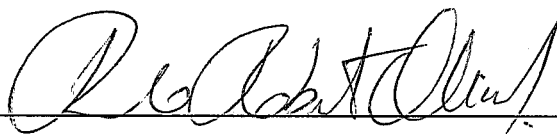


Algoritmos de Região de Confiança para Otimização com Restrições de Igualdade

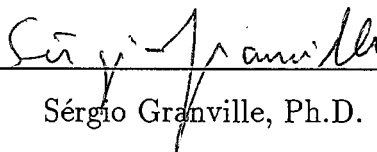
Sonia Maria Moreira

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 OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

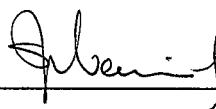
Aprovado por :



Paulo Roberto Oliveira, D.Ing.
(Presidente)



Sérgio Granville, Ph.D.



Susana Schreimberg de Makler, D.Sc.

RIO DE JANEIRO, RJ – BRASIL

Abril de 1994

MOREIRA, SONIA MARIA

Algoritmos de Região de Confiança para Otimização com Restrições de Igualdade [Rio de Janeiro] 1994.

IX, 80 p., 29,7cm (COPPE/UFRJ, M.Sc., Engenharia de Sistemas e Computação, 1994)

Tese – Universidade Federal do Rio de Janeiro, COPPE.

1 – Programação Não-Linear

2 – Região de Confiança

3 – Otimização com Restrições de Igualdade

4 – Subproblema de Celis-Dennis-Tapia

I. COPPE/UFRJ

II. Título (série).

*Ao Rafa,
que com sua alegria de viver
imortalizou seu sorriso em minha mente
e a saudade em meu coração.*

In memorium.

À Maria das Dôres, minha mãe.

À Diana, minha irmã.

À Vilma de Ochoa, minha mãe hondurenha.

Agradecimentos

Este trabalho é o produto final de uma importante etapa da minha vida. Na caminhada até aqui houve alguns momentos difíceis e muitos de alegria, e só me foi possível chegar à reta final, porque no decorrer do percurso encontrei pessoas que me tomaram pela mão e me conduziram. É a essas pessoas que gostaria de agradecer e dividir a alegria desta conquista.

Ao professor Paulo Roberto, pela orientação e paciência.

À minha família, em especial à minha mãe, pelo incentivo e carinho que sempre me deram.

À Maria Nilda e Sebastião Carvalho, pela acolhida e amizade.

Aos tios João Nogueira Maia, Valmir Marçal e Tia Sônia, pela amizade, carinho e certeza de sempre contar com seu apoio.

Às “irmãs” Natália e Sandra pelas valiosas horas de convivência e aprendizado.

Ao Paulinho, amigo certo em todas as horas.

Ao Ivan, pelo precioso auxílio nos momentos finais.

A José Arica pelos valiosos ensinamentos, e à Gudélia, sempre com palavras ternas de incentivo.

A Sandra Regina, Lina, Erivaldo, Neiva, Edu, Chris, Ricardo, Assis, Afonso, Lise e amigos do programa de Engenharia Civil por tantos momentos de alegria.

Aos amigos Satuf e Regina pela força e “torcida”.

Às secretárias Cláudia e Ana Paula sempre tão atenciosas.

Ao CNPq, CAPES e FAPERJ, pelo apoio financeiro.

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

Algoritmos de Região de Confiança para Otimização com Restrições de Igualdade

Sonia Maria Moreira

Abril, 1994

Orientador : Paulo Roberto Oliveira

Programa : Engenharia de Sistemas e Computação

Realizamos um estudo comparativo de algoritmos de Região de Confiança para problemas de Programação Não-Linear com restrições de igualdade. Os algoritmos estudados são classificados de acordo com a estratégia que utilizam para resolver a possível incompatibilidade entre as restrições do subproblema (de Região de Confiança) obtido. Analisamos as técnicas de resolução desses subproblemas, e as vantagens e desvantagens de um método em relação a outro. Propomos um algoritmo variante, dentro de uma das classes estudadas, cuja característica interessante é que o subproblema a ser resolvido em cada iteração é simples.

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

Trust Region Algorithms for Equality Constrained Optimization

Sonia Maria Moreira

April, 1994

Thesis Supervisor : Paulo Roberto Oliveira
Department : Computing and Systems Engineering

This work presents a comparative study of trust region algorithms for nonlinear programming problems with equality constraints. We classify the algorithms, according to their strategy to solve a possible incompatibility among the constraints of the associated subproblems.

The solution techniques for the aforementioned subproblems, and also the advantages and drawbacks of each method with respect to each other, are investigated. A variant method is proposed. An interesting characteristic of the algorithm proposed is that the associated subproblem to be solved in each iteration is reasonably simple.

Índice

1	Introdução	4
2	Métodos de Região de Confiança para Otimização Irrestrita	6
2.1	Introdução	6
2.2	O Passo de Região de Confiança	8
2.2.1	O Método Hook Step	13
2.2.2	O Método Dogleg	15
2.2.3	O Duplo Dogleg	18
3	Região de Confiança para Problemas Restritos	21
3.1	Introdução	21
3.1.1	Solução para o Problema (NEP)	23
3.2	Solução de (NEP) via Região de Confiança	25
4	O Algoritmo de Vardi	27
4.1	Introdução	27
4.1.1	Um Passo de Quase Newton	27
4.2	Descrição do Algoritmo	30

4.3	Resultados de Convergência	33
4.4	Variantes do Algoritmo de Vardi	36
4.4.1	O Algoritmo de Byrd, Schnabel e Shultz	36
4.4.2	O Algoritmo de Zhang e Zhu	40
4.4.3	Proposta de uma Variante do Algoritmo de Vardi	44
5	O Algoritmo de Celis-Dennis-Tapia	48
5.1	Introdução	48
5.2	Solução do Subproblema SPCDT	51
5.3	Aceitação do Passo e Atualização da Região de Confiança	52
5.4	O Algoritmo CDT	55
5.5	Variantes do Algoritmo CDT	56
5.5.1	Algoritmo de Zhang para solução de SPCDT	56
5.5.2	O Algoritmo de Powell e Yuan	60
5.5.3	O Algoritmo de Mahmoud El-Alem	64
6	Conclusões	67

Figuras

2.1	Região de Confiança	7
2.2	Curva $d(\mu)$	10
2.3	Aproximação 1 para $\ d(\mu) \ $	14
2.4	Aproximação 2 para $\ d(\mu) \ $	14
2.5	Curva Dogleg	16
2.6	Curva Duplo Dogleg	18
2.7	Curvas Dogleg e Duplo Dogleg	20
3.1	Efeito Maratos	25
4.1	Subproblema de Vardi	28
4.2	Direção de Byrd-Schnabel-Shultz	36
5.1	O Passo CDT	50
5.2	O Problema de Powell e Yuan	61

Capítulo 1

Introdução

Neste trabalho é feito um estudo comparativo de algoritmos encontrados na literatura, que usam a metodologia de Região de Confiança para solucionar o problema de Otimização Não-Linear Restrito

$$\begin{array}{ll} \min_x & f(x) \\ \text{s.a} & c_i(x) = 0, i = 1, \dots, m \end{array}$$

onde f e $c_i, i = 1, \dots, m$ são funções duas vezes continuamente diferenciáveis. Os algoritmos estudados são divididos em duas categorias principais segundo a maneira como resolvem o problema de incompatibilidade entre restrições, que aparece quando se usa Região de Confiança para Otimização Restrita. Dentro de cada uma dessas categorias são estudados alguns algoritmos variantes, observando qual a motivação para as mudanças propostas e as vantagens/desvantagens de cada uma. Propomos uma variante de um algoritmo dentro de uma dessas classes.

O Capítulo 2 apresenta a metodologia de Região de Confiança para Otimização Irrestrita, abordando os métodos de *Hook Step*, *Dogleg* e *Duplo Dogleg*. No terceiro capítulo é visto como aplicar a estratégia de Região de Confiança para problemas com restrições de Igualdade, observando as dificuldades inerentes a esse caso e quais as principais propostas de solução encontradas na literatura. Nesse ponto, é feita a divisão dos algoritmos em duas categorias: os que usam a solução proposta por *Vardi* ([Var85]) e os que usam a solução de *Celis* ([Cel85]). O Capítulo 4 trata da metodologia de *Vardi* e algumas variantes de seu algoritmo, enquanto que o Capítulo 5 faz o mesmo para a estratégia de *Celis*. Uma

proposta de algoritmo variante, dentro da classe de Vardi, encontra-se no quarto capítulo. As conclusões e sugestões para trabalhos futuros aparecem no Capítulo 6.

Capítulo 2

Métodos de Região de Confiança para Otimização Irrestrita

2.1 Introdução

Considere o problema de minimização de uma função não linear de n variáveis reais:

$$\min f(x), x \in \mathbb{R}^n. \quad (2.1)$$

Métodos de Região de Confiança constituem uma importante classe de métodos com convergência global para a solução de problemas de Otimização. A diferença fundamental entre os métodos de Região de Confiança e os de Busca Linear está no modo como eles combinam o uso do modelo quadrático com a determinação da direção de busca.

Nos métodos de Busca Linear, minimiza-se um modelo quadrático da função objetivo, em torno do ponto corrente x_k , para encontrar uma direção de busca d_k ; calcula-se, então, um tamanho de passo $\alpha_k > 0$ tal que $f(x_k + \alpha_k d_k) < f(x_k)$, através de um modelo unidimensional (por exemplo, cúbico ou quadrático). Veja, por exemplo, [DS89].

A idéia básica dos métodos de Região de Confiança é estimar uma região onde se pode confiar num modelo quadrático como uma boa aproximação de uma função. Dada essa estimativa, toma-se a direção (com comprimento dado) que minimiza tal modelo. O passo obtido será aceito ou rejeitado e a Região de Confiança atualizada, dependendo de

quão bem o modelo representa a função naquela região.

Quando se faz busca linear, o modelo de aproximação é usado para determinar uma direção e a partir daí, encontra-se um passo que obtenha uma melhora de alguma função mérito no novo ponto. Na metodologia de Região de Confiança, primeiro se determina um tamanho máximo de passo (região onde se confia que o modelo aproxima bem a função) e então determina-se a direção que fornece uma melhora no valor da função.

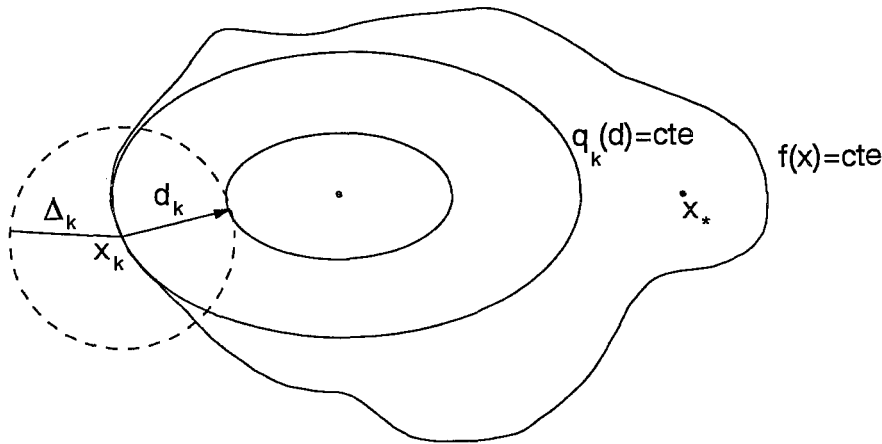


Figura 2.1: Região de Confiança

Dado um ponto x_k , uma cota máxima de passo Δ_k e uma aproximação $q_k(d)$ da função objetivo em torno de x_k , numa iteração em um algoritmo de Região de Confiança para otimização irrestrita, o passo d_k será obtido solucionando o subproblema

$$\begin{aligned} \min_{d \in \mathbb{R}^n} q_k(d) &= \nabla f_k^T d + \frac{1}{2} d^T B_k d \\ \text{s.a. } \|d\| &\leq \Delta_k, \end{aligned} \quad (2.2)$$

onde $\nabla f_k = \nabla f(x_k)$ é o gradiente da função f no ponto x_k , $B_k = \nabla^2 f(x_k)$ ou uma aproximação dela e $\|\cdot\|$ é a norma euclidiana.

Os métodos de Região de Confiança têm origem com os trabalhos de *Levenberg* ([Lev44]) e *Marquardt* ([Mar63]) para problemas de mínimos quadrados não lineares. A primeira abordagem para o problema geral foi feita por [GQT66]. *Powell* ([Pow70]) e *Thomas* ([Tho75]) aplicaram a teoria de convergência global em Regiões de Confiança para

situações gerais com iterações de Quase Newton. *Reinsh* ([Rei71]) e *Hebden* ([Heb73]) fizeram observações computacionais importantes para a implementação do método. Em 1978 *Moré* ([Mor78]) voltou a trabalhar com o problema de mínimos quadrados não lineares. *Fletcher* ([Fle81]) e *Sorensen* ([Sor82]) provaram propriedades de convergência global para algoritmos de Região de Confiança usando informações de segunda ordem. *Gay* ([Gay81]) acrescentou caracterizações teóricas e observações computacionais para o problema de minimização irrestrita com Hessiana indefinida. *Shultz, Schnabel e Byrd* ([SSB85]) apresentaram uma generalização para o trabalho de convergência global de *Sorensen*.

2.2 O Passo de Região de Confiança

O problema 2.2 é o subproblema básico dos métodos com estratégia de Região de Confiança. A solução de 2.2 é dada pelo teorema a seguir, que foi provado em [GQT66].

Teorema 2.2.1 *Seja $\nabla_k = \nabla f(x_k) \in \mathfrak{R}^n$, $B_k \in \mathfrak{R}^{n \times n}$ matriz simétrica, $\Delta_k > 0$. Seja $\lambda_1 \in \mathfrak{R}$ o menor autovalor de B_k e $\nu_1 \in \mathfrak{R}^n$ seu autovetor correspondente. Então, se B_k é definida positiva e $\| B_k^{-1} \nabla f_k \| \leq \Delta_k$, $d_k = B_k^{-1} \nabla f_k$ é a única solução de 2.2. Caso contrário, a solução para 2.2 satisfaz $\| d_k \| = \Delta_k$ e*

$$(B_k + \mu_k I) d_k = -\nabla f_k \quad (2.3)$$

para algum $\mu_k \geq 0$, onde $B_k + \mu_k I$ é, no mínimo, semi-definida positiva. Além disso, ou $B_k + \mu_k I$ é definida positiva e

$$d_k = -(B_k + \mu_k I)^{-1} \nabla f_k \quad (2.4)$$

para o único $\mu_k > \max\{0, -\lambda_1\}$ para o qual $\| d_k \| = \Delta_k$; ou $\mu_k = -\lambda_1$ e

$$d_k = -(B_k - \lambda_1 I)^\dagger \nabla f_k + \omega \nu_1, \quad (2.5)$$

onde \dagger denota a pseudoinversa e $\omega \in \mathfrak{R}$ é escolhido tal que $\| d_k \| = \Delta_k$.

Se B_k é definida positiva, a solução deve ser dada por 2.4; o caso 2.5 ocorre somente se B_k é indefinida e $\| (B_k + \mu_k I)^{-1} \nabla f_k \| < \Delta_k$ para todo $\mu_k > -\lambda_1$.

O último caso, (2.5), é denominado “caso difícil” por *Moré e Sorensen* ([MS83]).

Do teorema acima (equação 2.4) vemos que, mesmo que B_k seja definida positiva, o passo de Região de Confiança não é sempre na direção de Newton $-B_k^{-1}\nabla f_k$. Pode-se mostrar que para todo $\mu \geq \max\{0, -\lambda_1\}$ a função

$$\|d(\mu)\| = \|(B_k + \mu I)^{-1}\nabla f_k\|$$

é monotonicamente decrescente em μ . Esta propriedade é facilmente verificada quando B_k é definida positiva. Neste caso, $d(\mu) = (B_k + \mu I)^{-1}\nabla f_k$ pode ser escrito como $d(\mu) = \sum \alpha_i v_i / (\lambda_i + \mu)$, onde $\{(\lambda_i, v_i)\}$ é o conjunto de auto-valores e auto-vetores de B_k , estes supostos auto-normais. De $\|d(\mu)\|^2 = \sum \alpha_i^2 / (\lambda_i + \mu)^2$ se obtém diretamente a monotonia decrescente ([MS83]). Isto implica que $\|d(\mu)\| = \Delta_k$ tem uma única solução, que deve ser solução de 2.2.

Uma outra propriedade interessante da estratégia de Região de Confiança é o comportamento de $d(\mu)$ para $\mu \rightarrow \infty$ (ou $\Delta \rightarrow 0$). Tem-se

$$\begin{aligned} d_k &= -(B_k + \mu I)^{-1}\nabla f_k \\ &= -\frac{1}{\mu} \left(\frac{B_k}{\mu} + I \right)^{-1}\nabla f_k \end{aligned}$$

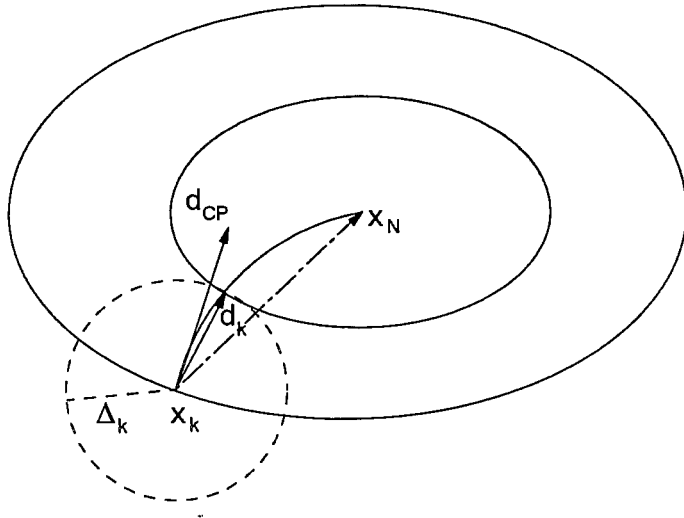
para μ suficientemente grande

$$\begin{aligned} d_k &= \frac{-1}{\mu} \left(I - \frac{B_k}{\mu} + \frac{B_k^2}{\mu^2} - \dots \right) \nabla f_k \\ \Rightarrow d_k &\rightarrow -\frac{\nabla f_k}{\mu} \end{aligned}$$

logo, quando $\Delta_k \rightarrow 0, \mu \rightarrow \infty, -(B_k + \mu I)^{-1}\nabla f_k \rightarrow -\frac{1}{\mu}\nabla f_k$, isto é, d_k tende à direção oposta à do gradiente. Para estudos mais detalhados ver [DS83].

Desta forma vemos que para Δ_k pequeno, a direção fornecida é aproximadamente a de descida máxima (Cauchy); enquanto que, quando Δ_k aumenta, a direção de Região de Confiança se aproxima, e por fim se torna, a direção de Newton (fig. 2.2).

Note duas propriedades atrativas dos métodos de Região de Confiança: a primeira é esta que acabamos de citar, de que passos pequenos são dados na direção de descida máxima, em geral, a melhor direção para modelos locais pouco confiáveis; enquanto que o passo de Newton é usado, em geral, próximo do ótimo, quando ele está dentro da Região de Confiança e B_k é definida positiva e deste modo, preserva-se a rápida convergência local. A segunda é que os métodos de Região de Confiança trabalham naturalmente com matrizes Hessianas indefinidas.

Figura 2.2: Curva $d(\mu)$

Entretanto, o passo de Região de Confiança, descrito pelo Teorema 2.2.1 é difícil de calcular, principalmente porque não há uma fórmula exata que forneça o único $\mu_k \geq \max\{0, \lambda_1\}$ tal que $\| (B_k + \mu_k I)^{-1} \nabla f_k \| = \Delta_k$. Assim, implementações eficientes de algoritmos de Região de Confiança devem determinar μ_k aproximadamente, através de um processo iterativo. Veremos mais adiante os principais métodos que solucionam 2.2 de forma aproximada. Vejamos, agora, um esquema global de um método de Região de Confiança para Otimização Irrestrita, incluindo ajuste do raio de confiança Δ_k , e as propriedades de convergência de métodos que usam este esquema solucionando 2.2 exatamente. Essas propriedades ainda são válidas quando se calcula o passo aproximado por métodos eficientes.

Uma vez calculada a direção d_k , num algoritmo de Região de Confiança, deve-se verificar se $x_k + d_k$ é uma próxima iterada melhor. Ela não será satisfatória se o modelo quadrático não estiver representando bem a função dentro da Região de Confiança. Neste caso, o raio de confiança é diminuído e o passo recalculado. Caso contrário, $x_k + d_k$ torna-se a próxima iterada $x_{k+1} = x_k + d_k$ e um novo raio é calculado. A seguir, um algoritmo conceitual de Região de Confiança ([DS89]).

