

Universidade Federal de Santa Catarina
Curso de Pós-Graduação em Matemática e
Computação Científica

Minimização de Quadráticas Convexas em
Caixas sobre Variedades Afins,
um sub-problema de PQS

Roger Behling

Orientador: Clóvis Caesar Gonzaga

Florianópolis

Mai de 2006

Universidade Federal de Santa Catarina

Curso de Pós-Graduação em Matemática e

Computação Científica

Minimização de Quadráticas Convexas em

Caixas sobre Variedades Afins,

um sub-problema de PQS

Dissertação submetida ao Curso de Pós-Graduação em Matemática e Computação Científica, do centro de Ciências Físicas e Matemáticas da Universidade Federal de Santa Catarina, para a obtenção do grau de Mestre em Matemática, com Área de concentração em Otimização.

Roger Behling

Florianópolis

Maior de 2006

“Agrada-te do Senhor, e ele satisfará aos desejos do teu coração”

Salmo 37-4

Agradecimentos

Eu gostaria de agradecer a Deus e a seu filho Jesus Cristo que proporcionam tantas oportunidades para que eu me torne uma pessoa melhor.

Agradeço a minha noiva Nina, que com seu carinho, abraços e beijos, me incentiva a estudar e me motiva a viver uma vida digna e com bastante amor.

Sou muito grato aos meus pais, Hans e Marilise Behling, pelo sempre auxílio financeiro e pela educação de ponta que me ofereceram.

Ao meu irmão e melhor amigo Ronan Behling por sempre ser boa companhia e fonte de auto-estima.

Agradeço ao meu professor orientador Clovis Caesar Gonzaga pela maneira humana e científica de ensinar e pelo modo apaixonante como encara a Matemática, a história da Matemática e em particular sua pesquisa em Otimização.

Obrigado à banda New Rose por servir de ponte entre arte e ciência na minha vida.

Obrigado a todos os meus amigos, professores e familiares que estão passando, e que já passaram pela minha vida deixando um pouco de si.

Resumo

Neste trabalho nós estudamos alguns métodos de Programação não Linear restrita e irrestrita dando ênfase ao problema que dá título a esta dissertação. No primeiro capítulo são estudados e enunciados métodos como os de Cauchy, Newton, Armijo, Região de Confiança e Dog Leg. No segundo, estudamos Programação Quadrática Seqüencial (PQS) pelo método de Restauração Inexata, que executa em cada iteração um passo de viabilidade e um de otimalidade. Nosso objetivo específico foi tratar do passo de otimalidade, conhecido como passo tangente do PQS, que na nossa proposta consiste em minimizar uma quadrática convexa numa caixa sobre uma variedade afim. Neste sentido, o terceiro capítulo surge para tratar do problema de barreira com o objetivo de definir centro analítico de um poliedro e trajetória central primal. Conceitos de muita importância para resolver, no último capítulo, o problema de minimização de uma quadrática convexa numa caixa sobre uma variedade afim. No tratamanto deste, utilizamos um método de pontos interiores primal-dual de trajetória central, em que nossa escolha de um ponto inicial primal-dual viável é original, representando um novo resultado em Matemática.

Abstract

In this work we study some constrained and unconstrained nonlinear programming methods, giving emphasis to the problem which gives the title to this dissertation. In the first chapter we state and study some methods like Cauchy, Newton, Armijo, Trust Region and Dog Leg. In the second study Sequential Quadratic Programming (SQP) through the Inexact Restoration method, which executes in each iteration, a restoration step and an optimality step. Our specific objective was working with the optimality step, known as tangent step in SQP, which, in our proposal consists in a convex quadratic minimization in a box intersected with an affine space. The third chapter comes to treat the Barrier problem with the objective of defining analytic center and primal central paths. These concepts are very important to solve, in the last chapter, the problem of minimization a convex quadratic function in an affine subspace with box constrains. We use a primal-dual interior point method which follows the central path, in which our choice of an initial feasible primal-dual point is original, representing a new result in Mathematics.

Lista de Figuras

1- Armijo com Modelo Quadrático.....	19
2- Ponto de Cauchy.....	20
3- Trajetória Dog Leg.....	25
4- Passo normal d	31
5- Minimização da função objetivo no núcleo de A em z	32
6- Passos do PQS.....	34
7- O Passo Normal e Tangente como um único Passo.....	41
8- Figura 8.....	44
9- Formiga.....	45
10- Barreira Infinita.....	46
11- Figura 11.....	49
12- Figura 12.....	49
13- Vizinhanças da Trajetória Central primal-dual C	71
14- Mudança de Escala.....	72
15- Máximo Declive na Mudança de Escala.....	73
16- Trajetória Central Primal e a Direção Afim-Escala.....	77
17- Trajetória Ótima de Região de Confiança.....	79
18- Ponto Primal inicial.....	81
19- Algoritmo de Pontos Interiores de Trajetória Central.....	84
20- Entre o Passo de Centralização e o de Newton e o de Newton Completo.....	86

Lista de Algoritmos

Capítulo 1

1.1 Busca de Armijo.....	15
1.2 Método de Cauchy com Busca de Armijo.....	16
1.3 Método de Newton puro.....	17
1.4 Armijo com modelo quadrático.....	19
1.5 Região de Confiança.....	22
1.6 Método Dog Leg.....	25

Capítulo 2

2.1 PQS com restauração inexata.....	35
2.2 O Passo Normal e Tangente simultaneamente.....	41

Capítulo 3

3.1 Método de Barreira Logarítmica.....	50
---	----

Capítulo 4

4.1 Ponto primal-dual inicial.....	80
4.2 Pontos Interiores.....	85

Sumário

Agradecimentos	iii
Resumo	iv
Abstract	v
Lista de Figuras	vi
Lista de Algoritmos	vii
Introdução	11
1. Região de Confiança.....	14
1.1 Busca de Armijo.....	15
1.2 Método de Cauchy.....	16
1.3 Método de Newton.....	17
1.4 Armijo com Modelo Quadrático.....	18
1.5 O Passo de Região de Confiança.....	21
1.6 Trajetória Dog Leg.....	23
2. Programação Quadrática Seqüencial.....	26
2.1 O Problema.....	27
2.2 Condições de KKT.....	28
2.3 Função Lagrangeano.....	28
2.4 Objetivos do Método PQS.....	29
2.5 Viabilização (Passo Normal).....	29
2.6 Otimalidade (Passo Tangente).....	31
2.7 Redução do Modelo Quadrático do Lagrangeano.....	33

2.8 O Algoritmo PQS.....	35
2.9 Algumas considerações sobre o PQS.....	38
2.9.1 A Hessiana do Lagrangeano.....	38
2.9.2 O Passo Tangente como redução da aproximação quadrática da função objetivo com a hessiana perturbada.....	39
2.10 O Passo Normal e Tangente calculados simultaneamente.....	40
3. Funções Barreira.....	43
3.1. O Problema.....	44
3.2 Barreira Infinita.....	45
3.3 Barreira Logarítmica.....	46
3.4. Problemas Penalizados.....	47
3.5 O Algoritmo.....	50
3.6 Pontos Centrais e Trajetória Central.....	50
3.7 Centro Analítico.....	51
3.8 Convergência.....	54
3.9 Multiplicadores de KKT.....	57
4. Minimização de Quadráticas Convexas em Caixas.....	58
4.1 O Problema.....	59
4.2 Um Ponto Inicial Primal-Dual Viável.....	62
4.2.1 Condições de KKT relaxadas.....	62
4.2.2 Vizinhanças da Trajetória Central.....	67
4.2.3 Mudança de Escala.....	71

4.2.4 Um novo ponto primal-dual viável.....	77
4.3 Método de Pontos Interiores de Trajetória.....	81
4.3.1 Cálculo do Passo de Newton.....	82
4.3.2 Passo Tangencial do PQS numa caixa.....	86
Conclusão.....	89
Referências Bibliográficas.....	90

Introdução

Este trabalho, de forma geral, consiste num enriquecimento e num primeiro contato com a pesquisa na área de otimização da minha parte. A princípio nosso objetivo se detinha a trabalhar com propriedades de métodos de filtro. Contudo, em dezembro de 2005 achamos o artigo de Ferris, Wathen e Armand [5] que nos motivou a mudar o curso de nossa pesquisa. O artigo em questão descreve a solução de um problema de complementaridade com recursos de memória limitada. O problema provém de simulações físicas, em geral choques ou colisões, que ocorrem em vídeo games e deve ser resolvido em pouco tempo e com pouca memória disponível. Neste contexto, Ferris, Wathen e Armand precisavam resolver problemas de minimização de quadráticas convexas em caixas de maneira eficiente, de modo que um de seus focos era minimizar $\frac{1}{2}x^T Ax + v_0^T x$ restrito a uma caixa $[l, h]$.

Até o começo de dezembro de 2005 tínhamos feito um estudo sobre Região de Confiança, Programação Quadrática Sequencial (PQS) e Funções Barreira. Então, em meio ao contexto do método PQS resolvemos abordar especificamente o Passo Tangente, que por sua vez se resume a minimização de uma quadrática, no nosso caso convexa, sobre uma variedade afim. Devido a várias propriedades da formulação do problema do Passo Tangente do PQS concluímos que o mesmo deveria ser abordado num contexto de Região de Confiança.

Sabemos que uma bola na norma infinito, que é um caso particular de caixa, é muito maior que uma bola na norma euclidiana e que uma caixa é uma restrição muito comum em problemas práticos. Deste modo obtivemos então um problema similar ao de Ferris,

Wathen e Armand [5], exceto pelo fato de estarmos sujeitando a minimização em questão a uma variedade afim.

Essa motivação nos propiciou o objetivo de trabalhar com um algoritmo de pontos interiores para minimizar uma quadrática convexa numa caixa sobre uma variedade afim. O artigo [5] apresentou um algoritmo tal qual, mas toda ênfase foi movida à Álgebra Linear da problemática, ao contrário da nossa análise que foi motivada em direção ao algoritmo de pontos interiores primal-dual que desenvolvemos. Deste modo acabamos nos focando bastante na obtenção de um par primal-dual inicial viável com certas propriedades, que culminou num resultado novo.

O primeiro dos quatro capítulos desta dissertação se destina a enunciar os algoritmos de Programação não Linear Irrestrita de Cauchy e de Newton, explicar a busca unidirecional feita por Armijo, introduzir o método de Região de Confiança e enunciar a Trajetória Dog Leg. Neste capítulo veremos o conceito de trajetória ótima de Região de Confiança que será muito importante quando propusermos um par primal-dual inicial viável para nosso problema alvo. Provas de convergência envolvendo os assuntos do primeiro capítulo se encontram em livros gerais de Programação não Linear como [3] e [21].

O segundo capítulo se destina a um estudo do método PQS ([20] e [22]) em meio a problemas de otimização restrita com restrições de igualdade. Apresentamos então o método, chamado de Restauração Inexata ([15] e [16]), que executa passos de viabilidade e de otimalidade. O de viabilidade, chamado de Passo Normal foi objeto de estudo de Francisco [7]. Já o de otimalidade, chamado de Passo Tangente foi alvo de estudo da nossa dissertação. Toda a análise deste segundo capítulo está sedimentada sobre as condições de Karush-Kunn-Tucker (KKT) ([3] e [21]).

O objetivo do terceiro capítulo, além de ser um estudo do método de Barreiras em termos de conceituação e convergência, é de definir centro analítico de um poliedro e

trajetória central primal. Definições que serão de extrema importância para o desenvolvimento do quarto capítulo.

O quarto, e último capítulo, destina-se à resolução do problema de minimizar uma quadrática convexa numa caixa sobre uma variedade afim. De princípio tratamos de reescrever este problema de modo a deixá-lo numa forma, a qual chamamos de padrão, que nos possibilitou uma melhor abordagem sob o aspecto das condições de KKT. Após reescrevê-lo propomos uma maneira de encontrar um ponto primal-dual viável, tendo como hipótese a saída de um ponto primal que é centro analítico do conjunto primal viável. Dividimos este estudo em duas partes. Uma quando o centro analítico mencionado for o vetor unitário, e a outra quando não for. Se não for, apresentamos uma mudança de escala. A partir de então nos prendemos ao desenvolvimento de um algoritmo de pontos interiores baseados em [10], [11] e [24] para enfim propor um passo de otimalidade do método PQS, ou seja, o Passo Tangente.

Capítulo 1

Região de Confiança

Métodos de Região de Confiança [19] são métodos que utilizam um modelo quadrático da função objetivo, visando encontrar uma redução deste modelo num conjunto fechado, que em geral é uma bola fechada. A expectativa é de conseguir uma redução do modelo e que esta represente uma redução na função objetivo. Por esse motivo usamos o termo Região de Confiança, já que procuramos uma região na qual o modelo quadrático da função objetivo a represente adequadamente, caso isso ocorra dizemos que há uma concordância entre a função objetivo e sua aproximação quadrática.

Mediante o estudo de problemas de mínimos quadrados, Levenberg, em 1944, e Marquardt, em 1963, instituíram pela primeira vez a idéia de procurar um conjunto em que o modelo quadrático fosse confiável. Já em 1966, o método de Região de Confiança, sob sugestão de Goldelf, Quandt e Trotter, foi aplicado a problemas de minimização irrestrita, [22].

Neste primeiro capítulo iremos expor brevemente algoritmos básicos de PNL (programação não linear) irrestrita, como o de Cauchy e de Newton ([2] e [20]). Também iremos falar sobre os métodos de Armijo (busca unidirecional [2] e [20]) e o método Dog Leg [13]. Com isso teremos algumas ferramentas para abordar, de início, problemas de região de confiança. Serão consideradas durante toda este capítulo funções objetivo $f_0 : R^n \rightarrow R$ de classe C^2 convexas. No entanto, deixaremos claro se as hipóteses de convexidade e de dupla diferenciabilidade não forem necessárias em algum momento.

