que $s_k = s_k^N$, para $k \geq k_0$ e logo a razão de convergência está dada pelo Teorema 3 visto anteriormente, no Capítulo III (referente a Preliminares Matemáticas).-

CAPITULO V

ESTRATÉGIAS PARA O CÁLCULO DO PASSOV.1. -PASSO LOCALMENTE RESTRITO ("HOOK" STEP)

Nossa primeira aproximação para calcular o passo no Algoritmo Região de Confiança (ALGORITMO 1, CAP IV), quando $\|s(0)\|_2 > \delta_c$ ($\Leftrightarrow \|s(0)\|_2 = \|(H_c + 0 I)^{-1} \nabla f(x_c)\|_2$), (já que quando ocorre que $\|s(0)\|_2 \leq \delta_c$, o problema é resolvido pelo Lema 1, CAP .IV), é achar um passo que dependa de μ , do tipo:

$$s(\mu) = -(H_c + \mu I)^{-1} \nabla f(x_c)$$

tal que $\|s(\mu)\|_2 \cong \delta_c$ e em seguida efetuar um passo de prova $x_+ = x_c + s(\mu)$.

Aqui se discutirá um algoritmo para encontrar uma solução aproximada μ_c , da equação escalar:

$$\phi(\mu) = \|s(\mu)\|_2 - \delta_c = 0 \quad (V.1)$$

(Na prática não se requer uma solução totalmente exata).

A idéia para resolver (V.1); é recorrer a um modelo local do problema em questão. Para isto, a versão unidimensional, neste caso, corresponde a:

$$\phi(\mu) = | -(h_c + \mu)^{-1} g_c | - \delta_c$$

que sugere um modelo local da forma:

$$m_c(\mu) = \left[\alpha_c / (\beta_c + \mu) \right] - \delta_c$$

com dois parâmetros livres α_c e β_c (que corresponderão ao gradiente e Hessiana, respectivamente).

Observação: Antes de determinar a forma de achar α_c e β_c , notemos que introduzimos uma nova notação para relacionar-se

com a iteração sobre μ , que provém do modelo anterior e tem lugar dentro da x -iteração principal e cujo propósito é achar o μ_c que satisfaz (V.1). Usaremos diretamente os valores atuais das quantidades α_c, β_c, μ_c que são trocadas na iteração interior sobre μ e (de) tipo normal para los valores em curso de $\delta_c, x_c, \nabla f(x_c), H_c$ que vem da iteração exterior e que não são trocados durante a solução da equação escalar (V.1).

Assim se obtém μ_c como a última iteração interna que aproxima μ_c a μ_* ; que corresponde a solução exata de (V.1) e se usa x_c e μ_c para definir: $x_+ = x_c + s(\mu_c)$.

O modelo $m_c(\mu)$ tem, como já foi dito, dois parâmetros livres: α_c e β_c e tal como se faz na obtenção do Método de Newton, é razoável escolhe-los de tal maneira que satisfaçam as duas condições seguintes:

$$i) m_c(\mu_c) = \frac{\alpha_c}{\beta_c + \mu_c} - \delta_c = \phi(\mu_c) = \|s(\mu_c)\|_2 - \delta_c \quad (V.2) \quad (V.3)$$

$$ii) m'_c(\mu) = \frac{-\alpha_c}{(\beta_c + \mu_c)^2} = \phi'(\mu_c) = - \frac{s(\mu_c)^t (H_c + \mu_c I)^{-1} s(\mu_c)}{\|s(\mu_c)\|_2} \quad (**) \quad (V.4)$$

Demonstração da última igualdade (**) em (V.4):

Por hipótese sabemos que:

$$\phi(\mu) = \|s(\mu)\|_2 - \delta_c$$

De tal forma que derivando com respeito a μ ; tem-se:

$$\phi'(\mu) = \frac{d\phi}{d\mu} = \frac{1}{\|s(\mu)\|_2} \left[\frac{d}{d\mu} (H_c + \mu_c I)^{-1} \nabla f(x_c) \right] (H_c + \mu_c I)^{-1} \nabla f(x_c)$$

Seja: $A(\mu) = (H_c + \mu_c I)^{-1}$

$$\rightarrow I = A(\mu) (H_c + \mu_c I)$$

Derivando esta última expressão que diz respeito a μ tem-se que:

$$0 = A'(\mu) (H_c + \mu_c I) + A(\mu) I$$

$$\begin{aligned} \rightarrow A'(\mu) &= -A(\mu) (H_c + \mu_c I)^{-1} \\ &= -(H_c + \mu_c I)^{-2} \end{aligned}$$

De tal forma que substituindo em $\phi'(\mu)$; temos :

$$\begin{aligned} \phi'(\mu) &= -\frac{1}{\|s(\mu)\|_2} \left[\nabla f(x_c)^t (H_c + \mu_c I)^{-1} \nabla f(x_c) \right] \\ &= (1/\|s(\mu)\|_2) \left[s^t(\mu) (H_c + \mu_c I)^{-1} s(\mu) \right]_{\mu=\mu_c} \end{aligned}$$

Assim, resolvendo o sistema de equações (V.3) e (V.4); achamos que os valores de α_c e β_c são dadas por:

$$\alpha_c = \frac{(\phi(\mu_c) + \delta_c)^2}{\phi'(\mu_c)} \quad (V.5)$$

$$\beta_c = \frac{(\phi(\mu_c) + \delta_c)}{\phi'(\mu_c)} - \mu_c \quad (V.6)$$

Como já temos nosso modelo $m_c(\mu_c)$, se escolhe μ_+ de tal forma que $m_c(\mu_+) = 0$. Ou seja:

$$m_c(\mu_+) = \alpha_c / (\beta_c + \mu_+) - \delta_c = 0$$

Donde obtemos que:

$$\mu_+ = (\alpha_c / \delta_c) - \beta_c \quad (V.7)$$

A idéia agora é utilizar as expressões anteriores, para demonstrar que o valor de μ_+ pode ser escrito como:

$$\mu_+ = \mu_c - \left[\|\mathfrak{s}(\mu_c)\|_2 / \delta_c \right] \left[\phi(\mu_c) / \phi'(\mu_c) \right] \quad (V.8)$$

De fato:

Substituindo (V.5) e (V.6) em (V.7), tem-se que:

$$\begin{aligned} \mu_+ &= \frac{\frac{(\phi(\mu_c) + \delta_c)^2}{\phi'(\mu_c)}}{\delta_c} + \frac{\phi(\mu_c) + \delta_c}{\phi'(\mu_c)} + \mu_c \\ &= \mu_c + \frac{-(\phi(\mu_c + \delta_c))^2 + \delta_c(\phi(\mu_c) + \delta_c)}{\delta_c \phi'(\mu_c)} \\ &= \mu_c + \frac{-\phi^2(\mu_c) - \delta_c^2 - 2\phi(\mu_c)\delta_c + \delta_c\phi(\mu_c) + \delta_c^2}{\delta_c \phi'(\mu_c)} \\ &= \mu_c + \frac{-\phi^2(\mu_c) - \phi(\mu_c)\delta_c}{\delta_c \phi'(\mu_c)} \\ &= \mu_c + \frac{-\phi^2(\mu_c) - \phi(\mu_c) \left[\|\mathfrak{s}(\mu_c)\| - \phi(\mu_c) \right]}{\delta_c \phi'(\mu_c)} \quad (\text{de (V.1)}) \\ &= \mu_c + \frac{-\phi^2(\mu_c) - \phi(\mu_c) \|\mathfrak{s}(\mu_c)\|_2 + \phi^2(\mu_c)}{\delta_c \phi'(\mu_c)} \end{aligned}$$

$$\Rightarrow \mu_+ = \mu_c - \frac{\|\mathfrak{s}(\mu_c)\|_2}{\delta_c} \left[\frac{\phi(\mu_c)}{\phi'(\mu_c)} \right] \quad (V.8)$$

■ ■

(Logo este valor de μ_+ , resolve a equação $\phi(\mu)=0$, que corresponde a (V.1)).

Observação : A forma de (V.8) mostra que:

i) Quando $\mu_c < \mu_*$ (onde μ_* corresponde a solução exata da equação escalar (V.1)), estamos assumindo uma correção maior que aquela que o Método de Newton poderia dar, porém quando $\mu_c > \mu_*$, tem-se que o passo é maior que o passo Newton.

ii) Como μ_c tende a μ_* ; (V.8) chega a ser cada vez mais (ou parece cada vez mais) como o Método de Newton. De fato (V.8) é localmente q-quadráticamente convergente para μ_* .

Daremos alguns resultados que transformam (V.8) em um algoritmo que pode ser implementado computacionalmente. Para isto, convém ter-se presente algumas considerações:

i) Com relação ao valor de partida para μ na resolução de (V.1):

REINSCH (1971) demonstrou que se em (V.8) se começa com um valor inicial $\mu_0 = 0$, então a μ iteração converge monótonamente a μ_* . Entretanto, gostaríamos de ter um ponto inicial mais próximo, já que cada iteração de (V.8) resulta em um sistema linear. Para isto MORE em 1977, propôs a seguinte idéia:

Se o passo atual que restringe δ_c é p vezes o último valor da cota do passo δ_- , então $\mu_0 = \mu_-/p$ é usado como ponto de partida em (V.8).

Aqui se prefere uma estratégia diferente. Recordemos aonde estamos na iteração. Acabamos de achar um passo $x: s(\mu_-)$ de x_- a x_c e agora queremos x_+ . Obtivemos δ_c de δ_- . Já temos H_c e temos $(H_- + \mu_- I)$ de forma fatorada, e como o custo é da ordem de n^2 , calculamos:

$$\phi = \| s(\mu_-) \| - \delta_c \quad y \quad \phi' = \frac{s(\mu_-)^t (H_- + \mu_- I)^{-1} s(\mu_-)}{\| s(\mu_-) \|}$$

Analogamente como fizemos com a fórmula (V.8), usamos:

$$\mu_0 = \mu_- - \left[\|\phi(\mu_-)\| / \delta_c \right] (\phi / \phi') \quad (\text{V.9})$$

para achar o valor de μ , obtido de μ_- , o modelo prévio e o novo raio de confiança δ_c .

ii) Para a geração e atualização das cotas superior e inferior (u e I_+) para μ_+ , digamos que estas cotas são usadas em (V.8) para "confinar" (obrigar que) μ_+ a estar no intervalo $[I_+, u]$.

Como a iteração de Newton aplicada a (V.1) sempre alcança (ou melhor dizendo: "chega a") a μ_* , faz-se:

$$I_0 = - \frac{\phi(0)}{\phi'(0)}$$

Logo se calcula: $\mu_+^N = \mu_c - \phi(\mu_c) / \phi'(\mu_c)$, juntamente com cada cálculo de (V.8) e atualizamos a menor cota para $I_+ = \max\{I_c, \mu_+^N\}$; onde I_c é a menor cota atual. Também, como:

$$\delta_c = \|(H_c + \mu_+ I)^{-1} \nabla f(x_c)\|_2 < \|\nabla f(x_c)\|_2 / \mu_*$$

(já que H_c é definida positiva e $\mu_* > 0$); tomamos $\|\nabla f(x_c)\|_2 / \delta_c$, como nossa cota superior inicial u_0 para μ_* . Logo, em cada iteração se $\phi(\mu_c) < 0$, atualizamos a cota superior para: $u_+ = \min\{u_c, \mu_c\}$; onde u_c é a cota superior atual.

Se em qualquer iteração, μ_+ não está em $[I_+, u_+]$, seguimos MORE (1977-1978), para escolher μ_+ como:

$$\mu_+ = \max \left\{ (I_+ * u_+)^{1/2}, 10^{-3} u_+ \right\} \quad (V.10)$$

onde o segundo termo que aparece em (V.10) é uma segurança dos valores próximos de 0 e I_+ . Na prática estas cotas são usadas com bastante freqüência no cálculo de μ_0 . Em particular (V.10) é usada para definir μ_0 sempre que (V.9) não possa ser usada, porque a iteração prévia usou o passo Newton.

iii) Finalmente não resolvemos (V.1) com grande exatidão, ao invés disto, fazemos $\|s(\mu)\|_2 \in [3/4 \delta_c, 3/2 \delta_c]$. O motivo disto é que (como se verá mais adiante) a região de confiança, nunca é aumentada ou diminuída por um fator menor que 2. Assim, se o raio da região de confiança em curso é δ_c , um anterior teve que ser maior ou igual a $2\delta_c$ ou menor ou igual a $\delta_c/2$. Logo consideramos que o valor atual de δ_c seja mais arbitrário dentro do intervalo $[3/4\delta_c, 3/2\delta_c]$, na qual ajusta a diferença em uma ou outra direção e parece razoável aceitar que $\|s(\mu)\|_2$ tenha qualquer valor neste intervalo. Algumas implementações, por exemplo MORE (1977) supõem que $\|s(\mu)\|_2 \in [0.9\delta_c, 1.1\delta_c]$ e não fica muito claro, se é ou não melhor.