ALGORITMO RCP - *Iteração de Região de Confiança*

Dados $f : \mathbb{R}^n \rightarrow \mathbb{R}$, x_k , $f(x_k)$, $g(x_k) = \nabla f(x_k)$, Hessiana (ou aproximação) B_k , raio de confiança $\Delta_k > 0$, $0 < \alpha_1 < \alpha_2 < 1$, $0 < \eta_1 < \eta_2 < 1 < \eta_3 < \eta_4$

aceito:=falso

REPITA

d_k := solução exata ou aproximada de 2.2

calcule $f(x_k + d_k)$

$ared_k := f(x_k + d_k) - f(x_k)$

$pred_k := g(x_k)^T d_k + \frac{1}{2} d_k^T B_k d_k$

Se $ared_k \leq \alpha_1 pred_k$ então

aceito := verdade

$x_+ := x_k + d_k$

Se $ared_k \leq \alpha_2 pred_k$ então

$\Delta_+ := \text{aumenta-delta}(\Delta_k)$ (* $\Delta_+ \in [\eta_3, \eta_4] \Delta_k$ *)

senão $\Delta_+ := \Delta_k$ (* ou $\Delta_+ \in [\eta_1, 1] \Delta_k$ *)

senão $\Delta_k := \text{diminui-delta}(\Delta_k)$ (* $\text{novo} \Delta_k \in [\eta_1, \eta_2] \text{velho} \Delta_k$ *)

ATE QUE aceito = verdade.

Note que o passo será aceito se a redução real ($ared$) for suficientemente menor que a prevista pelo modelo ($pred$). Para isso, testa-se a condição

$$ared \leq \alpha_1 pred, \quad (2.6)$$

onde $\alpha_1 = 10^{-4}$ é o valor típico usado. Se a inequação 2.6 não for satisfeita, diminui-se o raio e soluciona-se o mesmo modelo, agora dentro de uma região menor. Caso contrário, testa-se uma outra condição, a saber:

$$ared \leq \alpha_2 pred, \quad (2.7)$$

onde $\alpha_2 = 0.75$ é o valor típico. Essa condição indica quão bem a função está representada pelo modelo na região. Se a desigualdade 2.7 é satisfeita, o raio de Região de Confiança é aumentado; senão, toma-se $\Delta_{k+1} = \Delta_k$. O fato de se aumentar o raio é porque em iterações anteriores ele pode ter-se tornado pequeno. Isto é devido a que a função não

era bem representada pelo modelo quadrático, sendo necessário tomar regiões pequenas, e atualmente se está num ponto em que a função pode ser bem aproximada por tal modelo. Todavia, este procedimento de aumentar o raio de confiança, apesar de trazer vantagens práticas, não influi nas propriedades teóricas do método.

O teorema, a seguir, fornece as principais propriedades de convergência de um método baseado na iteração dada pelo Algoritmo RCP, no caso em que 2.2 é resolvido exatamente. Ele mostra convergência a pontos que satisfazem às condições de primeira ordem para mínimo local do problema original 2.2, e, se são usadas derivadas de segunda ordem exatas, obtém-se condições necessárias de segunda ordem. O teorema diz também que, próximo de um mínimo local x_* com Hessiana $H(x_*)$ definida positiva, assintoticamente, a restrição de Região de Confiança torna-se inativa, de modo que são dados somente passos de Newton e assim a convergência quadrática local é mantida. Resultados semelhantes podem ser encontrados em várias referências, incluindo [Fle81], [Gay81] e [Sor82].

Teorema 2.2.2 *Suponha $f : \mathbb{R}^n \rightarrow \mathbb{R}$ duas vezes continuamente diferenciável e limitada inferiormente. Para $x_0 \in \mathbb{R}^n$ e $\beta_1, \beta_2 > 0$, seja $H(x)$ uniformemente contínua e satisfazendo $\|H(x)\| \leq \beta_1$ para todo x com $f(x) \leq f(x_0)$. Seja $\{x_k\}$ a sequência de pontos produzida pelas iterações do Algoritmo RCP partindo de x_0 , usando $B_k = H(x_k)$ ou qualquer aproximação simétrica com $\|B_k\| \leq \beta_2$, em cada iteração e d_k solução exata de 2.2. Então,*

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

Se além disso, cada $B_k = H(x_k)$ (a hessiana é efetivamente usada), então para qualquer ponto limite x_ da sequência $\{x_k\}$, $\nabla f(x_*) = 0$ e $H(x_*)$ é, no mínimo, semi-definida positiva. Mais ainda, se cada $B_k = H(x_k)$, então, se x_k converge para x_* , $H(x_*)$ é definida positiva e $H(x)$ é Lipschitz-contínua em torno de $H(x_*)$, a taxa de convergência é q -quadrática.*

Para resolver 2.2 há duas relaxações usualmente utilizadas: uma aproximação linear da equação não linear em μ 2.3 (ou equivalente), para utilizar o método de Newton; ou uma aproximação poligonal (plana) da curva (em \mathbb{R}^n) $\mu \rightarrow d(\mu)$, para $\mu \geq 0$. Correspondendo a cada uma destas relaxações temos duas classes importantes de métodos, respectivamente: os métodos de *passo ótimo local aproximado* (*Hook Step*) e os métodos *Dogleg*. Nas próximas seções será visto um resumo de cada uma destas classes de métodos.

2.2.1 O Método Hook Step

Hebden ([Heb73]) e *Moré* ([Mor78]) desenvolveram um procedimento para encontrar $\mu > 0$ tal que

$$\| d(\mu) \| = \| (B_k + \mu I)^{-1} \nabla f_k \| \approx \Delta_k, \quad (2.8)$$

onde B_k é definida positiva e $\| B_k^{-1} \nabla f_k \| > \Delta_k$. Note que o que se procura ao tentar resolver 2.8 é encontrar um zero da função unidimensional

$$\phi(\mu) = \| d(\mu) \| - \Delta_k$$

Para tanto, aplica-se o método de Newton para determinar zeros de funções. *Hebden* ([Heb73]) e *Reinsch* ([Rei71]) observaram que a função $\| d(\mu) \|^2$ é uma função racional com polos de segunda ordem no conjunto $\{ -\lambda_j : \lambda_j \text{ é autovalor de } B_k \}$. Assim sendo, seria mais eficiente aplicar o método (de linearização local) de Newton para a função

$$\psi(\mu) = \frac{1}{\Delta_k} - \frac{1}{\| d(\mu) \|}$$

que não tem polos e é “quase” linear perto da solução. Temos que

$$\mu_+ = \mu - \frac{\psi(\mu)}{\psi'(\mu)}$$

ou seja,

$$\mu_+ = \mu - \left(\frac{1}{\Delta_k} - \frac{1}{\| d(\mu) \|} \right) / \left(\frac{-d(\mu)^T (B_k + \mu I)^{-1} d(\mu)}{\| d(\mu) \|^3} \right)$$

Em termos de $\phi(\mu) = \| d(\mu) \| - \Delta_k$, obtemos:

$$\begin{aligned} \mu_+ &= \mu - \frac{\| d(\mu) \| - \Delta_k}{\frac{\Delta_k}{\| d(\mu) \|} \phi'(\mu)} \\ &= \mu - \frac{\| d(\mu) \| \phi(\mu)}{\Delta_k \phi'(\mu)} \end{aligned}$$

Portanto, a diferença entre usar ϕ e ψ está no termo $\| d(\mu) \| / \Delta_k$, que modificará a inclinação da reta de aproximação, constituindo um fator de aceleração na busca do valor desejado. Com efeito, se $\| d(\mu_k) \| > \Delta_k$, como $\| d(\mu) \|$ é monótona decrescente temos: $\mu_k < \mu_*$, onde μ_* é a solução de $\| d(\mu_*) \| = \Delta_k$. Assim, $\frac{\| d(\mu_k) \|}{\Delta_k} > 1$, e teremos a situação representada pela figura 2.3. Por outro lado, se $\| d(\mu_k) \| < \Delta_k$, teremos $\mu_k > \mu_*$ e a iteração se comporta como na figura 2.4.

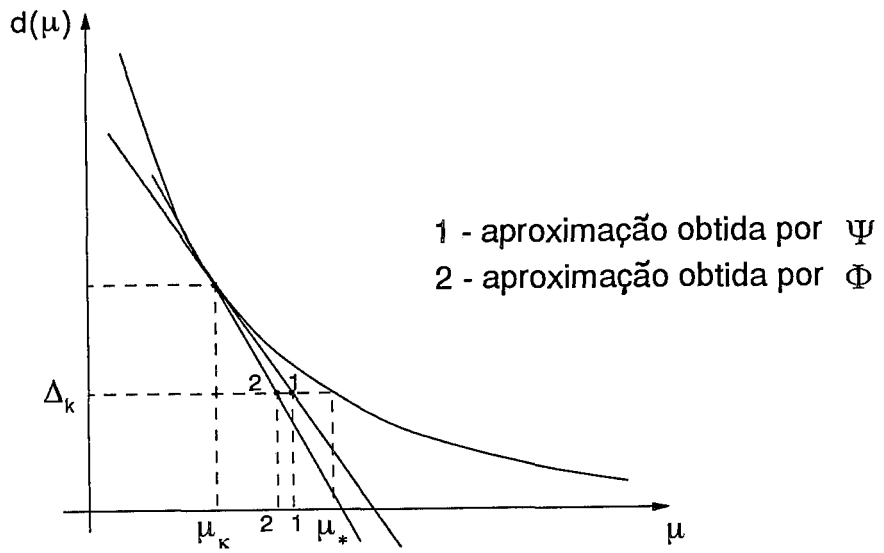


Figura 2.3: Aproximação 1 para $\| d(\mu) \|$

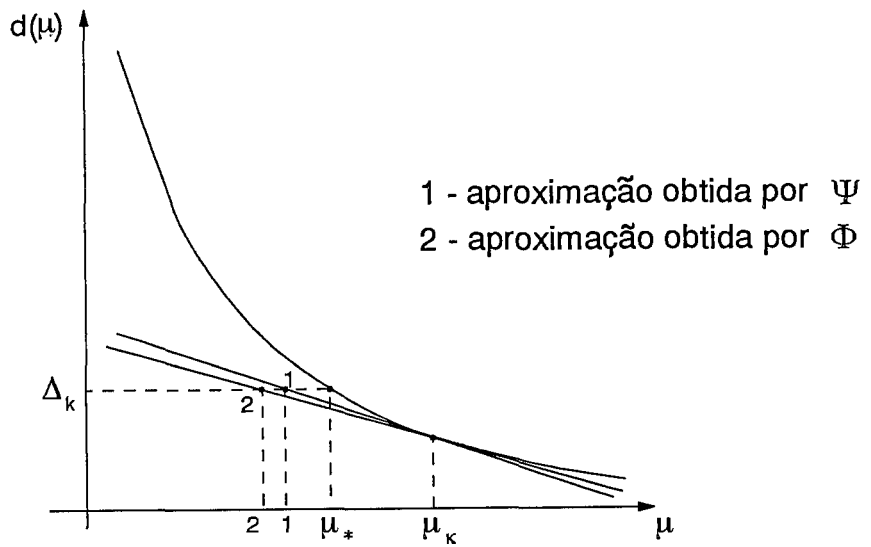


Figura 2.4: Aproximação 2 para $\| d(\mu) \|$

Assim sendo, uma solução aproximada para 2.2 é encontrada através de um algoritmo baseado no método de Newton em μ

$$\mu_+ = \mu - \frac{\|d(\mu)\| (\|d(\mu)\| - \Delta_k)}{\Delta_k d(\mu)^T (B_k + \mu I)^{-1} d(\mu)},$$

onde $d(\mu) = -(B_k + \mu I)^{-1} \nabla f_k$. Tipicamente, a restrição de confiança é relaxada satisfazendo $\|d_k\| \in [0.9, 1.1] \Delta_k$; com isto, normalmente um ou dois passos de Hook Step são necessários para cada iteração do algoritmo.

Vários autores, como *Gay* ([Gay81]), *Sorensen* ([Sor82]) e *Moré e Sorensen* ([MS83]) fazem generalizações de métodos tipo Hook Step estendendo para o caso em que B_k não é definida positiva. *Moré e Sorensen* apresentam um algoritmo eficiente e provam que a solução aproximada de 2.2 reduz o modelo quadrático $q_k(d)$ em, no mínimo, γ vezes a quantidade que a solução exata reduziria; onde $\gamma > 1$ é uma constante fixada. Eles mostram que seu método também requer, geralmente 1 ou 2 μ -iterações por iteração global e que mantém as propriedades de convergência do Teorema 2.2.2. As idéias expostas aqui foram baseadas em *Dennis e Schnabel* ([DS89]).

2.2.2 O Método Dogleg

O Método de Dogleg original foi proposto por *Powell* ([Pow70]). Como no Método Hook Step, ele também encontra uma solução aproximada para 2.2. Entretanto, ao invés de determinar um ponto $x_{k+1} = x_k + d_k$ sobre a curva $d(\mu)$ tal que $\|x_{k+1} - x_k\| \approx \Delta_k$, o método de Powell aproxima a curva $d(\mu)$ pela poligonal obtida ligando x_k , o ponto de Cauchy¹ x^{CP} e o ponto de Newton x^N . Então x_{k+1} é escolhido como sendo o ponto sobre esta poligonal tal que $\|x_{k+1} - x_k\| = \Delta_k$, exceto se $\|B_k^{-1} \nabla f_k\| < \Delta_k$, em cujo caso x_{k+1} é o ponto de Newton (fig. 2.5). Iremos supor daqui até o final desta seção que B_k é definida positiva. Para justificar que o método é bem definido devemos mostrar que $\|x^{CP} - x_k\| \leq \|x^N - x_k\|$ e que a interseção da curva Dogleg com a esfera de confiança é única. Com efeito,

$$\|x^{CP} - x_k\| \leq \|x^N - x_k\| \iff \lambda_* \|\nabla f_k\| \leq \|B_k^{-1} \nabla f_k\|$$

¹O ponto de Cauchy é dado pelo mínimo do modelo quadrático na direção de descida mínima $-\nabla f_k$ a partir de x_k , ou seja, $x^{CP} = x_k - \lambda_* \nabla f_k$, onde $\lambda_* = \operatorname{argmin} \lambda \|\nabla f_k\|^2 + \frac{1}{2} \lambda^2 \nabla f_k^T B_k \nabla f_k$, portanto $\lambda_* = \frac{\|\nabla f_k\|^2}{\nabla f_k^T B_k \nabla f_k}$.

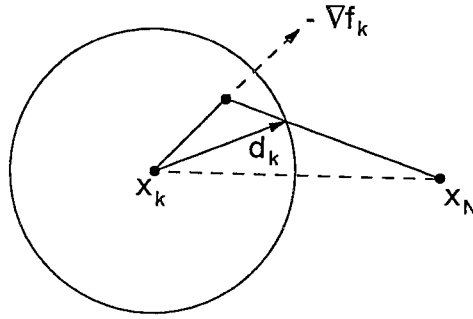


Figura 2.5: Curva Dogleg

mas

$$\begin{aligned} \lambda_* \|\nabla f_k\| &= \frac{\|\nabla f_k\|^3}{\nabla f_k^T B_k \nabla f_k} \\ &\leq \frac{\|\nabla f_k\|^4 \|B_k^{-1} \nabla f_k\|}{[\nabla f_k^T B_k \nabla f_k][\nabla f_k^T B_k^{-1} \nabla f_k]} \end{aligned} \quad (2.9)$$

e

$$\begin{aligned} \|\nabla f_k\|^4 &= [\nabla f_k^T \nabla f_k]^2 \\ &= [(\nabla f_k^T B_k^{\frac{1}{2}})(B_k^{-\frac{1}{2}} \nabla f_k)]^2 \\ &\leq [\|\nabla f_k^T B_k^{\frac{1}{2}}\| \|\nabla f_k^T B_k^{-\frac{1}{2}}\|]^2 \\ &= [(\nabla f_k^T B_k \nabla f_k)^{1/2} (\nabla f_k^T B_k^{-1} \nabla f_k)^{1/2}]^2 \\ &= (\nabla f_k^T B_k \nabla f_k)(\nabla f_k^T B_k^{-1} \nabla f_k) \end{aligned}$$

e segue-se que

$$\lambda_* \|\nabla f_k\| \leq \|B_k^{-1} \nabla f_k\|$$

Concluimos, assim, que a interseção da curva de Dogleg com a esfera de raio Δ_k é única.

Note que se tem duas situações possíveis: o ponto de *Cauchy* estar dentro da esfera de confiança ou cair fora dela. No segundo caso, tem-se

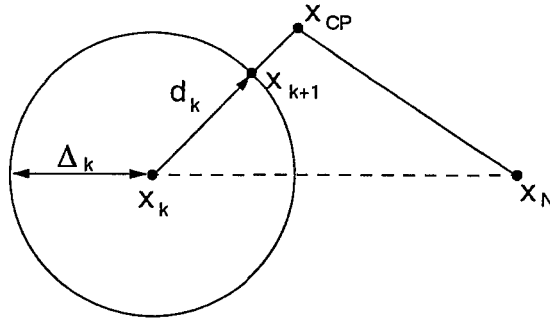
$$\frac{\|\nabla f_k\|^3}{\nabla f_k^T B_k \nabla f_k} \geq \Delta_k$$

e tomamos

$$d_k = -\Delta_k \frac{\nabla f_k}{\|\nabla f_k\|}$$

obtendo

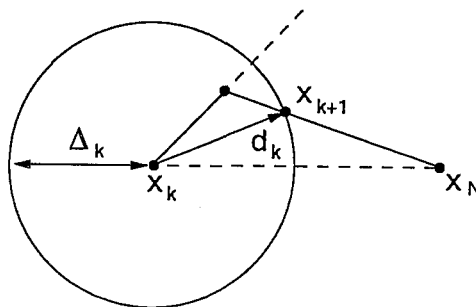
$$x_{k+1} = x_k - \Delta_k \frac{\nabla f_k}{\|\nabla f_k\|}$$



Por outro lado, se o ponto de Cauchy estiver dentro da Região de Confiança, temos

$$\frac{\|\nabla f_k\|^3}{\nabla f_k^T B_k \nabla f_k} \leq \Delta_k$$

e então calcula-se d_k tal que $\|d_k\| = \Delta_k$ e $x_{k+1} = x_k + d_k$ está sobre a linha que une x^{CP} e x^N .



Powell ([Pow70]) mostra que: se f é limitada inferiormente e a norma da matriz Hessiana da função e das matrizes que aproximam as hessianas nos pontos x_k são limitadas superiormente, então seu algoritmo obtém x_* tal que $\|\nabla f_*\| < \varepsilon$, para uma precisão ε dada.

2.2.3 O Duplo Dogleg

Dennis e Mei ([DM79]) fizeram uma melhora no método de *Powell* ([Pow70]), o *Duplo Dogleg*, que aproxima a curva $d(\mu)$ por uma poligonal obtida ligando os pontos $x_k, x^{CP}, \gamma x^N$ e x^N , onde $\gamma < 1$, escolhendo, então, o novo ponto como sendo o ponto sobre esta poligonal que satisfaz $\|d_k\| = \Delta_k$ (fig. 2.6).

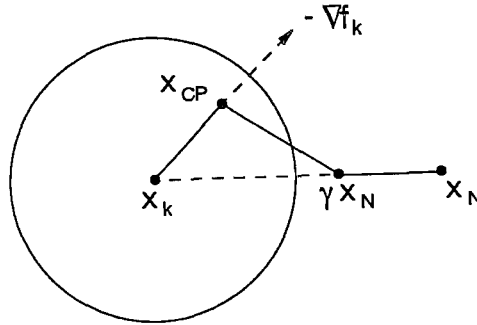


Figura 2.6: Curva Duplo Dogleg

Há duas propriedades importantes da curva Duplo Dogleg: primeiro, quando se caminha ao longo dela, a distância para x_k cresce monotonicamente. Desta forma, quando $\Delta_k \leq \|B_k^{-1} \nabla f_k\|$, há um único ponto x_+ sobre a poligonal tal que $\|x_+ - x_k\| = \Delta_k$, o que mostra que o processo está bem definido. Segundo, o modelo quadrático $q_k(d)$ decresce monotonicamente quando se vai de x_k em direção a x^N sobre a curva, isto torna o processo razoável. Para mais detalhes veja [DS83].