1.1 Busca de Armijo

Considere o problema de minimização de uma função $f_0 : R^n \rightarrow R$ de classe C^2 a partir de um ponto $x^k \in R^n$ numa direção $d \in R^n$. O método de busca de Armijo não tenta resolver o problema proposto, visa entretanto, uma boa redução de f_0 ao longo de d . Dados $x^k, d \in R^n$ e um número real $\rho > 0$ buscamos $\lambda \in [0, \rho]$ tal que $f_0(x^k + \lambda d) < f_0(x^k)$. Para isto usamos um modelo linear de f_0 , definido por

$$m(\lambda) = f_0(x^k) + \lambda \nabla f_0(x^k)^T d$$

Dado um valor $\lambda \geq 0$, definimos

$$ared = f_0(x^k) - f_0(x^k + \lambda d), \text{ redução efetiva de } f_0.$$

$$pred = m(0) - m(\lambda) = \lambda \nabla f_0(x^k)^T d, \text{ redução prevista pelo modelo.}$$

Algoritmo 1.1 Busca de Armijo

Dado $x^k, d \in R^n$, ρ , $\alpha \in (0,1)$, fixo, por exemplo $\alpha = 0,4$.

$$\lambda = \rho$$

Calcule

$$ared = f_0(x^k) - f_0(x^k + \rho.d)$$

$$pred = \rho \cdot \nabla f_0(x^k)^T d$$

Enquanto $ared < \alpha \cdot pred$

$$\lambda = 0,6\lambda$$

$$pred = \lambda \cdot \nabla f_0(x^k)^T d$$

$$ared = f_0(x^k) - f_0(x^k + \lambda.d)$$

Fim

1.2 Método de Cauchy

Seja f_0 como na seção 1.1. O método de Cauchy serve para abordar essencialmente problemas de minimização irrestrita. Ele utiliza a direção de máximo declive a partir de um ponto $x^k \in R^n$, ou seja, a direção dada pelo gradiente da função objetivo em x^k , no sentido oposto. Em seguida faz uma busca ao longo dessa direção. Não daremos muita ênfase à precisão na minimização unidirecional, pois nosso objetivo apenas se restringe a uma boa redução da função objetivo. Deste modo utilizaremos o método de Armijo para achar um ponto $x^{k+1} \in R^n$ tal que se tenha uma certa redução da função objetivo sobre a direção em questão. A seguir temos o algoritmo de Cauchy com busca de Armijo.

Algoritmo 1.2 Método de Cauchy com Busca de Armijo

Dados $x^0 \in R^n$ e de parâmetros ρ, α para Armijo

$k = 0$;

Enquanto $\nabla f_0(x^k) \neq 0$

Faça $d = -\nabla f_0(x^k)$;

Calcule $\bar{\lambda}$, passo advindo do método de Armijo na busca unidirecional, se existir;

Se não existir solução, PARE com fracasso;

$x^{k+1} = x^k + \bar{\lambda}d$;

Faça $k = k + 1$.

O Algoritmo de Cauchy é globalmente convergente. Contudo, possui apenas convergência linear (veja [20]).

1.3 Método de Newton

Novamente estamos num contexto de minimização irrestrita em que objetivamos um ponto \bar{x} tal que $\nabla f_0(\bar{x}) = 0$ (condição necessária de otimalidade). Sob esta necessidade o método de Newton propõe a resolução do sistema de equações $\nabla f_0(x) = 0$. Isto é proposto através de uma aproximação linear para gradientes, em que forçamos essa aproximação a ser zero, como segue

$$\nabla f_0(x^k + d) \cong \nabla f_0(x^k) + H(x^k)d = 0$$

Em que $H(x^k)$ denota a Hessiana de f_0 em x^k .

Algoritmo 1.3 Método de Newton puro

Dado $x^0 \in R^n$

$k = 0$;

Enquanto $\nabla f_0(x^k) \neq 0$

Resolva $H(x^k)d = -\nabla f_0(x^k)$ (se possível)

$x^{k+1} = x^k + d$;

Faça $k = k + 1$.

O passo de Newton só é efetuado se a função a ser minimizada for convexa, ou seja se $H(x^k)$ for semi definida-positiva.

O método de Newton não possui convergência global ([20]), apenas local. No entanto localmente a convergência é quadrática. No caso da minimização irrestrita de uma função quadrática convexa o método de Newton resolve o problema em uma iteração.

1.4 Armijo com Modelo Quadrático

Considere o modelo linear de f_0 a partir de um ponto x fixo

$$l(x+h) = f_0(x) + \nabla f_0(x)^T (x+h)$$

Dado um valor $\lambda > 0$, usando Armijo com modelo linear, testamos o ponto $x + \lambda d$ da seguinte maneira:

$$\begin{aligned} pred &= l(x) - l(x + \lambda d) \text{ redução prevista pelo modelo} \\ ared &= f_0(x) - f_0(x + \lambda d) \text{ redução efetiva} \end{aligned}$$

O ponto $x + \lambda d$ é aceito se $ared \geq \alpha \cdot pred$, com $\alpha \in (0,1)$, um valor fixo. O primeiro ponto que satisfaz $ared = \alpha \cdot pred$ é chamado Ponto de Armijo.

O parágrafo anterior se refere ao Armijo com modelo linear exposto na seção 1.1. Entretanto, se conhecemos a hessiana de f_0 em x , ou uma aproximação da mesma, podemos utilizar um modelo quadrático de f_0 . Então, para $d \in R^n$, $\lambda \in R$ defina,

$$m(x + \lambda d) = f_0(x) + \lambda \nabla f_0(x)^T d + \frac{1}{2} \lambda^2 d^T H(x) d$$

Observe que para $d \in R^n$ dado, esta função é ou constante, ou ilimitada inferiormente (se $d^T H(x) d < 0$), ou possui um único minimizador. No último caso, o minimizador é dado pelo valor de λ tal que

$$m'(x + \lambda d) = \nabla f_0(x)^T d + \lambda d^T H(x) d$$

Desta maneira achamos o minimizador do modelo quadrático m a partir de x na direção d , dado por

$$\lambda_{\min} = \frac{-\nabla f_0(x)^T d}{d^T H(x) d}$$

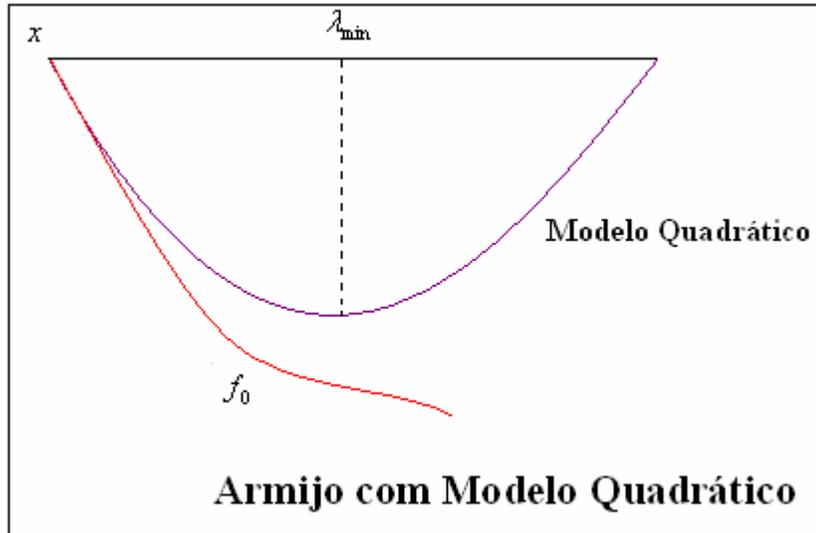


Figura 1

A figura 1 mostra que não vale a pena utilizar $\rho > \lambda_{\min}$. Nesse sentido, vamos propor então um algoritmo bem parecido com o Armijo com modelo linear, tendo como única distinção o fato de agora usarmos o modelo quadrático de f_0 . Para o algoritmo a seguir não é necessária a hipótese de convexidade.

Algoritmo 1.4 Armijo com modelo quadrático

Dado $x^k, d \in \mathbb{R}^n$, $\rho > 0$, $\alpha \in (0,1)$, fixo, por exemplo $\alpha = 0,4$.

Se $d^T H(x)d > 0$, calcular λ_{\min} e fazer $\lambda_c = \min\{\rho, \lambda_{\min}\}$

Senão $\lambda = \rho$

Calcule

$$ared = f_0(x^k) - f_0(x^k + \lambda d)$$

$$pred = m(x^k) - m(x^k + \lambda d)$$

Enquanto $ared < \alpha \cdot pred$

$$\lambda = 0,6\lambda$$

$$pred = m(x^k) - m(x^k + \lambda d)$$

$$ared = f_0(x^k) - f_0(x^k + \lambda d)$$

Fim

Resultado: λ_c

Ponto de Cauchy: Dados um ponto $x^k \in R^n$, um modelo quadrático $d \mapsto m(x^k + d)$ em torno de x^k e um valor $\Delta > 0$, definimos como ponto de Cauchy x_C^k ao minimizador do modelo quadrático ao longo da direção de Cauchy, dada pro $d = -\nabla f(x^k)$.

O cálculo é feito por

$$x_C^k = x^k - \lambda_C \cdot \nabla f_0(x^k).$$

Imediatamente segue que $\lambda_{\min} = +\infty$ se $\nabla f_0(x^k)^T H(x^k) \nabla f_0(x^k) \leq 0$, caso contrário,

$$\lambda_{\min} = \frac{\nabla f_0(x^k)^T \nabla f_0(x^k)}{\nabla f_0(x^k)^T H(x^k) \nabla f_0(x^k)}.$$

A figura a seguir ilustra a definição do Ponto de Cauchy.

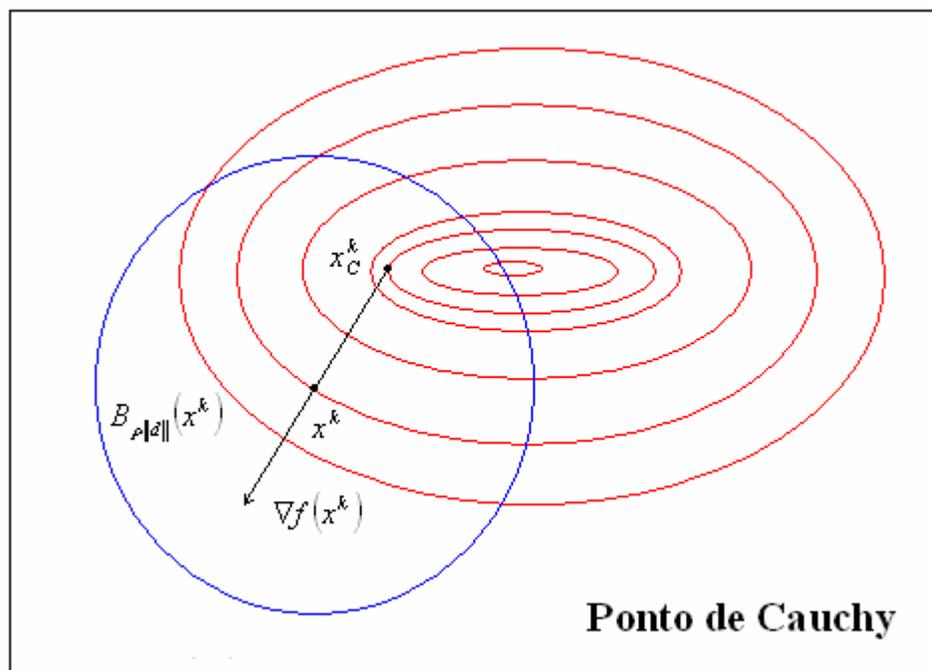


Figura 2

Ponto de Newton: Se $H(x^k)$ for semi definida-positiva (caso convexo), definimos o Ponto de Newton, x_N , como

$$x_N = x^k + d_N^k$$

em que d_N^k é dado pela resolução do sistema a seguir se existir solução. Se $H(x^k)$ é definida positiva, então existe uma única solução d_N^k . Senão, escolhamos a solução de mínima norma do sistema

$$H(x^k)d_N = -\nabla m(x^k).$$

O ponto x_N é chamado de Ponto de Newton porque advém do método de Newton aplicado a m em x^k . Os pontos de Cauchy e de Newton são de extrema importância na construção do método Dog Leg.

Esta seção foi baseada nas notas de aula de PNL de Gonzaga [8].

1.5 O Passo de Região de Confiança

O passo de Região de Confiança, ao contrário do passo do método de Armijo não utiliza uma direção dada. A cada iteração o método minimiza o modelo quadrático de f_0 , em torno do ponto x^k , na bola $B_\Delta(x^k)$, em que $\Delta > 0$ é o de raio de confiança. É importante frisar que nos métodos de região de confiança não necessitamos de convexidade do modelo quadrático de f_0 .

A princípio vamos supor que dispomos de uma rotina para minimizar o modelo quadrático de f_0 numa bola (uma forte suposição já que este é o grande foco desta dissertação). Supondo que esta minimização tenha sido feita, comparamos a redução do modelo (*pred*) com a redução efetiva da função (*ared*). Suponha também que um valor real fixo $\alpha \in (0,1)$ seja dado. Se $\frac{ared}{pred} > \alpha$, então o ponto advindo da minimização do modelo quadrático é aceito. Caso contrário, se $\frac{ared}{pred} < \alpha$, temos que diminuir o raio de confiança Δ e repetir o processo. Veja que α representa uma taxa percentual mínima que queremos garantir em relação a queda do modelo e a redução da função objetivo.

Sabemos, devido ao Cálculo Diferencial e Integral, que para algum raio $\Delta > 0$ suficientemente pequeno o modelo quadrático de uma função duplamente continuamente diferenciável é uma boa aproximação para a função objetivo na bola $B_\Delta(x^k)$, mesmo sem convexidade. Deste modo não precisamos nos preocupar com a existência de um $\Delta > 0$ tal que m represente bem f_0 . O algoritmo completo do Passo de Região de Confiança é descrito a seguir

Algoritmo Região de Confiança

Dado $x^0 \in R^n$, $\Delta > 0$, $\alpha, \beta \in (0,1)$ fixos.

Enquanto $\|\nabla f_0(x^k)\| \neq 0$

Construa o modelo $m(x^k + h) = \nabla f_0(x^k)^T h + \frac{1}{2} h^T H(x^k) h$

Calcule d , solução de minimizar $m(h)$
s.a. $\|h\| \leq \Delta$

$pred = -m(x^k + d)$
 $ared = f_0(x^k) - f_0(x^k + d)$

Se $ared < \alpha \cdot pred$ (aceitar o passo)

$$x^{k+1} = x^k + d$$

Senão (rejeitar o passo e reduzir Δ)

$$x^{k+1} = x^k$$

$$\Delta = \beta \cdot \Delta$$

Fim

Fim

Muitas vezes, quando o modelo quadrático representa muito bem a função objetivo na região de confiança, se opta em aumentar o raio da mesma na iteração seguinte.

1.6 Trajetória Dog Leg

O método Dog Leg é um método utilizado para solucionar de forma aproximada problemas de região de confiança. Ele é, no entanto, definido apenas num contexto de minimização convexa. Dados os pontos x^k (ponto de partida), x_C (ponto de Cauchy) e x_N (ponto de Newton, bem definido por causa da hipótese de convexidade) e um raio $\Delta > 0$ definimos a trajetória Dog Leg como sendo a poligonal ABC em que A é o ponto x^k , B o ponto de Cauchy x_C e C o ponto de Newton x_N .