Veremos agora um exemplo onde se aplica e explicita o que foi dito anteriormente:

EXEMPLO 1: Interessa-nos resolver o problema:

$$\begin{aligned} \min m_c(x_c + s) &= f(x_c) + \nabla f(x_c)^t s + (1/2)s^t H_c s, \\ \text{s. a. } \|s\|_2 &\leq \delta_c \end{aligned}$$

$$\text{onde : } f(x_1, x_2) = x_1^4 + x_1^2 + x_2^2 ; \delta_c = 1/2, \mu_- = 0$$

$$x_c = (1, 1)^t ; H_c = \nabla^2 f(x_c)$$

$$\nabla f(x_c) = (4x_1^3 + 2x_1, 2x_2) = (6, 2)^t$$

$$\|\nabla f(x_c)\|_2 = \sqrt{40} = 6.325.$$

$$H_c = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix}_{x_c} = \begin{bmatrix} 14 & 0 \\ 0 & 2 \end{bmatrix}$$

$$\begin{aligned} s_c^N &= -H_c^{-1} \nabla f(x_c) = - \begin{bmatrix} 1/14 & 0 \\ 0 & 1/2 \end{bmatrix} \begin{bmatrix} 6 \\ 2 \end{bmatrix} = \\ &= - \begin{bmatrix} 6/14 \\ 1 \end{bmatrix} = \begin{bmatrix} -3/7 \\ -1 \end{bmatrix} \end{aligned}$$

$$\Rightarrow \|s_c^N\|_2 = \sqrt{(-3/7)^2 + (-1)^2} \approx 1.088$$

Como se deve ter que $\|s(\mu)\|_2 \in [3/4\delta_c, 3/2\delta_c] = [3/8, 3/4]$ e aqui ocorre que $\|s_c^N(\mu)\|_2 > 3/2\delta_c$, (o que significa que o passo Newton é grande demais), então deve-se buscar a seguir algum $\mu > 0$ tal que :

$$\begin{aligned} s(\mu_i) &= \|(H_c + \mu_i I)^{-1} \nabla f(x_c)\|_2 \in [3/4\delta_c, 3/2\delta_c] = [3/8, 3/4] = \\ &= [0.375, 0.750] ; \end{aligned}$$

com $i = 0, 1, 2, \dots$

Para isto: Como $\mu_0 = 0$, então calculamos:

$$\begin{aligned} I_0 &= - \frac{\phi(0)}{\phi'(0)} = \frac{-\|s(0)\|_2 + \delta_c}{\frac{s(0)^t H_c^{-1} s(0)}{\|s(0)\|_2}} = \frac{-1.088 + 0.500}{-0.513} = \\ &= \frac{-0.588}{-0.472} = 1.250. \end{aligned}$$

$$u_0 = \frac{\|\nabla f(x_c)\|_2}{\delta_c} = \frac{6.325}{0.500} = 12.649$$

$$\begin{aligned} \mu_0 &= \text{Máx} \left\{ (I_0 * u_0)^{1/2}, 10^{-3} u_0 \right\} = \text{Máx} \left\{ (1.250 * 12.649)^{1/2}, 10^{-3} * 12.649 \right\} \\ &= \text{Máx} \left\{ 3.97, 0.012 \right\} = 3.97. \end{aligned}$$

$$\text{Logo: } s(\mu_0) = -(H_c + \mu_0 I)^{-1} \nabla f(x_0) = - \begin{bmatrix} 14+3.97 & 0 \\ 0 & 2+3.97 \end{bmatrix}^{-1} \begin{bmatrix} 6 \\ 2 \end{bmatrix}$$

$$= - \begin{bmatrix} 0.056 & 0 \\ 0 & 0.168 \end{bmatrix} \begin{bmatrix} 6 \\ 2 \end{bmatrix} = \begin{bmatrix} -0.334 \\ -0.335 \end{bmatrix}$$

$$= (-0.334, -0.335)^t$$

$$\Rightarrow \|s(\mu_0)\|_2 = \sqrt{(-0.334)^2 + (-0.335)^2} \approx 0.473$$

Como $\|s(\mu_0)\|_2 \approx 0.473 \in [0.375, 0.750]$, fazemos então:

$$\begin{aligned} \mu_c &= \mu_0 \text{ e } x_+ = x_c + s(\mu_c) = (1, 1)^t + (-0.334, -0.335)^t \\ &= (0.666, 0.665)^t \end{aligned}$$

que corresponde a nossa solução aproximada do problema:

$$\begin{aligned} \text{mín } m_c(x_c + s) &= f(x_c) + \nabla f(x_c)^t s + (1/2) s^t H_c s \\ \text{s.a.: } \|s\|_2 &\leq 1/2 \end{aligned}$$

Observação: Aqui não se utilizou a iteração dada pela fórmula (V.8) para achar o real $\mu > 0$, que resolve a equação

$$\phi(\mu) = 0.$$

A título de ilustração, digamos que se trabalhamos com a fórmula mencionada (onde $\mu_0 = 3.97$) se terá neste caso:

$$\mu_1 = \mu_0 - \frac{s(\mu_0)}{\delta_c} \left[\frac{\phi(\mu_0)}{\phi'(\mu_0)} \right]; \left[\approx 3.97 - \frac{0.473}{0.500} \left[\frac{\phi(\mu_0)}{\phi'(\mu_0)} \right] \right]$$

onde:

$$a) \phi(\mu_0) = \|s(\mu_0)\|_2 - \delta_c = \|s(3.97)\|_2 - 0.5$$

$$s(\mu_0) = (-0.334, -0.335)^t \Rightarrow \|s(\mu_0)\|_2 = 0.473$$

$$\Rightarrow \phi(\mu_0) = 0.473 - 0.500 = -0.027.$$

$$b) \phi'(\mu_0) = - \frac{s(\mu_0) (H_c + \mu_0 I)^{-1} s(\mu_0)}{\|s(\mu_0)\|_2} =$$

$$= \frac{-[-0.334 \quad -0.335] \begin{bmatrix} 0.056 & 0 \\ 0 & 0.168 \end{bmatrix} \begin{bmatrix} -0.334 \\ -0.335 \end{bmatrix}}{0.437} =$$

$$= -0.025 / 0.473 = -0.053$$

$$\Rightarrow \mu_1 \approx = 3.97 - \frac{0.473}{0.500} \left[\frac{-0.027}{-0.053} \right] \approx 3.49.$$

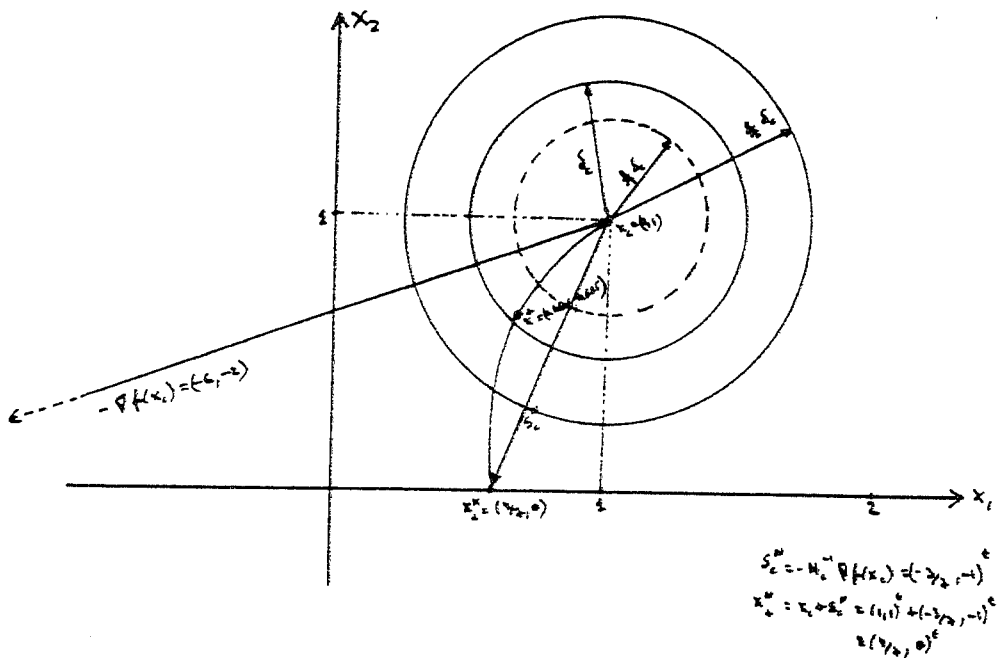
$$\text{e tal que : } s(\mu_1) = -(H_c + \mu_1 I)^{-1} \nabla f(x_c)$$

$$= - \begin{bmatrix} 14+3.49 & 0 \\ 0 & 2+3.49 \end{bmatrix}^{-1} \begin{bmatrix} 6 \\ 2 \end{bmatrix} =$$

$$= (-0.343, -0.364)^t$$

$$\rightarrow \|s(\mu_1)\|_2 = 0.5003624.$$

GRÁFICO:



V. 2. -O PASSO DOGLEG E DUPLO DOGLEG. -

Vejamos agora outra implementação do Modelo Região de Confiança, que é uma modificação do Algoritmo Região de Confiança feito por POWELL (1970). Esta modificação também encontra uma solução aproximada do problema:

$$\begin{aligned} \min m_c(x_c + s) &= f(x_c) + \nabla f(x_c)^t s + (1/2) s^t H_c s \\ \text{s. a. } \|s\|_2 &\leq \delta_c \end{aligned}$$

Entretanto, em vez de encontrar um ponto $x_+ = x_c + s(\mu_c)$; sobre a curva $s(\mu)$, tal que $\|x_+ - x_c\| = \delta_c$, aproxima esta curva por uma função linear em pedaços, que conecta o chamado "ponto de Cauchy" (C.P.) que corresponde ao mínimo do modelo quadrático $m_c(x_c + s)$ na direção de descida mais profundo $(-\nabla f(x_c))$, com a direção de Newton para $m_c(x_c + s)$.

Observação:

POWELL pensou que quando se requer passos menores que o passo de Newton, eles devem estar na direção de descida mais profunda: $-\nabla f(x_c)$, em vez da direção de Newton. Seu algoritmo guarda um parâmetro de região de confiança $\delta_c \in \mathbb{R}^+$, o qual se ajusta durante cada iteração. Assim x_+ é escolhido para que seja o único ponto sobre a "curva dogleg" que conecta x_c , C.P., e N (aparece mais adiante) tal que $\|x_+ - x_c\|_2 = \delta_c$ e onde C.P. é como já se disse o ponto de Cauchy e $N = x_c - H_c^{-1} \nabla f(x_c)$. -

GRAFICAMENTE tem-se:

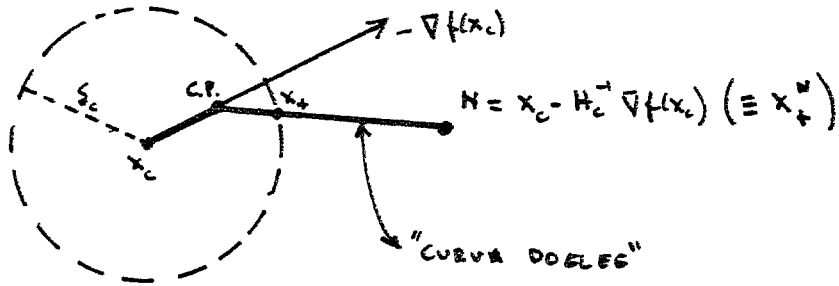


FIGURA 1

A modificação "dupla dogleg" (feita por DENNIS e MEI em 1979), muito usada na prática, leva o passo para direção de Newton, para incluir um passo ao longo de C.P. até a direção de Newton. A distância $\|C.P. - x_c\|_2$ é sempre menor que a distância $\|N - x_c\|_2$ (e algumas vezes na prática, resulta bastante menor)

GRAFICAMENTE tem-se:

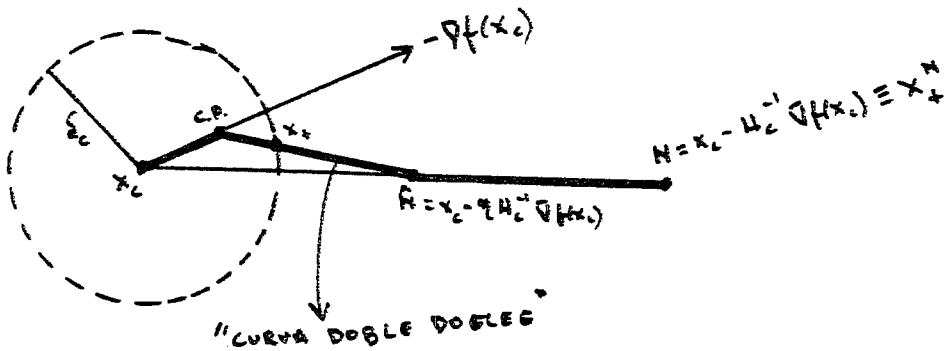


FIGURA 2

Aqui também se escolhe x_+ como o ponto no arco poligonal ("duplo dogleg") tal que $\|x_+ - x_c\|_2 = \delta_c$, salvo que $\|H_c^{-1} \nabla f(x_c)\|_2 < \delta_c$, em cujo caso x_c é o ponto de Newton.

A maneira específica de escolher a curva "duplo dogleg" faz com que ela tenha duas propriedades importantes:

1) Como se avança ao longo da "curva linear" em pedaços de x_c a C.P., a \hat{N} a x_+^N (melhor dizendo: de x_c até x_+^N , passando por C.P. e \hat{N}), a distância desde x_c aumenta monotonamente. (Mais adiante veremos que o ponto \hat{N} corresponde a $\hat{N} = x_c - \eta H_c^{-1} \nabla f(x_c)$, para algum $0 < \eta \leq 1$).

Assim para qualquer $\delta \leq \|H_c^{-1} \nabla f(x_c)\|_2$ existe um único ponto x_+ sobre a curva tal que $\|x_+ - x_c\|_2 = \delta$. Isto faz com que o processo esteja bem definido.

2) O valor do modelo quadrático $m_c(x_c + s)$, decresce monotonamente quando s vai ao longo da curva de x_c até x_+^N , passando por C.P. e \hat{N} . Isto faz com que o processo seja razoável.

O ponto C.P. (ver FIGURA 2) está resolvendo o problema:

$$\min m_c \left[x - \lambda \nabla f(x_c) \right] = f(x_c) - \lambda \|\nabla f(x_c)\|_2^2 + 1/2 \lambda^2 \nabla f(x_c)^t H_c \nabla f(x_c)$$

$$\lambda \in \mathbb{R}$$

Para isto:

$$\nabla_{\lambda} m_c (x - \lambda \nabla f(x_c)) = 0 \Leftrightarrow -\|\nabla f(x_c)\|_2^2 + \lambda \nabla f(x_c)^t H_c \nabla f(x_c) = 0$$

$$\rightarrow \lambda = \frac{\|\nabla f(x_c)\|_2^2}{\nabla f(x_c)^t H_c \nabla f(x_c)} \equiv \lambda_*$$

Logo C.P. = $x_c - \lambda_* \nabla f(x_c)$; e se :

$$\delta_c \leq \lambda_* \|\nabla f(x_c)\|_2 = \frac{\|\nabla f(x_c)\|_2^3}{\nabla f(x_c)^t H_c \nabla f(x_c)}, \text{ o algoritmo}$$

assume um passo de longitude δ_c na direção de descida mais profundo:

$$x_+ = x_c - \lambda_* \nabla f(x_c) ; \text{ onde } \lambda_* = \delta_c / \|\nabla f(x_c)\|_2$$

$$\text{Ou seja : } x_+ = x_c - \left[\delta_c / \|\nabla f(x_c)\|_2 \right] \nabla f(x_c)$$

*** Para que a curva "duplo dogleg" satisfaça a primeira propriedade estabelecida anteriormente, devemos mostrar que o ponto de Cauchy (C.P.) não está mais longe de x_c , do que o ponto de Newton x_+^N $\left[\text{isto é: P.d. } \|s^{C.P.}\|_2 \leq \|s^N\|_2 \text{ (ou que: } \|C.P. - x_c\|_2 \leq \|x_+^N - x_c\|_2) \right]$.

$$\text{Para isto: Seja } s^{C.P.} = -\lambda_* \nabla f(x_c)$$

$$s^N = -H_c^{-1} \nabla f(x_c)$$

$$\begin{aligned} \text{então : } \|s^{C.P.}\|_2 &= \lambda_* \|\nabla f(x_c)\|_2 = \frac{\|\nabla f(x_c)\|_2^3}{\nabla f(x_c)^t H_c \nabla f(x_c)} \leq \\ &\leq \frac{\|\nabla f(x_c)\|_2^3}{\nabla f(x_c)^t H_c \nabla f(x_c)} \frac{\|\nabla f(x_c)\|_2 \|H_c^{-1} \nabla f(x_c)\|_2}{\nabla f(x_c)^t H_c^{-1} \nabla f(x_c)} \end{aligned}$$

$$= \frac{\|\nabla f(x_c)\|_2^4}{\underbrace{\left[\nabla f(x_c)^t H_c \nabla f(x_c) \right] \left[\nabla f(x_c)^t H_c^{-1} \nabla f(x_c) \right]}_{\gamma}} \|s^N\|_2$$

$$= \gamma \|s^N\|_2 \quad (\text{V.11})$$

Se H_c é definida positiva, implica que $\gamma \leq 1$. Logo tem-se que:

$$\|s^{C.P.}\|_2 \leq \gamma \|s^N\|_2 \leq \|s^N\|_2$$

O ponto \hat{N} na curva "duplo dogleg", é escolhido agora e tem a forma: $\hat{N} = x_c - \eta H_c^{-1} \nabla f(x_c)$, para η tal que: $\gamma \leq \eta \leq 1$, e $m_c(x)$ decresce monotonamente ao longo da linha que vai desde C.P. a \hat{N} (V.12)

Como sabemos que $m_c(x)$ decresce monotonamente desde x_c a C.P. e desde \hat{N} a x_+ , e somando-se ainda a afirmação anterior (V.12), tem-se então que $m_c(x)$ decresce monotonamente ao longo de toda a curva "duplo dogleg".