Considere $d_{\gamma N}, d^N$ e d^{CP} tais que $\gamma x^N = x_k + d_{\gamma N}, x^N = x_k + d^N, x^{CP} = x_k + d^{CP}$, respectivamente. Do que foi visto na seção anterior, temos que

$$\|d^{CP}\| \leq \eta \|d^N\| \leq \|d^N\|$$

onde

$$\eta = \frac{\|\nabla f_k\|^4}{(\nabla f_k^T B_k \nabla f_k)(\nabla f_k^T B_k^{-1} \nabla f_k)}.$$

Isto justifica a primeira propriedade do parágrafo anterior. O ponto γx^N sobre a curva Duplo Dogleg é escolhido para algum γ satisfazendo:

1. $\eta \leq \gamma \leq 1$;

2. o modelo decresça monotonicamente ao longo da reta de x^{CP} a γx^N .

Temos que $q_k(d)$ decresce monotonicamente de x_k a x^{CP} e de γx^N a x^N (pela escolha de γ). Se a condição (2) for satisfeita teremos, então, que o modelo decresce monotonicamente ao longo de toda a curva Duplo Dogleg.

Para satisfazer (2), γ deve ser escolhido tal que a derivada direcional de $q_k(d)$ seja negativa em todo ponto do segmento de reta que liga x^{CP} a γx^N . A representação deste segmento é

$$x(\lambda) = x_k + d^{CP} + \lambda(\gamma d^N - d^{CP})$$

para $0 \leq \lambda \leq 1$. A derivada direcional de $q_k(d)$ ao longo deste segmento em um ponto arbitrário $x(\lambda)$ é dada por

$$\begin{aligned} D(\lambda) &= \nabla q_k(x(\lambda) - x_k)^T (x_{\gamma N} - x^{CP}) \\ &= [\nabla f_k + B_k(d^{CP} + \lambda(\gamma d^N - d^{CP}))]^T (\gamma d^N - d^{CP}) \\ &= [\nabla f_k + B_k(w + \lambda v)]^T v \\ &= [\nabla f_k + B_k w]^T v + \lambda v^T B_k v \end{aligned}$$

onde $v = -\gamma B_k^{-1} \nabla f_k + \lambda_* \nabla f_k$ e $w = -\lambda_* \nabla f_k$. Como B_k é definida positiva, $v^T B_k v > 0$, o que implica que $D(\lambda)$ é função crescente de λ . Assim, precisamos provar somente que $D(1)$ é negativo para obter que $D(\lambda) < 0$ para todo $0 \leq \lambda \leq 1$. Com efeito,

$$\begin{aligned} D(1) &= (\nabla f_k + B_k d^{CP})^T (\gamma d^N - d^{CP}) + (\gamma d^N - d^{CP})^T B_k (\gamma d^N - d^{CP}) \\ &= -\gamma \nabla f_k B_k^{-1} \nabla f_k + \lambda_* \nabla f_k^T \nabla f_k + \gamma^2 \nabla f_k B_k^{-1} \nabla f_k - \gamma \lambda_* \nabla f_k^T \nabla f_k \\ &= (1 - \gamma)(-\gamma \nabla f_k^T B_k^{-1} \nabla f_k + \lambda_* \nabla f_k^T \nabla f_k) < 0 \end{aligned}$$

para qualquer $\eta \leq \gamma \leq 1$, pois $\eta = \|\nabla f_k\|^4 / (\nabla f_k^T B_k \nabla f_k)(\nabla f_k^T B_k^{-1} \nabla f_k)$ e $\lambda_* = \|\nabla f_k\|^2 / (\nabla f_k^T B_k \nabla f_k)$.

Os métodos Dogleg tomam passos pequenos na direção de descida máxima, quando Δ_k é suficientemente pequeno; tomam passos de Newton quando Δ_k é grande; e uma combinação linear dessas duas direções nos outros casos. Devido ao uso de passos na direção de descida máxima quando $\Delta_k \leq \|x^{CP} - x_k\|$ e do descréscimo monotônico do modelo quadrático ao longo de toda sua curva, os métodos Dogleg sempre obtêm, no mínimo, tanto descréscimo no modelo quanto o melhor passo na direção de descida máxima de tamanho máximo Δ_k . Assim, se garante convergência global de primeira ordem, ou seja, a sequência de gradientes nas iteradas converge para zero.

Algoritmos que utilizam as técnicas de Dogleg e Duplo Dogleg apresentam melhor performance que aqueles que usam Hook Step. Sendo, ainda, o Duplo Dogleg melhor que o Dogleg, é o método efetivamente aplicado em algoritmos de Região de Confiança.

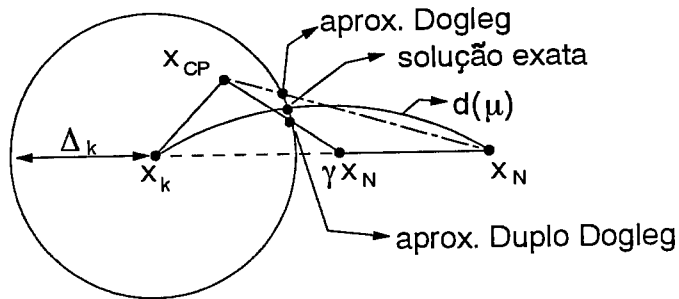


Figura 2.7: Curvas Dogleg e Duplo Dogleg

Capítulo 3

Região de Confiança para Problemas Restritos

3.1 Introdução

O problema que será estudado neste trabalho é o de minimização de uma função não linear, sujeita a restrições não lineares de igualdade

$$\begin{aligned} (NEP) \quad & \min && f(x) \\ & s.a && c^i(x) = 0 \\ & && i = 1, \dots, m \end{aligned}$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $c^i : \mathbb{R}^n \rightarrow \mathbb{R}$, $i = 1, \dots, m$ são funções duas vezes continuamente diferenciáveis.

Vejamos a notação e algumas definições que serão usadas aqui:

- $g(x)$ é o vetor gradiente de $f(x)$ e $\nabla^2 f(x)$ sua matriz Hessiana;
- $c(x) = (c^1(x), \dots, c^m(x))$ é o vetor de restrições, $a^i(x) \in \mathbb{R}^n$ o gradiente de $c^i(x)$ e $\nabla^2 c^i(x)$ sua matriz Hessiana;

- $A(x)$ é a matriz $n \times m$ formada pelos vetores coluna $a^i(x)$, com $i = 1, \dots, m$, suposta de posto completo $\forall x$;
- uma solução de (NEP) é denotada por x_* e o multiplicador de Lagrange correspondente por λ_* ;
- para qualquer x , seja $Z = Z(x)$ uma matriz $n \times (n - m)$ cujas colunas formam uma base ortonormal para o espaço nulo de $A(x)^T$, ou seja, $A(x)^T Z(x) = 0$ e $Z(x)^T Z(x) = I$, onde I é a matriz identidade;
- $\| \cdot \|$ é a norma euclidiana em \mathfrak{R}^n ;
- a função *Lagrangeana* associada ao problema (NEP) é dada por

$$L(x, \lambda) = f(x) + \lambda^T c(x);$$

e a função *Lagrangeano Aumentado* é

$$L_A(x, \lambda, \rho) = f(x) + \lambda^T c(x) + \frac{1}{2} \rho c(x)^T c(x)$$

onde $\lambda \in \mathfrak{R}^m$ é o vetor dos multiplicadores de Lagrange e ρ é um parâmetro de penalidade não negativo;

- $B = \nabla^2 L(x, \lambda)$ ou uma aproximação dela; se $B = \nabla^2 L(x, \lambda)$, então $Z^T B Z$ é dita a *Hessiana Reduzida*;
- a estimativa de primeira ordem do vetor dos multiplicadores de Lagrange, $\lambda(x)$, é dada por

$$\lambda(x) = -[A(x)^T A(x)]^{-1} A(x)^T g(x)$$

(isto é obtido diretamente das condições de otimalidade para (NEP));

- se B é uma matriz simétrica, $\nu_1(B)$ denota seu menor autovalor;
- sub-índices denotam o valor da função em um ponto da sequência de iteradas (i. é, $g_k = g(x_k)$, $f_k = f(x_k)$, $A_k = A(x_k)$), enquanto que expoentes denotam a componente particular de um vetor ou uma matriz (por ex., a_k^i é a i -ésima coluna de $A(x_k)$);
- x é um ponto estacionário do problema (NEP) se satisfaz à condição

$$\| c(x) \| + \| P(x)g(x) \| = 0$$

onde $P(x)$ é o operador de projeção dado por

$$P(x) = I - A(x)[A^T(x)A(x)]^{-1}A^T(x).$$

3.1.1 Solução para o Problema (NEP)

Métodos bastante usados para resolver o problema (NEP) são os baseados em *Programação Quadrática Sequencial* (SQP). Eles geralmente envolvem a minimização iterativa de um modelo quadrático em torno do ponto corrente x_k , sujeito a aproximações lineares das restrições. Este modelo é obtido, por exemplo, através da aproximação de primeira ordem das condições necessárias de otimalidade de

$$\begin{aligned} \min \quad & f(x_k + d) \\ \text{s.a} \quad & c^i(x_k + d) = 0 \\ & i = 1, \dots, m \quad , \end{aligned}$$

onde $g(x_k + d) + \sum \lambda_i \nabla c_i(x_k + d) = 0$ é substituído por $g_k + \nabla^2 f_k d + \sum_i \lambda_i [a_k + \nabla^2 c_k^i d] = 0$. Ou seja, em x_k resolve-se o problema

$$\begin{aligned} (QP) \quad & \min_{d \in \mathbb{R}^n} \quad g_k^T d + \frac{1}{2} d^T B_k d \\ & \text{s.a} \quad c_k + A_k^T d = 0. \end{aligned}$$

Uma solução d_k^{QP} do problema acima é chamada *direção (SQP)* e o multiplicador de Lagrange associado λ_k^{QP} . Um elemento que se faz claramente importante na formulação de (QP) é a escolha de B_k . Podemos analisar duas situações: quando as derivadas segundas exatas de $f(x)$ e $c(x)$ são calculáveis e quando não são. No primeiro caso, a escolha “ideal” óbvia para B_k , perto da solução, será a própria Hessiana da Lagrangeana $\nabla^2 L(x_k, \lambda_k)$, onde x_k e λ_k são aproximações correntes para x_* e λ_* . Com esta escolha, o método (SQP) “puro”, definido por

$$x_{k+1} = x_k + \alpha_k d_k,$$

(com $\alpha_k = 1$ e d_k direção de busca) produzirá convergência local quadrática em x e λ . Se $B_k = \nabla^2 L(x_k, \lambda_k)$, e a Hessiana Reduzida $Z_k^T B_k Z_k$ for definida positiva, então (QP) tem um único minimizador d_k^{QP} . Contudo, se $Z_k^T B_k Z_k$ for indefinida, o problema (QP) não terá solução limitada (ver [GMSW89]). No caso em que as derivadas segundas não

são calculáveis, a estratégia mais comum para se obter B_k é usar aproximações de quase Newton da Hessiana da função Lagrangeana, através de métodos como BFGS, DFP, etc.

Note duas dificuldades no problema (QP): as restrições linearizadas podem ser inconsistentes se A_k não tiver posto completo; e a direção de busca pode “explodir” em tamanho se as linhas de A_k são “quase” linearmente dependentes. Muitos autores sugerem modificações na estrutura básica do método SQP a fim de vencer estas dificuldades (SQP-relacionados) (Veja, por ex., [Fle81],[Fle85]). Os métodos SQP e SQP-relacionados são os métodos gerais mais eficientes para a solução de (NEP).

Uma das maiores dificuldades em Otimização Restrita não Linear é decidir se o novo ponto gerado $x_+ = x_k + d_k$ é uma melhor aproximação da solução que o anterior x_k . Diante disto, é importante se ter uma maneira de medir quando uma estimativa é “melhor” que a anterior. No caso de minimização irrestrita, uma medida de progresso natural é fornecida pela própria função objetivo. Parece razoável exigir, neste caso, um decréscimo na função objetivo em cada iteração e impor a condição $f(x_{k+1}) < f(x_k)$, $\forall k \geq 0$. Um método que usa esta condição é dito *Método de Descida*.

Para se solucionar um problema restrito, como era de se esperar, a situação se complica. É praticamente impossível, nesse caso, gerar uma sequência de iteradas viáveis com valores decrescentes da função objetivo. Diversas estratégias muito usadas não mantêm a viabilidade (i.é, geram pontos não necessariamente viáveis); depois, para decidir se o novo ponto obtido é uma “melhor” aproximação para a solução do problema, usam uma função (dita *Função Mérito*) que, de alguma maneira, equilibra o conflito entre o desejo de decrescer a função objetivo e satisfazer às restrições. As funções mérito mais comuns são: a *Função Penalidade Exata*

$$\phi(x) = f(x) + \sum_{i=1}^m \mu^i |c^i(x)|$$

onde μ^i 's são constantes positivas chamadas *coeficientes de penalidade*; e a função *Lagrangeano Aumentado*, definida na seção anterior (que também é exata).

O interesse nos métodos (SQP) surgiu com a descoberta de que d_k^{QP} é direção de descida para as funções mérito que acabamos de descrever, garantindo nestes casos, sob condições razoáveis, a convergência global para pontos estacionários (mínimo ou sela). Os métodos SQP e SQP-Relacionados associados a funções mérito são os métodos gerais mais eficientes para solução de (NEP).

Uma observação importante é que o uso de funções Mérito não diferenciável tipo ϕ , pode causar dificuldades na convergência. Uma delas é o chamado “Efeito Maratos” que foi notado por *Maratos* ([Mar78]), *Chamberlain* ([CLHP82]), e num contexto de Região de Confiança por *Yuan* ([Yua84]). Essa dificuldade aparece quando os pesos μ_i são muito grandes e assim a função ϕ é mais influenciada pelas restrições $c(x)$ que pela função objetivo, ficando portanto com curvas de nível alongadas muito parecidas com as das restrições (fig. 3.1). As direções de descida d_k^{QP} para ϕ podem tender a tangenciar a superfície das restrições. Desta forma, a condição $\phi(x_{k+1}) < \phi(x_k)$ requer que os passos perto da solução sejam bem pequenos, e assim, a convergência superlinear não ocorrerá.

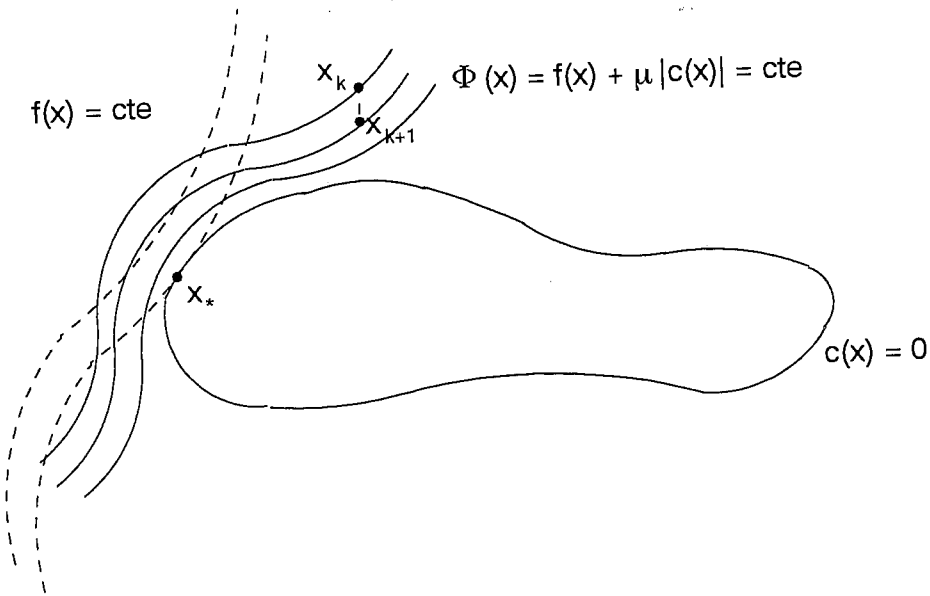


Figura 3.1: Efeito Maratos

3.2 Solução de (NEP) via Região de Confiança

Para solucionar o problema (NEP) usando a metodologia de Região de Confiança, a maneira mais natural é acrescentar a restrição de confiança ao problema (QP), ou seja, na k -ésima iteração resolve-se um subproblema do tipo

$$\begin{aligned} \min_{d \in \mathbb{R}^n} \quad & g_k^T d + \frac{1}{2} d^T B_k d \\ \text{s.a} \quad & c_k + A_k^T d = 0 \\ & \|d\| \leq \Delta_k. \end{aligned}$$

Uma dificuldade aparece claramente no modelo acima: as restrições $c_k + A_k^T d = 0$ e $\|d\| \leq \Delta_k$ podem ser incompatíveis. Para resolvê-la, há duas metodologias descritas na literatura: uma delas (proposta por Vardi ([Var85]) é fazer uma relaxação das restrições linearizadas, ou seja, substitui-se $c_k + A_k^T d = 0$ por $\alpha c_k + A_k^T d = 0$, onde $0 < \alpha \leq 1$ é um parâmetro escolhido tal que $\{d; c_k + A_k^T d = 0\} \cap \{d; \|d\| \leq \Delta_k\} \neq \emptyset$ e que deve tender para 1. A outra (introduzida por Celis ([CDT85]) é forçar progressivamente a viabilidade através do controle de $\|c_k + A_k^T d\|$ solucionando o subproblema

$$\begin{aligned} \min_{d \in \mathbb{R}^n} \quad & g_k^T d + \frac{1}{2} d^T B_k d \\ \text{s.a} \quad & \|c_k + A_k^T d\| \leq \theta_k \\ & \|d\| \leq \Delta_k, \end{aligned}$$

onde θ_k é uma constante positiva. Assim, o modelo quadrático está sujeito a um decréscimo na $\|c_k + A_k^T d\|$, controlando mais diretamente a viabilidade das restrições aproximadas. Nos próximos capítulos veremos mais detalhes de cada uma dessas metodologias.

Para decidir se o novo ponto obtido $x_+ = x_k + d_k$ será aceito, faz-se um procedimento análogo ao que foi descrito para Otimização Irrestrita, usando agora uma função Mérito ϕ como em (SQP). Calcula-se a *redução real* na função Mérito quando se vai de x_k para $x_+ = x_k + d_k$ e compara-se com a *redução prevista* pelo modelo local. Se for obtido um decréscimo suficiente para ϕ , x_+ é aceito como nova iterada x_{k+1} e atualiza-se o raio de confiança; senão, x_+ é rejeitado e a Região de Confiança diminuída. Veremos alguns algoritmos de Região de Confiança onde serão discutidas essas questões, entre outras, e como resolvê-las.

Capítulo 4

O Algoritmo de Vardi

4.1 Introdução

Vardi ([Var85]) foi o primeiro a propor um algoritmo de Região de Confiança para problemas com restrições de igualdade. Seu algoritmo combina o uso de direções de quase Newton com a estratégia de Região de Confiança. É usada uma função Penalidade Exata como função Mérito e para evitar as dificuldades de velocidade de convergência local, causadas por esse tipo de função, Vardi propõe que, quando se estiver próximo de um mínimo, seja usada a função Lagrangeana convencional no lugar da função Penalidade Exata. Com isso, ele obtém convergência global e convergência local superlinear para seu algoritmo. A fim de resolver o problema da incompatibilidade entre as restrições, Vardi faz uma relaxação das restrições linearizadas de modo a garantir que a interseção entre o conjunto viável e a Região de Confiança seja não vazia (fig. 4.1). Nas seções seguintes serão vistos detalhes do algoritmo de Vardi.