Queremos então minimizar uma quadrática numa bola, sendo que esta minimização não precisa ter boa precisão, pois já como comentamos, o método consiste numa aproximação. O método Dog Leg irá nos fornecer o ponto de Newton, se este estiver dentro da bola $B_\Delta(x^k)$, ou o ponto advindo da intersecção da trajetória Dog Leg com esta bola. Veja que já mostramos como calcular os pontos de Newton e de Cauchy. Já em relação a intersecção mencionada a pouco, basta fazer algumas contas sob a análise de dois casos. O primeiro é o caso em que o ponto de Cauchy é exterior a bola ou está na fronteira da bola $B_\Delta(x^k)$ e o segundo é quando o x_C é interior a $B_\Delta(x^k)$.

- (i) Ponto de Cauchy fora ou sobre a bola:

Neste caso basta dar um passo do tamanho do raio sobre a direção de Cauchy, ou seja,

$$x^{k+1} = x^k - \Delta \frac{\nabla f_0(x^k)}{\|\nabla f_0(x^k)\|}.$$

- (ii) Ponto de Cauchy dentro da bola:

Veja que a poligonal ABC pode ser parametrizada da seguinte maneira

$$ABC = \left\{ \begin{array}{l} tx_C, \text{ se } t \in [0,1] \\ x_C + (t-1)(x_N - x_C), \text{ se } t \in (1,2] \end{array} \right\}.$$

Fazendo algumas manipulações concluímos que basta considerar a raiz positiva (é fácil de ver que uma sempre é negativa) do polinômio

$$p(\lambda) = \|x_C\|^2 - \Delta^2 + 2\lambda x_C^T d + \lambda^2 \|d\|^2$$

com $\lambda = t - 1$ e $d = x_N - x_C$.

A raiz desejada é dada por

$$\bar{\lambda} = \frac{-2x_C^T d + \sqrt{(2x_C^T d)^2 - 4\|d\|^2(\|x_C\|^2 - \Delta^2)}}{2\|d\|^2}$$

Seja $\bar{t} = \bar{\lambda} + 1$. Temos então que

$$x^{k+1} = x_C + \bar{t}d .$$

Vale frisar que a trajetória Dog Leg é formada por dois subconjuntos do R^n , os quais são os segmentos AB e BC . Veja que tanto AB quanto BC são conjuntos pertencentes a espaços de dimensão 1 do R^n . Com isso temos que o método Dog Leg define uma trajetória em um subespaço de dimensão de no máximo 2 em R^n .

Dizemos que a trajetória ótima de um problema de região de confiança numa bola $B_\Delta(x^k)$, é o conjunto dos minimizadores da aproximação quadrática da função objetivo nas bolas $B_\delta(x^k)$ para $0 \leq \delta \leq \Delta$. Veja então que esta, a trajetória ótima, é um conjunto que pode ter dimensão n . Com isto concluímos que o método Dog Leg é de fato um método que busca apenas uma leve aproximação na resolução do problema de minimização de quadráticas convexas em bolas. O método é utilizado em geral sob bolas definidas pela norma euclidiana, devido à simplicidade das manipulações envolvidas na parametrização de ABC nesse contexto.

A figura 3 a seguir ilustra a idéia do método Dog Leg

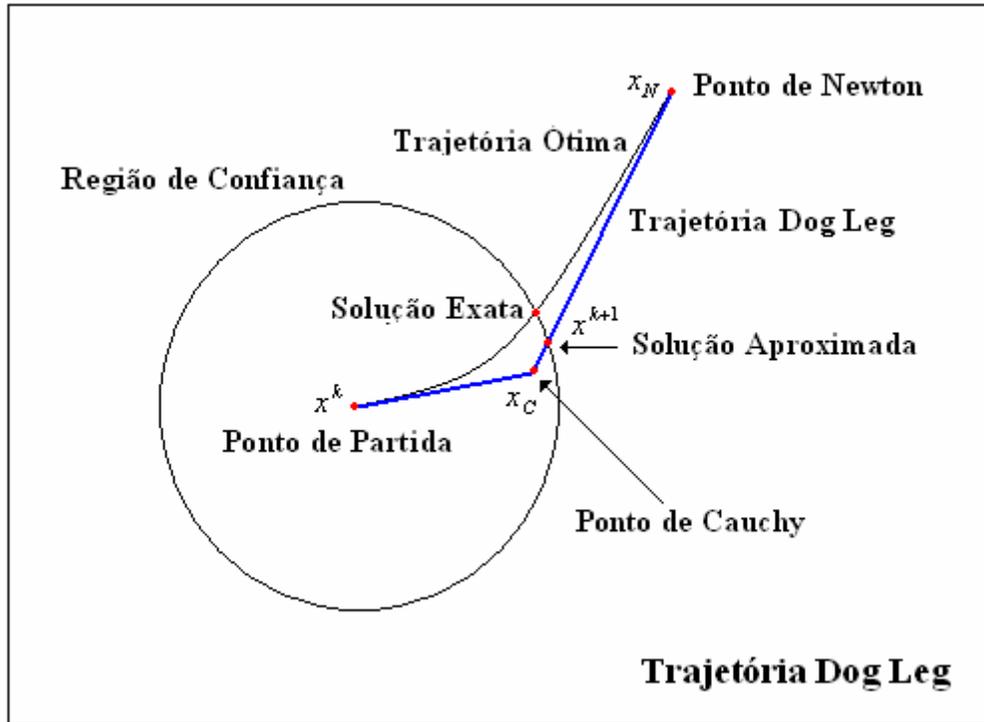


Figura 3

Com base no estudo deste sub-capítulo vamos apresentar um algoritmo para o Método Dog Leg. Veja mais em [13].

Algoritmo 1.6 Método Dog Leg

Dado $x^k \in R^n$, $\Delta > 0$.

Calcule o Ponto de Cauchy x_C

Se $\|x_C - x^k\| > \Delta$

$$x^{k+1} = x^k - \Delta \frac{\nabla f_0(x^k)}{\|\nabla f_0(x^k)\|}$$

Senão calcule x_N

Calcule também $\bar{\lambda} = \frac{-2x_C^T d + \sqrt{(2x_C^T d)^2 - 4\|d\|^2(\|x_C\|^2 - \Delta^2)}}{2\|d\|^2}$

$$x^{k+1} = x_C + (\bar{\lambda} + 1)d$$

Fim

Capítulo 2

Programação Quadrática Seqüencial

A família de métodos conhecida por Programação Quadrática Seqüencial (PQS) originou-se da tese de R.B. Wilson de 1963. O método veio, contudo, a ser mais popularizado a partir da metade da década de 70 com o estudo de Garcia, Mangasarian, Han e Powell [24]. O PQS surgiu então para tratar de problemas de otimização restrita. A idéia básica desse método consiste, a cada iteração, em abordar subproblemas de minimização de quadráticas. A cada passo o PQS irá trabalhar em função de dois objetivos, os quais são viabilização ou restauração [7] e otimização da função objetivo do problema. O PQS puro, visto brevemente na última seção deste capítulo, visa a restauração e a otimização em questão simultaneamente. Nós, entretanto, trabalharemos esses objetivos através de dois passos explicitamente separados. Métodos dessa natureza são chamados de restauração inexata.

Neste capítulo introduziremos então o método PQS com restauração inexata, bem como vários aspectos a ele relacionados, como convergência, interpretações algébricas e geométricas, e a equivalência local com o método de Newton. Daremos também a devida ênfase a um sub-problema de região de confiança que aparece no PQS, que é o de minimizar o modelo quadrático do Lagrangeano da função objetivo numa bola. Sub-problema que é na verdade o problema alvo desta dissertação. E este está relacionado ao passo tangente do PQS. É importante dizer que o contexto da minimização visada neste capítulo se refere a problemas de otimização restrita com restrições de igualdade.

2.1 O Problema

Vamos considerar o problema que segue

$$\begin{aligned} &\text{minimizar } f_0(x) \\ &\text{s.a. } f(x) = 0 \end{aligned} \tag{2.1}$$

Durante todo este capítulo considere as seguintes hipóteses:

Suponha $f_0 : R^n \rightarrow R$ uma função de classe C^2
 $x \mapsto f_0(x)$

Suponha também

$$\begin{aligned} &f : R^n \rightarrow R^m \\ &x \mapsto f(x) = (f_1(x), f_2(x), \dots, f_m(x)) \end{aligned}$$

em que $m \leq n$ e $f_i : R^n \rightarrow R$, $i = 1, \dots, m$ são funções de classe C^2 .

Denotaremos a derivada, ou matriz Jacobiana, de f num ponto $x \in R^n$ por $A(x)$, isto é,

$$A^T(x) = \left[\nabla f_1(x) \quad \nabla f_2(x) \quad \dots \quad \nabla f_m(x) \right] \tag{2.2}$$

Vamos supor durante todo este capítulo que $A(x)$ tem posto completo para todo $x \in R^n$, ou seja, $\text{rank}(A(x)) = m \quad \forall x \in R^n$.

2.2 Condições de KKT

As condições de otimalidade de 1ª ordem relacionadas a problemas tais como (2.1), também conhecidas como condições de KKT (Karush-Kunn-Tucker) são dadas por

$$\begin{aligned} \nabla f_0(x) + A^T(x)\lambda &= 0 \\ f(x) &= 0 \end{aligned} \quad (2.3)$$

Temos que se $x^* \in R^n$ é solução ótima de (2.1) então $\exists \lambda^* \in R^m$ tal que o par (x^*, λ^*) satisfaz (2.3) (condição necessária de otimalidade).

2.3 Função Lagrangeano

A função Lagrangeano associada ao problema (2.1) é dada por

$$\begin{aligned} L: R^n \times R^m &\rightarrow R \\ (x, \lambda) &\mapsto L(x, \lambda) = f_0(x) + \lambda^T f(x) \end{aligned} \quad (2.4)$$

Observe que as condições $\nabla_x L(x, \lambda) = 0$ e $f(x) = 0$ equivalem as condições de KKT de (2.1).

A Hessiana do Lagrangeano, em relação a x , no ponto (x, λ) será denotada por

$$W(x, \lambda) = \nabla_{xx}^2 L(x, \lambda). \quad (2.5)$$

Veja que tanto $\nabla_x L(x, \lambda)$ como $W(x, \lambda)$ estão bem definidas para todo $(x, \lambda) \in R^n \times R^m$ já que supomos anteriormente $f_i, i = 1, \dots, m$ funções de classe C^2 .

2.4 Objetivos do Método PQS de Restauração Inexata

O método PQS puro, que consiste em resolver um problema quadrático em se que faz um passo, que é restauração e ao mesmo de otimização, é exposto na seção 2.10. Entretanto, vamos centrar nosso estudo nos métodos conhecidos como métodos de restauração inexata, que iremos apresentar adiante, num contexto de minimização com restrições de igualdade. Esses métodos introduzidos por Martinez e Pilotta [16] fazem em cada iteração dois passos separados, um de restauração e outro de otimização.

No passo de restauração queremos, a partir de um ponto dado $x \in R^n$, dar um passo de restauração que consiste aproximar-se do conjunto viável $\Omega = \{x \in R^n; f(x)=0\}$ (a proximidade ao conjunto viável será medida por uma norma $\|f(x)\|$). No entanto, não nos bastará apenas uma melhora na viabilidade. O que objetivamos afinal é um bom aumento percentual em relação à viabilidade, ou seja, encontrar um ponto $x+d$ tal que $\|f(x+d)\| \ll \|f(x)\|$.

No passo de otimização estamos interessados em reduzir f_0 a partir de $z = x+d$, sem, no entanto, perder muito da viabilidade conquistada no passo de viabilização, o que nos dá a intuição de procurar alguma redução da função f_0 no espaço tangente às restrições em z . Através de exemplos geométricos bem elementares será fácil de concluir que não é bom minimizar f_0 nesse espaço tangente. Veremos que uma alternativa bem sucedida será reduzir o Lagrangeano em x a partir de z , para algum λ fixo, pois a função Lagrangeano leva em consideração a curvatura da restrição f . No entanto não reduziremos diretamente o Lagrangeano, e sim, um modelo quadrático do mesmo numa bola, que equivale a um subproblema de região de confiança.

2.5 Viabilização (Passo Normal)

Dado $x \in R^n$ tal que $f(x) \neq 0$ queremos encontrar um ponto $z = x + d$ tal que $\|f(x + d)\| \ll \|f(x)\|$.

Observe então que a aproximação linear de f no ponto z em função de x e d é dada por

$$f(z) = f(x + d) \approx f(x) + A(x)d.$$

Queremos $z = x + d$ mais perto do conjunto viável $\Omega = \{x \in R^n; f(x) = 0\}$. Para isto é razoável procurar d que satisfaça

$$f(x) + A(x)d = 0 \text{ (Passo de Newton)}$$

ou seja 1ª

$$A(x)d = -f(x). \tag{2.6}$$

Veja que $A(x) \in R^{m \times n}$. Lembre também que estamos supondo $\text{rank}(A(x)) = m$. Temos então duas situações para (2.6):

(i) $m = n$

Neste caso conclui-se facilmente que o sistema (2.6) possui solução, e ela é única, dada por $d = -A^{-1}(x)f(x)$.

(ii) $m < n$

Sob estas circunstâncias temos pelo menos uma solução, mas em geral teremos infinitas. No entanto, da Álgebra Linear, podemos concluir que existe uma única solução de (2.6) de norma mínima.

Devemos ter

$$d \perp N(A(x)),$$

ou equivalentemente,

$$d \in I(A^T(x)).$$

Da Álgebra Linear concluímos que

$$d = -A^T(x)(A(x)A^T(x))^{-1} f(x)$$

Obtemos então o passo d , chamado de Passo Normal, por meio de um passo de Newton bem definido já que $\text{rank}(A(x)) = m$ para todo x em R^n . A figura a seguir ilustra o raciocínio usado no desenvolvimento do Passo Normal.