Para satisfazer (V.12), η deve ser escolhido tal que a derivada direcional ao longo da reta que une C.P. com \hat{N} seja negativa para cada ponto sobre este segmento de reta: Para isto:

Parametrizemos este segmento de reta por:

$$x_+(\lambda) = x_c + s^{C.P.} + \lambda (\eta s^{\hat{N}} - s^{C.P.}); \text{ com } 0 \leq \lambda \leq 1$$

Se

Observemos que: a) Se $\lambda = 0 \Rightarrow x_+(\lambda) = x_+(0)$
 $= x_c + s^{C.P.}$
 (estamos em C.P.)

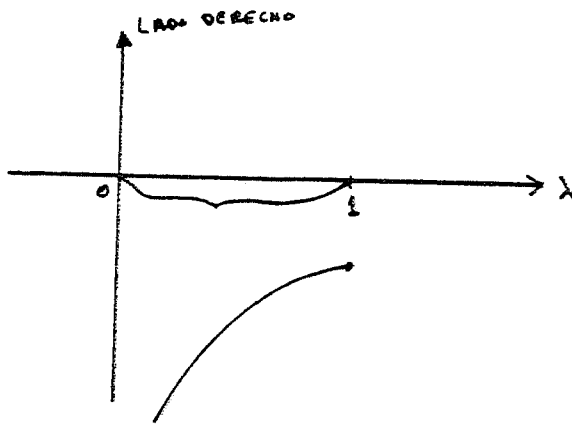
b) Se $\lambda = 1 \Rightarrow x_+(1) = x_c + s^{C.P.} + \eta s^{\hat{N}} - s^{C.P.}$
 (estamos em \hat{N})

A derivada direcional de m_c ao longo deste segmento de reta em $x_+(\lambda)$ é:

$$\begin{aligned} \nabla m_c(x_+(\lambda))^t (\eta s^N - s^{C.P.}) &= \left[\nabla f(x_c) + H_c(s^{C.P.} + \lambda(\eta s^N - s^{C.P.})) \right]^t * \\ &* (\eta s^N - s^{C.P.}) = \\ &= (\nabla f(x_c) + H_c s^{C.P.})^t (\eta s^N - s^{C.P.}) + \lambda (\eta s^N - s^{C.P.})^t H_c (\eta s^N - s^{C.P.}) \end{aligned} \quad (V.13)$$

Como H_c é definida positiva, o lado direito de (V.13) é uma função monotonamente crescente de λ . Portanto, só se necessita exigir que (V.13) seja negativa para $\lambda=1$; para que o intervalo $0 \leq \lambda \leq 1$ também seja negativa.

Graficamente, o que dissemos anteriormente pode ser representado por:



Assim:

$$\begin{aligned} (V.13) \Leftrightarrow & \left[\nabla f(x_t) \right]^t \left[\eta s^N - s^{C.P.} \right] + \left[H_c s^{C.P.} \right]^t \left[\eta s^N - s^{C.P.} \right] + \\ & + \lambda (\eta s^N - s^{C.P.})^t H_c \left[\eta s^N - s^{C.P.} \right] - \lambda (s^{C.P.})^t H_c \left[\eta s^N - s^{C.P.} \right] \end{aligned} \quad (V.14)$$

Como além disso: $s^N = -H^{-1} \nabla f(x_c)$, substituindo em (V.14) tem-se:

$$(V.14) \Leftrightarrow \left[\nabla f(x_c) \right]^t \left[\eta s^N - s^{C.P.} \right] + \eta (-\nabla f(x_c))^t \left[\eta s^N - s^{C.P.} \right] =$$

$$= (1 - \eta) \left[\nabla f(x_c) \right]^t \left[\eta s^N - s^{G.P.} \right] < 0 \quad (V.15)$$

Por outro lado, sabe-se que:

$$s^N = - H_c^{-1} \nabla f(x_c)$$

e de (V.11) : $s^{G.P.} = \gamma s^N$, então:

$$\begin{aligned} \eta s^N - s^{G.P.} &= \eta \left[- H_c^{-1} \nabla f(x_c) \right] - \gamma \left[- H_c^{-1} \nabla f(x_c) \right] \\ &= (\eta - \gamma) H_c^{-1} \nabla f(x_c) \end{aligned}$$

Assim em (V.15) tem-se:

$$(1 - \eta) (\nabla f(x_c))^t \left[\eta s^N - s^{G.P.} \right] = (1 - \eta) (\gamma - \eta) \left[\nabla f(x_c)^t H_c^{-1} \nabla f(x_c) \right] < 0$$

que se cumpre para qualquer $\eta \in (\gamma, 1)$.

De tal forma que \hat{N} pode ser escolhido como qualquer ponto na direção de Newton: $x_c + \eta s^N$, onde $\gamma \leq \eta \leq 1$, dado por (V.11).

Algumas Observações:

i) POWELL originalmente escolheu $\eta = 1$, em cujo caso se tem a curva simples "dogleg".

ii) Apesar de todo o cálculo anterior, muitos testes computacionais mostraram uma "velha" preferência pela direção de Newton, o que parece melhorar o algoritmo.

iii) DENNIS e MEI (1979) sugerem escolher $\eta = 0.8\gamma + 0.2$, produzindo a curva "duplo dogleg" (ver FIGURA 2, deste Capítulo V).

As escolhas de C. P. e \hat{N} especificam completamente a curva "duplo dogleg". A escolha do ponto x_+ na curva tal que $\|x_+ - x_c\|_2 = \delta_c$, é então um problema algébrico barato, como se observa no exemplo que se dará a seguir.

Note que o custo do algoritmo completo é somente da ordem de n^2 operações aritméticas, depois que s_c^N foi calculado.

EXEMPLO 2.- Vejamos o que ocorre em uma iteração para o problema:

$$\begin{aligned} \text{Mín } m_c(x_c + s) \\ \text{s. a. : } \|s\|_2 \leq \delta_c \end{aligned}$$

onde: $m_c(x_c + s) = f(x_c) + \nabla f(x_c)^t s + (1/2) s^t \nabla^2 f(x_c) s$
com os seguintes dados:

$$\begin{aligned} f: \mathbb{R} \longrightarrow \mathbb{R} / f(x_1, x_2) &= x_1^4 + x_1^2 + x_2^2 \\ x_c &= (1, 1)^t \\ \nabla f(x_c) &= (4x_1^3 + 2x_1, 2x_2)^t \Big|_{x_c} = (6, 2)^t \end{aligned}$$

$$\nabla^2 f(x_c) = H_c = \begin{bmatrix} 14 & 0 \\ 0 & 2 \end{bmatrix} \Rightarrow H_c^{-1} = \begin{bmatrix} 1/14 & 0 \\ 0 & 1/2 \end{bmatrix}$$

$$\delta_c = 0.75.$$

a) Tem-se que:

$$s_c^N = -H_c^{-1} \nabla f(x_c) = (-3/7, -1)^t$$

$$\|s_c^N\|_2 = 1.088.$$

b) Como ocorre que $\|s_c^N\|_2 > \delta_c = 0.75$, o algoritmo em consequência calcula o passo no Ponto de Cauchy (C. P.):

Ou seja, se calcula $s^{C.P.}$, recordando que C. P. é o mínimo do modelo quadrático: $m_c(x_c + s)$, na direção de descida mais profundo ($-\nabla f(x_c)$).

Para isto sabemos que:

$$\begin{aligned} s^{C.P.} &= -\lambda_* \nabla f(x_c) = \frac{\|\nabla f(x_c)\|_2^2}{\nabla f(x_c)^t H_c \nabla f(x_c)} \nabla f(x_c) \\ &= -(40/512) \begin{bmatrix} 6 \\ 2 \end{bmatrix} = \begin{bmatrix} -0.469 \\ -0.156 \end{bmatrix} \end{aligned}$$

$$\Rightarrow \|s^{C.P.}\|_2 \approx 0.494$$

c) Como $\|s^{C.P.}\|_2 < \delta_c = 0.75$, calcula-se o passo ao ponto \hat{N}

Para isto:

$s^{\hat{N}} = \eta s_c^N$, onde $\eta = 0.8\gamma + 0.2$ (DENNIS e MEI)

$$\text{com } \gamma = \frac{\|\nabla f(x_c)\|_2^4}{\left[\nabla f(x_c)^t H_c \nabla f(x_c)\right] \left[\nabla f(x_c)^t H_c^{-1} \nabla f(x_c)\right]}$$

(ver página , desta seção)

$$= \frac{40^2}{(512)(32/7)} \approx 0.684.$$

$$\Rightarrow \eta = 0.8 * 0.684 + 0.2 = 0.747$$

$$s^{\hat{N}} = \eta s_c^N = \begin{bmatrix} -0.320 \\ -0.747 \end{bmatrix}$$

Recordemos que o ponto \hat{N} é dado por $\hat{N} = x_c - \eta H_c^{-1} \nabla f(x_c)$ que, em nosso caso, corresponde a: $\hat{N} = (0.680, 0.253)^t$.

d) Como $\|s^N\|_2 = 0.813 > \delta_c = 0.75$, então o passo "duplo dogleg", deve estar na reta que une C.P. com \hat{N} , para o qual $\|s_c\|_2 = \delta_c$. Isto significa dizer que:

$$s_c = s^{C.P.} + \lambda (s^{\hat{N}} - s^{C.P.}); \text{ com } 0 < \lambda < 1.$$

O valor de λ é encontrado resolvendo-se a equação:

$$\|s^{C.P.} + \lambda (s^{\hat{N}} - s^{C.P.})\|_2 = \delta_c$$

donde se obtém: $\lambda \approx 0.867$.

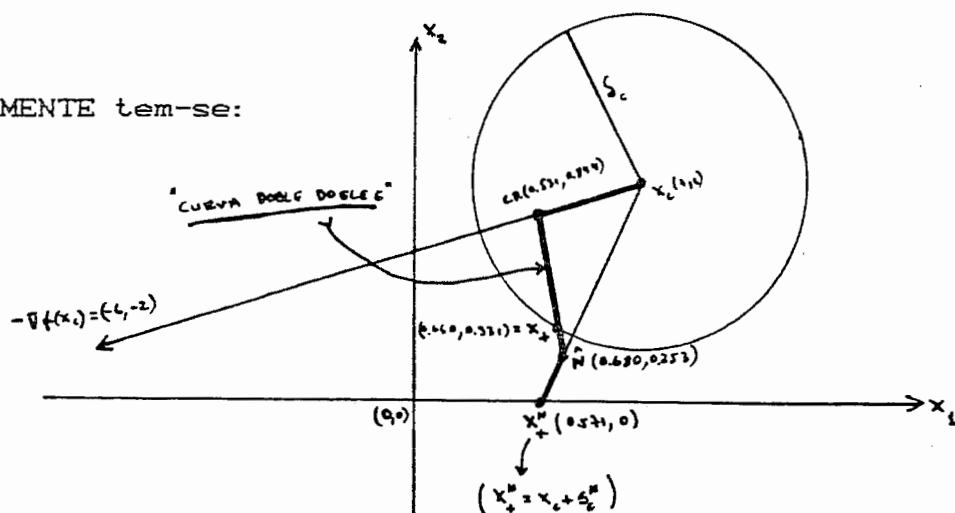
Logo:

$$\begin{aligned} s_c &= s^{C.P.} + \lambda (s^{\hat{N}} - s^{C.P.}) = (-0.469, -0.156)^t + \\ &+ 0.867 \left[(-0.320, -0.747)^t - (-0.469, -0.156)^t \right] \approx \\ &\approx \begin{bmatrix} -0.340 \\ -0.669 \end{bmatrix} \end{aligned}$$

e) Finalmente:

$$x_+ = x_c + s_c \approx (1, 1)^t + (-0.340, -0.669)^t = (0.660, 0.331)^t$$

GRAFICAMENTE tem-se:



V.3. -ATUALIZAÇÃO DA REGIÃO DE CONFIANÇA. -

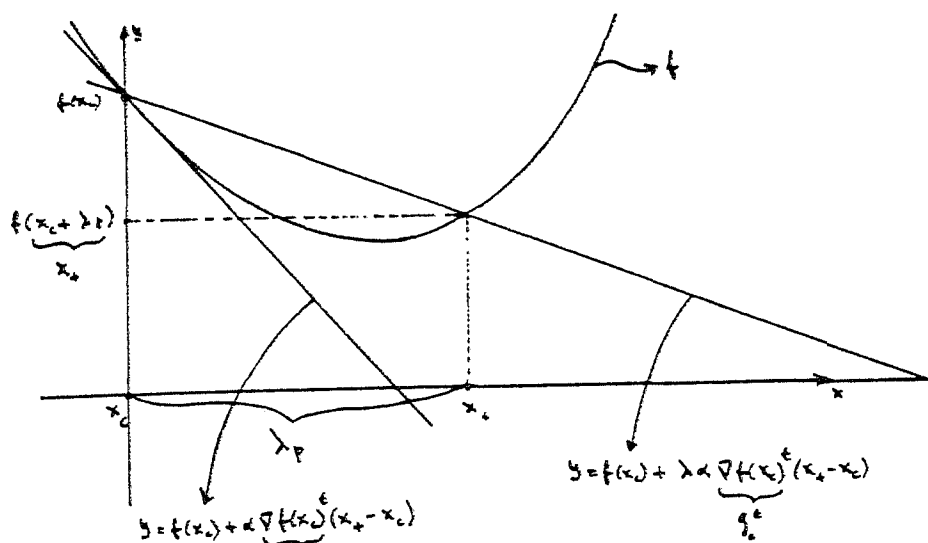
Agora estamos interessados, em escolher se o ponto x_+ , achado usando os métodos vistos anteriormente, é ou não um "ponto aceitável" (ou "iterado satisfatório"). A seguir veremos algumas idéias básicas para elucidar a questão.