4.1.1 Um Passo de Quase Newton

Deseja-se solucionar o problema de minimização não linear (NEP). Pelas condições de otimalidade para (NEP), se (x_*, λ_*) é tal que $\nabla L(x_*, \lambda_*) = 0$, $c(x_*) = 0$ e $\nabla^2 L(x_*, \lambda_*)$ é definida positiva, então x_* é mínimo local de (NEP). Assim sendo, a fim de encontrar

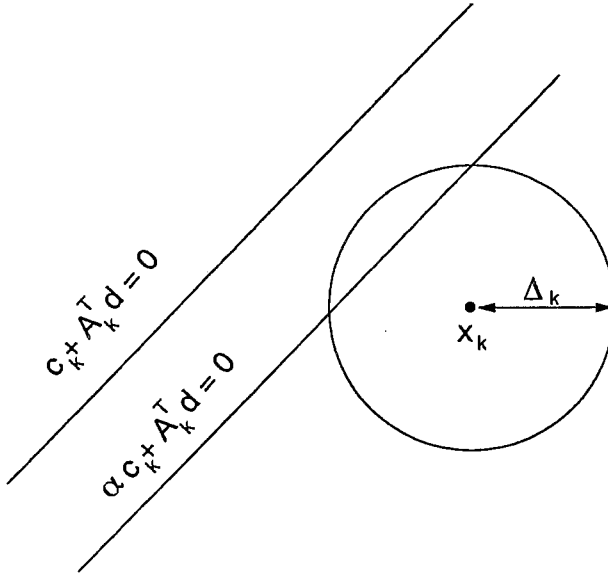


Figura 4.1: Subproblema de Vardi

uma solução para o problema original devemos resolver

$$\begin{aligned} \min \quad & L(x, \lambda) \\ \text{s.a} \quad & c(x) = 0, \end{aligned} \tag{4.1}$$

onde $L(x, \lambda) = f(x) + \lambda^T c(x)$ é a função Lagrangeana associada ao problema (NEP). O gradiente de L é dado por

$$\nabla L(x, \lambda) = \begin{pmatrix} \nabla_x L(x, \lambda) \\ \nabla_\lambda L(x, \lambda) \end{pmatrix} = \begin{pmatrix} \nabla f(x) + A(x)\lambda \\ c(x) \end{pmatrix}$$

onde

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{pmatrix} \text{ e } A = \begin{pmatrix} \frac{\partial c_1(x)}{\partial x_1} & \dots & \frac{\partial c_m(x)}{\partial x_1} \\ & \ddots & \\ \frac{\partial c_1(x)}{\partial x_n} & \dots & \frac{\partial c_m(x)}{\partial x_n} \end{pmatrix}.$$

A matriz Hessiana de L será denotada por

$$\nabla^2 L(x, \lambda) = \begin{pmatrix} \nabla_{xx}^2 L(x, \lambda) & \nabla_{x\lambda}^2 L(x, \lambda) \\ \nabla_{\lambda x}^2 L(x, \lambda) & \nabla_{\lambda\lambda}^2 L(x, \lambda) \end{pmatrix} = \begin{pmatrix} \nabla^2 f(x) + \sum_{i=1}^m \lambda_i \nabla^2 c_i(x) & A \\ A^T & 0 \end{pmatrix},$$

com

$$\nabla^2 f(x) = \left[\frac{\partial^2 f}{\partial x_k \partial x_l}(x) \right]_{k,l=1,\dots,n} \text{ e } \nabla^2 c_i(x) = \left[\frac{\partial^2 c_i}{\partial x_k \partial x_l}(x) \right]_{k,l=1,\dots,n} \quad i = 1, \dots, m.$$

Suponha que para todo $z \in \mathfrak{R}^n$ tal que $A(x_*)^T z = 0$, tem-se $z^T \nabla_{xx}^2 L(x_*, \lambda_*) z > 0$ (lembre-se que, por hipótese, A tem posto completo). Admita também, por enquanto, que $c(x)$ tem posto completo $\forall x \in \mathfrak{R}^n$; mais adiante, será visto como lidar com o caso em que essa condição não é satisfeita. Considere aproximações definidas positivas $B \approx \nabla_{xx}^2 L(x, \lambda)$, então, aproximações para a matriz $\nabla^2 L(x, \lambda)$ serão do tipo:

$$\bar{B} = \begin{pmatrix} B & A(x) \\ A(x)^T & 0 \end{pmatrix} \approx \nabla^2 L(x, \lambda) = \begin{pmatrix} \nabla_{xx}^2 L(x, \lambda) & A(x) \\ A(x)^T & 0 \end{pmatrix}.$$

Com isso, um passo de quase Newton para 4.1 é dado por

$$\begin{pmatrix} d \\ \Delta \lambda \end{pmatrix} = -\bar{B}^{-1} \nabla L(x_k, \lambda_k) = - \begin{pmatrix} B_k & A_k \\ A_k^T & 0 \end{pmatrix}^{-1} \begin{pmatrix} g_k + A_k \lambda_k \\ c(x_k) \end{pmatrix}.$$

No processo iterativo, B_k é atualizada a cada iteração através da fórmula BFGS. A matriz $\nabla_{xx}^2 L(x, \lambda)$ não é necessariamente definida positiva, mas o uso de aproximações definidas positivas garantem direções de descida sem interferir na convergência local. Os métodos de quase Newton, normalmente, depois de determinar a direção d fazem uma busca linear para determinar um $0 \leq \alpha \leq 1$ tal que $x_{k+1} = x_k + \alpha d$, $\lambda_{k+1} = \lambda_k + \alpha \Delta \lambda_k$ satisfazem um certo critério (como um decréscimo na função penalidade). Neste caso, porém, isto não será feito uma vez que é usada a metodologia de Região de Confiança.

Uma maneira equivalente de calcular o passo de Newton, descrito acima, é resolver o problema de minimização da função Lagrangeana com multiplicador λ_k , sujeito às restrições originais linearizadas no ponto x_k , ou seja, solucionar o modelo de Programação Quadrática

$$\begin{aligned} \min_d \quad & L(x_k, \lambda_k) + \nabla_x L(x_k, \lambda_k)^T d + \frac{1}{2} d^T B d \\ \text{s.a} \quad & c_k + A_k^T d = 0 \end{aligned} \quad (4.2)$$

Isto é, essencialmente, o modelo (SQP) descrito em 3.1.1. Desta forma, tem-se que d é a solução de 4.2 e $\Delta \lambda$ o vetor dos multiplicadores de Lagrange correspondentes às restrições lineares. Isto pode ser visto facilmente observando as condições de otimalidade para 4.2.

4.2 Descrição do Algoritmo

Na seção anterior obtivemos o modelo quadrático que será usado para solucionar (NEP). Acrescentando a restrição de Região de Confiança ao problema 4.2, temos

$$\begin{aligned} \min_d \quad & L(x_k, \lambda_k) + \nabla_x L(x_k, \lambda_k)^T d + \frac{1}{2} d^T B_k d \\ \text{s.a} \quad & c_k + A_k^T d = 0 \\ & \|d\| \leq \Delta_k \end{aligned} \quad (4.3)$$

Para resolver a possível incompatibilidade das restrições, substitui-se $c_k + A_k^T d = 0$ por $\alpha c_k + A_k^T d = 0$, onde α depende de Δ_k e é escolhido de modo que $\{d; \alpha c_k + A_k^T d = 0\} \cap \{d; \|d\| \leq \Delta_k\} \neq \emptyset$.

Das condições de otimalidade para um mínimo local d de 4.3, devemos ter

$$\begin{aligned} (B_k + \mu I)d + A_k \Delta \lambda_k + \nabla_x L(x_k, \lambda_k) &= 0 \\ \alpha c_k + A_k^T d &= 0, \end{aligned}$$

onde $\Delta \lambda$ é o vetor dos multiplicadores de Lagrange das restrições linearizadas e μ é o multiplicador da restrição de confiança. Isto motiva o seguinte passo de quase Newton para o problema 4.3:

$$\begin{pmatrix} d(\mu) \\ \Delta \lambda(\mu) \end{pmatrix} = - \begin{pmatrix} B + \mu I & A_k \\ A_k^T & 0 \end{pmatrix}^{-1} \begin{pmatrix} \nabla_x L(x_k, \lambda_k) \\ \alpha(\mu) c_k \end{pmatrix}, \quad (4.4)$$

onde

$$\alpha(\mu) = \min \left\{ 1, \max \left\{ \frac{1}{\mu}, \frac{1}{\mu^2} \frac{\|B\| \|\bar{Q} \nabla f(x)\|}{\|T^{-T} \Pi^T h(x)\|} \right\} \right\}$$

e μ é escolhido de modo que $\|d(\mu)\| \leq \Delta_k$; Q é uma matriz ortogonal, Π uma matriz de permutação e T matriz $m \times n$ triangular superior não singular, obtidas da decomposição de $A(x)$ tais que

$$QA(x)\Pi = \begin{pmatrix} T \\ 0 \end{pmatrix}.$$

O teorema a seguir, provado por *Vardi*, mostra que $d(\mu)$ e $\Delta \lambda(\mu) + \lambda_k$ não dependem de λ_k ; isto é interessante porque nos diz que o modelo quadrático da próxima iteração não depende da escolha do multiplicador de Lagrange do subproblema anterior.

Além disso, o teorema estabelece que $\|d(\mu)\|$ é uma função monótona decrescente de $\mu \geq 0$, repetindo, em essência, a relação de $\|d(\cdot)\|$ com o raio Δ_k . Este fato é também importante para estudo da convergência. O teorema e mostra, ainda, o comportamento assintótico de $d(\mu)$ e $\Delta\lambda(\mu) + \lambda_k$ quando $\mu \rightarrow \infty$.

Teorema 4.2.1 *Considere o passo $(d(\mu), \Delta\lambda(\mu))$ definido em 4.4. Então $d(\mu)$ e $\lambda + \Delta\lambda(\mu)$ não dependem de λ para todo $\mu \geq 0$. Quando $\mu \rightarrow \infty$, $(\lambda + \Delta\lambda(\mu)) \rightarrow \tilde{\lambda}_+ \equiv [A(x)^T A(x)]^{-1}[c(x) - A(x)^T \nabla f(x)]$ e $d(\mu) - \frac{1}{\mu} \nabla L(x, \tilde{\lambda}_+) \rightarrow 0$. Finalmente, $\|d(\mu)\|$ é uma função monótona decrescente de $\mu \geq 0$.*

A aceitação de um novo ponto é, como foi visto no caso irrestrito, determinada através de uma função mérito, no caso, a Penalidade Exata

$$PL(x) = f(x) + \sum_{i=1}^m \rho_i |c_i(x)|, \quad (4.5)$$

onde os ρ_i 's são constantes fixas chamadas *coeficientes de penalidade*. A seguir, uma descrição do algoritmo de Vardi.

ALGORITMO I - Conceitual

PASSO 1 : Inicialize com $x_0, \lambda_0, \Delta_0, B_0, k = -1$;

PASSO 2 : $k = k + 1$;

PASSO 3 : Encontre μ_k (tente primeiro $\mu_k = 0$) e $d_k, \Delta\lambda_k$ tal que

$$\begin{pmatrix} d_k \\ \Delta\lambda_k \end{pmatrix} = - \begin{pmatrix} B_k + \mu_k I & A_k \\ A_k^T & 0 \end{pmatrix}^{-1} \begin{pmatrix} g_k + A_k \lambda_k \\ \alpha(\mu_k) c_k \end{pmatrix} \quad (4.6)$$

$$\|d_k\| \leq \Delta_k, \mu_k \geq 0 \text{ e } \mu_k(\|d_k\| - \Delta_k) = 0;$$

PASSO 4 : Cheque se $PL(x_k + d_k) < PL(x_k)$. Se não, reduza Δ_k e volte para o passo 3;

PASSO 5 : Cheque a convergência. Se não verificada, continue;

PASSO 6 : Compare a aproximação quadrática da Lagrangeana L com seu valor real e de seus gradientes em $(x_k + d_k, \lambda_k + \Delta\lambda_k)$, dependendo, aumente ou diminua Δ_{k+1} ;

PASSO 7 : Calcule $y_k = \nabla_x L(x_k + d_k, \lambda_k + \Delta\lambda_k) - \nabla_x L(x_k, \lambda_k + \Delta\lambda_k)$ e atualize $B_{k+1} = BFGS(B_k, d_k, y_k)$;

PASSO 8 : $x_{k+1} = x_k + d_k$; $\lambda_{k+1} = \lambda_k + \Delta\lambda_k$: vá para o passo 2.

Para garantir que o algoritmo está bem definido é importante mostrar que, em cada iteração, haverá um (raio suficientemente pequeno ou) $\mu > 0$ suficientemente grande, tal que a função Penalidade $PL(x)$ 4.5 diminuirá. Isto é estabelecido pelo teorema a seguir, provado em [Var85].

Teorema 4.2.2 *Se $(d, \Delta\lambda)$ é calculado como em 4.4 e $\nabla L(x, \tilde{\lambda}_+) \neq 0$, então existe $\mu > 0$ suficientemente grande tal que $L(x + d(\mu), \lambda + \Delta\lambda(\mu)) < L(x, \lambda + \Delta\lambda(\mu))$. Além disso, suponha que $\forall i = 1, \dots, m$*

$$\rho_i > |(\tilde{\lambda}_+)_{i}| = |[(A(x)^T A(x))^{-1}(c(x) - A(x)^T \nabla f(x))]_i| \quad (4.7)$$

então, existe $\mu > 0$ suficientemente grande tal que $PL(x + d(\mu)) < PL(x)$, onde PL é definida por 4.5.

Uma escolha semelhante para os ρ_i 's é feita por Han em [Han77].

Enfatizamos duas boas propriedades do algoritmo de Vardi :

1. quando o raio de confiança Δ é suficientemente grande, a restrição de Região de Confiança torna-se inativa, o passo de quase Newton completo é tomado e $\mu = 0$. Quando Δ é muito pequeno, e conseqüentemente, μ é muito grande, $d \approx -\frac{1}{\mu} \nabla_x L(x, \tilde{\lambda}_+)$ que é uma direção de descida máxima na minimização de $L(x, \tilde{\lambda}_+)$. Como foi estabelecido no Teorema 4.2.1, $\tilde{\lambda}_+$ não depende de B e tem sido usado, inclusive, em outros algoritmos para Otimização com restrições de igualdade (por ex. [Tap77]);
2. se for conhecida uma cota para o lado direito de 4.7, pode-se fixar os parâmetros ρ_i 's na função Penalidade e não há necessidade de trocá-los no processo de convergência.

4.3 Resultados de Convergência

Para estabelecer o teorema de convergência de seu algoritmo, Vardi faz uma mudança no procedimento de escolha de μ_k , em cada iteração, da seguinte forma: admite que no passo 3 do Algoritmo I, ao invés de se tomar μ_k tal que $\|d_k(\mu_k)\| \leq \Delta_k$, escolhe-se μ_k que minimize a função Penalidade $PL(x_k, d_k(\mu_k))$ (isto é possível devido ao Teorema 4.2.2). Assim, pode-se substituir os passos 3 e 4 do Algoritmo I por:

Ache $\mu_k, d_k(\mu_k)$ e Δ_k tal que

$$PL(x_k + d_k(\mu_k)) = \min_{\mu: \mu \geq 0} PL(x_k + d_k(\mu))$$

(Isto será denotado por “busca μ -exata”). Com este procedimento, Vardi obtém um algoritmo semelhante ao de Han ([Han77]) e, assim sendo, o teorema que estabelece as propriedades de convergência global de seu algoritmo é similar ao estabelecido por Han.

Teorema 4.3.1 *Suponha que o Algoritmo I é executado com busca μ -exata e existem dois números positivos α e β tais que $\alpha z^T z \leq z^T B_k z \leq \beta z^T z$ para cada k e qualquer $z \in \mathbb{R}^n$. Admita, também, que $\rho_i, i = 1, \dots, m$ satisfazem a desigualdade 4.7 em todos os pontos x_k gerados pelo Algoritmo e o conjunto $\{x : f(x) + \sum \rho_i |c_i(x)| < f(x_0) + \sum \mu_i |c_i(x_0)|\}$ é limitado. Então, a sequência $\{x_k\}$ se mantém limitada, e se \tilde{x} é um ponto de acumulação qualquer, $c(\tilde{x}) = 0$ e existe $\tilde{\lambda} \in \mathbb{R}^m$ tal que $g(\tilde{x}) + A(\tilde{x})\tilde{\lambda} = 0$.*

prova: [Var85].

O teorema acima garante, portanto, que qualquer ponto de acumulação da sequência $\{x_k\}$ gerada pelo Algoritmo I é um ponto de Kuhn-Tucker para (NEP).

Será visto, agora, as propriedades locais do Algoritmo I. Suponha que $\lim\{(x_k, \lambda_k)\} = (x_*, \lambda_*)$ é uma solução local e que $d_k \neq 0 \forall k$. Admite-se, por enquanto, que $\nabla_{xx}^2 L(x_*, \lambda_*)$ é uma matriz definida positiva; o caso contrário será discutido posteriormente.

Foi visto no Capítulo 3 que o uso de funções penalidade do tipo PL pode interferir na convergência local, ocorrendo que, mesmo quando se está próximo da solução, o uso de PL pode impedir a convergência superlinear (Veja por ex. [CLHP82]). Para solucionar esta dificuldade, Vardi propõe uma mudança na função Mérito quando se

está numa vizinhança de pontos viáveis, substituindo $PL(x)$ pela função Lagrangeana $L(x, \lambda) = f(x) + \lambda^T c(x)$. No entanto, quando não se estiver numa vizinhança de um ponto viável, será usada a função Penalidade PL , a fim de se manter as propriedades globais do Algoritmo. Assim sendo, para estudar as propriedades locais, usa-se o Algoritmo II, dado a seguir, que é uma variante do Algoritmo I em que $L(x, \lambda)$ é usada no lugar da função $PL(x)$.

ALGORITMO II

PASSO 1 : Inicialize com $x_0, \lambda_0, \Delta_0, B_0, k = 1$;

PASSO 2 : $k = k + 1$;

PASSO 3 e 4 : Determine Δ_k tal que, o passo d_k , calculado pela fórmula 4.6, satisfaz

$$L(x_k + d_k, \lambda_k + \Delta\lambda_k) < L(x_k, \lambda_k + \Delta\lambda_k) \quad (4.8)$$

(Tente primeiro $\Delta_k = \|(\bar{B})^{-1}\nabla L(x_k, \lambda_k)\|$, isto é, verifique se o passo de Quase Newton satisfaz à condição acima);

PASSO 5 : Cheque a convergência. Se não for verificada, continue;

PASSO 6 : Calcule $y_k = \nabla_x L(x_k + d_k, \lambda_k + \Delta\lambda_k) - \nabla_x L(x_k, \lambda_k + \Delta\lambda_k)$ e atualize $B_{k+1} = BFGS(B_k, d_k, y_k)$;

PASSO 7 : $x_{k+1} = x_k + d_k; \lambda_{k+1} = \lambda_k + \Delta\lambda_k$; vá para o passo 2;

Os principais instrumentos usados por Vardi, para estabelecer os resultados de convergência local de seu algoritmo, estão em [BDM73]. A seguir, são dados os resultados mais importantes da teoria de convergência local para o Algoritmo II. Antes, defina $M = [\nabla_{xx}^2 L(x_*, \lambda_*)]^{1/2}$ e $\|A\|_M = \|MAM\|_F$, onde $\|d\|_F = \sqrt{\sum_{ij} d_{ij}^2}$.

Teorema 4.3.2 *Existem ϵ e δ positivos tais que, se*

$$\|(x_0, \lambda_0) - (x_*, \lambda_*)\| < \epsilon \text{ e } \|B_0^{-1} - \nabla_{xx}^2 L(x_*, \lambda_*)^{-1}\|_M < \delta,$$

então $\forall k \Delta_k = \|(\bar{B}_k)^{-1}\nabla L(x_k, \lambda_k)\|$ é aceito. Além disso, se $(x_{k+1}, \lambda_{k+1}) = (x_k, \lambda_k) - (\bar{B}_k)^{-1}\nabla L(x_k, \lambda_k)$ e $B_{k+1} = BFGS(B_k, d_k, y_k)$, então $\|(x_{k+1}, \lambda_{k+1}) - (x_, \lambda_*)\| < \epsilon$ e*

$\| B_{k+1}^{-1} - \nabla_{xx}^2 L(x_*, \lambda_*)^{-1} \| < 2\delta, \forall k$. Finalmente, $\| (x_{k+1}, \lambda_{k+1}) - (x_*, \lambda_*) \| \leq t \| (x_k, \lambda_k) - (x_*, \lambda_*) \|$ para algum $0 < t < 1$, de tal forma que $\{(x_k, \lambda_k)\}$ converge para (x_*, λ_*) linearmente.

O teorema acima nos diz que, na vizinhança de uma solução local do problema (NEP), o Algoritmo II toma passos de Quase Newton e converge linearmente para a solução (x_*, λ_*) .

Teorema 4.3.3 *Existem ϵ e δ positivos tais que, se*

$$\| (x_0, \lambda_0) - (x_*, \lambda_*) \| < \epsilon \text{ e } \| B_0^{-1} - \nabla_{xx}^2 L(x_*, \lambda_*)^{-1} \|_M < \delta$$

e o Algoritmo II é seguido, então a convergência superlinear em $\{(x_k, \lambda_k)\}$ é atingida.

Para o caso em que $\nabla_{xx}^2 L(x_*, \lambda_*)$ não é definida positiva, não se pode falar em $\| B_0^{-1} - \nabla_{xx}^2 L(x_*, \lambda_*)^{-1} \|_M < \delta$ no Teorema 4.3.2. Note, porém, que a hipótese de que $\forall z$ tal que $A(x_*)^T z = 0$, $z^T \nabla_{xx}^2 L(x_*, \lambda_*) z > 0$, em termos da decomposição Q-R

$$Q_* A(x_*)^T = \begin{pmatrix} T \\ 0 \end{pmatrix}$$

e da partição $Q_* = \begin{bmatrix} \bar{Q}_* \\ \bar{Q}_* \end{bmatrix}$ é equivalente a dizer que $\bar{Q}_* \nabla_{xx}^2 L(x_*, \lambda_*) \bar{Q}_*$ é uma matriz definida positiva. De acordo com [Pow78], é suficiente supor que $(\bar{Q}_0 B \bar{Q}_0)^{-1}$ (\bar{Q}_0 da decomposição Q-R de $A(x_0)$) está próxima de $(\bar{Q}_* \nabla_{xx}^2 L(x_*, \lambda_*) \bar{Q}_*)$ para se obter convergência local rápida.