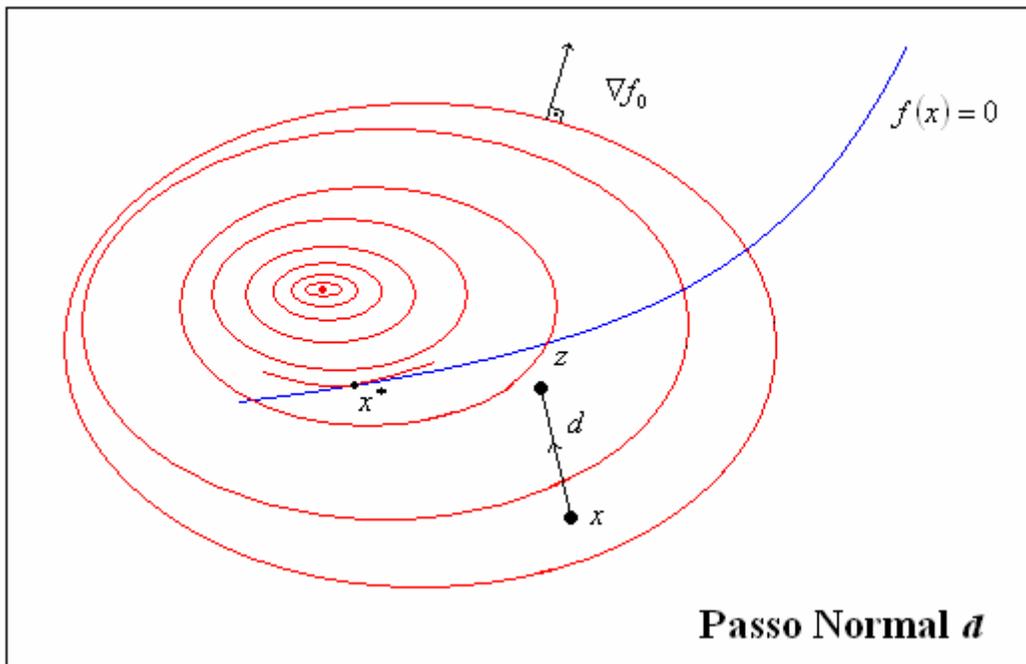


Figura 4

Note que a rigor deveríamos nos preocupar mais com o Passo Normal, já que estamos dando um passo de Newton, pois sabemos da não convergência do método de Newton. Para remediar isto poderíamos sujeitar nosso passo a uma região de confiança em torno do ponto x . Entretanto, não entraremos em detalhes em relação a isto. Um tratamento detalhado do passo normal encontra-se na tese de doutoramento de Francisco [7].

2.6 Otimalidade (Passo Tangente)

Dado $z \in R^n$ queremos determinar um ponto $\bar{x} = z + h$ tal que $f_0(\bar{x}) < f_0(z)$ sem que a viabilidade de \bar{x} piore muito em relação a de z . Para não piorar muito a viabilidade sujeitamos a minimização de f_0 ao espaço tangente as restrições em z . Esse espaço é determinado pelo núcleo da matriz Jacobiana de f em z , dado por

$$N(A(z)) = \{h \in R^n; A(z)h = 0\}. \quad (2.7)$$

É fácil perceber que minimizar f_0 ao longo de $N(A(z))$ não vai funcionar, como ilustra a figura a seguir

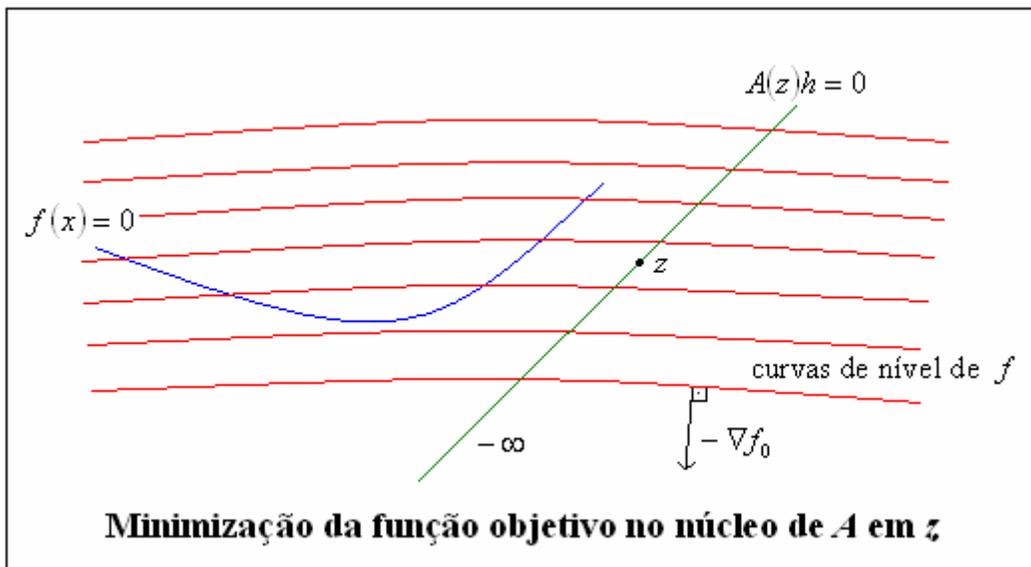


Figura 5

O procedimento que vamos adotar e justificar posteriormente, é de reduzir o Lagrangeano associado ao problema (2.1) em relação a x , usando λ fixo, um estimador dos multiplicadores ótimos de KKT, que será denotado durante todo este capítulo por λ^* (que pode ser dado, por exemplo, pela resolução de um problema de mínimos quadrados). Mas,

iremos na verdade reduzir uma aproximação quadrática do Lagrangeano restrito ao núcleo de $A(z)$. O passo tangente é muito importante no nosso estudo pois, mediante ele, iremos formular o problema que dá título a esta dissertação.

2.7 Redução do Modelo Quadrático do Lagrangeano

Considere novamente um ponto dado $z \in R^n$. Suponha também $\lambda \in R^m$ um estimador para o multiplicador ótimo λ^* . Já comentamos que se um ponto x^* é ótimo local de (2.1) então existe $\lambda^* \in R^m$ que satisfaz as condições de KKT (2.3). Veja que este multiplicador é único já que estamos supondo $\text{rank}(A(x)) = m$, para todo $x \in R^n$.

Considere então o modelo quadrático do Lagrangeano (dado em (2.4)) em z e λ , dado por

$$m(h) = L(z, \lambda) + \nabla_x L(z, \lambda)^T h + \frac{1}{2} h^T W(z, \lambda) h. \quad (2.8)$$

Como já mencionado, estamos interessados na abordagem do problema a seguir, relacionado ao passo tangente do PQS,

$$\begin{aligned} &\text{minimizar } m(h) \\ &\text{s.a. } A(z)h = 0. \end{aligned} \quad (2.9)$$

Aplicando as condições de KKT em (2.9) temos

$$\begin{aligned} \nabla m(h) + A^T(z)\mu &= 0 \\ A(z)h &= 0 \end{aligned}.$$

Através do desenvolvimento de $\nabla m(h)$ e algumas simplificações obtemos

$$\begin{aligned} \nabla f_0(z) + A^T(z)(\lambda + \mu) + W(z, \lambda)h &= 0 \\ A(z)h &= 0 \end{aligned} \quad (2.10)$$

Veja que (2.10) é um sistema de $m+n$ equações e $m+n$ incógnitas. Mediante a Álgebra Linear conseguimos uma solução exata ou aproximada para (2.10), solução que chamaremos de (h, μ) . O passo $h \in R^n$, se for aceito, será chamado de Passo Tangente,

enquanto $\lambda + \mu$ será o estimador de Lagrange a ser utilizado na próxima iteração do passo tangente do PQS. Se o passo $h \in R^n$ for bom em termos de uma função de mérito ou de um filtro aceitamos o passo. Note, contudo, que h é um passo de Newton que pode ser muito ruim. Para contornar este problema podemos propor (2.9) numa região de confiança. Esta região pode ser uma bola na norma euclidiana, por exemplo. Neste sentido teríamos o método Dog Leg para resolver este problema de forma aproximada. Entretanto, queremos propor algo mais exato nesta dissertação. Estudaremos então o passo tangente (que será abordado no capítulo 4) tendo como foco o problema (2.9) restrito a bola $B_\Delta(0)$ na norma infinito, que é equivalente ao conjunto $E = \{h \in R^n; -\Delta e \leq h \leq \Delta e, \Delta > 0\}$.

A princípio obtemos então o ponto $\bar{x} = z + h$ e o estimador $\lambda + \mu$ sem nos preocuparmos muito com as características de \bar{x} em termos de viabilidade e otimalidade. Comentaremos adiante, sob forma de proposição, do porque do uso do estimador $\lambda + \mu$ no método PQS. Observe a seguir uma figura que ilustra a idéia básica do de um passo do método PQS, o Passo Normal e o Passo Tangente. $A(z)h = 0$

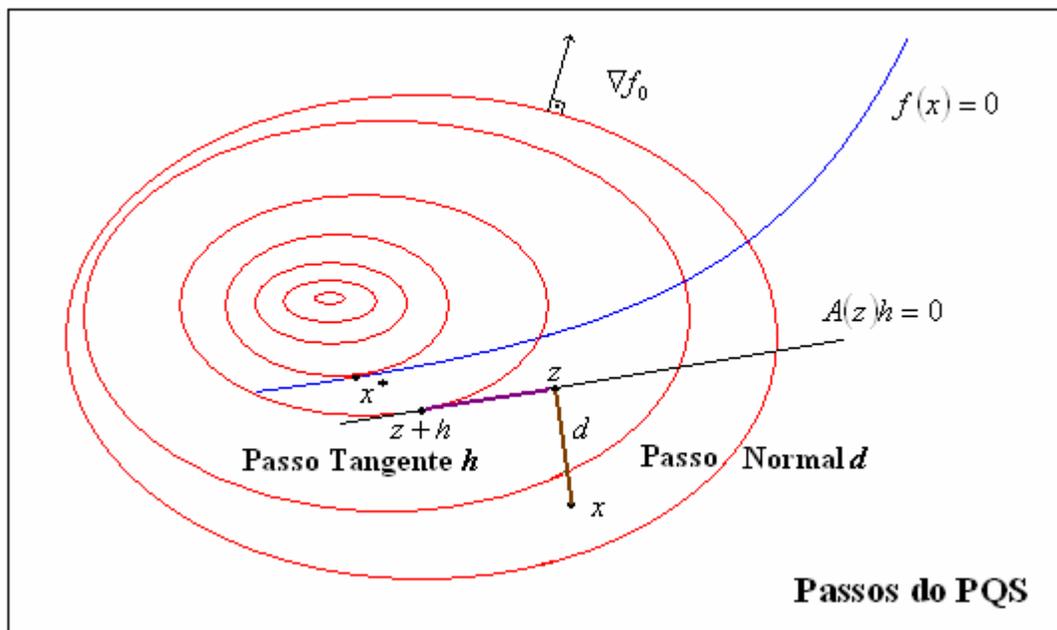


Figura 6

2.8 O Algoritmo PQS

Algoritmo 2.1 PQS com restauração inexata

Dado $x^0 \in R^n$, $\lambda^0 \in R^m$ e $\mu^0 = 0$

para $k = 1, 2, 3, \dots$

Calcule $f(x^k)$ e $A(x^k)$

Resolva (2.6) com $x = x^k$ para obter d^k (Passo Normal)

Faça $z^k = x^k + d^k$

Calcule $\nabla f_0(z^k)$, $W(z^k, \lambda^k)$ e $A(z^k)$

% (Sem grandes complicações poderíamos utilizar $A(x^k)$ para não ter um custo computacional tão elevado)

Resolva (2.10) com $z = z^k$, $\lambda = \lambda^k$ para obter $(h, \mu) = (h^k, \mu^k)$ (Passo Tangente)

Faça $x^{k+1} = z^k + h^k$ e $\lambda^{k+1} = \lambda^k + \mu^k$

se “convergência” PARE com solução aproximada (x^{k+1}, λ^{k+1}) .

Fim

LEMA 2.1: Dada uma matriz $A \in R^{m \times n}$ com posto completo e uma seqüência $(\lambda^k)_{k \in N}$ com $\lambda^i \in R^m$ tal que $A^T \lambda^k \rightarrow 0$ então $\lambda^k \rightarrow 0$.

Demonstração:

Da Álgebra Linear temos que se A tem posto completo a matriz AA^T é inversível. Sabemos também que a operação de multiplicação de uma matriz real B fixada, por uma variável, é uma operação contínua. Se tivermos então uma seqüência $(y^k)_{k \in N}$ com $y^k \rightarrow 0$, a seqüência $By^k \rightarrow 0$, respeitadas as dimensões para as quais a multiplicação em questão faz sentido. Sendo assim, concluímos que:

$$\|\lambda^k\| = \|(AA^T)^{-1} AA^T \lambda^k\| = \|((AA^T)^{-1} A) A^T \lambda^k\| \rightarrow 0$$

já que $(AA^T)^{-1}$ existe e por hipótese $A^T \lambda^k \rightarrow 0$. ■

LEMA 2.2: Suponha $(x^k), (z^k)$ e (λ^k) seqüências geradas pelo Algoritmo 2.1. Suponha também que $x^k \rightarrow x^*$ e $z^k \rightarrow x^*$ (em que x^* é uma solução ótica local do problema (2.1)) e que $(\lambda^k)_{k \in N}$ é uma seqüência limitada. Então existe um único multiplicador de Lagrange λ^* associado a x^* e, além disso, $\lambda^k \rightarrow \lambda^*$.

Demonstração:

Unicidade de λ^* :

Como x^* é solução ótica de (2.1) sabemos que existe pelo menos multiplicador ótico de Lagrange que satisfaz a primeira condição de KKT

$$\nabla f_0(x^*) + A^T(x^*)\lambda^* = 0.$$

Como $\text{rank}(A) = m$, concluímos que λ^* é dado de maneira única por

$$\lambda^* = -(AA^T)^{-1}(x^*)A(x^*)\nabla f_0(x^*).$$

Este é o primeiro momento em que o fato de supormos $\text{rank}(A) = m$ é relevante, ou mesmo necessário.

Convergência de λ^k :

Pela definição do passo h , passo tangente do PQS, temos que

$$\begin{aligned} \nabla f_0(z^k) + A^T(z^k)\lambda^k + W(z^k, \lambda^k)h^k &= 0 \\ A(z^k)h^k &= 0 \end{aligned}, \quad \forall k \in N$$

Como $f_i, i = 0, 1, \dots, m$ são funções de classe C^2 temos a convergência das seguintes seqüências

$$\begin{aligned} \nabla f_0(z^k) &\rightarrow \nabla f_0(x^*) \\ A^T(z^k) &\rightarrow A^T(x^*) \end{aligned}$$

Além disso a seqüência $(h^k)_{k \in N}$, definida como no Algoritmo 2.1, converge para zero, já que $(z^k)_{k \in N}$ é convergente, e em particular é seqüência de Cauchy.

Podemos concluir também, devido à convergência de $(z^k)_{k \in N}$, a limitação de $(\lambda^k)_{k \in N}$ e a continuidade de W e que $(W(z^k, \lambda^k))_{k \in N}$ é uma seqüência limitada. Para isto basta lembrar que uma função contínua leva conjuntos compactos em compactos.