A condição mais importante para que x_+ seja efetivamente um "ponto aceitável", é que se cumpra a desigualdade:

$$f(x_+) \leq f(x_c) + \alpha g_c^t (x_+ - x_c) \quad (V.16)$$

onde: $g_c = \nabla f(x_c)$; (ou uma aproximação), α é uma constante tal que $\alpha \in (0, 1/2)$ (em algumas aplicações se escolhe geralmente: $\alpha = 10^{-4}$)

GRAFICAMENTE tem-se:



1. - Se x_+ não é um ponto aceitável, isto é, não satisfaz a desigualdade (V.16), reduz-se a região de confiança a um fator que está entre $1/10$ e $1/2$ e volta-se a solução

aproximada do problema de minimização, usando alguma das técnicas do passo vistas anteriormente (ou seja, pelo passo ótimo localmente restringido "HOOK", ou pelo método "dogleg" ou "duplo dogleg").

O fator de redução que se denotará por λ_* , se determina mediante a seguinte técnica: Se modela $f(x_c + \lambda(x_+ - x_c))$ usando para isto o modelo quadrático $m_q(\lambda)$, que ajusta: $f(x_c)$, $f(x_+)$ e a derivada direcional de f em x_c , na direção $(x_+ - x_c)$, que denotaremos por: $g_c^t(x_+ - x_c)$. A seguir fazemos que o novo raio de confiança δ_c , seja o mínimo deste modelo, o que ocorre quando:

$$\lambda_* = \frac{-g_c^t(x_+ - x_c)}{2 \left[f(x_+) - f(x_c) - g_c^t(x_+ - x_c) \right]}$$

Como obtemos este resultado ?

De fato:

i) Define-se:

$$\hat{f}(\lambda) = f(x_c + \lambda(x_+ - x_c))$$

$$\Rightarrow \hat{f}'(\lambda) = \underbrace{\nabla f[x_c + \lambda(x_+ - x_c)]^t}_{g_c} (x_+ - x_c)$$

$$\Leftrightarrow \hat{f}'(\lambda) = g_c^t(x_+ - x_c)$$

ii) Desejamos ajustar $f(x_c)$ e $f(x_+)$, para isto, λ deve tomar os valores 0 e 1 respectivamente. Ou seja:

$$\text{Se } \lambda = 0 \Rightarrow \hat{f}(0) = f(x_c) \text{ e } \hat{f}'(0) = \nabla f(x_c)^t(x_+ - x_c) \quad (\text{V.17})$$

$$\text{Se } \lambda = 1 \Rightarrow \hat{f}(1) = f(x_+) \quad (\text{V.18})$$

iii) Modelamos $\hat{f}(\lambda)$; por uma quadrática $\hat{m}_q(\lambda)$, que satisfaça (V.17) e (V.18).

Para isto:

$$\hat{m}(\lambda) = A\lambda^2 + B\lambda + C = \hat{f}(\lambda) \rightarrow \hat{f}'(\lambda) = 2A\lambda + B$$

$$\text{Se } \lambda = 0 \Rightarrow C = \hat{f}(0)$$

$$\text{Se } \lambda = 1 \Rightarrow A + B + C = \hat{f}(1) \quad (\text{V.19})$$

$$\text{Além disso } \hat{f}'(0) = B$$

$$\text{Logo em (V.19)} \Rightarrow A = \hat{f}(1) - \hat{f}'(0) - \hat{f}(0)$$

Então:

$$\hat{m}_q(\lambda) = \left[\hat{f}(1) - \hat{f}'(0) - \hat{f}(0) \right] \lambda^2 + \hat{f}'(0)\lambda + \hat{f}(0)$$

Ou seja:

$$\hat{m}_q(\lambda) = \left[f(x_+) - \underbrace{\nabla f(x_c)^t (x_+ - x_c)}_{g_c^t(x_+ - x_c)} - f(x_c) \right] \lambda^2 + \nabla f(x_c)^t (x_+ - x_c) \lambda + f(x_c)$$

iv) Agora minimizamos esta função $\hat{m}_q(\lambda)$. Para isto:

$$\nabla_{\lambda} \hat{m}_q(\lambda) = 0 \Leftrightarrow 2\lambda \left[f(x_+) - \nabla f(x_c)^t (x_+ - x_c) - f(x_c) \right] + \nabla f(x_c)^t (x_+ - x_c) = 0$$

$$\Rightarrow \lambda = \frac{-\nabla f(x_c)^t (x_+ - x_c)}{2 \left[f(x_+) - \nabla f(x_c)^t (x_+ - x_c) - f(x_c) \right]}$$

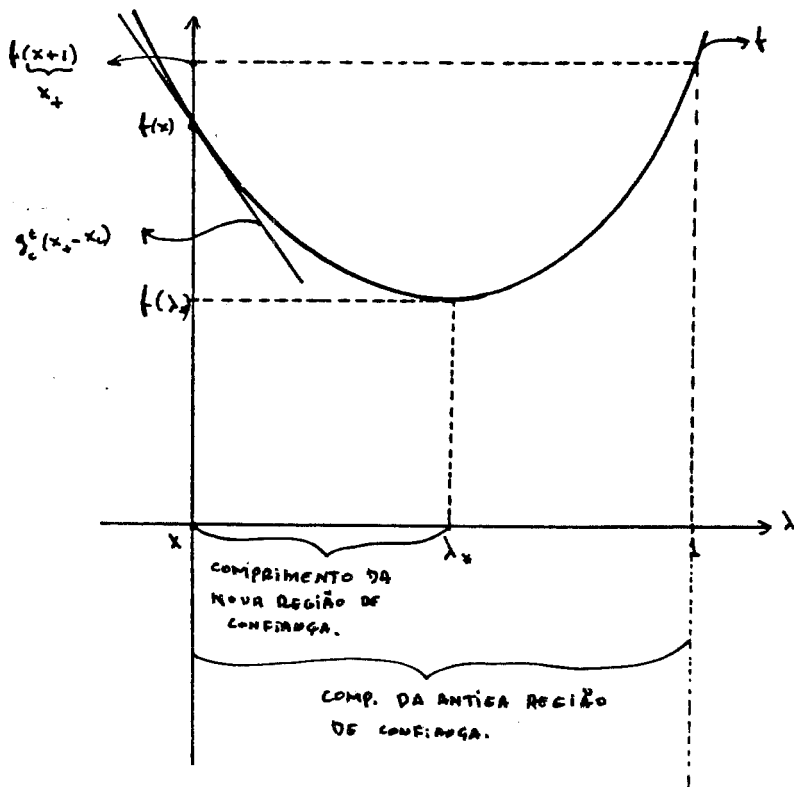
$$= \frac{-g_c^t(x_+ - x_c)}{2 \left[f(x_+) - \nabla f(x_c)^t (x_+ - x_c) - f(x_c) \right]} = \lambda^*$$

Então, o novo raio de confiança δ_+ tem forma:

$$\delta_+ = \lambda_* \|x_+ - x_c\|_2$$

Observação: Se ocorre que este δ_+ é tal que $\notin [\delta_c/10, \delta_c/2]$, então deixa-se ele ficar o mais próximo possível, do ponto final deste intervalo; para assim se evitar erros de precisão finita, especialmente quando o gradiente é calculado usando-se diferenças finitas.

GRAFICAMENTE tem-se:



2. -Suponhamos agora que efetivamente encontrado um ponto x_+ que satisfaz (V.16).

Se x_+ é um passo de Newton desde x_c , então achamos o passo. Atualizamos δ , construímos o novo modelo e continuamos com a iteração seguinte. Entretanto, se $(x_+ - x_c)$ não é o passo Newton, consideramos primeiro ensaiar um passo maior desde x_c (ou dilatar o passo), usando o modelo geral. A justificativa para fazer isto, é evitar ter que avaliar o gradiente (e a Hessiana) em x_+ , a qual representa freqüentemente maior custo, em grandes problemas.

A razão para dilatar o passo, é que a região de confiança pode chegar a se contrair durante o curso do algoritmo e se pode necessitar que ela seja aumentada. Isto ocorre quando se tem uma região onde a cota do passo foi pequena, já que a função não estava bem representada por uma quadrática qualquer e entramos em uma região onde aí sim a função está bem representada por uma função quadrática.

Para se decidir se se faz ou não um passo maior desde x_c , se compara a redução real definida por:

$$\Delta f = f(x_+) - f(x_c)$$

com a redução já dita definida por:

$$\Delta f_{\text{pred}} = m_c(x_+) - f(x_c)$$

e se aceita a x_+ como o iterado, a menos que:

i) O ajuste seja tão bom, como por exemplo que:

$$|\Delta f_{\text{pred}} - \Delta f| \leq 0.1 |\Delta f|$$

que se suspeite que δ_c seja uma aproximação do raio em que m_c representa adequadamente a f , ou

ii) Que a redução real (efetiva) de f seja tão grande

como a presença de curvatura negativa e que portanto esteja implicado um decrescimento rapidamente contínuo em $f(x)$, ou seja:

$$f(x_+) \leq f(x_c) + \nabla f(x_c)^t (x_+ - x_c)$$

Em ambos os casos se aproveitam (ou guardam) x_+ e $f(x_+)$, mas em vez de mover-se diretamente para x_+ , primeiro dobramos δ_c e se calcula um novo x_+ , usando nosso modelo geral. Se (V.16) não é satisfeito para o novo x_+ , volta-se então ao último "passo bom" já calculado. Mas se se satisfaz, se considera dobrar novamente. (Na prática, assim se pode evitar um número significativo de cálculos do gradiente (ou de chamados do gradiente))

OBSERVAÇÃO:

Uma situação interessante que se apresenta é quando a variação do passo deve contrair-se e logo dilatar-se. Isto ocorre quando o algoritmo passa perto de um ponto que parece ser um mínimo. Os passos de Newton se acurtam, o algoritmo faz uso destes passos de Newton e este funciona (bem) como se estivesse convergindo. Então o algoritmo descobre uma saída: os passos de Newton se alargam e o algoritmo muda. A referida conduta é desejável, já que um problema perturbado pode efetivamente ter um mínimo real neste "ponto de distração". Neste caso gostaríamos de ser capazes de aumentar rapidamente o valor de δ_c . -

Suponhamos agora que estamos satisfeitos que x_+ seja nosso iterado seguinte e se necessita atualizar δ_c para δ_+ .

Daremos aqui três alternativas para obter δ_+ . A saber: dobrar, dimidiar ou deixar igual a variação admitida (cota) do passo.

As verdadeiras condições são de certo modo arbitrárias. O que é importante é que: se nosso modelo quadrático está prognosticando bem a função, aumentamos a região de

confiança. Porém, se prognostica pobremente, diminuímos a região de confiança.

Então: a) Se o modelo quadrático prognosticou a redução real da função, claramente, isto é: se $\Delta f \leq 0.75 \Delta f_{pred}$, então:

$$\delta_+ = 2\delta_c$$

b) Se o modelo sobre-estimou muito o decrescimento em $f(x)$, por exemplo: $\Delta f > 0.1 \Delta f_{pred}$ então: $\delta_+ = (1/2)\delta_c$.

c) Se não; deixa-se: $\delta_+ = \delta_c$.

EXEMPLO 3.

Seja $f(x_1, x_2) = x_1^4 + x_1^2 + x_2^2$

onde: $x_c = (1, 1)^t$

$$\nabla f(x_c) = (6, 2)^t$$

$$H_c = \nabla^2 f(x_c) = \begin{bmatrix} 14 & 0 \\ 0 & 2 \end{bmatrix}$$

$$\delta_c = 0.5$$

e suponhamos, do EXEMPLO 1, visto no Cap. V.1, que o passo s_c seja:

$$s_c = (-0.334, -0.335)^t$$

Logo :

$$\begin{aligned} x_+ &= x_c + s_c = (1, 1)^t + (-0.334, -0.335)^t \\ &= (0.666, 0.665)^t \end{aligned}$$

Agora desejamos:

- i) Decidir se x_+ é ou não, um "ponto aceitável".
 ii) Atualizar a região de confiança.

Com efeito: a) Para que x_+ seja um ponto satisfatório, recordemos que deve satisfazer-se a desigualdade:

$$f(x_+) \leq f(x_c) + \alpha \nabla f(x_c)^t (x_+ - x_c); \text{ com } \alpha = 10^{-4} \quad (*)$$

$$\text{Para isto : } f(x_+) = (0.666)^4 + (0.666)^2 + (0.665)^2 \approx 1.083$$

$$f(x_c) = 1^4 + 1^2 + 1^2 = 3$$

$$f(x_c) + \alpha \nabla f(x_c)^t (x_+ - x_c) = 3 + 10^{-4} [6 \quad 2] \begin{bmatrix} 0.666 - 1 \\ 0.665 - 1 \end{bmatrix}$$

$$\approx 2.9997.$$

Como $1.083 \leq 2.9997$ (ou seja: se satisfaz a desigualdade $(*)$) então x_+ é efetivamente um ponto satisfatório.

b) Para atualizar a região de confiança devemos decidir se procuramos ou não um passo maior na iteração em curso. Para isto, vamos usar o teste da comparação:

$$\left| \frac{\Delta f_{\text{pred}}}{\Delta f} - 1 \right| \leq 0.1$$

$$\text{Como } \Delta f = f(x_+) - f(x_c) = 1.083 - 3 = -1.917$$

$$e \quad \Delta f_{\text{pred}} = m_c(x_+) - f(x_c) = f(x_c) + \nabla f(x_c)^t s_c + \frac{1}{2} s_c^t H_c s_c - f(x_c)$$

$$\Delta f_{\text{pred}} = 2.673 + 1/2[-0.334 \quad -0.335] \begin{bmatrix} 14 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} -0.334 \\ -0.335 \end{bmatrix}$$

$$\approx -1.781$$

Então:

$$\left| \frac{\Delta f_{\text{pred}}}{\Delta f} - 1 \right| = \left| \frac{-1.781}{-1.917} - 1 \right| = 0.071 < 0.1$$

De tal modo que dobramos o raio de confiança, ou seja, fazemos $\delta_c = 1.0$ e voltamos ao algoritmo de passo ótimo localmente restrito ("HOOK"), e temos que:

i) Nosso x_c é agora: $x_c = (0.666, 0.665)^t$ (que corresponde ao antigo x_+)

$$\Rightarrow \nabla f(x_c) = (4x_1^3 + 2x_1, 2x_2)^t \Big|_{x_c} = (2.514, 1.33)^t$$

$$H_c = \nabla^2 f(x_c) = \begin{bmatrix} 12x_1^2 + 2 & 0 \\ 0 & 2 \end{bmatrix}_{x_c} = \begin{bmatrix} 7.323 & 0 \\ 0 & 2 \end{bmatrix}$$

$$H_c^{-1} = \begin{bmatrix} 0.137 & 0 \\ 0 & 0.5 \end{bmatrix}$$

$$\begin{aligned} \text{ii) Logo } s_c^N &= -H_c^{-1} \nabla f(x_c) = - \begin{bmatrix} 0.137 & 0 \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} 2.514 \\ 1.330 \end{bmatrix} = \\ &= \begin{bmatrix} -0.344 \\ -0.665 \end{bmatrix} \end{aligned}$$

$$\Rightarrow \|s_c^N\|_2 = 0.749$$

Como $\|s_c^N\| = 0.749 < 1.0$, o algoritmo "HOOK" (antes assinalado) seleciona o passo Newton s_c^N .