Na próxima seção, serão vistos algoritmos variantes que usam a metodologia de Vardi.

4.4 Variantes do Algoritmo de Vardi

4.4.1 O Algoritmo de Byrd, Schnabel e Shultz

Byrd, Schnabel e Shultz ([BSS87]) propõem uma variante do Algoritmo de Vardi em que o subproblema,

$$\begin{aligned} \min_{d \in \mathbb{R}^n} \quad & q_k(d) = g_k^T d + \frac{1}{2} d^T B_k d \\ \text{s.a} \quad & \alpha c_k + A_k^T d = 0 \\ & \| d \| \leq \Delta_k \end{aligned} \tag{4.9}$$

é reduzido a um problema do tipo Região de Confiança Irrestrita, através de uma projeção no espaço tangente às restrições. A direção d é decomposta em

$$d = \alpha v_k + Z_k u \tag{4.10}$$

onde Z_k é uma base ortonormal para o espaço nulo de A_k^T , $u \in \mathbb{R}^{n-m}$ é a componente de d no espaço nulo de A^T e $v_k \in Im A$ é a componente no espaço imagem de A . Substituindo

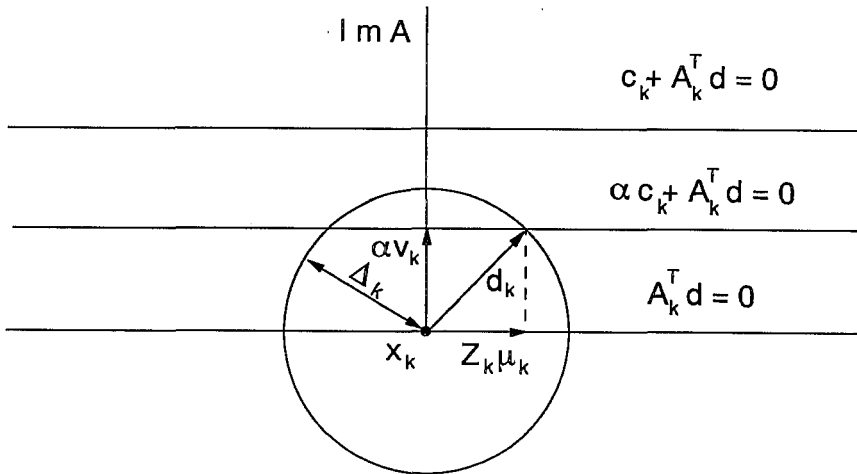


Figura 4.2: Direção de Byrd-Schnabel-Shultz

4.10 na equação $\alpha c_k + A_k^T d = 0$ obtemos

$$v_k = -A_k(A_k^T A_k)^{-1} c_k. \tag{4.11}$$

Em termos da decomposição 4.10 o subproblema 4.9 é equivalente a

$$\begin{aligned} \min_u \quad & \tilde{g}_k^T Z_k u + \frac{1}{2} u^T Z_k^T B_k Z_k u \\ \text{s.a} \quad & \| u \|^2 \leq \Delta_k^2 - (\alpha \| v_k \|^2), \end{aligned}$$

com $\tilde{g}_k = g_k + B_k \alpha v_k$, cuja solução pode ser obtida pelos métodos descritos no Capítulo 2.

A escolha de α está ligada ao controle da componente v_k . Observe que a restrição de Região de Confiança requer, em particular, que

$$\alpha \|v_k\| < \Delta_k.$$

Para garantir que é obtido um progresso significativo, no sentido de satisfazer às restrições, a cada iteração exige-se também que, quando $\alpha < 1$, o passo no espaço imagem das restrições não seja muito pequeno com relação à Região de Confiança, ou seja,

$$\alpha \|v_k\| \geq \theta \Delta_k,$$

onde $\theta > 0$ é uma constante. A reunião dessas duas condições fornece o seguinte intervalo para α :

$$\frac{\theta \Delta_k}{\|v_k\|} \leq \alpha \leq \frac{\Delta_k}{\|v_k\|}.$$

Este intervalo é usado para a escolha de α no passo (5) do algoritmo, descrito a seguir.

ALGORITMO B-S-S

Dados $\eta \in (0, 1)$, $\theta \in (0, 1]$, $\xi \in (0, \theta)$, $\tau_1, \tau_2 \in (0, 1)$, $\rho > 0$ e $\mu \in \mathfrak{R}^n$ com $\mu > 0$.

PASSO 1: Faça $k = 0$ e entre com x_0 ;

PASSO 2: Calcule f_k, g_k, c_k, A_k, Z_k e $B_k \in \mathfrak{R}^{n \times n}$;

PASSO 3: Se $Z_k^T g_k = 0$ e $c_k = 0$, então PARE;

PASSO 4: Tome $\Delta \geq \Delta_{k-1}$ inicial, e (qualquer) $\Delta_0 > 0$ se $k = 0$;

PASSO 5: Calcule $v_k = -A_k(A_k^T A_k)^{-1} c_k$;

PASSO 6: Escolha $\alpha_k \in [\min\{1, \theta(\Delta / \|v_k\|)\}, \min\{1, \Delta / \|v_k\|\}]$;

PASSO 7: Calcule $\tilde{\Delta} = (\Delta^2 - \alpha_k^2 \|v_k\|^2)^{\frac{1}{2}}$;

PASSO 8: Encontre o passo u_k que resolve

$$\begin{aligned} \min_u \quad & (g_k + \alpha_k B_k v_k)^T Z_k u + \frac{1}{2} u^T Z_k^T B_k Z_k u \\ \text{s.a} \quad & u^2 \leq \tilde{\Delta}^2 \end{aligned}$$

e faça

$$d_k = \alpha_k v_k + Z_k u_k;$$

PASSO 9: Calcule $\tilde{\mu}_k = -(A_k^T A_k)^{-1} A_k^T (g_k + \frac{1}{2} \alpha_k B_k v_k)$; para cada i , se $\mu^i < |\tilde{\mu}_k^i| + \rho$, então faça $\mu^i = |\tilde{\mu}_k^i| + 2\rho > |\tilde{\mu}_k^i| + \rho$;

PASSO 10: Calcule $pred_k = -g_k^T d_k + \frac{1}{2} d_k^T B_k d_k + \sum_{i=1}^m \mu^i \{ |c_k^i| - |c_k^i + (a_k^i)^T d_k| \}$;

PASSO 11: Calcule $f_+ = f(x_k + d_k)$ e $c_+ = c(x_k + d_k)$;

PASSO 12: Calcule $ared_k = f_k - f_+ - \sum_{i=1}^m \mu^i \{ |c_k^i| - |c_+^i| \}$;

PASSO 13: Se

$$\frac{ared_k}{pred_k} \geq \eta$$

então faça $x_{k+1} := x_k + d_k$; $k := k + 1$ e vá para o passo 2; senão, se $\|v\| \leq \xi \Delta$ então

$$ared_k = f_k - f(x_k + d_k + w) - \sum_{i=1}^m \mu^i \{ |c_k^i| - |c^i(x_k + d_k + w)| \},$$

onde $w = -A_k (A_k^T A_k)^{-1} c_+$.

Se $\frac{ared_k}{pred_k} \geq \eta$ então faça $x_{k+1} = x_k + d_k + w$, $k = k + 1$ e vá para o passo 2; faça $\Delta := \tau \Delta$ para algum $\tau \in [\tau_1, \tau_2]$ e vá para o passo 5.

Note que a condição $\|v\| \leq \xi \Delta$, no passo (13), garante que a correção de segunda ordem w é tomada somente se a componente v_k é pequena com relação a Δ , e a condição $\xi \in (0, \theta)$, imposta no início, implica que a correção é feita somente se as restrições linearizadas não estão relaxadas, ou seja, se se está próximo da região viável. Em resumo, o algoritmo não faz uma correção de segunda ordem a cada iteração, mas somente quando ocorrem os seguintes fatos:

- a direção d não decresce a função Mérito suficientemente (i. é, quando $\frac{ared}{pred} < \eta$);
- as restrições não estão relaxadas e $\|v_k\|$ é, no máximo, uma fração fixa de Δ_k .

A função mérito usada aqui é :

$$\phi(x) = f(x) + \sum_{i=1}^m \rho^i | c^i(x) |$$

onde os ρ^i 's são pesos positivos. Estes pesos serão aumentados, se necessário, em qualquer iteração, mas poderão eventualmente permanecer fixos. Byrd, Schnabel e Shultz fazem uma hipótese sobre o tamanho dos μ^i 's que envolve as quantidades

$$\tilde{\rho}_k = -(A_k^T A_k)^{-1} A_k^T (g_k + \frac{1}{2} \alpha B_k v_k)$$

que expressam o equilíbrio entre a mudança na função objetivo e a mudança nas restrições. Hipóteses semelhantes a estas são necessárias em algoritmos que fazem busca linear na função Mérito (veja, por ex., [Han77]).

A fim de evitar as dificuldades na convergência local causadas pelo uso de funções Mérito não diferenciáveis do tipo $\phi(x)$ (Efeito Maratos por ex.), Byrd, Schnabel e Shultz fazem uma correção de segunda ordem $w_k = -A_k(A_k^T A_k)^{-1}c(x_k + d_k)$ e mostram que isso é suficiente para superar as possíveis dificuldades. Para o estabelecimento dos resultados de convergência de seu algoritmo, impõem (além das hipóteses padrão relacionadas com limites e continuidade do problema (NEP)) condições sobre a componente u da direção d , análogas às condições dadas em [SSB85] para o caso irrestrito. Em resumo, essas condições dizem que:

- o passo no espaço nulo de A^T fornece um decréscimo suficiente no modelo (condição necessária em todos os resultados de convergência);
- u decrescerá o modelo significativamente se uma direção de curvatura negativa está presente (condição necessária para o resultado de convergência global de segunda ordem);
- se o minimizador exato do modelo sujeito às restrições linearizadas está dentro da Região de Confiança, então essa será a direção tomada (condição necessária para o resultado de convergência local) (Veja [BSS87]).

Com isso, Byrd, Schnabel e Shultz provam convergência local superlinear e convergência global com condições de segunda ordem (eles foram os primeiros a obter tal resultado para um algoritmo de otimização não linear restrito). As principais propriedades estabelecidas por eles são:

- o algoritmo está bem definido, no sentido em que cada iteração terminará com um passo aceitável depois de um número finito de iterações;
- qualquer ponto de acumulação x_* da sequência de iteradas $\{x_k\}$, gerada pelo Algoritmo B-S-S, satisfaz às condições necessárias de primeira ordem para solução de (NEP), isto é, $c(x_*) = 0$ e $Z(x_*)^T g(x_*) = 0$;
- a sequência de iteradas, se convergir, converge para um ponto que satisfaz às condições necessárias de segunda ordem para um mínimo local de (NEP);
- usando uma boa aproximação da Hessiana da Lagrangeana, a correção de segunda ordem permite provar que a restrição de Região de Confiança tornar-se-á inativa, na vizinhança de um ponto que satisfaz às condições suficientes de segunda ordem para uma solução de (NEP). Assim, se a sequência de iteradas chega suficientemente perto de tal ponto, ela convergirá q-quadraticamente.

4.4.2 O Algoritmo de Zhang e Zhu

Zhang e Zhu ([ZZ90]) usam a estratégia de Região de Confiança para obter convergência global para o método (tipo Quase-Newton) de *Nocedal e Overton* ([NO85]). A principal diferença entre o método de Nocedal e Overton e os métodos de Quase-Newton usuais é que, no primeiro, a matriz atualizada em cada iteração é uma aproximação da matriz quadrada $Z(x)^T B(x, \lambda) Z(x)$ de ordem t , onde $B(x, \lambda) = \nabla_{xx}^2 L(x, \lambda)$, enquanto que, nos métodos Quase-Newton são atualizadas aproximações da matriz $\nabla_{xx}^2 L(x, \lambda)$. Em cada iteração, o algoritmo de Nocedal e Overton parte de uma decomposição de A_k do tipo $A_k = [Y_k Z_k][R_k 0]^T$ e resolve as equações lineares

$$\begin{aligned} R_k^T p_k^Y &= -c_k \\ B_k^T p_k^Z &= -Z_k^T g_k, \end{aligned}$$

obtendo p_k^Y e p_k^Z ; faz $d_k^Y = Y_k p_k^Y$, $d_k^Z = Z_k p_k^Z$ e toma $x_{k+1} = x_k + d_k$, onde $d_k = d_k^Y + d_k^Z$. O algoritmo termina quando

$$\|c_k\| + \|Z_k^T g_k\| \leq \epsilon,$$

onde ϵ é uma tolerância de convergência. A matriz B_k é atualizada usando as fórmulas DFP ou BFGS. Veja [NO85] para detalhes.

Uma das vantagens desse método é que o tamanho da matriz atualizada B_k pode ser reduzido consideravelmente. Além disso, para a análise da convergência é preciso supor somente que a Projeção Bilateral $Z_*^T B_* Z_*$ é definida positiva, onde $B_* = B(x_*, \lambda_*)$. Como a direção d_k é obtida via decomposição nos subespaços complementares $N(A^T)$ e $Im(A)$, o subproblema é decomposto em dois subproblemas menores, tornando os cálculos mais simples. Diante disso, o método é promissor, principalmente para problemas grandes.

Nocedal e Overton discutiram a convergência local e provaram que se o ponto inicial x_0 está suficientemente próximo da solução x_* , e se $B_0 \approx Z_*^T B_* Z_*$, então a sequência de iteradas $\{x_k\}$ gerada por seu algoritmo converge para x_* . Sob a hipótese

$$\lim_{k \rightarrow \infty} \frac{\| (B_k - Z_*^T B_* Z_*) p_k^Z \|}{\| d_k \|} = 0$$

e algumas outras condições, é provada a convergência q-superlinear passo dois, ou seja

$$\lim_{k \rightarrow \infty} \frac{\| x_{k+1} - x_* \|}{\| x_{k-1} - x_* \|} = 0.$$

Zhang e Zhu provam que o uso de Região de Confiança não interfere na convergência q-superlinear do método de Nocedal e Overton.

O subproblema que Zhang e Zhu resolvem é do tipo que aparece em minimização irrestrita e para resolver o problema de incompatibilidade das restrições é usada a técnica de Vardi. Em cada iteração é resolvido o subproblema quadrático

$$\begin{aligned} \min_{p^Z} \quad & (Z_k^T g_k)^T p^Z + \frac{1}{2} (p^Z)^T B_k p^Z \\ \text{s.a} \quad & \| p^Z \| \leq \Delta_k \end{aligned} \quad (4.12)$$

obtido através da projeção da direção de busca no espaço tangente às restrições. Se p_k^Z é a solução de 4.12, seja

$$d_k^Z = Z_k p_k^Z \quad (4.13)$$

e

$$d_k^Y = \alpha_k Y_k u_k, \quad (4.14)$$

onde $u_k = -R_k^{-T} c_k$ (isto porque $c_k + A_k^T v_k = c_k - R_k^T Y_k^T Y_k R_k^{-T} c_k = 0$). O valor de α_k é determinado pela seguinte minimização unidimensional

$$\begin{aligned} \epsilon_k &= \min \| c_k + \alpha A_k^T Y_k u_k \|^2 \\ &= \min_{0 \leq \alpha \leq \Delta_k / \| u_k \|} (1 - \alpha)^2 \| c_k \|^2 \\ &= (1 - \alpha_k)^2 \| c_k \|^2. \end{aligned}$$

Claramente, α_k é dado por

$$\alpha_k = \begin{cases} 1, & \text{se } c_k = 0 \\ 1, & \text{se } c_k \neq 0 \text{ e } \frac{\Delta_k}{\|u_k\|} \geq 1 \\ \Delta_k / \|u_k\|, & \text{se } \frac{\Delta_k}{\|u_k\|} < 1 \end{cases} \quad (4.15)$$

Note que $\alpha_k \in (0, 1]$, e que se $c_k = 0$, $d_k^Y = 0$ qualquer que seja o valor de α_k , assim, por conveniência, foi definido $\alpha_k = 1$ se $c_k = 0$. O Algoritmo usa o Lagrangeano Aumentado $L_A(x, \lambda, \eta)$ como função Mérito (ϕ), e o parâmetro η é atualizado a cada iteração.

Algoritmo Z-Z

PASSO 0: Escolha $\epsilon_1 > 0, 0 < \eta_1 < \eta_2 < 1, \delta' > 0, \Delta_* > 0, \nu > 1, 0 < \beta < \frac{1}{2}$. Tome $x_0 \in \mathfrak{R}^n, \Delta_0 > 0$, e uma matriz definida positiva B_0 (pode-se tomar $B_0 = I$). Faça $k = 0, j = 1, \epsilon = \epsilon_1$.

PASSO 1: Calcule f_k, c_k, g_k, A_k . Faça uma decomposição de A_k

$$A_k = \begin{bmatrix} Y_k & Z_k \end{bmatrix} \begin{bmatrix} R_k \\ 0 \end{bmatrix}.$$

Obtenha $\lambda_k = R_k^{-1} Y_k g_k$.

PASSO 2: Se

$$\|c_k\| + \|Z_k^T g_k\| = 0 \quad (4.16)$$

então pare. Caso contrário, se

$$\|c_k\| + \|Z_k^T g_k\| < \epsilon \quad (4.17)$$

então faça $\epsilon := \epsilon_{j+1} = \min\{\|c_k\| + \|Z_k^T g_k\|, \epsilon_j / \nu\}$ e $j := j + 1$.

PASSO 3: Resolva o subproblema 4.12 para obter sua solução primal p_k^Z . Faça $d_k^Z = Z_k p_k^Z$.

PASSO 4: Obtenha α_k de 4.15 e determine d_k^Y por 4.14. Faça $d_k = d_k^Z + d_k^Y$.

PASSO 5: Calcule f, c, g, A, λ , em $x_k + d_k^Z$ e $x_k + d_k$ respectivamente.

PASSO 6: Calcule

$$\begin{aligned} pred_k &= (Z_k^T g_k) p_k^Z + \frac{1}{2} (p_k^Z)^T B_k p_k^Z + \frac{1}{2} [\lambda(x_k + d_k^Z) - \lambda_k] A_k^T d_k \\ &- [\lambda(x_k + d_k) - \lambda_k]^T (c_k + \frac{1}{2} A_k^T d_k) + \rho_k (\|c_k + A_k^T d_k\|^2 - \|c_k\|^2). \end{aligned}$$

Se

$$pred_k \leq -\beta \epsilon \Delta_k, \quad (4.18)$$

então vá para o Passo 7. Caso contrário, faça $\Delta_k := \Delta_k / \nu$ e vá para o Passo 3.

PASSO 7: Obtenha $ared_k = \phi(x_k + d_k, \rho_k) - \phi(x_k, \rho_k)$ e calcule $\xi_k = ared_k / pred_k$.

PASSO 8: Faça

$$x_{k+1} = \begin{cases} x_k + d_k & \text{se } \xi_k > 0 \\ x_k & \text{caso contrário,} \end{cases}$$

$$\lambda_{k+1} = \begin{cases} \lambda(x_k + d_k) & \text{se } \xi_k > 0 \\ \lambda_k & \text{caso contrário,} \end{cases}$$

$$\Delta_{k+1} = \begin{cases} \Delta / \nu & \text{se } \xi_k < \eta_1 \\ \nu \Delta_k & \text{se } \xi_k > \eta_2 \text{ e } \nu \Delta_k \leq \Delta_* \\ \Delta_k & \text{outro caso} \end{cases}$$

PASSO 9: Atualize B_k por DFP, BFGS ou PSB para obter B_{k+1} . Faça $k = k + 1$ e volte para o Passo 2.

Zhang e Zhu mostram que a sequência $\{x_k\}$ obtida pelo Algoritmo Z-Z tem ponto de acumulação x_* que satisfaz às condições de Kunh-Tucker ($c_* = 0, Z_*^T g_* = 0$), e mostram também que, depois de um número finito de iterações, a condição 4.17 será satisfeita, não importando quão pequeno seja ϵ . Na prática, ao invés de 4.16, o Algoritmo deve parar quando 4.17 for satisfeita. O valor típico do parâmetro β , no Passo 6, é 10^{-6} e se a condição 4.18 não for satisfeita o raio de Região de Confiança é reduzido e d_k e $pred_k$ recalculados. Zhang e Zhu provam ainda que, depois de decrescer Δ_k um número finito de vezes, a desigualdade 4.18 será satisfeita, o que torna o processo bem definido. Sob hipóteses usuais, são provadas propriedades de convergência global e de convergência local superlinear passo dois para o Algoritmo Z-Z.