Temos então que $W(z^k, \lambda^k)h^k \rightarrow 0$ pois

$$\|W(z^k, \lambda^k)h^k\| \leq \|W(z^k, \lambda^k)\| \|h^k\| \leq M \|h^k\|$$

Veja que

$$\lim_{k \rightarrow \infty} \nabla f_0(z^k) + A^T(z^k)\lambda^k + W(z^k, \lambda^k)h^k = 0 = \nabla f_0(x^*) + A^T(x^*)\lambda^*$$

Da Análise temos que se o limite da soma de duas parcelas é zero e o limite de uma das parcelas também é zero, o limite da outra existirá e será zero. Por meio desta observação e de algumas manipulações obtemos

$$\lim_{k \rightarrow \infty} A^T(z^k)\lambda^k - A^T(x^*)\lambda^* = 0$$

Podemos reescrever este limite da seguinte forma

$$\lim_{k \rightarrow \infty} (A^T(z^k) - A^T(x^*))\lambda^k - A^T(x^*)(\lambda^* - \lambda^k) = 0$$

Veja que $\|(A^T(z^k) - A^T(x^*))\lambda^k\| \leq \|A^T(z^k) - A^T(x^*)\| \|\lambda^k\| < \|A^T(z^k) - A^T(x^*)\| M$

Logo

$$\lim_{k \rightarrow \infty} (A^T(z^k) - A^T(x^*))\lambda^k = 0$$

Pelo mesmo argumento de antes, importado da Análise, concluímos que

$$\lim_{k \rightarrow \infty} A^T(x^*)(\lambda^k - \lambda^*) = 0$$

Pelo Lema 2.1 temos então que $\lim_{k \rightarrow \infty} (\lambda^k - \lambda^*) = 0$, ou seja,

$$\lambda^k \rightarrow \lambda^*$$

■

Poderíamos fazer uma discussão sobre qual a hipótese mais fraca que devemos estabelecer sobre a Hessiana do Jacobiano para que quando $z^k \rightarrow x^*$ tenhamos $(\lambda^k)_{k \in N}$ uma seqüência limitada. Nesse caso teríamos que se $z^k \rightarrow x^*$ então $\lambda^k \rightarrow \lambda^*$, fato que se concluiria de modo trivial através dos Lemas 2.1 e 2.2. No entanto, esta discussão não é muito relevante porque as proposições anteriores tinham como único objetivo justificar a utilização dos estimadores de Lagrange do PQS. Veja que não há problema algum se temos, por exemplo, $z^k \rightarrow x^*$ com $(\lambda^k)_{k \in N}$ uma seqüência não convergente, pois afinal chegamos à solução ótima do problema. Concluimos então que $\lambda + \mu$ é um bom palpite para ser utilizado como estimador de Lagrange no contexto do método PQS. Poderíamos também, a cada passo, encontrar um estimador de Lagrange mediante o método de quadrados mínimos.

2.9 Algumas considerações sobre o PQS

2.9.1 A Hessiana do Lagrangeano

Em problemas práticos teremos que, efetivamente, calcular a Hessiana do Lagrangeano, $W(x, \lambda)$. Vamos supor, no contexto deste trabalho, que temos um bom software para cálculo de derivadas, os quais são cada vez mais comuns e eficientes. Dispomos também de métodos tipo BFGS que fazem cálculos aproximados de Hessianas.

Na prática as matrizes Hessianas que aparecem são em geral esparsas. Dessa maneira é necessário saber trabalhar bem com esparsidade. Nesse trabalho cálculos tais quais não serão abordados.

2.9.2 O Passo Tangente como redução da aproximação quadrática da função objetivo com a hessiana perturbada

Seja m , $\bar{x} \in R^n$, $\bar{\lambda} \in R^m$ fixos. Considere o problema

$$\begin{aligned} \text{minimizar } & f_0(\bar{x}) + \nabla^T f_0(\bar{x})h + \frac{1}{2}h^T W(\bar{x}, \bar{\lambda})h \\ \text{s.a. } & A(\bar{x})h = 0. \end{aligned}$$

Veja que o problema acima é, trivialmente, equivalente ao problema

$$\begin{aligned} \text{minimizar } & \nabla^T f_0(\bar{x})h + \frac{1}{2}h^T W(\bar{x}, \bar{\lambda})h \\ \text{s.a. } & A(\bar{x})h = 0. \end{aligned} \tag{2.11}$$

As condições de KKT para o problema (2.11) são

$$\begin{aligned} \nabla f_0(\bar{x}) + W(\bar{x}, \bar{\lambda})h + A^T(\bar{x})\hat{\lambda} &= 0 \\ A(\bar{x})h &= 0. \end{aligned} \tag{2.12}$$

Lembre que as condições de KKT associadas ao Passo Tangente do PQS (2.10) são

$$\begin{aligned} \nabla f_0(z) + A^T(z)(\lambda + \mu) + W(z, \lambda)h &= 0 \\ A(z)h &= 0. \end{aligned}$$

Se fizermos $\bar{x} = z$ e $\bar{\lambda} = \lambda$ então a condição (2.10) e a condição (2.12) são exatamente a mesma e teríamos $\hat{\lambda} = \lambda + \mu$. Dessa maneira conseguimos uma outra interpretação para o passo tangente do PQS. Esse passo pode ser visto então, mediante as conclusões anteriores, como uma redução da aproximação quadrática da função objetivo, em que a hessiana utilizada é a hessiana do Lagrangeano aplicada ao par (z^k, λ^k) .

2.10 O Passo Normal e Tangente calculados simultaneamente

(PQS puro)

Ao invés de efetuar dois passos independentes, um de viabilização e outro de otimização, podemos simplesmente efetuar um passo que viabilize e otimize ao mesmo tempo. Nisto que consiste o chamado PQS puro. Deste modo as condições de KKT do problema quadrático que deve ser resolvido a cada passo do PQS puro são:

$$\begin{aligned}\nabla f_0(x^k) + A^T(x^k)(\lambda^k + \mu^k) + W(x^k, \lambda^k)h^k &= 0 \\ A(x^k)h^k &= -f(x^k).\end{aligned}\tag{2.13}$$

O sistema (2.13) representa então as condições de KKT para o problema de minimização da aproximação quadrática do Lagrangeano de f_0 restrita ao conjunto $A(x^k)h^k = -f(x^k)$ (que não é, em geral, um espaço vetorial). Este poderia ser o principal problema do trabalho, já que ele, numa região de confiança, em que a região fosse definida por uma bola na norma infinito, seria um problema de minimização de uma quadrática convexa numa caixa sobre uma variedade afim. É claro que poderíamos experimentar o método Dog Leg, introduzido no capítulo 1, mas já foi comentado antes que ele não resolveria o problema de modo eficiente. No entanto vamos nos ater ao PQS com restauração inexata, nos focando no passo tangente. Mesmo assim considere o algoritmo sem região de confiança, baseado na resolução de (2.13).

Suponha que saímos então de um ponto inicial x^0 e de um estimador de Lagrange λ^0 . A cada passo, resolvendo o sistema (2.13) encontramos h^k e μ^k . O próximo ponto será dado por $x^{k+1} = x^k + h^k$. Já estimador de Lagrange associado será dado por $\lambda^{k+1} = \lambda^k + \mu^k$. O processo todo é análogo ao da seção 2.6. A única distinção fica por conta de estarmos restringindo a minimização a outro conjunto.

Considere a figura a seguir para ilustrar a idéia de viabilizar e otimizar ao mesmo tempo.

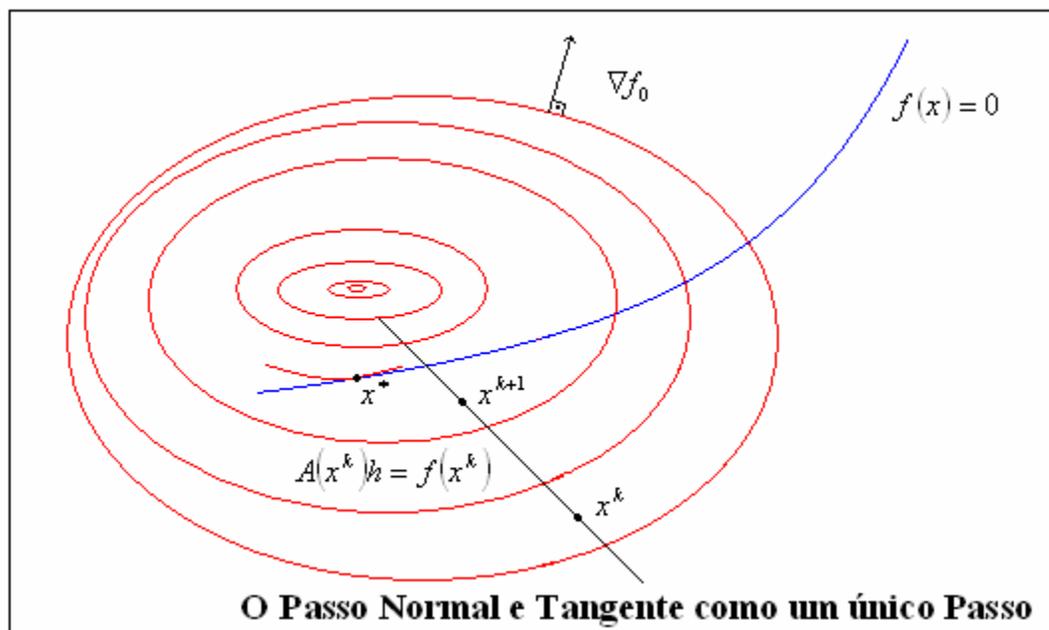


Figura 7

Algoritmo 2.2 O Passo Normal e Tangente simultaneamente, PQS puro

Dado $x^0 \in R^n$, $\lambda^0 \in R^m$ e $\mu^0 = 0$

para $k = 1, 2, 3, \dots$

Calcule $f_0(x^k)$, $\nabla f_0(x^k)$, $f(x^k)$ e $A(x^k)$

Resolva (2.13) para obter h^k e μ^k

Faça $x^{k+1} = x^k + h^k$ e $\lambda^{k+1} = \lambda^k + \mu^k$

se “convergência” PARE com solução aproximada (x^{k+1}, λ^{k+1}) .

Fim

Para provar algum tipo de convergência necessitaríamos abordar o problema numa região de confiança, e além disso utilizar uma função de mérito ou um filtro [13] para decidir se o ponto calculado pelo algoritmo 2.2 deve ou não ser aceito.

Capítulo 3

Funções Barreira

Neste capítulo introduziremos o conceito de Função Barreira, em especial da Barreira Infinita e da Barreira Logarítmica, a mais utilizada. O estudo apresentado adiante é essencial para que se abra caminho a conceitos tais como centro analítico e trajetória central, indispensáveis na abordagem prevista para o capítulo 4.

O método de Barreiras consiste em abordar problemas gerais de minimização restrita resolvendo uma seqüência de problemas de minimização irrestrita. Não iremos efetivamente utilizar esse método nesta dissertação, mas usufruir de conceitos, os brevemente mencionados no parágrafo anterior, para nos encaminharmos ao problema de minimização de quadráticas convexas em variedades afins. Mesmo assim, construímos o método, abordando aspectos tais como de convergência. Além disso, foi exposto um exemplo bem simples de minimização em R que ilustra o funcionamento do método de barreiras.

3.1 O Problema

Considere o problema

$$\begin{aligned} &\text{minimizar } f_0(x) \\ &\text{s.a. } f(x) \leq 0 \end{aligned} \tag{3.1}$$

Suponha $f_0 : R^n \rightarrow R$ uma função de classe pelo menos C^1 convexa.

Suponha também $f : R^n \rightarrow R^m$ tal que $f(x) = (f_1(x), f_2(x), \dots, f_m(x))$ em que $f_i : R^n \rightarrow R$, $i = 1, \dots, m$ são funções de classe pelo menos C^1 .

Vamos assumir as seguintes hipóteses:

- Seja $\Omega = \{x \in R^n; f(x) \leq 0\}$ tal que o seu interior seja não vazio e dado por $\overset{\circ}{\Omega} = \{x \in R^n; f(x) < 0\}$.

A figura 8 mostra um exemplo em que o conjunto Ω é um intervalo que é dado por $\{x \in R^n; f(x) \leq 0\}$, com f em C^∞ , tal que $\overset{\circ}{\Omega} = \emptyset$.

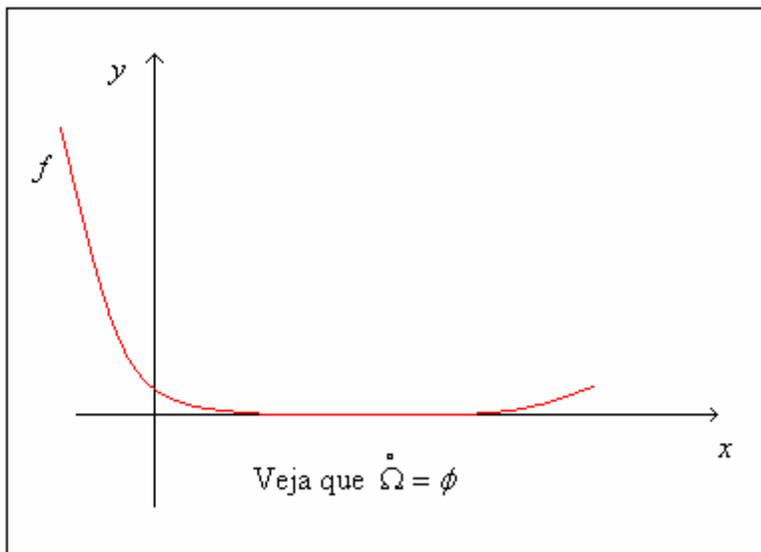


Figura 8

- Queremos também que o fecho de $\overset{\circ}{\Omega}$ seja o próprio Ω . Note que isto, para um conjunto qualquer é uma hipótese, e não uma proposição, como mostra a figura a seguir.

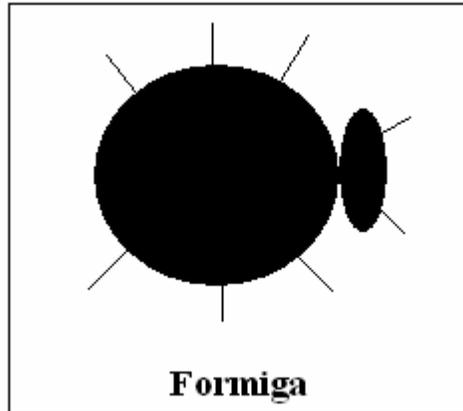


Figura 9

- Ω limitado.
- Um ponto $x^0 \in \overset{\circ}{\Omega}$.

3.2 Barreira Infinita

Considere a seguinte função, obviamente não bem definida

$$P_{\infty} = \begin{cases} 0, & \text{se } x \in \Omega \\ +\infty, & \text{se } x \in \mathbb{R}^n - \Omega \end{cases} .$$

Observe que o problema

$$\begin{aligned} &\text{Minimizar } f_0(x) && (3.2) \\ &\text{s.a. } x \in \Omega \end{aligned}$$

é equivalente ao problema

$$\begin{aligned} &\text{Minimizar } f_0(x) + P_{\infty}(x) && (3.3) \\ &\text{s.a. } x \in \mathbb{R}^n . \end{aligned}$$

Essa equivalência se deve, de modo trivial, ao fato da função $f_0 + P_\infty$ coincidir com f_0 em Ω e valer infinito fora de Ω . Passamos então de um problema restrito para um irrestrito, que a princípio é intratável. A figura a seguir ilustra essa equivalência mediante um problema em R^2 .