OBSERVAÇÃO: Como uma forma de reforçar o que já estudamos, efetuamos uma nova iteração. Tem-se agora que:

$$\begin{aligned} x_{++} &= x_c + s_c = (0.666, 0.665)^t + (-0.334, -0.665)^t = \\ &= (0.332, 0.000)^t \end{aligned}$$

a') Devemos decidir agora se este ponto x_{++} é ou não satisfatório. Para isto:

$$f(x_{++}) = (0.332)^4 + (0.332)^2 + (0.000)^2 \approx 0.122$$

$$f(x_c) = (0.666)^4 + (0.666)^2 + (0.665)^2 = 1.083$$

$$\begin{aligned} f(x_c) + \alpha \nabla f(x_c)^t (x_{++} - x_c) &= 1.083 + 10^{-4} [2.514 \quad 1.33] * \\ & * \begin{bmatrix} 0.332 - 0.666 \\ 0.000 - 0.665 \end{bmatrix} \end{aligned}$$

$$\approx 1.082.$$

Logo como $f(x_{++}) \leq f(x_c) + \alpha \nabla f(x_c)^t (x_{++} - x_c)$, então o ponto $x_{++} = (0.332, 0.000)^t$ é efetivamente um ponto aceitável.

b') Por analogia ao que se faz em b), para atualizar a região de confiança devemos testar se:

$$\left| \frac{\Delta f_{\text{pred}}}{\Delta f} - 1 \right| \leq 0.1$$

Com efeito:

$$\text{Como } \Delta f_{\text{pred}} = \nabla f(x_c)^t s_c + (1/2) s_c^t H_c s_c$$

$$\begin{aligned}
&= [2.514 \quad 1.390] \begin{bmatrix} -0.344 \\ -0.665 \end{bmatrix} + (1/2)[-0.344 \quad -0.665] * \\
&\quad * \begin{bmatrix} 7.323 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} -0.344 \\ -0.665 \end{bmatrix} \\
&= -0.874.
\end{aligned}$$

$$e \quad \Delta f = f(x_{++}) - f(x_c) = 0.122 - 1.083 = -0.961.$$

$$\begin{aligned}
\text{então: } \left| \frac{\Delta f_{\text{pred}}}{\Delta f} - 1 \right| &= \left| \frac{-0.874}{-0.961} - 1 \right| \\
&= 0.091 < 0.1
\end{aligned}$$

Então, novamente dobramos a região de confiança, tendo agora $\delta_c = 2.0$.

CAPITULO VISOBRE UMA EXTENSÃO DO MÉTODO GERAL DE REGIÃO DE CONFIANÇA (REFERIDO A ESTRATÉGIAS DE BUSCA DO PASSO PARA MATRIZ HESSIANA DEFINIDA NEGATIVA).

(NOVOS ALGORITMOS QUE USAM CURVATURA NEGATIVA)

A idéia aqui é apresentar algumas estratégias de seleção do passo que usam informação de segundo ordem, para resolver o problema:

$$\begin{array}{l} \text{mín } f(x): \mathbb{R}^n \longrightarrow \mathbb{R} \\ x \in \mathbb{R}^n \end{array}$$

As referidas estratégias estão idealizadas, no sentido que elas podem usar o valor próprio mais negativo da matriz Hessiana e uma direção de curvatura suficientemente negativa, que denotaremos por q_k , sem especificar detalhadamente como estas quantidades serão obtidas. Entretanto, desenvolveremos ao finalizar o Capítulo, um exemplo particular para fixar algumas idéias que aqui se expõem.

Recordemos que o Algoritmo Região de Confiança gera uma sucessão de pontos $\{x_k\}$ tal que:

a) Em uma k -ésima iteração, o modelo quadrático da função objetivo f , perto do ponto x_k , denotamos por:

$$\phi_k(w) = f_k + g_k^t w + 1/2 w^t B_k w$$

donde $w \in \mathbb{R}^n$ e $B_k \in \mathbb{R}^{n \times n}$ é uma matriz simétrica.

b) Se efetua uma "iteração menor" (a qual possivelmente há que se repetir), que usa o raio da região de confiança Δ_k e a informação contida no modelo quadrático, para calcular um passo p_k , que denotaremos por:

$$p_k (\Delta_k) = p(g_k, B_k, \Delta_k)$$

e logo compara a "redução real" da função objetivo:

$$\text{ared}_k(\Delta_k) = f_k - f(x_k - p_k(\Delta_k))$$

[que é equivalente a notação do Capítulo V.3, quando escrevemos $\Delta f = f(x_+) - f(x_c)$] com a "redução prevista" (ou "prognosticada") pelo modelo quadrático:

$$\text{pred}_k(\Delta_k) = f_k - \psi_k(p_k(\Delta_k)) \quad (\text{VI.1})$$

[que também é equivalente a: $\Delta f_{\text{pred}} = m_c(x_+) - f(x_c)$, do mesmo Capítulo V.3].

c) Se a redução é satisfatória, então pode tomar-se o passo p_k , (questão tratada com detalhe no Capítulo anterior referente a Atualização da Região de Confiança).

Observação:

Outra maneira de escrever (VI.1) é:

$$\text{pred}(g, B, \Delta) = -g^t p(g, B, \Delta) - (1/2) p(g, B, \Delta)^t B p(g, B, \Delta)$$

Com esta nova notação podemos escrever três condições que uma estratégia de seleção de passo deve satisfazer. Elas são:

CONDIÇÃO 1: Existem constantes positivas \bar{c}_1 e σ_1 , tais que: para todo $g \in \mathbb{R}^n$; para toda matriz $B \in \mathbb{R}^{n \times n}$ e para todo $\Delta > 0$:

$$\text{pred}(g, B, \Delta) \geq \bar{c}_1 \|g\| \min \left\{ \Delta, \frac{\sigma_1 \|g\|}{\|B\|} \right\}$$

Isto significa dizer que: o passo deve dar um decréscimo suficiente do modelo quadrático.

CONDIÇÃO 2: Existe uma constante $\bar{c}_2 > 0$, tal que para todo $g \in \mathbb{R}^n$, para toda matriz simétrica $B \in \mathbb{R}^{n \times n}$ e para todo $\Delta > 0$:

$$\text{pred}(g, B, \Delta) \geq \bar{c}_2 (-\lambda_1(B)) \Delta^2 .$$

$\lambda_1(B)$: menor valor próprio da matriz B)

Isto significa dizer que: quando $H(x)$ ($=B$) é indefinida (Recordemos que $H(x)$ é indefinida, quando $[x^t H(x) x] [y^t H(x) y] < 0$, para algum $x, y \in \mathbb{R}^n$) o passo dá um bom decréscimo do modelo quadrático, quando na direção existe uma curvatura suficientemente negativa.

CONDIÇÃO 3: Se B é definida positiva e $\| -B^{-1}g \| \leq \Delta$;
então:

$$p(g, B, \Delta) = -B^{-1}g$$

Isto significa dizer que: se a Hessiana é definida positiva e o passo Newton está dentro da Região de Confiança, então se escolhe o passo Newton (denotado aqui por $p(g, B, \Delta) = -B^{-1}g$).

A) Uma primeira estratégia de seleção de passo nos mostra como uma busca linear que usa informação de segundo ordem, pode ser estendida ao caso indefinido de maneira natural.

(Esta estratégia aparece como algoritmo em GILL e MURRAY (1972))

Notas: i) Em continuação, denotaremos

$$B_k = H(x_k) = \nabla^2 f(x_k)$$

ii) Recordemos que: $k(A) = \|A\| \|A^{-1}\|$ (dependendo da norma) e assim: $k_2(A) = \|A\|_2 \|A^{-1}\|_2 = \sigma_1(A) / \sigma_n(A)$

$$= \frac{\text{maior auto valor de } A}{\text{menor auto valor de } A}$$

ALGORITMO 1 (BUSCA LINEAR INDEFINIDA DO PASSO)

Seja $k \gg 1$;

$$k \leq 1/\varepsilon \text{ (da máxuma)}$$

a) Quando $\lambda_1(B_k) \geq 0$ (menor auto valor de B_k)
e $k_2(B_k) \leq k$

Se $\|B_k^{-1} g_k\| \leq \Delta$; então:

$$\text{o passo } p_k(\Delta) = -B_k^{-1} g_k \quad \left\{ \Leftrightarrow s_k(\Delta) = - \left[\nabla^2 f(x_k) \right]^{-1} \nabla f(x_k) \right\}$$

$$\text{Senão: } p_k(\Delta) = - \left(\Delta / \|B_k^{-1} g_k\| \right) B_k^{-1} g_k$$

b) Quando $\lambda_1(B_k) < 0$ (menor auto valor de B_k) ou $k_2(B_k) > k$;
se escolhe um $\alpha_k \in \mathbb{R}$ tal que $(B_k + \alpha_k I)$ seja definida positiva
e tal que: $k_2(B_k + \alpha_k I) = k$.

Aqui o passo $p_k(\Delta)$ se escolhe de tal maneira que:

bi) Se $\|(B_k + \alpha_k I)^{-1} g_k\| \geq \Delta$; então :

$$p_k(\Delta) = - \left(\Delta / \|(B_k + \alpha_k I)^{-1} g_k\| \right) (B_k + \alpha_k I)^{-1} g_k$$

bii) Se $\|(B_k + \alpha_k I)^{-1} g_k\| < \Delta$ e $\lambda_1(B_k) \geq 0$

$$\text{então: } p_k(\Delta) = -(B_k + \alpha_k I)^{-1} g_k$$

biii) Se $\|(B_k + \alpha_k I)^{-1} g_k\| < \Delta$ e $\lambda_1(B_k) < 0$
então o passo $p_k(\Delta)$ é:

$$p_k(\Delta) = -(B_k + \alpha_k I)^{-1} g_k + \xi q_k$$

donde: i) q_k deve satisfazer:

$$q_k^t B_k q_k \leq c_4 \lambda_1(B_k) \|q_k\|^2 \quad \text{com } c_4 \in (0,11)$$

(ou seja: q_k deve estar em uma direção suficientemente negativa)

ii) $\xi \in \mathbb{R}$; tal que $\|p_k(\Delta)\| = \Delta$ e

$$\text{sgn}(\xi) = -\text{sgn}\left(g_k^t (B_k + \alpha_k I)^{-1} g_k\right)$$

Observação: A convergência do Algoritmo geral, que utiliza esta estratégia para achar o passo, pode ser vista em G. A. SCHULTZ , R. B. SCHNABEL e R. H. BYRD (1985).

B) As duas estratégias de seleção de passo que vêm a seguir, estão baseadas na estratégia do passo dogleg para o caso indefinido:

A primeira destas estratégias não é implementada, já ela usa o pseudo inverso e do auto valor mais negativo e de seu vetor próprio de B_k correspondente. Inclue-se aqui somente para motivar a segunda estratégia, que é similar a primeira, mas que pode ser implementada.

1ª ESTRATÉGIA: ALGORITMO 2 (PASSO "DOGLEG" INDEFINIDO A)

a) Quando $\lambda_1(B_k)$ (menor auto-valor de B_k) > 0 , então:

$$p_k = \text{argmín} \left\{ g_k^t w + 1/2 w^t B_k w : \|w\| \leq \Delta; w \in [g_k, B_k^{-1} g_k] \right\}$$

[donde $\text{argmín} \langle \% \rangle =$ solução do problema:

$$\left. \begin{array}{l} \text{mín} \left\{ g_k^t w + 1/2 w^t B_k w \right\} \\ \text{s. a. : } \|w\| \leq \Delta \end{array} \right]$$

b) Quando $\lambda_1(B_k) \leq 0$; apresentam-se dois casos:

i) Se g_k não é ortogonal ao espaço nulo de $(B_k - \lambda_1 I)$ ó $\|(B_k - \lambda_1 I)^+ g_k\| \geq \Delta$ então:

$$p_k(\Delta) = \text{argmín} \left\{ g_k^t w + (1/2) w^t B_k w : \|w\| = \Delta ; w \in [g_k, v_k] \right\}$$

onde $B_k v_k = \lambda_1 v_k$;

ii) Se g_k é ortogonal ao espaço nulo de $(B_k - \lambda_1 I)$, então o passo p_k é:

$$p_k(\Delta) = -(B_k - \lambda_1 I)^+ g_k + \xi v_k$$

onde ξ é escolhido tal que $\|p_k(\Delta)\| = \Delta$

OBSERVAÇÕES: 1) O cálculo do passo mediante o uso deste algoritmo, pode ser substituído na parte a), pelo passo "dogleg" ou "duplo dogleg", visto anteriormente no Capítulo V.2.

2) Note que minimizar o modelo quadrático sobre um subespaço bidimensional, mistura a execução do chamado "Algoritmo Optimal", quando $n=2$ e que consiste em tomar como passo $p_k(\Delta_k)$:

$$p_k(\Delta_k) = \operatorname{argmín} \left\{ f(x_k) + g_k^t w + (1/2)w^t B_k w ; \|w\| \leq \Delta \right\}$$

ou de forma equivalente; resolver um polinômio de quarto grau em uma variável; o que significa que seu custo computacional é bastante baixo.

2ª ESTRATÉGIA : ALGORITMO 3 (PASSO "DOGLEG" INDEFINIDO B)

(a) Quando $\lambda_1(B_k) > 0$ (menor auto-valor de B_k) se trabalha fazendo uso da mesma idéia que no ALGORITMO 2; desta seção [se pode inclusive usar diretamente o passo "dogleg" ou "duplo dogleg", como foi indicado anteriormente na observação 1)].

(b) Quando $\lambda_1(B_k) \leq 0$, então:

-escolhe-se um α_k , tal que: α_k

$$\alpha_k \in \left[-\lambda_1(H(x_k)) , c_3 \max \left\{ |\lambda_1(H(x_k))| , \lambda_n(H(x_k)) \right\} \right]$$

com $c_3 > 1$.

-define-se $r_k = -(B_k + \alpha_k I)^{-1} g_k$ e o passo $p_k(\Delta)$, escolhe-se de tal modo que:

bi) Se $\|r_k\| \geq \Delta$

$$\rightarrow p_k(\Delta) = \operatorname{argmín} \left\{ g_k^t w + (\alpha/2) w^t B_k w : \|w\| \leq \Delta ; w \in [g_k, r_k] \right\}$$

bii) Se $\|r_k\| < \Delta \Rightarrow p_k(\Delta) = r_k + \xi q_k$; onde ξ e q_k se escolhem como no ALGORITMO 1; parte biii).