O algoritmo de Zhang e Zhu tem algumas características em comum com o algoritmo de Powell e Yuan ([PY91]) (seção 5.6.1): por ex., ambos usam função penalidade diferenciável (Lagrangeano Aumentado) para controlar a convergência, entretanto, os subproblemas resolvidos por cada um deles são distintos. Enquanto Powell e Yuan resolvem um subproblema quadrático com duas restrições quadráticas (que exige uma técnica especial para solução), Zhang e Zhu resolvem um subproblema quadrático semelhante ao que se tem em Otimização Irrestrita. O teorema de convergência global de Powell e Yuan mostra que seu algoritmo deve parar em um ponto x_k com d_k correspondente igual a zero. Embora isso garanta $g_k + A_k^T \lambda_k = 0$ para um vetor de multiplicadores λ_k , não exclui a possibilidade de $c_k \neq 0$, assim, o ponto de parada x_k pode não ser viável. Essa terminação não ocorre no algoritmo de Zhang e Zhu (veja [ZZ90]). Por outro lado, o algoritmo de Powell e Yuan tem uma taxa de convergência q-superlinear passo um, enquanto que Zhang e Zhu garantem somente taxa q-superlinear passo dois e, para isso, precisam ainda de um cálculo extra no ponto $x_k + d_k^Z$ (que Powell e Yuan não fazem).

As hipóteses impostas por Zhang e Zhu para análise de convergência são semelhantes às usadas em [BSS87]; apesar de ambos obterem um subproblema tipo Região de Confiança Irrestrita, é difícil fazer uma comparação clara entre seus algoritmos (como foi feita para Powell e Yuan). Isto porque Nocedal e Overton, de que se originou o trabalho de Zhang e Zhu, utilizam a decomposição do espaço através de $N(A^T)$ e $Im(A)$ para a representação do gradiente de f , ao passo que Byrd, Schnabel e Shultz o fazem para a decomposição da direção d .

4.4.3 Proposta de uma Variante do Algoritmo de Vardi

O subproblema tratado por Vardi ([Var85]) é ¹

$$(PV) \quad \min_d \quad a_k^T d + \frac{1}{2} d^T B d$$

$$s.a \quad \alpha c_k + A_k^T d = 0$$

$$\frac{1}{2} \|d\|^2 \leq \frac{1}{2} \Delta_k^2,$$

onde $a_k = \nabla_x L(x_k, \lambda_k)$. Nossa proposta é reduzir o subproblema (PV) (que é quadrático com restrições lineares e uma quadrática) a um problema quadrático com somente res-

¹a restrição de confiança $\|d\| \leq \Delta_k$ é substituída pela desigualdade equivalente $\frac{1}{2} \|d\|^2 \leq \frac{1}{2} \Delta_k^2$, a fim de evitar problemas de não diferenciabilidade inerentes à primeira.

trições lineares, cuja solução pode ser obtida por métodos bem conhecidos. Para isso, fazemos uma penalização da restrição de Região de Confiança e obtemos o seguinte problema:

$$(Pt) \quad \min_d \quad a_k^T d + \frac{1}{2} d^T B d + \frac{t}{2} \|d\|^2 \\ \text{s.a} \quad \alpha c_k + A_k^T d = 0,$$

onde $t \geq 0$ é um número convenientemente escolhido. Note que (Pt) pode ser escrito como:

$$\min_d \quad a_k^T d + \frac{1}{2} d^T (B + tI) d \\ \text{s.a} \quad \alpha c_k + A_k^T d = 0,$$

É fácil ver que (Pt) e (PV) são equivalentes, comparando suas condições de otimalidade de primeira ordem. Com efeito, seja L_{PV} a Lagrangeana associada a (PV) dada por

$$L_{PV}(d, \lambda, \mu) = a_k^T d + \frac{1}{2} d^T B_k d + \lambda^T (\alpha c_k + A_k^T d) + \frac{\mu}{2} (\|d\|^2 - \Delta^2),$$

e seu gradiente em relação a d

$$\nabla_d L_{PV}(d, \lambda, \mu) = a_k + B_k d + A_k \lambda + \mu d,$$

logo, a condição $\nabla_d L_{PV}(d, \lambda, \mu) = 0$ nos dá

$$(B_k + \mu I) d = -(a_k + A_k \lambda). \quad (4.19)$$

Por outro lado, se L_{Pt} é a Lagrangeana associada a (Pt), temos

$$L_{Pt}(d, \lambda) = a_k^T d + \frac{1}{2} d^T (B_k + tI) d + \lambda^T (\alpha c_k + A_k^T d)$$

e

$$\nabla_d L_{Pt}(d, \lambda) = a_k + (B_k + tI) d + A_k \lambda.$$

Portanto, $\nabla_d L_{Pt}(d, \lambda) = 0$ implica

$$(B_k + tI) d = -(a_k + A_k \lambda). \quad (4.20)$$

As equações 4.19 e 4.20 são iguais se tomarmos $t = \mu$, onde μ é o multiplicador de Lagrange da restrição quadrática em (PV).

Uma solução $d = d(t)$ para o problema (Pt) (que depende obviamente do t escolhido) satisfaz a equação (ver [Gay81] e [Sor82])

$$(B + tI)d = -(a_k + A_k \lambda).$$

Suponha que a matriz B seja definida positiva, então $(B + tI)$ também o é, e

$$d = d(t) = -(B + tI)^{-1}(a_k + A_k \lambda). \quad (4.21)$$

Se $d(t)$ é definida por 4.21, $\|d(t)\|^2$ é uma função monótona decrescente de t . Com efeito, seja $\{\xi_i, \omega_i\}_{i=1, \dots, n}$ o conjunto de autovalores e autovetores de B . Então, $\{\xi_i + t, \omega_i\}_{i=1, \dots, n}$ e $\{\frac{1}{\xi_i + t}, \omega_i\}_{i=1, \dots, n}$ são, respectivamente, os conjuntos de autovalores e autovetores das matrizes $B + tI$ e $(B + tI)^{-1}$. Podemos escrever:

$$a_k + A_k^T \lambda = \sum_i \alpha_i \omega_i.$$

Assim, supondo os ω 's ortonormais, temos:

$$\begin{aligned} \|d(t)\|^2 &= \|-(B + tI)^{-1}(a_k + A_k^T \lambda)\|^2 \\ &= \|(B + tI)^{-1} \sum_i \alpha_i \omega_i\|^2 \\ &= \left\| \sum_i \frac{\alpha_i}{\xi_i + t} \omega_i \right\|^2 \\ &= \sum_i \frac{\alpha_i^2}{(\xi_i + t)^2}. \quad \square \end{aligned}$$

Além disso, se $d(t)$ é solução de (Pt) devemos ter

$$\|d(t)\|^2 = \Delta_k^2,$$

(veja [DS83]). O que nos fornece a seguinte relação entre o raio de Região de Confiança Δ_k e o parâmetro t :

$$\Delta_k^2 = \sum_i \frac{\alpha_i^2}{(\xi_i + t)^2}.$$

Se B não for definida positiva, a escolha de $t > -\min \xi_i$ garante os resultados acima, bem como justifica o algoritmo. A seguir, é dado um algoritmo para determinar $d(t)$.

Algoritmo I

PASSO 0: Inicie com $t \in (t_{min}, t_{max})$.

PASSO 1: Resolva (Pt) para obter $d(t)$.

PASSO 2: Se $\|d(t)\| \leq \Delta_k$ pare; senão, calcule um novo t (por Dicotomia, por exemplo) e vá para o Passo 1.

Considerando a função Penalidade Exata

$$PL(x) = f(x) + \sum_i \eta_i |c_i(x)|$$

como função Mérito, temos o algoritmo:

ALGORITMO - Conceitual

PASSO 0: Inicie com $x_0, \Delta_0, k = -1$.

PASSO 1: Faça $k = k + 1$.

PASSO 2: Determine $d_k = d(t)$ através do **Algoritmo I**.

PASSO 3: Se $PL(x_k + d_k) < PL(x_k)$, vá para o Passo 4; senão, faça $\Delta_k = \frac{1}{2}\Delta_k$ e vá para o Passo 2.

PASSO 4: Se convergiu, pare; senão, faça $x_{k+1} = x_k + d_k$, atualize a matriz B_k para B_{k+1} usando a fórmula BFGS e vá para o Passo 1.

Uma idéia similar a que expomos aqui é feita por *Zhang* em [Zha92] para o subproblema de Celis-Dennis-Tapia (veja Cap. 5).

Capítulo 5

O Algoritmo de Celis-Dennis-Tapia

5.1 Introdução

Será visto neste capítulo o método de Região de Confiança para a solução de (NEP) proposto por Maria Rosa Celis em sua tese de Doutorado. *Celis* ([Cel85]) introduz uma nova (diferente da proposta por Vardi) técnica para resolver a incompatibilidade entre as restrições linearizadas e a Restrição de Confiança. Ela sugere que se substitua o subproblema de Região de Confiança padrão, associado a (NEP)

$$\begin{aligned} \min_d \quad & g^T d + \frac{1}{2} d^T B d \\ \text{s.a} \quad & c + A^T d = 0 \\ & \| d \| \leq \Delta \end{aligned}$$

por um outro um pouco diferente. A motivação para a escolha deste novo subproblema é a seguinte: suponha que se quer resolver $c(x) = 0$ usando Região de Confiança. Tem-se um ponto corrente x_k e um limite Δ_k do tamanho do passo que se quer dar a partir de x_k . Em cada iteração k , a direção é calculada resolvendo

$$\begin{aligned} \min_d \quad & \frac{1}{2} \| c_k + A_k^T d \|^2 \\ \text{s.a} \quad & \| d \| \leq \Delta_k. \end{aligned} \tag{5.1}$$

É usado $\| c_k + A_k^T d \|^2$, ao invés do modelo de Newton de $\| c(x_k + d) \|^2$ dado por

$$c_k^T c_k + 2c_k^T A_k^T d + d^T (A_k A_k^T + \nabla^2 c_k c_k) d_k.$$

As duas expressões diferem somente pelo termo $\nabla^2 c_k c_k$, incluído pelo modelo de Newton na matriz Hessiana ($A_k A_k^T + \nabla^2 c_k c_k$), mas omitido em $\|c_k + A_k^T d\|^2$. Entretanto, como se está resolvendo o problema $c(x) = 0$, tem-se que na solução x_* , $\nabla^2 c(x_*) c(x_*) = 0$. Desta forma, usando $\|c_k + A_k^T d\|^2$ tem-se as mesmas propriedades de convergência que se teria caso fosse usado o modelo de Newton, com a vantagem de se evitar o uso de informações de segunda ordem de $c(x)$.

Tomando o passo de descida máxima para o modelo 5.1 (ou seja, o passo de Cauchy d^{CP}), sob hipóteses razoáveis, pode-se demonstrar convergência de primeira ordem. Para formalizar esse resultado, Celis aplica a teoria de *Shultz, Schnabel e Byrd* ([SSB85]) para o problema de minimização da função $F(x) = \frac{1}{2}c(x)^T c(x)$, cujo teorema principal é dado a seguir.

Teorema 5.1.1 *Considere o passo d_k em cada iteração dado por $d_k = x_{k+1} - x_k$ e a redução prevista*

$$pred_k(d) = -\nabla F(x_k)^T d - \frac{1}{2}d^T B_k d,$$

onde $\nabla F(x_k) = A_k^T c_k$, $B_k = A_k A_k^T$, $k = 0, 1, 2, \dots$. Seja $c : \mathbb{R}^n \rightarrow \mathbb{R}^m$ duas vezes continuamente diferenciável, $F(x)$ como definida acima, e a Hessiana de $F(x)$ limitada no espaço todo. Seja $\{x_k\}$, $x_k \in \mathbb{R}^n$, $k = 0, 1, 2, \dots$ uma sequência de pontos gerada pela aplicação de um algoritmo de Região de Confiança padrão ao problema de minimização de $F(x)$, com ponto inicial $x_0 \in \mathbb{R}^n$. Se em cada iteração a direção d_k satisfaz

$$pred_k(d_k) \leq pred_k(d^{CP})$$

e B_k é uniformemente limitada em k , então, $\nabla F(x_k)$ converge para zero.

Com base no teorema acima, são escolhidos passos que satisfazem à condição

$$\|c_k + A_k^T d\| \leq \|c_k + A_k^T d^{CP}\|.$$

Isto significa que o passo tomado deve ser, no mínimo, tão linear viável quanto o passo de Cauchy d_{CP} para minimização de $\frac{1}{2}c^T(x)c(x)$. Agora, considere o conjunto

$$D_{CP} = \{d : \|d\| \leq \Delta_k \text{ e } \|c_k + A_k^T d\| \leq \|c_k + A_k^T d^{CP}\|\},$$

ou seja, D_{CP} é o conjunto dos passos de x_k que estão dentro da Região de Confiança e dão, no mínimo, tanto decréscimo na norma euclidiana dos resíduos das restrições linearizadas,

quanto o passo de Cauchy. Escolhendo qualquer ponto em D_{CP} , em cada iteração, será gerada uma sequência $\{x_k\}$ que, sob hipóteses razoáveis, deverá convergir para um ponto viável. Celis tira vantagem dessa liberdade tomando uma direção d que minimiza algum modelo quadrático da função objetivo sobre D_{CP} (fig. 5.1).

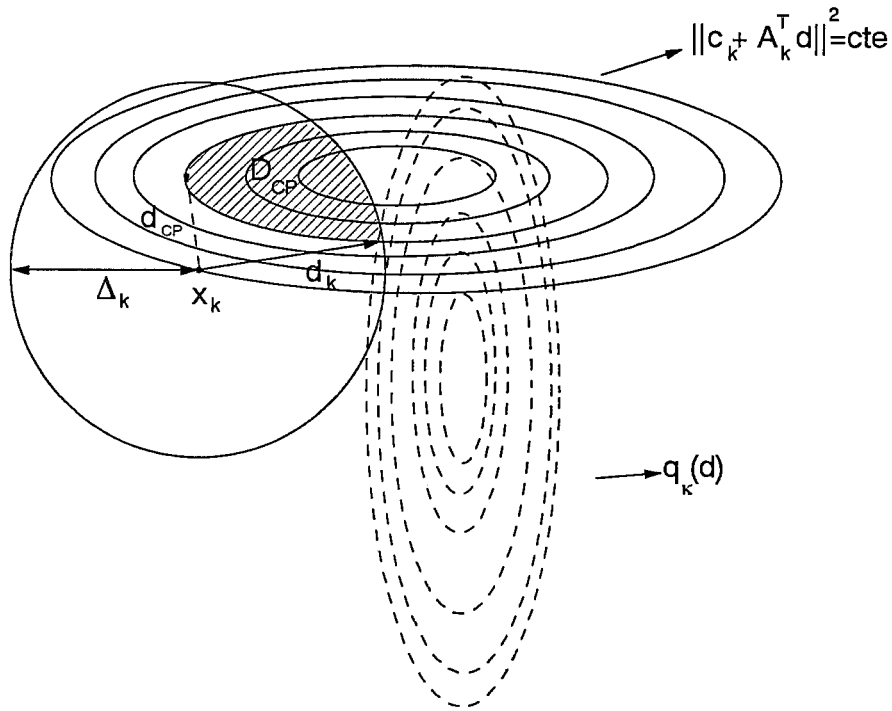


Figura 5.1: O Passo CDT

Assim sendo, o passo é calculado solucionando o problema

$$\begin{aligned}
 (SPCDT) \quad & \min \quad q_k(d) = a_k^T d + \frac{1}{2} d^T B_k d \\
 & \text{s.a} \quad \|d\| \leq \Delta_k \\
 & \quad \quad \|c_k + A_k^T d\| \leq \theta_k,
 \end{aligned}$$

onde $a_k \in \mathfrak{R}^n$, $B_k \in \mathfrak{R}^{n \times n}$ é simétrica não singular, $\theta_k = \|c_k + A_k^T d^{CP}\|$. O problema acima é conhecido como *Subproblema de Celis-Dennis-Tapia* (devido a um artigo publicado por eles) e sua solução é dada pelo teorema:

Teorema 5.1.2 *Se d_* resolve o subproblema SPCDT, então d_* é uma solução para a equação*

$$(B + \mu I + \eta AA^T)d_* = -(a + A\eta c)$$

com $\mu, \eta \geq 0$, $\|d_*\| \leq \Delta$, $\mu(\Delta - \|d_*\|) = 0$, $\|c + A^T d_*\| \leq \theta$ e $\eta(\theta - \|c + A^T d_*\|) = 0$.

Na próxima seção será discutido o algoritmo para determinar uma solução de (SPCDT).

5.2 Solução do Subproblema SPCDT

A fim de determinar uma solução de (SPCDT), Celis procede da seguinte maneira: verifica primeiro se a solução d_k^{QP} do subproblema quadrático

$$\begin{aligned} \min_d \quad & g_k^T d + \frac{1}{2} d^T B_k d \\ \text{s.a} \quad & c_k + A_k^T d = 0 \end{aligned}$$

está dentro da Região de Confiança, se $\|d_k^{QP}\| \leq \Delta_k$, então $d = d_k^{QP}$ é a solução de (SPCDT). Supondo que $\|d_k^{QP}\| > \Delta_k$, são consideradas três possibilidades: somente a Região de Confiança é ativa; somente a restrição de θ -viabilidade é ativa; ambas as restrições são ativas. Para determinar se somente a restrição $\|d\| \leq \Delta_k$ é ativa, Celis encontra a solução d_Δ do problema de Região de Confiança Irrestrita

$$\begin{aligned} \min_d \quad & g_k^T d + \frac{1}{2} d^T B_k d \\ \text{s.a} \quad & \|d\| \leq \Delta_k, \end{aligned} \tag{5.2}$$

usando as técnicas já conhecidas para este caso. Se d_Δ é tal que

$$\|c + A^T d_\Delta\| \leq \theta_k,$$

então, a restrição de viabilidade não é ativa e d_Δ é a solução de (SPCDT). Caso contrário, é solucionado o problema

$$\begin{aligned} \min_d \quad & g_k^T d + \frac{1}{2} d^T B_k d \\ \text{s.a} \quad & \|c_k + A_k^T d\| \leq \theta_k \end{aligned} \tag{5.3}$$

para determinar se somente a restrição de viabilidade é ativa na solução. Celis resolve esse problema através de uma técnica tipo Hook Step, obtendo a solução d_θ . Se $\|d_\theta\| \leq \Delta_k$,

então d_θ é a solução de (SPCDT). Senão, é o caso em que ambas as restrições são ativas e será necessário resolver o problema (SPCDT) completo; novamente, Celis usa o método Hook Step. Entretanto, ela sugere também o uso de métodos Dogleg, propondo uma curva Dogleg e uma tipo Duplo Dogleg (para detalhes [Cel85]). O algoritmo a seguir, define uma iteração para determinar uma solução aproximada de (SPCDT), resumindo o que foi exposto acima.

ALGORITMO I: *Solução Aproximada de (SPCDT).*

PASSO 1: Determine uma solução aproximada d_Δ para o problema 5.2.

PASSO 2: Se $\|c + A^T d_\Delta\| \leq \theta$ então $d := d_\Delta$; PARE.

PASSO 3: Determine uma solução aproximada d_θ para o problema 5.3.

PASSO 4: Se $\|d_\theta\| \leq \Delta$ então $d := d_\theta$; PARE.

PASSO 5: Determine uma solução aproximada d para o problema (SPCDT).

5.3 Aceitação do Passo e Atualização da Região de Confiança

A função Mérito usada por Celis é o Lagrangeano Aumentado

$$L_A(x, \lambda, \eta) = f(x) + \lambda^T c(x) + \frac{1}{2} \eta c(x)^T c(x),$$

onde o parâmetro η é determinado pela solução de (SPCDT). Com o uso de uma função Mérito diferenciável são evitados problemas de velocidade de convergência que aparecem quando se usa funções não diferenciáveis (por ex. o Efeito Maratos).