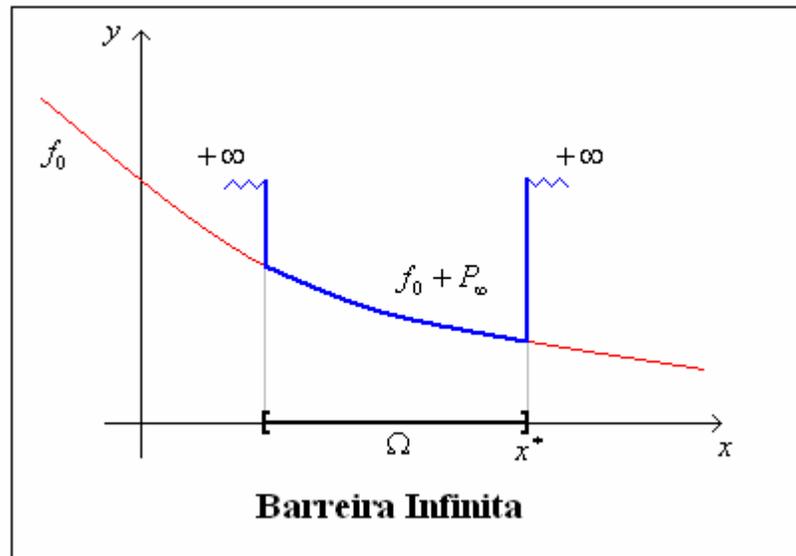


Figura 10

3.3 Barreira Logarítmica

A função Barreira Logarítmica, que iremos estudar agora, se comporta de forma parecida com a Barreira Infinita. Foram desenvolvidas várias funções barreira, mas a Logarítmica é a mais usada.

A função Barreira Logarítmica associada ao problema (3.1) é dada pela função $p : \overset{\circ}{\Omega} \rightarrow R$, tal que,

$$p(x) = -\sum_{i=1}^m \log(-f_i(x)). \quad (3.4)$$

Veja que, por outro lado, $p(x) = -\log \prod_{i=1}^m (-f_i(x))$. Olhando dessa forma podemos concluir que se o valor de qualquer restrição estiver muito próximo de zero, ou seja, x próximo da fronteira de Ω , o produto fica próximo de zero e com isto, $p(x)$ fica grande.

Agora, adotaremos um procedimento clássico, introduzido por Fiacco e McCormick [2], que consiste em resolver uma série de problemas, que chamaremos de P_μ , que se aproximam do problema de Barreira Infinita quando μ se aproxima de zero. Esta seqüência de problemas P_μ , por mais curioso que possa parecer, são problemas essencialmente restritos, que serão no entanto tratados com algoritmos de otimização irrestrita. Vamos comentar mais a respeito após a introdução formal do problema P_μ .

3.4 Problemas Penalizados

Seja $\mu \in R_+^*$, considere o problema P_μ a seguir

$$\begin{aligned} \text{Minimizar } & f_0(x) + \mu \cdot p(x) & (3.3) \\ \text{s.a. } & x \in \overset{\circ}{\Omega} \end{aligned}$$

A princípio o problema parece tão, ou mais difícil que o problema original (3.1). Iremos, no entanto ignorar a restrição $x \in \overset{\circ}{\Omega}$ e tratar (3.3) mediante algum algoritmo de minimização irrestrita (temos vários no Capítulo 1). Escolhemos então uma seqüência $(\mu_k)_{k \in N}$ com $\mu_k > 0$, $\forall k \in N$, e $\mu_k \rightarrow 0$. Resolvemos então, um por um, os problemas P_{μ^k} . Chamaremos de x_{μ^k} a solução que encontramos ao minimizar $F_{\mu^k}(x) = f_0(x) + \mu \cdot p(x)$ em R^n . Esses pontos serão chamados de pontos centrais e o conjunto $A = \{x_\mu \in R^n; \mu \in (0, +\infty)\}$ é chamado de Trajetória Central. Devido às hipóteses tomadas ao abordar o problema (3.1), é possível concluir que Trajetórias Centrais, no nosso

contexto, serão curvas contínuas. Vejamos agora um exemplo ilustrando a idéia de Fritsch [24].

Exemplo: Considere o Problema

$$\begin{aligned} & \text{Minimizar } \frac{(x-5)^2}{5} & (3.4) \\ & \text{s.a. } -x \leq 0 \text{ e } x-3 \leq 0. \end{aligned}$$

Ilustramos, mediante gráficos, que aparecem a seguir, a resolução dos problemas P_1 e $P_{0,2}$ associados a (3.4).

$$\begin{aligned} (P_1) \quad & \text{Minimizar } f_0(x) + \mu \cdot p(x) \\ & \text{s.a. } x \in \overset{\circ}{\Omega}, \end{aligned}$$

ou seja,

$$\begin{aligned} & \text{Minimizar } \frac{(x-5)^2}{5} - \log(x) - \log(3-x) & (3.5) \\ & \text{s.a. } -x < 0 \text{ e } x-3 < 0. \end{aligned}$$

Para resolver P_1 desconsideramos a condição $x \in \overset{\circ}{\Omega}$. Partimos então de um ponto $x^0 \in \overset{\circ}{\Omega}$, por exemplo, $x^0 = 1$ e minimizamos $\frac{(x-5)^2}{5} - \log(x) - \log(3-x)$ em R usando qualquer algoritmo de minimização irrestrita que seja convergente.

O mesmo procedimento é tomado para resolver $P_{0,2}$. A partir daí obtemos os resultados expressos nos gráficos a seguir. Veja que é fácil perceber que a solução do problema (3.4) é $x^* = 3$.

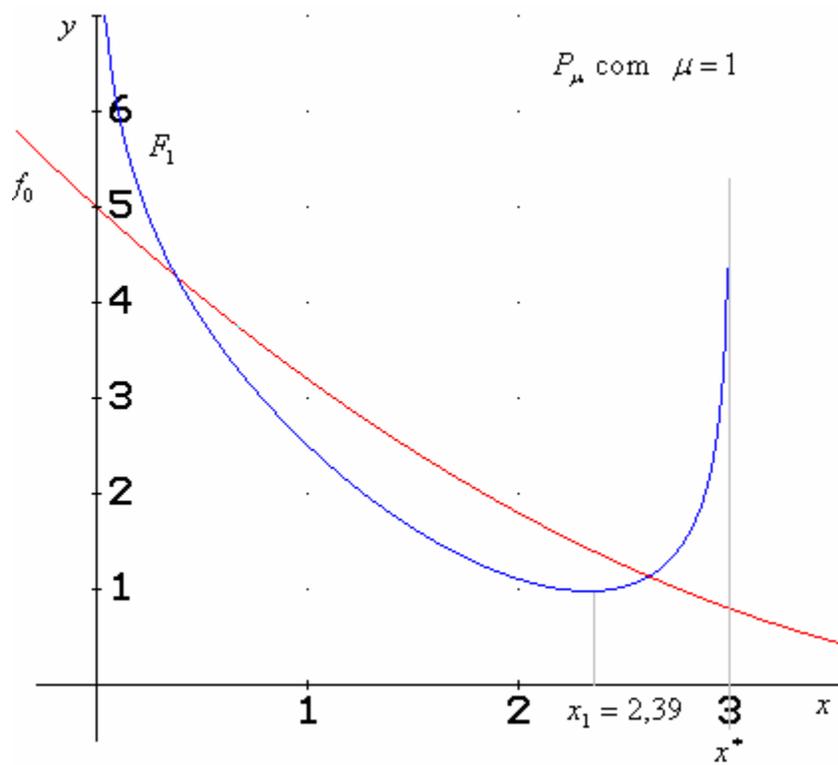


Figura 11

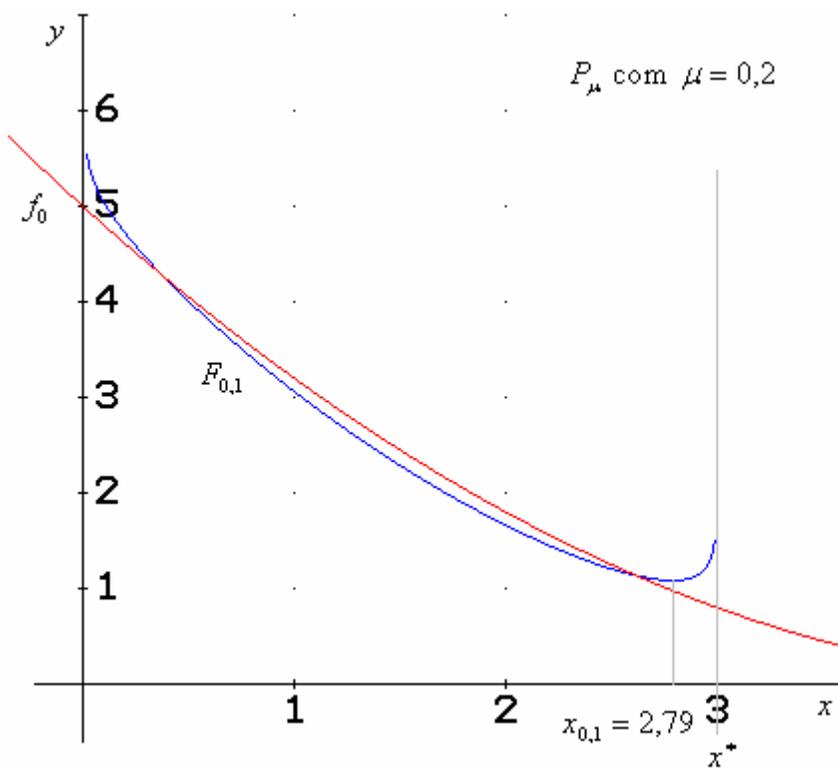


Figura 12

A solução de (3.4) é o limite, se existir, das soluções dos problemas P_{μ^k} .

3.5 O Algoritmo

Antes de enunciar o algoritmo observe que algumas dificuldades são encontradas em problemas gerais. Precisamos, a princípio de um ponto viável x^0 , cuja obtenção nem sempre é trivial. Sabemos também que não é possível calcular logaritmos de números negativos, neste sentido devemos garantir que a cada passo seja encontrado um ponto viável. Discussões relacionadas a estes empecilhos serão estudadas no próximo sub-capítulo.

Algoritmo 3.1 Método de Barreira Logarítmica

Dados $\alpha \in (0,1)$, $x^0 \in \overset{\circ}{\Omega}$ e $\mu^0 > 0$
para $k = 1,2,3,\dots$

Resolva P_{μ^k} mediante uma algoritmo de minimização irrestrita obtendo $x^k = x_{\mu^k}$.

O algoritmo reinicia em x^k .

$$\mu^k = \alpha \mu^{k-1}$$

Fim

3.6 Pontos Centrais e Trajetória Central

O método de Barreiras consiste então em resolver uma seqüência de problemas. Para cada valor $\mu > 0$ temos o problema P_μ dado em (3.3). Os pontos que resolvem P_μ são chamados de Pontos Centrais. Já o conjunto

$$A = \left\{ x_\mu \in R^n; x_\mu \in \arg \min_{x \in \overset{\circ}{\Omega}} \{ f_0(x) + \mu p(x) \} \right\},$$

é chamado de Trajetória Central. Veja, no entanto, que x_μ pode ser um conjunto não necessariamente unitário e com isso a Trajetória Central pode não ser um caminho contínuo no R^n .

3.7 Centro Analítico

Dado um poliedro $\Omega \in R^n$, chamamos de Centro Analítico de Ω qualquer minimizador da função Barreira Logarítmica em Ω .

LEMA 3.1: Seja $\Omega \in R^n$ um poliedro limitado. Nessas condições sempre existe um ponto que seja Centro Analítico de Ω , e ele é único.

Demonstração:

Sabemos que ao nos aproximarmos internamente da fronteira de Ω a função Barreira Logarítmica $p(\cdot)$ tende a $+\infty$ (lembre que fora de Ω ela nem sequer está definida). Como Ω é limitado, a função Barreira Logarítmica, que é uma função C^∞ , possui pelo menos um minimizador em $\overset{\circ}{\Omega}$. Por outro lado esse minimizador é único porque $p(\cdot)$ é estritamente convexa (Basta conferir que sua derivada segunda é positiva-definida. Isto está claro no capítulo 4).

■

LEMA 3.2: Seja $\Omega \subset R^n$ um poliedro limitado e P seu Centro Analítico. Então $P = \lim_{\mu \rightarrow \infty} P_\mu$ em que P_μ são Pontos Centrais de Ω .

Demonstração:

Seja $p(\cdot)$ a função Barreira Logarítmica. Considere então o seguinte problema:

$$\text{minimizar } f_0(x) + \mu.p(x)$$

$$\text{s.a. } x \in \overset{\circ}{\Omega}.$$

Veja que P_μ é solução do problema anterior. Para cada μ temos que o problema acima é equivalente, com $\mu \neq 0$, a

$$\text{minimizar } \frac{1}{\mu} (f_0(x) + \mu p(x)), \text{ s.a. } x \in \overset{\circ}{\Omega},$$

que é equivalente a

$$\text{minimizar } \frac{1}{\mu} f_0(x) + p(x), \text{ s.a. } x \in \overset{\circ}{\Omega}.$$

Observe que se $\mu \rightarrow \infty$ nosso problema se resume a minimizar $p(\cdot)$ em $\overset{\circ}{\Omega}$, já que f_0 é limitada em $\overset{\circ}{\Omega}$. Sabemos que este problema tem solução e é única pela proposição anterior. Portanto o centro analítico é o ponto da trajetória central que corresponde ao parâmetro $\mu = \infty$. Podemos ainda dizer, mesmo com o fato de P_μ não estar, em geral, univocamente definido, que, o Centro Analítico P de um conjunto satisfaz, $P = \lim_{\mu \rightarrow \infty} P_\mu$. ■

Deste modo o Centro Analítico destaca-se como sendo o ponto inicial da Trajetória Central.

Mais em frente nos depararemos com caixas no R^n , por isso definamos formalmente o conceito de Caixa. Chamamos de caixa no R^n qualquer conjunto $[u, w]$, tal que

$$[u, w] = \{x \in R^n; u \leq x \leq w\},$$

com $u, w \in R^n$ tais que $u < w$.

No último capítulo deste trabalho iremos abordar, como já foi frisado antes, problemas de região de confiança que consistem em minimizar quadráticas convexas em caixas. Trabalharemos com um algoritmo Primal-Dual que partirá de um ponto tal que o primal associado é o centro analítica da caixa na qual for sujeita a minimização. Neste contexto esclarecemos a seguir, sob forma de lema, quem é o centro analítico de uma caixa.