EXEMPLO: Seja $f: \mathbb{R} \longrightarrow \mathbb{R} / f(x) = x^3 + 6x^2$

ITER 1: Seja $x_0 = -5$ $\Delta_0 = 0.5$

Então: $\nabla f(x_0) = g_k = 15$; $\nabla^2 f(x_0) = -18$ e

$\lambda_1(B_k) = -18$ {menor auto valor da matriz
 $B_k = \nabla^2 f(x_0)$ }

i) Cálculo do passo:

Uso do Algoritmo 3 (parte b): já que $\lambda_1(B_k)$ é negativo, escolhe-se um α_k , tal que :

$$\alpha_k \in \left[-\lambda_1(B_k), c_3 \max\{|\lambda_1(B_k)|, \lambda_n(B_k)\} \right]$$

com $c_3 = 2.8$ (dado assim, para utilizar a parte: biii) do algoritmo 1) .Ou seja :

$$\alpha_k \in \left[-(-18), 2.8*18 \right] = \left[18, 50.4 \right],$$

Seja por exemplo $\alpha_k = 50.4$ (aqui interessa que $\alpha_k > 18$, para que $(B_k + \alpha_k I)^{-1}$ seja definida positiva)

Agora define-se r_k , onde:

$$r_k = -(B_k + \alpha_k I)^{-1} g_k$$

ou seja:

$$r_k = -(-18+50.4)*15 \approx -0.463$$

$$\rightarrow \|r_k\| = 0.463.$$

Como $\|r_k\| < \Delta_0 = 0.5$, então o passo $p_k(\Delta)$ é dado por:

$$p_k(\Delta) = r_k + \xi q_k ;$$

onde ξ e q_k são eleitos da seguinte maneira:

a) q_k : vetor-direcional (em \mathbb{R} pode ser -1 ou 1)

$$\begin{aligned}
 \text{-Se } q_k = -1 \Rightarrow \operatorname{sgn}(\xi) &= -\operatorname{sgn} \left(\begin{array}{c} q_k^t (B_k + \alpha_k I)^{-1} g_k \\ < 0 & > 0 \\ & < 0 \end{array} \right) \\
 &\Rightarrow \operatorname{sgn}(\xi) > 0
 \end{aligned}$$

Logo: $\xi q_k < 0$

$$\begin{aligned}
 \text{-Se } q_k = 1 \Rightarrow \operatorname{sgn}(\xi) &= -\operatorname{sgn} \left(\begin{array}{c} q_k^t (B_k + \alpha_k I)^{-1} g_k \\ > 0 & > 0 \\ & < 0 \end{array} \right) \\
 &\Rightarrow \operatorname{sgn}(\xi) > 0
 \end{aligned}$$

Logo: $\xi q_k < 0$

OBSERVAÇÃO: Em ambos os casos estamos na direção onde a curvatura é negativa.

b) Cálculo de ξ

Como $\xi \in \mathbb{R}$, tal que $\|p_k(\Delta)\| = \Delta$, tem-se que:

$$\begin{aligned}
 \|p_k(\Delta)\| &= |-(B_k + \alpha_k I)^{-1} g_k + \xi q_k| = |r_k + \xi q_k| = \\
 &= |-0.463 - \xi| = 0.5
 \end{aligned}$$

$$\Rightarrow \xi = 0.037$$

$$\text{Logo } p_k(\Delta) = r_k + \xi q_k = -0.463 + (-0.037) = -0.5$$

$$\Rightarrow x_+ = x_1 = x_0 + p_k(\Delta) = -5 + (-0.5) = -5.5$$

ii) Cálculo de ρ_k

$$\rho_k = \frac{f(x_1) - f(x_0)}{\psi_k(x_0)}$$

$$\text{Como } \psi_k(w) = \nabla f(x_0)^t w + (1/2) w^t \nabla^2 f(x_0) w$$

$$\psi_k(w) = -9 w^2 + 15 w$$

$$\psi_k(s_k) \Big|_{s_k=0.5} = \psi_k(p_k(\Delta)) = -9.75.$$

$$\text{Logo: } \rho_k = (15.125 - 25) / -9.75 \approx 1.0128$$

Então, como $\rho_k > 0.75$, tem-se que:

$$\Delta_1 = 2\Delta_0 = 2 * 0.5 = 1.0$$

ITER 2. Seja $x_1 = -5.5$

$$\Delta_1 = 1.0$$

$$\text{Logo: } \nabla^2 f(x_1) = -21 \quad \text{e} \quad g_k = \nabla f(x_1) = 24.75$$

$$\text{com } \lambda_1(B_k) < 0$$

i) Cálculo do passo:

Novamente usamos o Algoritmo 3 (parte b):

Como $\lambda_1(B_k) < 0$, escolhe-se analogamente

ao que foi visto em ITER 1:

$$\alpha_k \in \left[-(-21), 2.8 * 21 \right] = [21, 58.8]$$

Seja por exemplo $\alpha_k = 58.8$. Interessa-nos que $(B_k + \alpha_k I)^{-1} g_k$ seja definida positiva e além disso que usamos a parte biii) do Algoritmo 1.

-Define-se $r_k / r_k = -(B_k + \alpha_k I)^{-1} g_k$, que aqui corresponde a:

$$r_k = -(-21 + 58.8) * (24.75) \approx -0.655$$

$$\Rightarrow \|r_k\| = 0.655 < \Delta_1 = 1.0$$

Então o passo $p_k(\Delta)$ é dado por:

$$p_k(\Delta) = r_k + \xi q_k$$

Nota: Analogamente a ITER 1, tem-se também que: $\xi q_k < 0$.

Cálculo de ξ :

Sabemos que $\xi \in \mathbb{R} / \|p_k(\Delta)\| = \Delta$. Ou seja:

$$|r + \xi q_k| = |-0.655 - \xi| = 1.0$$

$$\rightarrow \xi = 0.345.$$

$$\text{Logo } p_k(\Delta) = r_k + \xi q_k = -0.655 - 0.345 = -1.0$$

$$\rightarrow x_+ = x_2 = x_1 + p_k(\Delta) = -5.5 + (-1.0) = -6.5$$

ii) Cálculo de ρ

$$\text{Como } \psi_k(w) = \nabla f(x_1)^t w + (1/2) w^t \nabla^2 f(x_1) w$$

$$\rightarrow \psi_k(s_k) = \psi_k(p_k(\Delta)) = -35.25$$

$$\text{Logo } \rho_k = \left[f(-6.5) - f(-5.5) \right] / \psi_k(s_k) = 1.028 > 1$$

$$\text{Como } \rho_k \geq 0.75 \rightarrow \Delta_2 = 2 \Delta_1 = 2.0$$

ITER 3: Seja $x_2 = -6.5$; $\Delta_2 = 2.0$

$$\text{Então: } \nabla^2 f(x_2) = -27 ; g_k = \nabla f(x_2) = 48.75$$

$$\lambda_1(B_k) = -27.$$

i) Cálculo do passo:

Uso do Algoritmo 3 (parte b). Como $\lambda_1(B_k)$ é negativo; escolhemos α_k tal que:

$$\alpha_k \in \left[-\lambda_1(B_k), c_3 * \max \left\{ |\lambda_1(B_k)|, \lambda_n(B_k) \right\} \right]$$

$$\text{Ou seja: } \alpha_k \in \left[27, 75.6 \right].$$

Escolhemos $\alpha_k = 75.6$, (pelas mesmas razões dadas nas iterações anteriores).

- Define-se $r_k / r_k = -(B_k + \alpha_k I)^{-1} g_k$; que neste caso corresponde a:

$$r_k = -(-27 + 75.6) * (48.75) = -1.003.$$

$$\rightarrow \|r_k\| = 1.003 < \Delta_2 = 2.0$$

Então o passo $p_k(\Delta)$ é dado por:

$$p_k(\Delta) = r_k + \xi q_k$$

Nota: Aqui também tem-se $\xi q_k < 0$

Cálculo de ξ

Sabemos que $\xi \in \mathbb{R} / \|p_k(\Delta)\| = \Delta$. Ou seja:

$$\begin{aligned} |r_k + \xi q_k| &= |-1.003 - \xi| = 2.0 \\ \rightarrow \xi &= 0.997 \end{aligned}$$

$$\text{Logo } p_k(\Delta) = r_k + \xi q_k = -1.003 - 0.997 = -2.0$$

$$\rightarrow x_+ = x_3 = x_2 + p_k(\Delta) = -6.5 - 2.0 = -8.5$$

ii) Cálculo de ρ_k

$$\begin{aligned} \text{Como } \psi_k(w) &= (1/2)w^t \nabla^2 f(x_c)w + \nabla f(x_c)^t w \\ &= -13.5 w^2 + 48.75 w \end{aligned}$$

$$\psi_k(s_k) = \psi_k(-2) = -151.5$$

Logo:

$$\rho_k = \left[f(x_3) - f(x_2) \right] / \psi_k(s_k) \Big|_{s_k = -2} \approx -201.7 / -151.5 \approx 1.33$$

De tal modo que, como $\rho_k > 0.75$

$$\rightarrow \Delta_3 = 2 \Delta_2 = 4.0.$$

Nota: Continua-se iterando desta maneira, até que a Hessiana seja definida positiva. Em particular, neste exemplo, isto não ocorrerá já que a função não está limitada inferiormente. -

Vejamos o que ocorre agora, se tomamos $x_0 = -3.5$; ponto situado a direita do ponto $x^* = -4$. Neste caso tem-se:

$$\begin{aligned} \text{ITER 1: Seja: } x_0 &= -3.5 & \Delta_0 &= 0.5 \\ \text{então: } f(x) &= x^3 + 6x^2 \Big|_{x_0} & &= 30.63 \end{aligned}$$

$$\nabla f(x) = 3x^2 + 12x \Big|_{x_0} = -5.25$$

$$\nabla^2 f(x) = 6x + 12 \Big|_{x_0} = -9 \rightarrow B_k = -9$$

$$\begin{aligned} f(x_0 + w) &= f(x_0) + \nabla f(x_0)w + (1/2) w^t \nabla^2 f(x_0)w \\ &= -9/2 w^2 - 5.25 w + 30.63 \end{aligned}$$

$$\rightarrow \psi_k(w) = -9/2 w^2 - 5.25 w$$

i) Cálculo do passo.

Como $\lambda_1(B_k) < 0$, escolhe-se α_k tal que:

$$\alpha_k \in \left[-(-9), c_3 \max \left\{ |\lambda_1(B_k)|, \lambda_n(B_k) \right\} \right]$$

usamos aqui e no que segue $c_3 = 1.2$.

$$\rightarrow \alpha_k \in [9, 10.8]$$

Escolhemos aqui $\alpha_k = 10.8$

$$\rightarrow r_k = - (B_k + \alpha_k I)^{-1} g_k \approx 2.92$$

Como $\|r_k\| = 2.92 \geq \Delta_0$; então o passo $p_k(\Delta)$ é dado por:

$$p_k(\Delta) = \operatorname{argmín} \left\{ g_k^t w + (1/2) w^t B_k w : \|w\| \leq \Delta, w \in [g_k, r_k] \right\}$$

Para resolver este problema, temos:

$$\begin{aligned} L(w, \lambda) &= -9/2 w^2 - 5.25 w + \lambda |w - 0.5| \\ &\rightarrow w = 0.5 \end{aligned}$$

Como w deve ser tal que $w \in [g_k, r_k] = [-5.25, 2.92]$

então $p_k(\Delta) = w = 0.5$.

$$\rightarrow x_1 = x_0 + p_k(\Delta) = -3.5 + 0.5 = -3.0$$

$$\alpha_k \in \left[-(-6) , c_s \max \left\{ \underbrace{|\lambda_1(B_k)| , \lambda_n(B_k)}_6 \right\} \right]$$

$$\Rightarrow \alpha_k \in [6 , 7.2]$$

Seja aqui $\alpha_k = 7.2$

$$\Rightarrow r_k = -(B_k + \alpha_k I)^{-1} g_k = -(-6 + 7.2) * (-9) = 7.5$$

Assim: $\| r_k \| = 7.5 > \Delta_1$; então o passo $p_k(\Delta)$ está dado por:

$$p_k(\Delta) = \operatorname{argmín} \left\{ g_k^t w + \alpha/2 w^t B_k w : \| w \| < \Delta : w \in [g_k, r_k] \right\}$$

Resolução deste problema:

$$L(w, \lambda) = -3 w^2 - 9 w + \lambda | w - 1.0 |$$

$$\Rightarrow w = 1.0$$

Como w está no intervalo $[g_k, r_k] = [-9, 7.5]$

então : $p_k(\Delta) = w = 1.0$

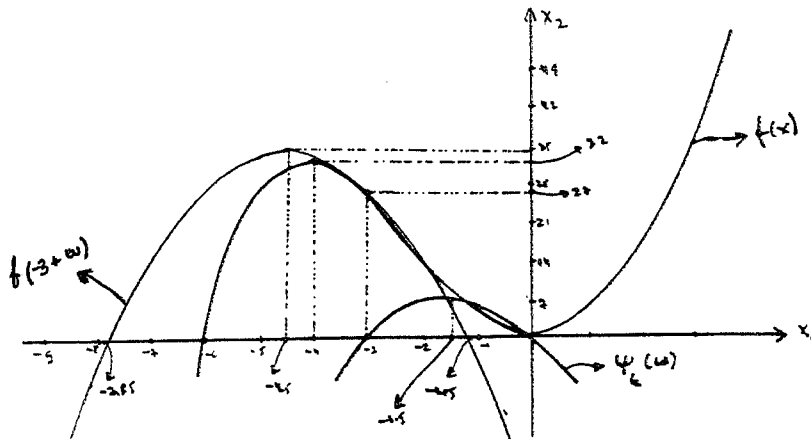
$$\Rightarrow x_2 = x_1 + p_k(\Delta) = -3.0 + 1.0 = -2.0$$

ii) Cálculo de ρ_k :

$$\rho_k = (f(x_2) - f(x_1)) / \psi_k(w) \Big|_{w=1} = 16 - 27 / -12 = 0.92 > 0.75$$

$$\Rightarrow \Delta_2 = 2 * \Delta_1 = 2 * 1.0 = 2.0$$

GRÁFICO:



ITER 3: Seja $x_2 = -2.0$ e suponhamos novamente (isto se faz a propósito) que $\Delta_2 = 1.0$ (e não como foi encontrado na iteração anterior (ITER 2)).

$$\begin{aligned} \text{Então: } f(x_2) &= 16 & \nabla^2 f(x_2) &= 0 = B_k \\ \nabla f(x_2) &= -12 \\ f(x_2+w) &= f(x_2) + \nabla f(x_2)^t w + (1/2)w^t B_k w \\ &= -12w + 16 \\ \Rightarrow \psi_k(w) &= -12w \end{aligned}$$

i) Cálculo do passo:

Como $\lambda_1(B_k) < 0$, escolhemos α_k , tal que :

$$\alpha_k \in \left[0, c_3 \max \left\{ |\lambda_1(B_k)|, \lambda_n(B_k) \right\} \right]$$