Um passo x_+ será aceito como próximo iterado se

$$L_A(x_+, \lambda_+, \eta_+) \leq L_A(x_k, \lambda_+, \eta_+) + \xi \nabla_x L_A(x_k, \lambda_+, \eta_+)^T d_k, \quad (5.4)$$

onde λ_+ e η_+ são, respectivamente, o vetor multiplicador e a constante de penalidade associados à direção d_k , e ξ é uma constante positiva no intervalo $(0, \frac{1}{2})$. Se x_+ não

satisfaz 5.4, a Região de Confiança é reduzida por um fator entre $\frac{1}{10}$ e $\frac{1}{2}$, e torna-se a calcular uma direção numa região menor. Por outro lado, uma vez aceita x_+ como nova iterada, Celis analisa se deve ou não aumentar a Região de Confiança. Para isto, compara a redução real

$$\Delta L_A \equiv L_A(x_+, \lambda_+, \eta_+) - L_A(x_k, \lambda_+, \eta_+), \quad (5.5)$$

com a redução prevista

$$\Delta(L_A)_{pred} \equiv \nabla_x L_A(x_k, \lambda_+, \eta_+)^T (x_+ - x_k) + \frac{1}{2} (x_+ - x_k)^T H_k (x_+ - x_k), \quad (5.6)$$

onde $H_k = B_k + \eta AA^T$ e $B_k = \nabla_{xx}^2 L(x_k, \lambda_k)$ ou alguma aproximação dela. Se o ajuste é tal que

$$| \Delta(L_A)_{pred} - \Delta L_A | \leq \beta | \Delta L_A |, \quad (5.7)$$

onde $\beta = 0.1$, ou a redução real em L_A é muito grande, isto é,

$$L_A(x_+, \lambda_+, \eta_+) \leq L_A(x_k, \lambda_+, \eta_+) + \nabla_x L_A(x_k, \lambda_+, \eta_+), \quad (5.8)$$

então, a Região de Confiança é aumentada e um novo x_+ é calculado, usando o modelo corrente sujeito a essa região maior. O valor de d_k é guardado, pois se a nova direção obtida não satisfizer 5.4, volta-se ao último valor aceito. Para atualizar o raio de Região de Confiança, Celis usa o mesmo procedimento da minimização irrestrita. O novo raio Δ_+ é obtido em função do raio anterior e de $\| d_k \|$. O parâmetro de penalidade é escolhido dentro do conjunto

$$N_k = \left\{ \eta : \begin{aligned} & -\frac{1}{2} \frac{\| \nabla_x L_A(x_k, \lambda_+, \eta_+) \|^4}{\nabla_x L_A(x_k, \lambda_+, \eta_+)^T H_k \nabla_x L_A(x_k, \lambda_+, \eta_+)} \\ & \geq \nabla_x L_A(x_k, \lambda_+, \eta_+)^T d_k + \frac{1}{2} d_k^T H_k d_k \end{aligned} \right\}$$

dependendo de quão bem o modelo estima a função mérito. A seguir, o algoritmo de aceitação do passo e atualização da Região de Confiança.

ALGORITMO II: *Aceitação de um Passo e Atualização da Região de Confiança.*

Seja $0 < \sigma_1 < \sigma_2 < 1$ e $0 < \tau_1 < \tau_2 < 1 < \tau_3$ constantes dadas. Defina N_k como no parágrafo anterior e

$$\begin{aligned} N_{dbl} &= \{ \eta : | \Delta(L_A)_{pred} - \Delta L_A | \leq \beta | \Delta L_A | \text{ ou } \Delta L_A \leq \sigma_2 \Delta(L_A)_{pred} \} \\ N_{inc} &= \{ \eta : \Delta L_A < \sigma_2 \Delta(L_A)_{pred} \} \\ N_{sme} &= \{ \eta : \sigma_1 \Delta(L_A)_{pred} < \Delta L_A < \sigma_2 \Delta(L_A)_{pred} \} \\ N_{dec} &= \{ \eta : L_A(x_+, \lambda_+, \eta_+) < L_A(x_k, \lambda_+, \eta_+) + \xi \nabla_x L_A(x_k, \lambda_+, \eta_+)^T d_k \}. \end{aligned}$$

PASSO 1: Se $c_k + A_k^T d_k = 0$, então,

- faça $\lambda_+ = \lambda^{QP}$
- determine N_k

senão

- faça $\lambda_+ = 0$ e
- $N_k = \{\eta_k\}$

PASSO 2: Se $N_k = \emptyset$, então

- escolha $\Delta_k \in [\tau_1 \|d_k\|, \tau_2 \|d_k\|]$
- calcule um novo d_k

PASSO 3: Determine N_{dec}

PASSO 4: Se $N_k \cap N_{dec} = \emptyset$ então

- escolha $\Delta_k \in [\tau_1 \|d_k\|, \tau_2 \|d_k\|]$
- calcule um novo d_k

PASSO 5: Aceite $x_+ = x_k + d_k$, escolha $\eta_+ \in N_k \cap N_{dec}$

PASSO 6: Determine N_{inc}

PASSO 7: Se $N_k \cap N_{inc} \neq \emptyset$ então

- escolha $\eta_+ \in N_k \cap N_{inc}$
- $\Delta_+ \in [\Delta_k, \tau_3 \Delta_k]$
- PARE

PASSO 8: Determine N_{sme}

PASSO 9: Se $N_k \cap N_{sme} \neq \emptyset$ então

- escolha $\eta_+ \in N_k \cap N_{sme}$
- $\Delta_+ = \Delta_k$

senão

- escolha $\Delta_k \in [\tau_1 \|d_k\|, \tau_2 \|d_k\|]$.

5.4 O Algoritmo CDT

O Algoritmo de Celis-Dennis-Tapia (CDT) toma um passo SQP sempre que $\|d_k^{QP}\| \leq \Delta$, a fim de manter a rápida convergência local. Quando um passo SQP completo não pode ser tomado, resolve-se o problema (SPCDT) obtendo o passo d_k . Como foi visto na seção anterior, a função Lagrangeano Aumentado é usada para determinar a aceitação do passo e para atualizar a Região de Confiança. A constante de penalidade η é obtida na solução de (SPCDT) (ela é o multiplicador de Lagrange da restrição de viabilidade) ou pelo procedimento descrito na seção anterior. Para a implementação preliminar de seu algoritmo, Celis usa informações de segunda ordem.

Em cada iteração o problema (QP) é definido por

$$\begin{aligned} \min_d \quad & g_k^T d + \frac{1}{2} d^T \nabla_{xx}^2 L(x_k, \lambda_k) d \\ \text{s.a} \quad & c_k + A_k^T d = 0 \end{aligned}$$

e o problema (SPCDT)

$$\begin{aligned} \min \quad & g_k^T d + \frac{1}{2} d^T B_k d \\ \text{s.a} \quad & \|c_k + A_k^T d\| \leq \theta_k \\ & \|d\| \leq \Delta_k, \end{aligned}$$

onde $B_k = \nabla_{xx}^2 L(x_k, \lambda_k)$ se $\theta_k = 0$ e $B_k = \nabla^2 f(x_k)$ caso contrário. Isto permite que se tenha passos tão próximos quanto possível do passo SQP, se as restrições linearizadas são satisfeitas. A escolha do multiplicador usado na matriz Hessiana $\nabla_{xx}^2 L(x_k, \lambda_k)$ depende da última direção tomada. Se d_{k-1} satisfaz às restrições linearizadas, escolhe-se λ_k como sendo o multiplicador associado ao passo SQP, isto é, $\lambda_k = \lambda^{QP}$. Essa escolha é motivada pelo fato de que λ^{QP} é o único multiplicador que permite que as restrições linearizadas sejam satisfeitas. Por outro lado, se d_{k-1} não satisfaz às restrições linearizadas, é escolhido o multiplicador pela fórmula

$$\lambda_k = (A_k^T A_k)^{-1} (c_k - A_k^T g_k),$$

que, na prática, é bem sucedida (veja, por ex., [MMLC72]). A estimativa inicial λ_0 também é escolhida dessa maneira. Finalmente, o raio de Região de Confiança inicial Δ_0 é escolhido como sendo uma percentagem do tamanho do passo de Cauchy para o problema 2.1. Essa escolha segue da estratégia sugerida por Powell ([Pow70]) para o caso irrestrito. Vejamos a descrição formal do algoritmo.

ALGORITMO CDT

Dado $x_0 \in \mathfrak{R}^n$, determine λ_0, Δ_0 . Para $k = 0, 1, 2, \dots$ até convergir faça:

PASSO 1: calcule $g_k, \nabla^2 f_k, A_k, \nabla_{xx}^2 L(x_k, \lambda_k)$;

PASSO 2: determine a solução d_k^{QP} (e λ_k^{QP}) do problema (QP);

PASSO 3: se $\|d_k^{QP}\| \leq \Delta_k$, então $d_k = d_k^{QP}$ e vá para 5;

PASSO 4: encontre uma solução aproximada d_k para o problema (SPCDT) pelo Algoritmo I;

PASSO 5: $x_+ = x_k + d_k$

PASSO 6: decida se x_+ é aceito e calcule um novo valor para Δ_k pelo Algoritmo II;

PASSO 7: se x_+ não é aceito, então vá para 4, caso contrário $x_{k+1} = x_+$ e $\Delta_{k+1} = \Delta_k$.

5.5 Variantes do Algoritmo CDT

Celis faz uma implementação preliminar de seu algoritmo, que embora necessite de um procedimento mais eficiente para solucionar o subproblema (SPCDT), fornece testes que são importantes para mostrar que o método é robusto. Vários autores propõem modificações para o método de Celis-Dennis-Tapia a fim de obter boas propriedades de convergência, mas o ponto crucial para que se consiga um algoritmo computacionalmente eficiente é um procedimento para solucionar o subproblema (SPCDT); isto é estudado por *Yuan* em [Yua90], e por *Zhang* em [Zha92], que propõe um algoritmo para solução de (SPCDT). Nas próximas seções serão estudadas algumas variantes do Algoritmo CDT, encontradas na literatura (incluindo o algoritmo de Zhang mencionado acima).

5.5.1 Algoritmo de Zhang para solução de SPCDT

Sem a restrição de θ -viabilidade o subproblema SPCDT é equivalente ao subproblema de otimização irrestrita

$$(SPI) \quad \min \quad g^T d + \frac{1}{2} d^T B d \\ \text{s.a} \quad \| d \| \leq \Delta.$$

Para o caso sem restrições, tem-se condições (fortes) de otimalidade necessárias e suficientes para uma solução global do subproblema irrestrito (veja Capítulo 2). Com base nessas condições foram construídos vários algoritmos eficientes (tipo Newton) para solução de (SPI), entretanto, quando a restrição $\| c + A^T d \|^2$ é acrescentada, uma condição de otimalidade necessária e suficiente similar não pode ser obtida, como foi provado em [Yua90]. A semi-definição positiva da matriz $B + \mu I + \eta A A^T$ (Teorema 5.1.2) pode não ser garantida para matrizes simétricas gerais B . Yuan ([Yua90]) dá exemplos demonstrando que a matriz $B + \mu I + \eta A A^T$ pode ter um autovalor negativo quando μ e η são únicos e pode até mesmo ter dois autovalores negativos numa situação desfavorável. A falta de condições de otimalidade necessárias e suficientes e de uma Hessiana da Lagrangeana (com relação a d) semi-definida positiva faz com que se torne mais difícil construir algoritmos eficientes para a solução do problema SPCDT. Yuan fornece condições suficientes para uma solução de SPCDT, sob a hipótese de que a Hessiana da Lagrangeana com relação a d é semi-definida positiva em d_* . A seguir é dado o teorema estabelecido por Yuan, com o resultado de unicidade acrescentado por Zhang em [Zha92].

Teorema 5.5.1 *A condição*

$$(B + \mu I + \eta A A^T) d_* = -(g + \eta A c) \quad (5.9)$$

onde

$$\mu = 0 \text{ e } \| d_* \| < \Delta \text{ ou } \mu \geq 0 \text{ e } \| d_* \| = \Delta \quad (5.10)$$

$$\eta = 0 \text{ e } \| c + A^T d_* \| < \theta \text{ ou } \eta \geq 0 \text{ e } \| c + A^T d_* \| = \theta \quad (5.11)$$

e a semi-definição positiva da matriz $(B + \mu I + \eta A A^T)$ são suficientes para d_* viável ser solução do problema SPCDT. Além disso, se $B + \mu I + \eta A A^T$ é definida positiva, então a solução é única.

Zhang ([Zha92]) propõe um método para solução de SPCDT que trata a restrição de θ -viabilidade sob uma perspectiva proximal e reduz o problema original a um problema tipo Região de Confiança Irrestrita, dado por

$$P(\eta) \quad \min_{d \in \mathbb{R}^n} \quad q(d) + \eta \psi(d) \\ \text{s.a} \quad \| d \| \leq \Delta,$$

onde $q(d) = g^T d + \frac{1}{2} d^T B d$ e $\psi(d) = \frac{1}{2} (\|c + A^T d\|^2 - \theta^2) \in \Re$. Uma solução para $P(\eta)$ será denotada por $d(\eta)$.

A estratégia de Zhang é motivada pela seguinte observação: se a restrição de θ -viabilidade do problema SPCDT é ativa (ele mostra que esse é o caso que interessa) e se é possível definir μ como uma função de η , tal que $\mu(\eta)$ sempre satisfaz à condição de viabilidade

$$\|d(\mu(\eta), \eta)\| \leq \Delta$$

e a condição de complementaridade 5.10 que é equivalente a

$$\mu(\eta) = 0 \text{ ou } \|d(\mu(\eta), \eta)\| - \Delta = 0,$$

então, a solução do problema SPCDT se reduz a resolver a condição de complementaridade 5.11, ou equivalentemente, solucionar a equação univariacional

$$\psi(\eta) = 0.$$

Observando as condições de otimalidade para $P(\eta)$, Zhang obtém a função $\mu(\eta)$ de maneira natural (veja [Zha92]). Ele mostra que $\psi(\eta) = \psi(d(\mu(\eta), \eta))$ e $q(d(\mu(\eta), \eta))$ são funções contínuas e monótonas; e que $\mu(\eta)$, e conseqüentemente, $d(\mu(\eta), \eta)$ e $\psi(\eta)$, são, no mínimo diferenciáveis por partes (condição necessária a fim de que se possa aplicar a método de Newton para a equação $\psi(\eta) = 0$.) Assim, Zhang reformula o problema SPCDT para o problema de encontrar um zero de uma função monotonicamente decrescente e, no mínimo, diferenciável por partes.

Na prática, substitui-se $\psi(\eta) = 0$ por $\Psi(\eta) = 0$, onde

$$\Psi(\eta) = \frac{1}{\|c + A^T d\|} - \frac{1}{\theta}.$$

A função $\Psi(\eta)$ ainda é monótona e possui as mesmas propriedades de $\psi(\eta)$. É fácil ver que $\Psi'(\eta) = \psi'(\eta) / \|c + A^T d(\eta)\|^3$, e que $\Psi(\eta)$, normalmente, é mais próxima de uma função linear do que $\psi(\eta)$.

Zhang implementa um algoritmo para solução de $\Psi(\eta)$ que é, basicamente, o método de Newton com uma forma de salvaguarda.

ALGORITMO Z-I - Solução de $\Psi(\eta) = 0$

Dada uma tolerância τ_1 , seja $\eta = 0$, $\eta_L = 0$ e $\eta_R = \infty$.

PASSO 1: Resolva $P(\eta)$ para obter $d(\eta)$ e calcule $\Psi(\eta)$.

PASSO 2: Se ($\eta = 0$ e $\Psi(\eta) > 0$) ou ($\| \| c + A^T d \| - \theta \| / \theta \leq \tau_1$), faça $d_* = d(\eta)$ e pare.

PASSO 3: Se $\Psi(\eta) < 0$, então $\eta_L = \eta$ e $\Psi_L = \Psi(\eta)$; senão $\eta_R = \eta$ e $\Psi_R = \Psi(\eta)$.

PASSO 4: Calcule o passo de Newton

$$\eta := \eta - \Psi(\eta) / \Psi'(\eta).$$

PASSO 5: Se $\eta \in (\eta_L, \eta_R)$, use o passo da *Regula Falsi*

$$\eta := \eta_L - \Psi(\eta_R - \eta_L) / \Psi_R - \Psi_L.$$

PASSO 6: Vá para o Passo 1.

A fim de solucionar $P(\eta)$ usando a estratégia Hook Step é necessário resolver a equação

$$\Upsilon \equiv \frac{1}{\| d(\mu, \eta) \|} - \frac{1}{\Delta} = 0$$

para μ , quando η é fixo. Para isso, Zhang utiliza o método de Newton como foi dado em [MS84], adicionando a restrição $\mu \geq \epsilon$ e usando um critério de parada prático. O algoritmo é dado a seguir.

ALGORITMO Z-II - Solução de $\Upsilon(\eta) = 0$ para um η fixo.

Dado $\eta, \epsilon \geq 0$ e $\tau_2 > 0$, seja $\mu \geq \epsilon$.

PASSO 1: Seja $H = B + \eta A A^T + \mu I$ e resolva $Hd = -g$ para d .

PASSO 2: Se ($\mu = \epsilon$ e $\| d \| \leq \Delta$) ou ($\| \| d \| - \Delta \| / \Delta \leq \tau_2$), então faça $d_* = d$ e pare.

PASSO 3: Calcule o passo de Newton restrito

$$\mu := \max(\epsilon, \mu + \frac{d^T d}{d^T H^{-1} d} \frac{\|d\| - \Delta}{\Delta}).$$

PASSO 4: Vá para o Passo 1.

Quando B é definida positiva, pode ser usado $\epsilon = 0$ no algoritmo acima. Entretanto, o ϵ um pouco maior que epsilon da máquina deverá ser utilizado quando B é somente semi-definida positiva, a fim de garantir que a matriz H no Passo 1 é numericamente não-singular. Do ponto de vista prático, o efeito de tal perturbação sobre a solução é insignificante.

Se ambas as restrições 5.10 e 5.11 são ativas o algoritmo de Zhang pode ser resumido (de forma bem simplificada) como o seguinte: o problema SPCDT se reduz ao sistema não-linear de duas equações com duas variáveis:

$$\Upsilon(\mu, \eta) = 0$$

$$\Psi(\mu, \eta) = 0.$$

A variável μ é eliminada solucionando a primeira equação para um η fixo. Depois, por substituição, o sistema é reduzido a única equação $\Psi(\mu(\eta), \eta) = 0$. Felizmente, ambas as equações (a primeira, para μ fixo, e a segunda, depois da substituição $\mu = \mu(\eta)$) são funções monótonas bem comportadas e fáceis de resolver.

Zhang compara seu algoritmo com o de *Yuan* em [Yua88] (único existente na época para cálculo de um passo de Região de Confiança para o problema SPCDT na norma euclidiana) e conclui que o seu é conceitualmente mais simples. Os experimentos computacionais indicam que o Algoritmo de Zhang é confiável, robusto e produz bons passos para problemas pequenos ($n \leq 20$).

5.5.2 O Algoritmo de Powell e Yuan

Powell e Yuan propõem uma escolha diferente para a cota θ no Algoritmo CDT. Eles tomam θ_k , na iteração k , como sendo o menor valor de $\|c_k + A_k^T d\|$ sujeito a $\|d\| < b\Delta_k$, para algum $b \in [b_2, b_1]$, onde b_1, b_2 são constantes pré-fixadas tais que $0 < b_2 \leq b_1 < 1$.

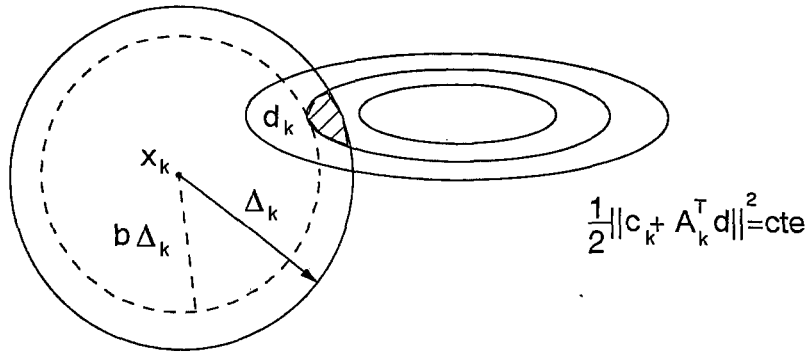


Figura 5.2: O Problema de Powell e Yuan

Ou seja, θ_k é um número que satisfaz

$$\min_{\|d\| \leq b_1 \Delta_k} \|c_k + A_k^T d\| \leq \theta_k \leq \min_{\|d\| \leq b_2 \Delta_k} \|c_k + A_k^T d\| .$$

Com isso obtêm a vantagem de que θ_k é zero em mais iterações, quando um mau condicionamento torna o passo de Cauchy pequeno (fig. 5.2).