LEMA 3.3: O Centro Analítico de uma caixa $[-u, u]$, com $u \in R^n$, é o ponto $P = 0$.

Demonstração:

Devemos mostrar que o minimizador da função Barreira Logarítmica é o ponto $P = 0$.

Barreira Logarítmica: $-\sum_{i=1}^m \log(-f_i(x))$. No nosso caso as f_i s correspondem à caixa $[-u, u]$, ou seja $f_i(x) = x_i - u_i$, se $i = 1, \dots, n$ e $f_i(x) = -x_{i-n} - u_{i-n}$, se $i = n + 1, \dots, 2n$, ou seja $m = 2n$ restrições. Então

$$\text{minimizar } -\sum_{i=1}^{2n} \log(-f_i(x))$$

ou ainda,

$$\text{minimizar } -\sum_{i=1}^n \log(-f_i(x)) - \sum_{i=n+1}^{2n} \log(-f_i(x))$$

Obtemos então o seguinte problema

$$\text{minimizar } -\sum_{i=1}^n \log(x_i - u_i) - \sum_{i=1}^n \log(-x_i - u_i)$$

equivalente ao problema

$$\text{minimizar } -\log \prod_{i=1}^n -(x_i + u_i)(x_i - u_i)$$

ou

$$\text{minimizar } -\log \prod_{i=1}^n (u_i^2 - x_i^2)$$

Como a função logaritmo é uma função crescente, é fácil ver que o que procuramos é na verdade um maximizador para $\prod_{i=1}^n (u_i^2 - x_i^2)$. Cada fator do produto em questão é

independente, logo basta que achemos o maximizador de cada um, que trivialmente é $x_i = 0$. ■

Um fato interessante ligado ao centro analítico é que ele independe da função objetivo com que se está trabalhando, mas qualquer outro ponto da Trajetória Central, por mais próximo que esteja do Centro Analítico depende. Por meio de uma simples translação e de uma abordagem similar a anterior podemos demonstrar que o Centro Analítico de uma caixa $[l, u]$ é o ponto $x = \frac{l+u}{2}$.

3.8 Convergência

Suponha que consigamos calcular a cada passo, x_{μ_k} , a solução (já comentamos ser única no nosso contexto) do problema P_{μ_k} (problema tratado como de minimização global irrestrito). Suponha também que \bar{x} é um ponto de acumulação da seqüência (x_{μ_k}) , seqüência gerada a partir de uma seqüência decrescente (μ_k) , tal que $\mu_k \rightarrow 0$. Sabemos então da existência de uma subseqüência de (x_{μ_k}) convergente para \bar{x} . Vamos chamá-la, sem confusão, da mesma maneira. Ou seja, suponha então que $x_{\mu_k} \rightarrow \bar{x}$.

LEMA 3.4: Considere uma seqüência (μ_k) tal que $\mu_k \in R_{++}$ com $x_{\mu_k} \rightarrow \bar{x}$, com \bar{x} um ponto arbitrário em $\hat{\Omega}$. Então $\lim_{k \rightarrow \infty} F_{\mu_k}(\bar{x}) = f_0(\bar{x})$.

Demonstração:

Lembre que $F_{\mu_k}(\bar{x}) = f_0(\bar{x}) - \mu_k \sum_{i=1}^m \log(-f_i(\bar{x}))$. Como o fator $\sum_{i=1}^m \log(-f_i(\bar{x}))$ é constante e $\mu_k \rightarrow 0$, o Lema é imediato. ■

LEMA 3.5: Se \bar{x} estiver na fronteira de Ω então os valores $F_{\mu_k}(x_{\mu_k})$ são decrescentes em relação a k , para $k \in N$ suficientemente grande.

Demonstração:

Considere $F_k = F_{\mu_k}$. Queremos mostrar que $F_{\mu_{k+1}}(x_{\mu_{k+1}}) \leq F_{\mu_k}(x_{\mu_k})$ para $k \in N$ suficientemente grande.

Dado $k \in N$ temos que

$$F_{\mu_{k+1}}(x_{\mu_{k+1}}) = f_0(x_{\mu_{k+1}}) - \mu_{k+1} \sum_{i=1}^m \log(-f_i(x_{\mu_{k+1}})) \leq f_0(x_{\mu_k}) - \mu_{k+1} \sum_{i=1}^m \log(-f_i(x_{\mu_k})) = F_{\mu_{k+1}}(x_{\mu_k}).$$

A desigualdade anterior é válida graças à própria definição de $x_{\mu_{k+1}}$. Lembrando que $x_{\mu_{k+1}}$ é a solução do problema $P_{\mu_{k+1}}$.

Por outro lado lembre que $\mu_{k+1} \leq \mu_k$ (construção). Devemos mostrar então que

$$F_{\mu_{k+1}}(x_{\mu_k}) = f_0(x_{\mu_k}) - \mu_{k+1} \sum_{i=1}^m \log(-f_i(x_{\mu_k})) \leq f_0(x_{\mu_k}) - \mu_k \sum_{i=1}^m \log(-f_i(x_{\mu_k})) = F_{\mu_k}(x_{\mu_k}).$$

Veja que a desigualdade anterior só depende da positividade da função $-\sum_{i=1}^m \log(-f_i(\cdot))$. Temos, contudo, \bar{x} na fronteira de Ω .

Neste caso sabemos que $-\sum_{i=1}^m \log(-f_i(x_{\mu_k})) \rightarrow +\infty$, quando $k \rightarrow \infty$, o que garante a positividade de $-\sum_{i=1}^m \log(-f_i(x_{\mu_k}))$ a partir de um certo $K \in N$.

■

TEOREMA 3.1: Suponha que x_{μ_k} converge para \bar{x} . Então \bar{x} é solução ótima do problema (3.2), ou seja, de minimizar f_0 em Ω .

Demonstração:

Dividiremos a demonstração em dois casos.

(i) Suponha \bar{x} na Fronteira de Ω .

Suponha por contradição que existe $\hat{x} \in \Omega$ tal que $f_0(\hat{x}) = f_0(\bar{x}) - \varepsilon$, para algum $\varepsilon > 0$. Veja que o que estamos supondo é equivalente a dizer que \bar{x} não é minimizador do problema (3.2).

$$\text{Temos em particular que } f_0(\hat{x}) < f_0(\bar{x}) - \frac{\varepsilon}{2}.$$

Observe que desde o começo deste capítulo estamos supondo que o fecho do interior de Ω é o próprio conjunto Ω . Esse fato nos garante que qualquer ponto de Ω pode ser aproximado por pontos do interior de Ω , devido à definição de fecho. Logo existe $\tilde{x} \in \overset{\circ}{\Omega}$ tal que $f_0(\tilde{x}) \leq f_0(\bar{x}) - \frac{\varepsilon}{4}$, já que f_0 é contínua.

Veja que $f_0(x_{\mu_k}) \leq f_0(x_{\mu_k}) + \mu_k p(x_{\mu_k}) = F_{\mu_k}(x_{\mu_k})$, devido a positividade de p a partir de um certo $K \in \mathbb{N}$. Por outro lado, pela definição de x_{μ_k} , $F_{\mu_k}(x_{\mu_k}) \leq F_{\mu_k}(\tilde{x})$. Logo,

$$f_0(x_{\mu_k}) \leq F_{\mu_k}(\tilde{x}) = f_0(\tilde{x}) + \mu_k p(\tilde{x})$$

Tomando limites teremos

$$\lim_{k \rightarrow \infty} f_0(x_{\mu_k}) \leq \lim_{k \rightarrow \infty} f_0(\tilde{x}) + \mu_k p(\tilde{x}).$$

Mediante o Lema 3.5 e a hipótese da convergência de x_{μ_k} para \bar{x} concluímos que

$$f_0(\bar{x}) \leq f_0(\tilde{x}), \text{ o que contradiz o fato de que } f_0(\tilde{x}) \leq f_0(\bar{x}) - \frac{\varepsilon}{4}.$$

Logo \bar{x} é solução ótima do problema (3.2).

(ii) Suponha $\bar{x} \in \overset{\circ}{\Omega}$

Temos então que $f_0(x^*) \leq f_0(x)$, $\forall x \in \Omega$. Por outro lado note que

$$F_{\mu_k}(x^k) \leq F_{\mu_k}(x^*), \text{ ou seja, } f_0(x^k) + \mu_k p(x^k) \leq f_0(x^*) + \mu_k p(x^*),$$

o que implica em

$$\lim_{k \rightarrow \infty} f_0(x^k) \leq \lim_{k \rightarrow \infty} f_0(x^*) + \mu_k (p(x^*) - p(x^k))$$

$$f_0(\bar{x}) \leq f_0(x^*).$$

Portanto $\bar{x} = x^*$.



3.9 Multiplicadores de KKT

Considere $x^k = x_{\mu_k}$. Devido ao fato de x^k ser minimizador global da função

F_{μ_k} temos que

$$\nabla F_{\mu_k}(x^k) = 0$$

$$\text{Por outro lado } \nabla F_{\mu_k}(x^k) = \nabla \left(f_0(x^k) + \mu_k \sum_{i=1}^m \log(f_i(x^k)) \right) = \nabla f_0(x^k) + \mu_k \sum_{i=1}^m \frac{-\nabla f_i(x^k)}{f_i(x^k)}.$$

Defina $\lambda^k = \frac{\mu_k}{-f_i(x^k)}$. Observe que $\lambda^k > 0$ já que x^k é viável e $\mu_k > 0$.

Temos então, para cada k

$$\nabla f_0(x^k) + \sum_{i=1}^m \lambda^k \nabla f_i(x^k) = 0, \quad \lambda^k > 0, \quad \mu_k = \lambda^k (-f_i(x^k)) \text{ e } x^k \in \Omega.$$

Veja que só falta a complementaridade. Entretanto, se tomarmos limites vemos que

$$0 = \lim_{k \rightarrow \infty} \mu_k = \lim_{k \rightarrow \infty} \lambda^k (-f_i(x^k)), \text{ o que implica que } \lim_{k \rightarrow \infty} \lambda^k f_i(x^k) = 0.$$

Finalmente concluímos que os multiplicadores de KKT são dados pelo limite

$$\bar{\lambda} = \lim_{k \rightarrow \infty} \frac{\mu_k}{-f_i(x^k)} \geq 0 \text{ (Multiplicador de Lagrange).}$$

Desta maneira são satisfeitas todas as condições de KKT, as quais são

$$\nabla f_0(\bar{x}) + \sum_{i=1}^m \bar{\lambda} \nabla f_i(\bar{x}) = 0$$

$$\bar{\lambda} \geq 0$$

$$\bar{x} \in \Omega$$

$$\bar{\lambda} f_i(\bar{x}) = 0$$

Capítulo 4

Minimização de Quadráticas Convexas em Caixas

Este é o capítulo principal desta dissertação, que visará justificar o estudo de todos os tópicos abordados até este momento. Iremos propor meios teóricos, inclusive utilizar um resultado novo (Teorema 4.1), para solucionar nosso problema alvo.

Trataremos de dar um outro formato ao problema de minimização de uma quadrática numa caixa. A partir de então, saindo do ponto primal que é o centro analítico de um poliedro construído através da caixa dada faremos uma análise bem específica para encontrar um ponto inicial primal-dual viável, para que possamos daí sim nos focar em um algoritmo de Pontos Interiores primais-duais. Com este algoritmo procuraremos seguir a Trajetória Central Primal-Dual em termos da vizinhança $N_{-\infty}$, que também será levemente explanada.

4.1 O Problema

$$\begin{aligned} & \text{minimizar } \frac{1}{2} x^T H x + c^T x \\ & \text{s.a. } \begin{aligned} & Ax = b \\ & x \geq 0 \\ & x \leq u, \quad u > 0 \end{aligned} \end{aligned} \quad (4.1)$$

em que $u, c, x \in R^n$, $b \in R^m$, $A \in R^{m \times n}$ com $\text{rank}(A^T) = n$ e $H \in R^{n \times n}$ uma matriz simétrica semi-positiva definida. Suporemos também que $x^0 = \frac{u}{2}$, centro geométrico e analítico da caixa $[0, u]$, é viável.

Introduzindo variáveis de folga w em (4.1) obtemos

$$\begin{aligned} & \text{minimizar } \frac{1}{2} x^T H x + c^T x \\ & \text{s.a. } \begin{aligned} & Ax = b \\ & x + w = u \\ & x \geq 0 \\ & w \geq 0 \end{aligned} \end{aligned}$$

Podemos ainda reescrever o problema anterior como

$$\begin{aligned} & \text{minimizar } \frac{1}{2} \begin{bmatrix} x \\ w \end{bmatrix}^T \begin{bmatrix} H & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ w \end{bmatrix} + \begin{bmatrix} c \\ 0 \end{bmatrix}^T \begin{bmatrix} x \\ w \end{bmatrix} \\ & \text{s.a. } \begin{bmatrix} A & 0 \\ I & I \end{bmatrix} \begin{bmatrix} x \\ w \end{bmatrix} = \begin{bmatrix} b \\ u \end{bmatrix} \\ & \begin{bmatrix} x \\ w \end{bmatrix} \geq 0 \end{aligned}$$

Fazendo $\bar{H} = \begin{bmatrix} H & 0 \\ 0 & 0 \end{bmatrix}$, $\bar{x} = \begin{bmatrix} x \\ w \end{bmatrix}$, $\bar{A} = \begin{bmatrix} A & 0 \\ I & I \end{bmatrix}$, $\bar{b} = \begin{bmatrix} b \\ u \end{bmatrix}$ e $\bar{c} = \begin{bmatrix} c \\ 0 \end{bmatrix}$ obtemos

$$\begin{aligned} & \text{minimizar } \frac{1}{2} \bar{x}^T \bar{H} \bar{x} + \bar{c}^T \bar{x} \\ & \text{s.a. } \bar{A} \bar{x} = \bar{b} \\ & \bar{x} \geq 0 \end{aligned} \tag{4.2}$$

com $\bar{x}, \bar{c} \in R^{2n}$, $\bar{H} \in R^{2n \times 2n}$, $\bar{b} \in R^{2n}$, $\bar{A} \in R^{2m \times 2n}$.

Veja que conseguimos escrever o problema (4.1), no formato primal de problemas de programação quadrática. Note, contudo, que isso nos forçou a dobrar o tamanho do problema, ou seja, passamos de um problema do R^n para um do R^{2n} . Observe, contudo, que \bar{H} , assim como H , também é simétrica semi-definida positiva e que $\text{rank}(\bar{A}) = 2m$. Optamos por trabalhar com o problema (4.1) no formato (4.2) por ser um formato padrão, que simplifica a exposição das condições de KKT, como veremos adiante.