↓
0

$$\Rightarrow \alpha_k = 0$$

$$\Rightarrow r_k = -(0 + 0)^{-1} g_k : \text{ não existe } (\#)$$

Nota: Neste caso, voltamos ao ponto $x_1 = -3$, como foi dito no Capítulo V, parágrafo V.3, e duplicamos $\Delta_1 = 1.0$; ou seja: seja $x_2 = x_1 = -3$ e $\Delta_2 = 2 * \Delta_1 = 2$.

$$\begin{aligned} \text{Com isto: } \nabla f(x_2) &= \nabla f(-3) = -9 \\ \nabla^2 f(x_2) &= -6 ; \Delta_2 = 2.0 \end{aligned}$$

i') Cálculo do passo:

Como $\lambda_1(B_k) \leq 0$; escolhe-se α_k tal que:

$$\alpha_k \in \left[-(-6), 1.2 * \max \left\{ |\lambda_1(B_k)|, \lambda_n(B_k) \right\} \right]$$

↓
6

$$\Rightarrow \alpha_k \in [6, 7.2]$$

$$\Rightarrow r_k = -(B_k + \alpha_k I)^{-1} g_k = 7.5$$

$$\text{Então, como } \|r_k\| = 7.5 > 2.0$$

$$\Rightarrow p_k(\Delta) = \operatorname{argmín} \left\{ g_k^t w + (1/2)w^t B_k w, \|w\| < \Delta, w \in [-9, 7.5] \right\}$$

Resolução deste problema:

$$L(w, \lambda) = -3 w^2 - 9 w + \lambda |w - 2|$$

$$\rightarrow w = 2.0$$

e como w está no intervalo :

$$[g_k, r_k] = [-9, 7.5]$$

então:

$$p_k(\Delta) = w = 2.0$$

$$\rightarrow x_3 = x_2 + p_k(\Delta) = -3 + 2 = -1$$

ii) Cálculo de ρ_k :

$$\rho_k = \frac{f(x_3) - f(x_2)}{\psi_k(w) \Big|_{v=2.0}} = \frac{5 - 27}{-30} = 0.7\bar{3} < 0.75$$

$$\left[\text{onde } \psi_k(w) = -3w^2 - 9w \right]$$

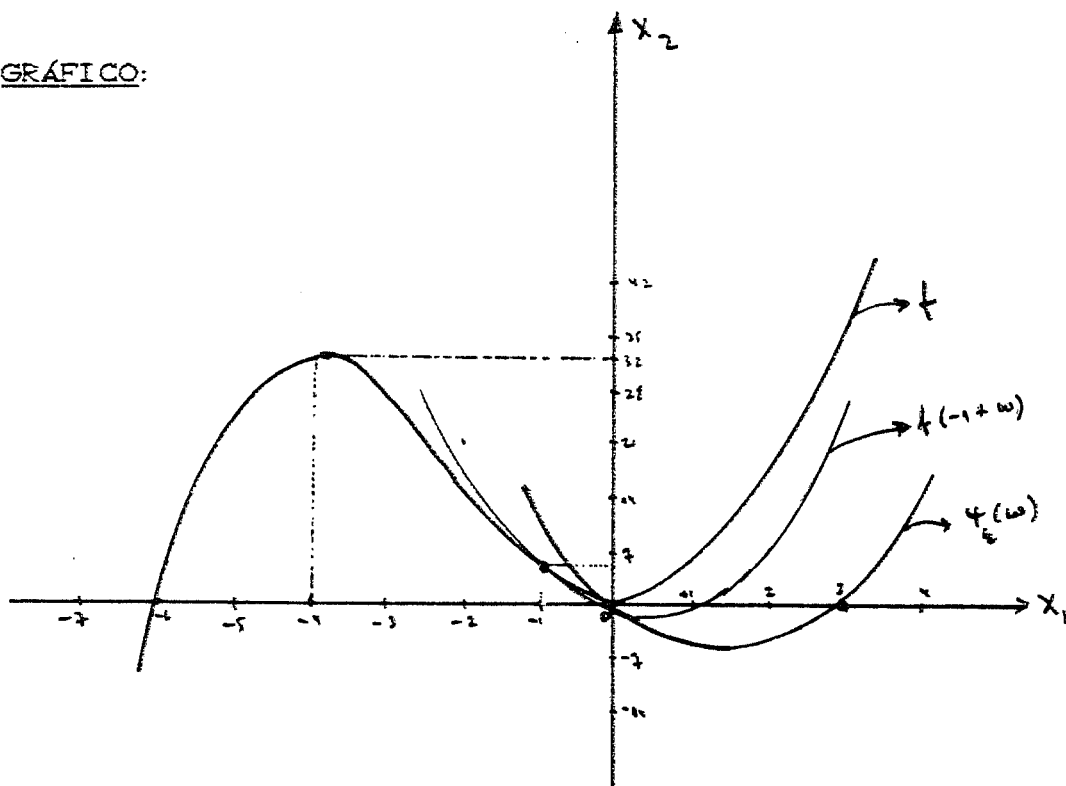
ITER 4: Seja $x_3 = -1.0$ $f(x_3) = 5.0$

$$\Delta_3 = 1.0$$

$\nabla f(x_3) = -9$ $\nabla^2 f(x_3) = 6 > 0$; e voltamos

ao caso em que a curvatura é positiva.

GRÁFICO:



CAPÍTULO VIIDESCRIÇÃO E EXEMPLO DO ALGORITMO DE REGIÃO DE CONFIANÇA

Sejam dados : x_k , H_k , Δ_k , y $0 \leq s < 0.25$

1) Problema (P_k):

$$\text{Mín } q_k(d) = \text{Mín } \psi_k(w) = \text{Mín } f(x_k) + \nabla f(x_k)^t w + (1/2) w^t \nabla^2 f(x_k) w$$

$$\|d\| \leq \Delta_k \quad \|w\| \leq \Delta_k$$

Resolução do Problema (P_k):

i) Obtém-se uma solução aproximada $s_k (\equiv d_k)$

ii) Calcula-se o critério $\rho_k \left[= \frac{f(x_k + s_k) - f(x_k)}{\psi_k(s_k)} \right]$

iii) Ver observação sobre a Resolução do Problema (P_k).

2) Atualização de x_k :

$$\text{Se } \rho_k \geq s \quad \rightarrow \quad x_{k+1} = x_k + s_k$$

$$\text{Se não :} \quad \rightarrow \quad x_{k+1} = x_k$$

3) Atualização da Região de Confiança:

$$\text{Se } \rho_k \geq 0.75 \quad \rightarrow \quad \Delta_{k+1} = 2 \Delta_k$$

$$\text{Se } \rho_k \geq 0.25 \quad \rightarrow \quad \Delta_{k+1} = (1/2) \Delta_k$$

$$\text{Se não :} \quad \Delta_{k+1} = \Delta_k$$

4) Atualização de H_k :

Fórmulas de quase-Newton.

Observação sobre a Resolução do Problema (P_k)

Recordemos que o problema (P_k) é :

$$(P_k) : \text{Mín } \psi_k(w) \\ \text{s. a. } \|v\| \leq \Delta_k$$

s_k : solução de (P_k)

- Si $\|s_k\| < \Delta_k$, então s_k é solução de :

$$\boxed{H_k s = -\nabla f(x_k)} \quad \left[\Leftrightarrow \nabla^2 f(x_k) s = -\nabla f(x_k) \right]$$

- Si $\|s_k\| = \Delta_k$, então s_k é solução de :

$$\boxed{(H_k + \mu I) s = -\nabla f(x_k)}$$

μ : Multiplicador de Lagrange associado a restrição equivalente:

$$\|s_k\|^2 = \Delta_k^2$$

Apresentamos a seguir um exemplo onde se aplica o que foi dito anteriormente:

$$\text{Seja } f: \mathbb{R} \longrightarrow \mathbb{R} / f(x) = x^3 + 6x^2.$$

Interessa-nos achar os valores extremos da referida função, e nos interessa especificamente achar um mínimo, iterando perto do ponto $x_0 = 0$.

Para isto temos:

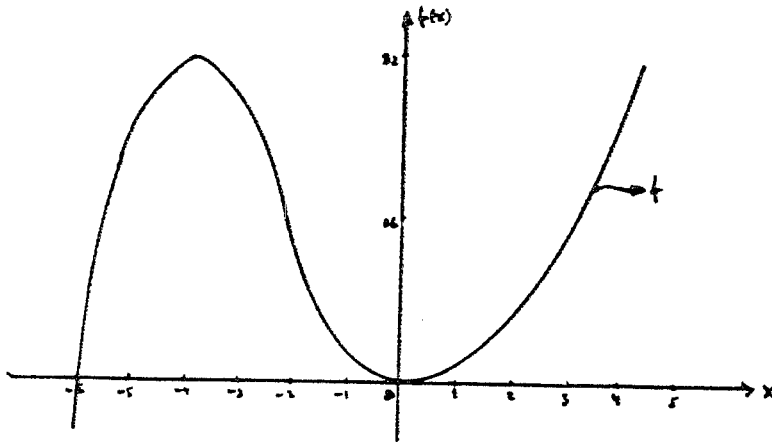
$$\begin{cases} \nabla f(x) = 3x^2 + 12x \\ \nabla^2 f(x) = 6x + 12 \end{cases}$$

$$\text{Mín em } (0,0) \rightarrow \nabla^2 f(x) |_{(0,0)} = 12$$

Máx em $(-4, 32) \rightarrow \nabla^2 f(x) |_{(0,0)} = -12$

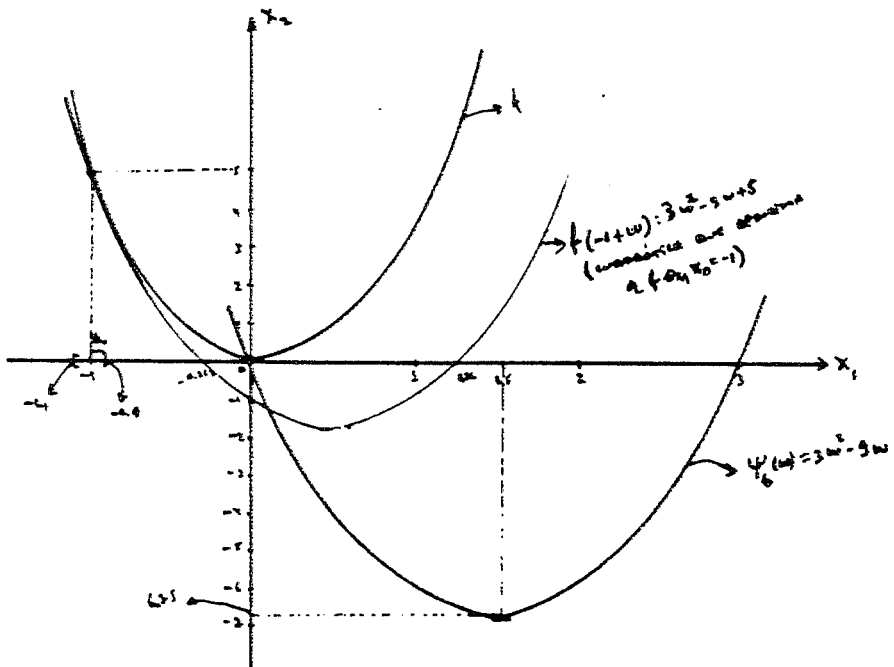
Ponto de inflexão : $(-2, 16)$

GRÁFICO:



ITERAÇÃO 1

GRÁFICO:



i) Seja $x_0 = -1$ $\Delta_0 = 0.1$

$$f(x_0 + w) = f(x_0) + \nabla f(x_0)^t w + (1/2) w^t \nabla^2 f(x_0) w$$

$$f(x_0) = f(-1) = 5 \quad \nabla f(x_0) = -9 \quad \nabla^2 f(x_0) = 6$$

$$\Rightarrow f(-1 + w) = 3w^2 - 9w + 5 \quad \Rightarrow \psi_k(w) = 3w^2 - 9w$$

ii) Mín $\psi_k(w) = \text{Mín } (3w^2 - 9w)$

$$\text{s. a. } \|v\| \leq \Delta_0 \quad \text{s. a. } \|v\| \leq 0.1$$

$$L(w, \lambda) = -9w + 3w^2 + \lambda |w - 0.1|$$

$$\nabla_w L(w, \lambda) = -9 + 6w + \lambda = 0$$

$$\nabla_\lambda L(w, \lambda) = |w - 0.1| = 0 \Rightarrow w = 0.1 = s_k : \text{ Solução do problema restrito.}$$

$$\Rightarrow \psi_k(w) \Big|_{v=0.1} = -0.87.$$

Nota: Observe que se a função $\psi_k(w)$ for minimizada sem a restrição, teríamos que:

$$\psi_k(w) = -9w + 3w^2 \Rightarrow \nabla \psi_k(w) = -9 + 6w = 0 \Rightarrow w = 1.5$$

iii) Calcula-se agora $\rho_k = \frac{f(x_0 + s_k) - f(x_0)}{\psi_k(s_k)} =$

$$= \frac{f(-1 + 0.1) - f(-1)}{\psi_k(0.1)} \approx 0.9988505 < 1$$

iv) Atualização de x_k :

$$\text{Como } \rho_k \geq s \text{ (onde } 0 \leq s \leq 0.25) \Rightarrow x_{k+1} = x_k + s_k$$

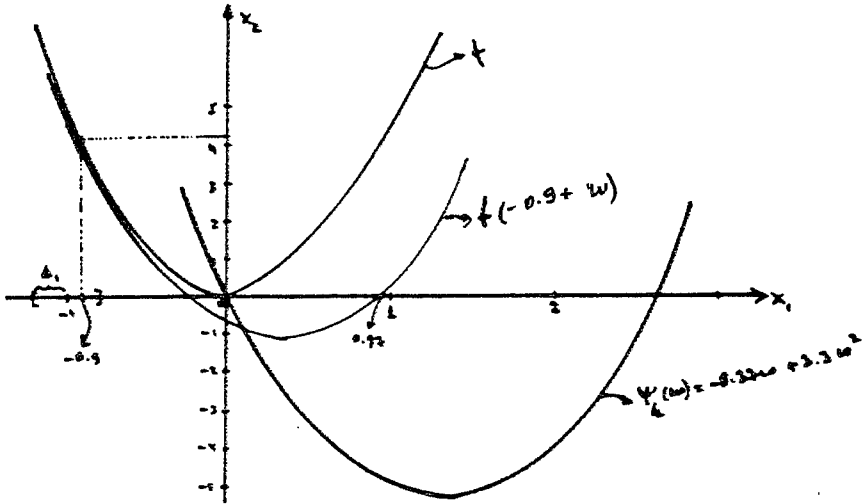
$$\Rightarrow x_1 = x_0 + 0.1 = -0.9$$

v) Atualização da região de confiança Δ_k

$$\text{Como } \rho_k \geq 0.75 \rightarrow \Delta_{k+1} = 2 \Delta_k \rightarrow \Delta_1 = 2\Delta_0 = 2 \cdot 0.1 = 0.2$$