Na k -ésima iteração, se x_k não satisfaz às condições de Kuhn-Tucker, o algoritmo de Powell e Yuan calcula uma tentativa de passo solucionando o subproblema (SPCDT) com o novo valor de θ_k , como foi definido acima. O multiplicador $\lambda(x) \in \mathbb{R}^m$ minimiza a soma dos quadrados dos resíduos das condições de Kuhn-Tucker (aproximação de primeira ordem):

$$\|g(x) - A(x)\lambda\|^2 .$$

A função Mérito (ϕ) usada é o Lagrangeano Aumentado. Para decidir a aceitação do passo, Powell e Yuan definem uma mudança M_k na função Mérito, dada por:

$$M_k = (g_k - A_k \lambda_k)^T d_k + \frac{1}{2} d_k^T B_k \hat{d}_k \quad (5.12)$$

$$- [\lambda(x_k + d_k) - \lambda_k]^T (c_k + \frac{1}{2} A_k^T d_k) + \eta_k (\|c_k + A_k^T d_k\|^2 - \|c_k\|^2),$$

onde η_k é escolhido tal que $M_k < 0$ e \hat{d}_k é a projeção ortogonal de d_k no espaço nulo de A_k^T ; e definem a razão

$$r_k = \frac{\phi_k(x_k + d_k) - \phi_k(x_k)}{M_k}$$

da mudança real para a mudança prevista. A direção d_k será aceita se $r_k > 0$, caso contrário, toma-se $x_{k+1} = x_k$. A escolha do próximo raio Δ_{k+1} depende de Δ_k , $\|d_k\|$ e r_k . Uma matriz B_{k+1} é definida e isso completa a iteração.

Na definição de M_k , o vetor \hat{d}_k aparece porque a condição usada por Powell e Yuan para as matrizes $\{B_k : k = 1, 2, 3, \dots\}$, a fim de forçar a convergência superlinear, não é que a razão $\|(B_k - B_*)d_k\| / \|d_k\|$ tende a zero quando $k \rightarrow \infty$ (com $B_* = \nabla_{xx}^2 L(x_*, \lambda_*)$), é a condição mais fraca: $\|\omega_k^T (B_k - W_*)d_k\| / \|d_k\| \rightarrow 0$ (onde ω_k é qualquer vetor normalizado tal que $A_k^T \omega_k = 0$). Com isso, eles provam convergência q-superlinear para seu algoritmo. O parâmetro de penalidade η_k é ajustado automaticamente e só é aumentado quando o passo de x_k para x_{k+1} fornece um aumento na parte Lagrangeana da função Mérito. Além disso, o uso de $\lambda(x_k + d_k)$, na expressão que define M_k , conduz a um ajuste conveniente do parâmetro de penalidade sem o cálculo de quaisquer derivadas segundas (ver [PY91] para detalhes). A seguir, uma descrição formal do algoritmo de Powell e Yuan.

ALGORITMO P-Y

PASSO 0: Dados $x_1 \in \mathfrak{R}^n$, $B_1 \in \mathfrak{R}^{n \times n}$, $\Delta_1 > 0$, $0 < b_2 \leq b_1 < 1$. Escolha $\eta_1 > 0$ e $\varepsilon > 0$ pequeno. Faça $k = 1$.

PASSO 1: Se $\|c_k\| + \|g_k - A_k \lambda_k\| \leq \varepsilon$ então PARE. Caso contrário, resolva o problema (SPCDT) para obter d_k .

PASSO 2: Calcule M_k pela fórmula 5.12. Se a desigualdade

$$M_k \leq \frac{1}{2} \eta_k (\|c_k + A_k^T d_k\| - \|c_k\|)$$

não é satisfeita, então aumente o valor de η_k como a seguir

$$\eta_k^{atual} = 2\eta_k^{anterior} + \max\{0, 2M_k^{anterior} / \|c_k\|^2 - \|c_k + A_k^T d_k\|^2\},$$

que garante que o novo valor de M_k satisfaz à desigualdade acima.

PASSO 3: Calcule a razão r_k . Faça

$$x_{k+1} = \begin{cases} x_k + d_k & \text{se } r_k > 0, \\ x_k & \text{caso contrário} \end{cases}$$

e

$$\Delta_{k+1} = \begin{cases} \max[\Delta_k, 4 \|d_k\|] & \text{se } r_k > 0.9, \\ \Delta_k & \text{se } 0.1 \leq r_k \leq 0.9 \\ \min[\Delta_k/4, \|d_k\|/2] & \text{se } r_k < 0.1. \end{cases}$$

Obtenha B_{k+1} . Faça $\eta_{k+1} = \eta_k$, $k = k + 1$ e vá para o PASSO 1.

Para o desenvolvimento de sua teoria de convergência, Powell e Yuan admitem as seguintes **Hipóteses**:

H1: Existe um conjunto convexo $\Omega \subset \mathfrak{R}^n$ fechado e limitado tal que x_k e $x_k + d_k$ estão em $\Omega \forall k$;

H2: $A(x)$ tem posto coluna completo para todo $x \in \Omega$;

H3: As matrizes $\{B_k : k = 1, 2, \dots\}$ são uniformemente limitadas em Ω .

É provado que o algoritmo termina em um número finito de iterações, com ε no passo 1, qualquer constante positiva. Powell e Yuan obtêm um limite superior para M_k , isto é importante para o teste que aumenta o valor de η_k , no sentido de garantir que a desigualdade do PASSO 2 será satisfeita em algum momento. Eles mostram ainda que a sequência $\{\eta_k : k = 1, 2, \dots\}$ é limitada (fato importante para o estabelecimento das propriedades de convergência do algoritmo). A seguir, é dado o resultado de convergência global para o Algoritmo P-Y.

Teorema 5.5.2 *Sob as Hipóteses H1, H2 e H3, o Algoritmo P-Y termina depois de um número finito de iterações. Em outras palavras, se removermos o teste de convergência do passo 1, então $d_k = 0$ para algum k ou o limite*

$$\lim_{k \rightarrow \infty} \inf [\|c_k\| + \|P_k g_k\|] = 0$$

é obtido, o qual garante que a sequência $\{x_k : k = 1, 2, \dots\}$ não é limitada longe de pontos estacionários do problema original (NEP).

Supondo $\varepsilon = 0$, Powell e Yuan provam a convergência superlinear de seu Algoritmo usando resultados provados por *Boggs, Tolle e Wang* em [BTW73] e por *Powell* em [Pow83], demonstrando que tais resultados continuam válidos quando se usa estratégia de Região de Confiança. A necessidade de calcular $\lambda(x)$ para qualquer x introduz sérias dificuldades no Algoritmo P-Y, se os gradientes das restrições se tornam linearmente dependentes. Para remediar essa situação, Powell e Yuan fazem uma mudança na definição de λ_k observando que as propriedades demonstradas continuam válidas depois da mudança. Apesar de ter sido desenvolvido a partir de um procedimento que foi testado computacionalmente, o algoritmo de Powell e Yuan não foi usado ainda para cálculos numéricos por apresentar procedimentos complicados para implementação, sendo uma das principais dificuldades a solução do subproblema CDT.

5.5.3 O Algoritmo de Mahmoud El-Alem

Mahmoud El-Alem ([EA91]) apresenta uma teoria de convergência global para uma variante do algoritmo de Celis-Dennis-Tapia, cujas principais diferenças em relação ao método original estão no esquema de atualização do parâmetro de penalidade e na escolha do limite θ no subproblema. El-Alem escolhe θ_k , na k -ésima iteração, como sendo

$$\theta_k = (1 - \tau) \|c_k\| + \tau \|c_k + A_k^T d_k^{CP}\|, \quad (5.13)$$

para algum $0 < \tau \leq 1$, onde d_k^{CP} é a direção de Cauchy. Essa escolha foi sugerida por *Celis et al.* em [CDM⁺89] e tem a vantagem de assegurar considerável liberdade no conjunto viável do subproblema, em comparação com a escolha original no Algoritmo CDT, permitindo ainda que cada iterada (obtida pela solução do subproblema) seja “puxada” em direção ao ponto ótimo de (NEP), ao invés de avançar muito na direção da viabilidade não linear às custas da otimalidade.

O algoritmo de El-Alem faz o seguinte: em cada iteração obtém a direção d_k^{QP} e o multiplicador λ_k^{QP} resolvendo o subproblema de Programação Quadrática

$$\begin{aligned} (SQP) \quad & \min_d \quad \nabla_x L(x_k, \lambda_k)^T d + \frac{1}{2} d^T B_k d \\ & s.a \quad c_k + A_k^T d = 0; \end{aligned}$$

se d_k^{QP} está dentro da Região de Confiança, i.é, $\|d_k^{QP}\| \leq \Delta_k$, então $\hat{d}_k = d_k^{QP}$ e $\Delta\hat{\lambda}_k = \Delta\lambda_k^{QP}$, onde \hat{d}_k é a tentativa de passo e $\Delta\hat{\lambda}_k$ o multiplicador de Lagrange correspondente. Caso contrário, se x_k não é viável, então o subproblema (SPCDT) é solucionado, com o valor de θ_k escolhido como foi descrito no parágrafo anterior; por outro lado, se x_k é viável, resolve-se o subproblema de Região de Confiança

$$\begin{aligned}
 (SPRC) \quad & \min_d \quad \nabla_x L(x_k, \lambda_k)^T d + \frac{1}{2} d^T B_k d \\
 & s.a \quad c_k + A_k^T d = 0 \\
 & \quad \quad \|d\| \leq \Delta_k.
 \end{aligned}$$

Note que, quando x_k é viável, $\theta_k = 0$ e o subproblema (SPCDT) se reduz ao subproblema (SPRC). A função Mérito escolhida é o Lagrangeano Aumentado. Através de experimentos numéricos, El-Alem percebeu que o algoritmo obtém melhor performance quando o parâmetro de penalidade é mantido o menor possível. Para o desenvolvimento da teoria de convergência global, é necessário que a sequência $\{\eta_k\}$, dos coeficientes de penalidade, seja não decrescente e que a redução prevista na função Mérito em cada iteração seja, no mínimo, tanto quanto uma fração do decréscimo de Cauchy na norma euclidiana das restrições linearizadas. Portanto, a idéia é manter o parâmetro η o menor possível, sujeito a essas duas condições. El-Alem parte de $\eta = 1$ e aumenta esse valor somente quando for necessário para satisfazer essas condições.

Observe que, se a tentativa de passo for obtida pela solução de (SPRC) ou (SPCDT), a escolha do multiplicador de Lagrange requer que seja resolvido o seguinte sistema linear:

$$A_k \Delta\hat{\lambda}_k = -(\nabla L_k + B_k \hat{d}_k).$$

Powell e Yuan ([PY91]) usam a estimativa do multiplicador de mínimo quadrado como fórmula de atualização do multiplicador de Lagrange. Com essa fórmula, o seguinte sistema linear tem que ser resolvido

$$A(x_k + \hat{d}_k) \Delta\hat{\lambda}_k = -(g(x_k + \hat{d}_k) - A(x_k + \hat{d}_k) \lambda_k).$$

A escolha de Powell e Yuan é mais cara, pois requer uma fatorização de $A(x_k + \hat{d}_k)$ em cada tentativa de passo, enquanto a escolha de El-Alem requer fatorização somente quando o algoritmo vai para um novo ponto depois de encontrar um passo aceitável. E se o passo é na direção d^{QP} o multiplicador é obtido sem nenhum custo extra, porque ele é o próprio multiplicador SQP.

Sob as mesmas hipóteses assumidas por Powell e Yuan, El-Alem estabelece a convergência global de seu algoritmo. Os principais resultados são enunciados abaixo.

Teorema 5.5.3 *Sob as Hipóteses enunciadas na seção anterior, em qualquer ponto (x_k, λ_k) gerado pelo algoritmo de El-Alem, ou a condição de terminação será satisfeita ou um passo aceitável será encontrado.*

O teorema acima diz, portanto, que o algoritmo está bem definido no sentido em que ele sempre encontra um passo aceitável, a partir de qualquer ponto que não satisfaça o critério de terminação. O resultado seguinte mostra que o algoritmo termina bem sucedido, ou seja ele gera uma sequência de iteradas convergindo para, no mínimo, um ponto estacionário do problema original.

Teorema 5.5.4 *Sob as hipóteses do teorema anterior, o algoritmo de El-Alem produz iteradas $\{x_k\}$ que satisfazem*

$$\liminf_{k \rightarrow \infty} [\|c_k\| + \|P_k g_k\|] = 0.$$

O propósito de El-Alem foi desenvolver uma teoria de convergência para uma variante do Algoritmo CDT, portanto, as considerações que ele faz sobre experimentos numéricos são similares às feitas por Powell e Yuan.

Capítulo 6

Conclusões

Vimos que uma das principais dificuldades encontradas, quando se usa Região de Confiança para Otimização Restrita, é que o subproblema obtido (que deverá ser resolvido a cada iteração) pode ser de difícil solução. São encontradas na literatura várias propostas para tratar essa questão, mas não se tem ainda uma que seja comprovadamente a melhor, sobretudo no que diz respeito ao aspecto computacional. Nossa proposta de algoritmo apresenta uma característica interessante, que é o fato de que o subproblema a ser resolvido a cada iteração é quadrático com restrições lineares. Como há métodos bem conhecidos e eficientes para solução de tais problemas, isso simplifica o trabalho computacional. Acreditamos, portanto, que se pode obter bons resultados com essa metodologia.

Trabalhos futuros que podem ser realizados a partir da proposta feita aqui são: testes computacionais e estabelecimento das propriedades de convergência do algoritmo.

Bibliografia

- [BDM73] C. G. Broyden, J. E. Dennis Jr., and J. J. Moré. On the local and superlinear convergence of Quasi-Newton methods. *J. Inst. Math. Anal*, 12:223–246, 1973.
- [Ber82] D. P. Bertsekas. Constrained optimization and Lagrange multiplier methods, 1982.
- [BSS87] R. H. Byrd, R. B. Schnabel, and G. A. Shultz. A trust region algorithm for nonlinearly constrained optimization. *SIAM J. Num. Anal.*, 24 , 5, 1987.
- [BTW73] P. T. Boggs, J. W. Tolle, and P. Wang. On the local convergence of Quasi-Newton methods for constrained optimization. *SIAM J. Control and Optimization*, 20:161–171, 1973.
- [CC80] T. F. Coleman and A. R. Conn. Second-order conditions for an exact penalty function. *Mathematical Programming*, 19:178–185, 1980.
- [CDM⁺89] M. R. Celis, J. E. Dennis, J. M. Martínez, R. A. Tapia, and K. Williamson. An algorithm based on a constrained trust region subproblem for nonlinear programming. in preparation, 1989.
- [CDT85] M. R. Celis, J. E. Dennis Jr., and R. A. Tapia. *A Trust Strategy for Nonlinear Equation Constrained Optimization*. SIAM Proceedings 20 in Numerical Optimization, 1985.
- [Cel85] M. R. Celis. *A Trust Region Strategy for Nonlinear Equality Constrained Optimization*. Tese de Doutorado, Rice University, 1985.
- [CLHP82] R. M. Chamberlain, C. Lemarechal, H. C. Hendersen, and M. J. D. Powell. The watch-dog technique for forcing convergence in algorithms for constrained

- optimization. *Mathematical Programming Study*, 16, 1982. in Algorithms for Constrained Optimization of Smooth Functions.
- [CR82] T. F. Coleman and Conn A. R. Nonlinear programming via an exact penalty function: asymptotic analysis. *Mathematical Programming*, 24:178–185, 1982.
- [DM79] J. E. Dennis Jr. and W. H. Mei. Two new unconstrained optimization algorithm which use function and gradient values. *J.O.T.A.*, 28:453–482, 1979.
- [DS83] J. E. Dennis Jr. and R. B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice-Hall, New Jersey, 1983.
- [DS89] J. E. Dennis Jr. and R. B. Schnabel. *A View Unconstrained Optimization*. Elsevier Science Publisher, North-Holland, 1989.
- [EA91] Mahmoud El-Alem. A global convergence theory for the Celis-Dennis-Tapia trust region algorithm for constrained optimization. *SIAM J. Num. Anal.*, 28:266–290, 1991.
- [Fle81] R. Fletcher. *Practical Methods of Optimization*, volume 2. John Wiley and Sons, New York, 1981.
- [Fle85] R. Fletcher. An l_1 penalty method for nonlinear constrained. *SIAM Num. Optimiz.*, pages 26–40, 1985.
- [Gay81] D. M. Gay. Computing optimal locally constrained steps. *SIAM J. Sci. Stat. Comp.*, 2:186–197, 1981.
- [GMSW89] P. E. Gill, W. Murray, M. A. Saunders, and M. H. Wright. *Constrained Nonlinear Programming*, volume 1. G. L. Nemhauser, North- Holland, 1989.
- [GQT66] S. M. Goldfeldt, R. E. Quandt, and H. F. Trotter. Maximization by quadratic hill-climbing. *Econometrica* 34, pages 541–551, 1966.
- [Han77] S. P Han. A globally convergent method for nonlinear programming. *J.O.T.A.*, 22:297–307, 1977.
- [Heb73] M. D. Hebden. An algorithm for minimization using exact second derivatives. Technical Report TP515, Atomic Energy Research Establishment, England, 1973.

- [Lev44] K. Levenberg. A method for the solution of certain nonlinear problems in least squares. *Quarterly of Applied Mathematics*, 2:164–168, 1944.
- [Mar63] W. D. Marquardt. An algorithm for least squares estimation of nonlinear parameters. *SIAM J. Appl. Math.*, 11:431–441, 1963.
- [Mar78] N. Maratos. *Exact Penalty Function for Finite- Dimensional and Control Optimization Problems*. Tese de Doutorado, University of London, 1978.
- [MMLC72] A. Miele, P. E. Moeseley, A. V. Levy, and G. M. Cogging. On the method of multipliers for mathematical programming problems. *J.O.T.A*, pages 1–33, 1972.
- [Mor78] J. J. Moré. *The Levenberg-Marquardt Algorithm: implementation and theory*. Springer-Verlag, Berlin, 1978.
- [MP82] D. Q. Mayne and E. Polak. A superlinearly convergent algorithm for constrained optimization problem. *Mathematical Programming Study*, 16:45–61, 1982. in *Algorithms for Constrained Optimization of Smooth Functions*.
- [MS83] J. J. Moré and D. C. Sorensen. Computing a trust region step. *SIAM J. Sci. Stat. Comp.*, 4:553–572, 1983.
- [MS84] J. J. Moré and D. C. Sorensen. Newton’s method. *MAA Studies in Mathematics*, 24:29–82, 1984.
- [NO85] J. Nocedal and M. L. Overton. Projected hessian updating algorithms for nonlinearly constrained optimization. *SIAM J. Num. Anal.*, 22:821–850, 1985.
- [Pow70] M. J. D. Powell. *A New Algorithm for Unconstrained Optimization*. Academic-Press, New York, 1970.
- [Pow78] M. J. D. Powell. *The Convergence of Variable Metric Methods for Constrained Optimization*. Nonlinear Programming 3. Academic Press, 1978.
- [Pow83] M. J. D. Powell. *Variable Metric Methods for Constrained Optimization*. Mathematical Programming. Springer-Verlag, 1983. in: A. Bachem, M. Grottschel and B. Korte, eds.

- [PY86] M. J. D. Powell and Y. Yuan. A trust region for equality constrained optimization. Technical report, University of Cambridge, 1986. Report DAMTP 1986/NA2.
- [PY91] M. J. D. Powell and Y. Yuan. A trust region for equality constrained optimization. *Mathematical Programming*, 49:169–211, 1991.
- [Rei71] C. H. Reinsh. Smoothing by spline functios II. *Numer. Math.*, 16:451–454, 1971.
- [Sor82] D. C. Sorensen. Newton’s method with a model trust region modification. *J. Numerical Analysis*, 19, 2, 1982.
- [SSB85] G. A. Shultz, R. B. Schnabel, and R. H. Byrd. A family of trust region based algorithms for unconstrained optimization with strong global convergence properties. *SIAM J. Num. Anal.*, 22:47–67, 1985.
- [Tap77] R. A. Tapia. Quasi-Newton methods for equality constrained optimization: equivalence of existing methods and a new implementation, 1977.
- [Tho75] S. W. Thomas. *Sequential Estimation Techniques for Quasi-Newton Algorithms*. Tese de Doutorado, Cornell University, Ithaca - New York, 1975.
- [Var85] A. Vardi. A trust region algorithm for equation constrained minimization: convergence properties and implementation. *SIAM J. Num. Anal.*, 22:575–591, 1985.
- [Yua84] Y. Yuan. An example of only linear convergence of trust region algorithms for non-smooth optimization. *IMA J. Num. Anal.*, 4:327–355, 1984.
- [Yua88] Y. Yuan. A dual algorithm for minimizing a quadratic function with two quadradratic constraints. Technical report, University of Cambridge, 1988. Technical Report NA3 DAMPT.
- [Yua90] Y. Yuan. On a subproblem of trust region algorithm for constrained optimization. *Mathematical Programming*, 47:53–63, 1990.
- [Zha92] Yin Zhang. Computing a Celis-Dennis-Tapia trust region step for equality constrained optimization. *Mathematical Programming*, 55:109–124, 1992.

- [ZZ90] Y. Zhang and D. T. Zhu. Projected Quasi-Newton algorithm with trust region for constrained optimization. *J.O.T.A.*, 67(2), 1990.