ITERAÇÃO 2

GRÁFICO:



i) Seja $x_1 = -0.9$ $\Delta_1 = 0.2$

$$f(x_1 + w) = f(x_1) + \nabla f(x_1)^T w + (1/2) w^T \nabla^2 f(x_1) w$$

$$f(x_1) = f(-0.9) = 4.131 \quad \nabla f(x_1) = \nabla f(-0.9) = -8.37$$

$$\nabla^2 f(x_1) = \nabla^2 f(-0.9) = 6.6$$

$$\rightarrow f(-0.9 + w) = 4.131 - 8.37w + 3.3 w^2$$

$$\rightarrow \psi_k(w) = -8.37 w + 3.3 w^2$$

ii) $\text{Mín } \psi_k(w) = -8.37w + 3.3 w^2$

$$\text{s. a. } \|w\| \leq 0.2$$

$$L(w, \lambda) = -8.37 w + 3.3 w^2 + \lambda |w - 0.2|$$

$$\nabla_w L(w, \lambda) = -8.37 + 6.6w + \lambda = 0$$

$$\nabla_\lambda L(w, \lambda) = w - 0.2 = 0 \rightarrow w = 0.2 (> 0) = s_k \text{ Solução do problema restrito.}$$

i) Seja $x_2 = -0.7$ $\Delta_2 = 0.4$

$$f(x_2 + w) = f(x_2) + \nabla f(x_2)^T w + (1/2) w^T \nabla^2 f(x_2) w$$

$$f(-0.7) = 2.597 \quad \nabla f(-0.7) = -6.93 \quad \nabla^2 f(-0.7) = 7.8$$

$$f(-0.7 + w) = 2.597 - 6.93 w + 1/2 * 7.8 w^2$$

$$\Rightarrow \psi_k(w) = 3.9 w^2 - 6.93 w$$

ii) Mín $\psi_k(w) : 3.9 w^2 - 6.93 w$

$$s. a. \quad \|v\| \leq 0.4$$

$$L(w, \lambda) = 3.9 w^2 - 6.93 w + \lambda |w - 0.4|$$

$$\nabla_w L(w, \lambda) = 7.8 w - 6.93 + \lambda = 0$$

$$\nabla_\lambda L(w, \lambda) = |w - 0.4| = 0 \quad \rightarrow \quad w = 0.4 (> 0) = s_k$$

Solução do problema restrito.

Nota: Mín $\psi_k(w)$ sem restrições; tem-se que :

$$\nabla \psi_k(w) = 7.8 w - 6.93 = 0 \quad \rightarrow \quad w = 0.8884615$$

$$\Rightarrow \psi_k(w) \Big|_{v=0.4} = -2.148$$

iii) Cálculo de ρ_k :

$$\rho_k = \frac{f(-0.7 + 0.4) - f(-0.7)}{\psi_k(w) \Big|_{0.4}} = 0.9702048 < 1$$

iv) Atualização de x_k

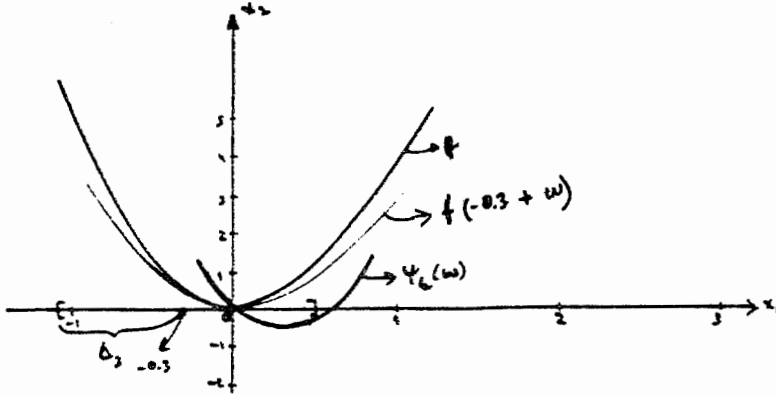
$$\text{Como } \rho_k \geq s \quad (0 \leq s \leq 0.25) \quad \rightarrow \quad x_3 = x_2 + s_k = -0.7 + 0.4$$

$$\rightarrow \quad x_3 = -0.3$$

v) Atualização da região de confiança Δ_k :

$$\text{Como } \rho_k \geq 0.75 \quad \rightarrow \quad \Delta_3 = 2 * \Delta_2 = 2 * 0.4 \quad \rightarrow \quad \Delta_3 = 0.8$$

ITERAÇÃO 4

GRÁFICO:

i) Seja $x_3 = -0.3$ $\Delta_3 = 0.8$

$$f(-0.3) = 0.513 \quad \nabla f(-0.3) = -3.33 \quad \nabla^2 f(-0.3) = 10.2$$

$$\begin{aligned} f(-0.3 + w) &= f(-0.3) + \nabla f(-0.3) w + \frac{1}{2} w \nabla^2 f(-0.3) w \\ &= 0.513 - 3.33w + 5.1w^2 \end{aligned}$$

$$\Rightarrow \psi_k(w) = 5.1 w^2 - 3.3 w$$

ii) $\text{Mín } \psi_k(w): 5.1 w^2 - 3.3 w$

$$\text{s. a. } \|w\| \leq 0.8$$

Resolvendo este problema tem-se que :

$$w = 0.3235 = s_k \quad \text{onde } s_k = 0.3235 \in \left[-1.1, 0.5 \right]_{\Delta_k}$$

$$\Rightarrow \psi_k(0.3235) = -0.5338235$$

iii) Cálculo de ρ_k

$$\rho_k = \frac{f(-0.3 + 0.3235) - f(-0.3)}{\psi_k(0.3235)} \approx 0.9547602 < 1$$

iv) Atualização de x_k

$$\begin{aligned} \text{Como } \rho_k \geq s \quad (0 \leq s < 0.25) &\Rightarrow x_4 = x_3 + s_k = \\ &= -0.3 + 0.3235 \\ &\Rightarrow x_4 = 0.0235 \end{aligned}$$

v) Atualização da Região de Confiança Δ_k

$$\text{Como } \rho_k \geq 0.75 \quad \Rightarrow \Delta_4 = 2 * \Delta_3 = 1.6$$

Observação: Outro modo de resolver o problema:

$$\begin{aligned} \text{Mín } \psi(w) &\equiv \text{Mín } (5.1 w^2 - 3.3 w) \\ \text{s. a. } \|v\| \leq 0.8 &\quad \text{s. a. } \|v\| \leq 0.8 \end{aligned}$$

Este problema tem uma solução do tipo:

$$w(\mu) = - \left[\nabla^2 f(x_c) + \mu I \right]^{-1} \nabla f(x_c)$$

a menos que $\|w(0)\| \leq \Delta_k = 0.8$; em cujo caso:

$$\begin{aligned} w(0) = s_c^N &= - \left[\nabla^2 f(x_c) \right]^{-1} \nabla f(x_c) \\ \text{é a solução.} \end{aligned}$$

Vejamos aqui:

$$w(0) = - \left[\nabla^2 f(-0.3) \right]^{-1} \nabla f(-0.3) =$$

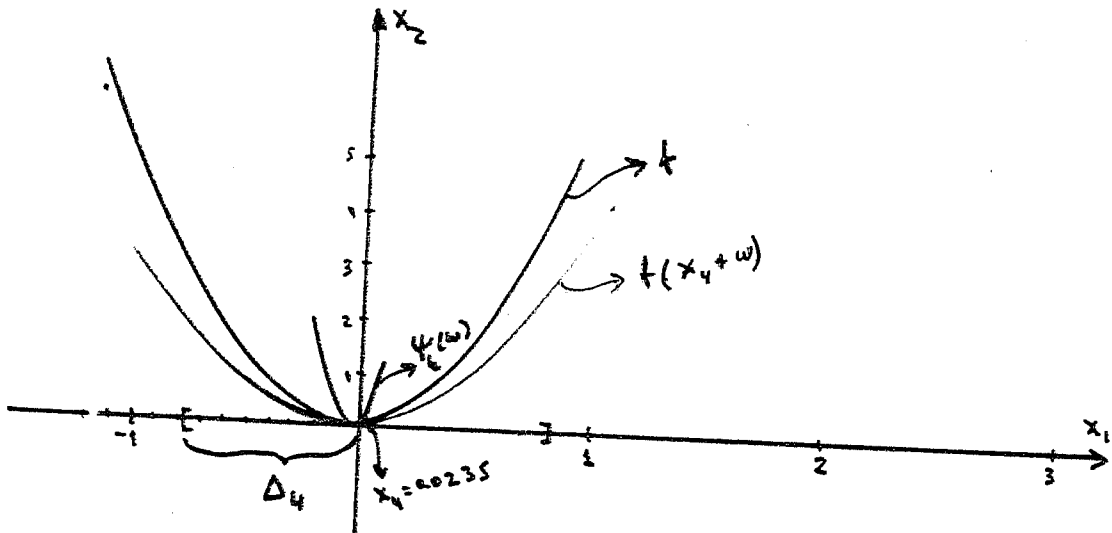
$$= -(10.2)^{-1} * (-3.3) \approx 0.3264705.$$

$$\text{onde } \|w(0)\| < 0.8 = \Delta_k$$

$$\begin{aligned} \Rightarrow w(0) = s_c^N &= 0.3264705 \text{ é a} \\ \text{solução (= } s_k \text{)} \end{aligned}$$

ITERAÇÃO 5

GRÁFICO:



- i) Seja $x_4 = 0.0235$ $\Delta_4 = 1.6$ (inclusive pode permanecer
e aqui faz-se assim: $\Delta_4 = \Delta_3 = 0.8$)

$$f(0.0235) = 3.33 \cdot 10^{-3} \quad \nabla f(0.0235) = 0.283$$

$$\nabla^2 f(0.0235) = 12.141$$

$$\Rightarrow f(x_4 + w) = 3.33 \cdot 10^{-3} + 0.283 w + (1/2)(12.141)w^2$$

$$\psi_k(w)$$

$$\Rightarrow \psi_k(w) = 6.0705 w^2 + 0.283 w$$

- ii) Mín $\psi_k(w) : 6.0705 w^2 + 0.283 w$
s. a. $\|v\| \leq 0.8$

$$\Rightarrow w \approx -0.0230 = s_k \quad \Rightarrow \psi_k(w) \Big|_{v=-0.0230} \approx -3.2287 \cdot 10^{-3}$$

- iii) Cálculo de ρ_k

$$\rho_k = \frac{f(0.0235 - 0.0230) - f(0.0235)}{\psi(-0.0230)} = 1.0308 > 1$$

(o qual significa entre outras coisas, que o mínimo que estávamos buscando é menor que : $x_4 = 0.0235$)

iv) Atualização de x_k

$$\begin{aligned} \text{Como } \rho_k \geq s \quad (0 \leq s \leq 0.25) &\rightarrow x_5 = x_4 + s_k \\ x_5 &= 0.0235 + (-0.0230) \\ x_5 &\approx 0.0005. \end{aligned}$$

v) Atualização da Região de Confiança Δ_k :

$$\text{Como } \rho_k \geq 0.75 \rightarrow \Delta_5 = 2 * \Delta_4 = 1.6$$

(onde novamente se pode deixar o mesmo $\Delta_5 = \Delta_4 = 0.8$)

REFERÊNCIAS BIBLIOGRÁFICAS.

- (1) BROYDEN, C. G. (1965) "A class of methods for solving nonlinear simultaneous equations", Math. Comp. 19, 577-593.
- (2) DENNIS, J. E. Jr y H. H. MEI (1979), "Two new unconstrained optimization algorithms which use function and gradient values", Journal of Optimization Theory and its Applications 28, 453-482.
- (3) DENNIS, J. E. y R. B. SCHNABEL (1983), "Numerical Methods for Unconstrained Optimization and Nonlinear Equations" (Prentice-Hall, Englewood Cliffs, N. J.).
- (4) FLETCHER, R. (1972) "An algorithm for solving linearly constrained optimization problems", Math. Programming 2, 133-165.
- (5) FLETCHER, R. (1980). Practical Methods of Optimization, Volume 1: Unconstrained Optimization, John Wiley and Sons.
- (6) FLETCHER, R. (1981). Practical Methods of Optimization, Volume 2: Constrained Optimization, John Wiley and Sons.
- (7) P. E. GILL y W. MURRAY (1972), "Quasi-Newton methods for unconstrained optimization", Journal Inst. Math. Appl, 9; 91-108.
- (8) GOLDFELD, S., QUANDT, R. y TROTTER, H. (1966). "Maximization by quadratic hill climbing", Econométrica 34, 541-551.
- (9) GRIFFITH, R. E. y STEWART, R. A. (1961). "A nonlinear programming technique for the optimization of continuous processing systems", Management Sci. 7, 379-392.
- (10) HEBDEN, M. D. (1973). "An algorithm for minimization using exact second derivatives", Atomic Energy Research Establishment, Report T. P. 515 Harwell, England.
- (11) LEVENBERG, K (1944) "A Method for the solution of certain nonlinear problems in least squares", Quart. Appl. Math. 2, 164-168.
- (12) MADSEN, K. (1975) "An algorithm for minimax solution of overdetermined systems of nonlinear equations", J. Inst. Math. Appl. 16, 321-328.

- (13) MARQUARDT, D.W. (1963). "An algorithm for least squares estimation of nonlinear parameters", Siam J. Appl. Math. 11, 431-441.
- (14) MORE, J.J. (1978). "The Levenberg-Marquardt algorithm: Implementation and theory, "Proceeding of the Dundee Conference on Numerical Analysis, G.A. Watson, ed., Springer-Verlag.
- (15) MORE J.J. y SORENSEN D.C. (1981). "Newton's Methods" T-R. Applied Mathematics Division. Argonne National Laboratory. Argonne Illinois.
- (16) MORE, J.J. y SORENSEN D.C. (1983). "Computing a trust region step", Siam Journal on Scientific and Statistical Computing 4, 553-572.
- (17) POWELL, M.J.D. (1970). "A hibryd method for nonlinear equations", Numerical Methods for Nonlinear Algebraic Equations, P. Rabinowitz, ed., Gordon and Breach.
- (18) REINSCH, C.H. (1967). "Smoothing by Spline functions", Numer. Math. 10, 177-183.
- (19) REINSCH, C.H. (1971). "Smoothing by Spline functions II, Numer. Math. 16, 451-454.
- (20) G.A. SCHULTZ, R.B. SCHNABEL y R.H. BYRD (1985). "A family of trust region based algorithms for unconstrained minimization with strong global convergence properties", Siam Journal Numer. Anal. vol 22, n^o1; 47-67.
- (21) SORENSEN, D.C. (1982), "Newton's Methods with a model trust region modification, Siam Journal on Numer. Anal. 19, 409-426.
- (22) ZOWE, J (1988), "The BT-Algorithm for minimizing a nonsmooth functional subject to linear constraints". T.R. University of Bayreuth, D-8580. Bayreuth, W.-Germany.