LEMA 4.1: O Centro Analítico do conjunto $\Omega = \{\bar{x} \in R^{2n}; \bar{A}\bar{x} = \bar{b}, \bar{x} \geq 0\}$ é o ponto

$$\begin{bmatrix} 0.5u \\ 0.5u \end{bmatrix}.$$

Demonstração:

Devemos mostrar que $\begin{bmatrix} 0.5u \\ 0.5u \end{bmatrix}$ é o minimizador da função barreira sob as restrições de desigualdade impostas nessa proposição.

Veja que as restrições de desigualdade em questão correspondem a $\bar{x} = \begin{bmatrix} x \\ w \end{bmatrix} \geq 0$ (lembre que $w = u - x$). Ou seja, temos $2n$ restrições de desigualdade, as quais são descritas melhor a seguir:

$f_i : R^{2n} \rightarrow R$, $f_1(\bar{x}) = x_1$, ..., $f_n(\bar{x}) = x_n$ e $f_{n+1}(\bar{x}) = w_1$, ..., $f_{2n}(\bar{x}) = w_n$, com $\bar{x} \in R^{2n}$.

Como $w = u - x$ temos que $f_{n+1}(\bar{x}) = u_1 - x_1$, ..., $f_{2n}(\bar{x}) = u_n - x_n$.

Minimização da Função Barreira:

$$\text{minimizar } - \sum_{i=1}^m \log(-f_i(\bar{x}))$$

$$\text{s.a. } Ax = b.$$

Mediante as definições anteriores temos que

$$\begin{aligned} - \sum_{i=1}^m \log(-f_i(\bar{x})) &= - \sum_{i=1}^{2n} \log(\bar{x}_i) = -(\log x_1 + \dots + \log x_n + \log(u_1 - x_1) + \dots + \log(u_n - x_n)) = \\ &= -\log(x_1(u_1 - x_1)x_2(u_2 - x_2)\dots x_n(u_n - x_n)) = -\log\left(\prod_{i=1}^n x_i(u_i - x_i)\right). \end{aligned}$$

Queremos então que $\prod_{i=1}^n x_i(u_i - x_i)$ seja máximo. Como os fatores

$x_1(u_1 - x_1), \dots, x_n(u_n - x_n)$ são independentes, basta maximizar cada um deles, o que é

trivial. Veja que o maximizador de $x_i(u_i - x_i)$ é único e é dado trivialmente por $x_i = \frac{u_i}{2}$.

Logo $w_i = u_i - \frac{u_i}{2} = \frac{u_i}{2}$. Com isso temos que $\arg \min \left\{ - \sum_{i=1}^m \log(-f_i(\bar{x})) \right\} = \left\{ \begin{bmatrix} 0.5u \\ 0.5u \end{bmatrix} \right\}$. Com

mais forte razão este ponto minimiza $p(\cdot)$ na interseção do conjunto $\{x \in R^{2n}; x \geq 0\}$ com a variedade definida por $Ax = b$, completando a demonstração. ■

Considere abandonada a notação \bar{H} , \bar{x} , para que a mesma não fique sobrecarregada. Façamos então, sem confusão, para a próxima seção, $H = \bar{H}$, $x = \bar{x}$, etc.

Podemos agir tranquilamente desta forma, pois trabalharemos, a partir de agora, somente com vetores em R^{2n} e R^{2m} .

4.2 Um Ponto Inicial Primal-Dual Viável

Observe que nosso problema, devido aos resultados da seção anterior, se resume a

$$\begin{aligned} & \text{minimizar } \frac{1}{2} x^T H x + c^T x \\ & \text{s.a. } \quad Ax = b \\ & \quad \quad x \geq 0. \end{aligned} \tag{4.3}$$

Seja $\Omega = \{x \in R^{2n}; Ax = b, x \geq 0\}$. Sabemos que o interior desse conjunto é dado por $\overset{\circ}{\Omega} = \{x \in R^{2n}; Ax = b, x > 0\}$. Considere também o conjunto $\Gamma := \{(x, s); x \in \Omega, s \geq 0\}$ e note que o interior do mesmo é dado por $\overset{\circ}{\Gamma} = \{(x, s); x \in \overset{\circ}{\Omega}, s > 0\}$.

Queremos, nesta seção, encontrar um ponto primal-dual $(x, s) \in R^{2n} \times R^{2n}$ em $\overset{\circ}{\Gamma}$. Teremos como hipótese que nosso ponto de partida primal, x , é o centro analítico de Ω . Objetivamos um ponto primal-dual em $\overset{\circ}{\Gamma}$ porque estaremos interessados em seguir a Trajetória Central primal-dual mediante um algoritmo de pontos interiores. Isto nos força a encontrar um primeiro par primal-dual em $\overset{\circ}{\Gamma}$ para que nosso algoritmo esteja bem definido. Além disso vamos querer um par em $\overset{\circ}{\Gamma}$ que satisfaça a primeira condição de KKT, como veremos no estudo a seguir.

4.2.1 Condições de KKT relaxadas

Considere as condições de KKT relaxadas (relaxadas no sentido de que estamos necessariamente em $\overset{\circ}{\Gamma}$ e a complementaridade não é satisfeita):

$$\begin{aligned} -Hx + A^T y + s &= c \\ Ax &= b \\ xs &= \mu e \\ x, s &> 0 \end{aligned} \tag{4.4}$$

em que xs denota o produto de x por s componente a componente.

Lembre que já falamos um pouco sobre trajetória central primal no Capítulo 3, seção 3.6, o que nos possibilitará deduzir que (4.4) corresponde às condições necessárias e suficientes para um ponto primal pertencer à trajetória central primal.

LEMA 4.2: Se um ponto pertence à Trajetória Central Primal de (4.3) então ele satisfaz (4.4), com algum $y \in R^{2m}$, $s \in R^{2n}$.

Demonstração:

Problema penalizado com a Barreira Logarítmica:

$$\begin{aligned} \text{minimizar } & \frac{1}{2} x^T H x + c^T x + \mu p(x) \\ \text{s.a. } & Ax = b \\ & x > 0 \end{aligned}$$

em que $p(x) = -\sum_{i=1}^{2n} \log(-x_i)$ e $\mu > 0$. Temos então que as condições de otimalidade para o problema penalizado são

$$\begin{aligned} c + Hx - A^T y - \mu \frac{1}{x} &= 0 \\ Ax &= b \\ x &> 0 \end{aligned}$$

Em que $\frac{1}{x}$ é o vetor dos inversos das componentes de x . Chame de $s_i = \frac{\mu}{x_i}$, então $sx = \mu e$ com $x, s > 0$. Com isso temos exatamente as condições (4.4), demonstrando a proposição. ■

Dessa maneira definimos então a Trajetória Central Primal-Dual de (4.3) como sendo o conjunto dos pontos $(x, s) \in R^{2n} \times R^{2n}$ que satisfazem (4.4). Chamamos de Ponto Central Primal-Dual associado ao parâmetro $\mu > 0$ um ponto $(x_\mu, s_\mu) \in R^{2n} \times R^{2n}$ que pertence a Trajetória Central Primal-Dual. No nosso caso, em que a função objetivo é quadrática convexa e Ω é um poliedro, é possível demonstrar que a Trajetória Central é uma curva contínua tal que para cada valor de $\mu > 0$ existe um único Ponto Central associado, veja Gonzaga [11]. Considere contudo, a proposição adiante que trata da existência e da unicidade do centro analítico de um poliedro.

LEMA 4.3: Se Ω é um poliedro limitado então $\arg \min \left\{ p(x) ; x \in \overset{\circ}{\Omega} \right\}$, em que $p(\cdot)$ é a Barreira Logarítmica, é um conjunto formado por um único ponto, ou seja, o Centro Analítico existe e é único.

Demonstração:

Sabemos que os conjuntos de nível de $p(\cdot)$ são limitados em Ω , caso contrário, contrariaríamos a hipótese de Ω ser limitado. Veja que $p(\cdot)$ é contínua num conjunto fechado, com isso temos que $p(\cdot)$ é contínua num conjunto de nível em Ω . Logo $p(\cdot)$ possui minimizador em $\overset{\circ}{\Omega}$.

Veja que $p(\cdot)$ é estritamente convexa em $\overset{\circ}{\Omega}$ como mostramos a seguir

Função Barreira:

$$x \in R_{++}^{2n} \mapsto p(x) = -\sum_{i=1}^{2n} \log(x_i)$$

Logo $\nabla p(x) = -\frac{1}{x}$ e $\nabla^2 p(x) = \text{diag}(x^{-2}) > 0$ (Hessiana Positiva-Definida). Portanto a

hessiana $\nabla^2 p(x)$ é definida positiva para todo $x \in \overset{\circ}{\Omega}$, e então $p(\cdot)$ possui minimizador único em $\overset{\circ}{\Omega}$. ■

Dado $x \in R^{2n}$ em $\overset{\circ}{\Omega}$, $s \in R_+^{2n}$ será dito dual viável se, e somente se, s for folga dual, o que para nós significará $P(s) = P(\nabla f(x))$ (equivalente a primeira condição de KKT vista em (4.4)), em que $P(\cdot)$ é a função projeção sobre $N(A)$ e $s \geq 0$ (dual viável interior se $s > 0$). Então, definimos o conjunto dos pares primais duais viáveis como sendo V , em que $(x, s) \in V$ se, e somente se, $(x, s) \in \Gamma$ e $P(s) = P(\nabla f(x))$. Deste modo é possível concluir que o conjunto dos pares primais duais viáveis é o conjunto $\overset{\circ}{V}$ tal que $(x, s) \in \overset{\circ}{V}$ se, e somente se, $(x, s) \in \overset{\circ}{\Gamma}$ e $P(s) = P(\nabla f(x))$.

Nossa busca por um par primal-dual em $\overset{\circ}{V}$ consistirá na abordagem de dois casos. O primeiro caso terá como hipótese $x = e \in R^{2n}$ sendo o centro analítico de Ω . O segundo caso corresponde a $x > 0$ qualquer, sendo que a resolução deste será baseada no primeiro caso.

1º Caso

Suponha então que $x = e$ com e sendo o centro analítico do poliedro Ω . Sabemos que e é centro analítico de Ω se, e somente se, é o minimizador da função Barreira

Logarítmica. Com isso, veja que a primeira condição de KKT aplicada ao problema de minimizar $p(\cdot)$ na variedade $Ax = b$ é $\nabla p(e) + A^T y = 0$, ou então, equivalentemente $\nabla p(e) \perp N(A)$, ou ainda, $P(\nabla p(e)) = 0$. Mas $\nabla p(e) = -\frac{1}{e} = -e$. Como P é uma projeção temos $P(e) = -P(-e)$. Logo $P(e) = 0$.

A proposição a seguir é similar ao que se descreve em [11]:

LEMA 4.4: Se e é centro analítico de Ω então, existe $\mu_0 \in R_+^*$ tal que para $\mu > \mu_0$ temos $(e, s) \in \overset{\circ}{V}$ com $s = P(\nabla f(e)) + \mu e$.

Demonstração:

Sabemos que $P(s) = P(\nabla f(e))$ já que e é um ponto primal viável.

Por outro lado,

$$P(s) = P(P(\nabla f(e)) + \mu e) = P(\nabla f(e)) + \mu P(e) = P(\nabla f(e)) + 0 = P(\nabla f(e)).$$

Deste modo $s = P(\nabla f(e)) + \mu e$ é folga dual.

Não sabemos, entretanto, a priori, se $s = \nabla f(e) + \mu e > 0$. Mas para $\mu > 0$ suficientemente grande teremos $s > 0$: basta tomar $\mu^0 = \min \{ \|\nabla f_i(e)\|, i = 1, \dots, n \}$ ou ainda $\mu > \|\nabla f(e)\|_\infty \geq \mu_0$. Logo $s = P(\nabla f(e)) + \mu e$ é folga dual viável. ■

A conclusão é a seguinte: Se $x = e$ for centro analítico de Ω então o par $(e, P(\nabla f(e)) + \mu e)$ é viável se $\mu > \|\nabla f(e)\|_\infty$. Ou seja, temos um jeito simples e prático para encontrar um ponto primal-dual inicial em $\overset{\circ}{V}$. Entretanto não nos contentaremos com este par primal-dual. Nesse sentido vamos procurar um par primal-dual tal que o primal associado não seja centro analítico de Ω

2º Caso

Será estudado nos sub-capítulos 4.2.2 e 4.2.3.

4.2.2 Vizinhanças da Trajetória Central Primal Dual

Neste sub-capítulo vamos adotar o termo proximidade, no sentido de decidir o quão perto um ponto de $\overset{\circ}{V}$ está da Trajetória Central primal-dual, ou mesmo de um ponto central. Definimos então a proximidade de um ponto $(x, s) \in \overset{\circ}{V}$ a um ponto (x_μ, s_μ) da Trajetória Central como sendo o valor

$$\delta(x, s, \mu) = \left\| \frac{xs}{\mu} - e \right\|, \text{ com } \mu > 0 \quad (4.5)$$

Dado um par $(x, s) \in \overset{\circ}{V}$, sabemos portanto calcular a proximidade $\delta(x, s, \mu)$ ao ponto central (x_μ, s_μ) . Dado um ponto $(x, s) \in \overset{\circ}{V}$, gostaríamos de definir uma proximidade de (x, s) à trajetória central (e não a um ponto central dado).

Seja $(x, s) \in \overset{\circ}{V}$ fixo. Se para algum valor de $\mu > 0$, $\delta(x, s, \mu)$ for um número próximo de zero, temos que (x, s) está próximo da Trajetória Central (grosseiramente falando, porque isso depende da norma).

Estaremos, portanto, interessados em procurar um valor de μ tal que a proximidade do ponto central (x_μ, s_μ) a (x, s) seja mínima. Devemos então tentar minimizar a proximidade $\left\| \frac{xs}{\mu} - e \right\|$ ou mesmo $\|xs - \mu e\|$ na variável μ . Após acharmos esse μ (que vamos provar que existe) saberemos automaticamente, mediante o valor da proximidade, se estamos de certa forma perto, ou não, da Trajetória Central. Mais adiante, na seção 4.3, ficará clara a motivação que nos leva a procurar tal μ .

(i) Seja (x, s) um ponto primal-dual viável queremos minimizar $\|xs - \mu e\|$ em μ :

$$\text{Note que } \arg \min_{\mu} \{\|xs - \mu e\|\} = \arg \min_{\mu} \{\|xs - \mu e\|^2\}$$

Basta então, achar o ponto crítico da função quadrática real ϕ dada por

$$\phi(\mu) = \|xs - \mu e\|^2 \text{ com } \mu > 0.$$

Para isso, façamos

$$\phi'(\mu) = \langle xs - \mu e, xs - \mu e \rangle' = (2n\mu^2 - 2x^T s \mu + (x^2)^T s^2)' = 4n\mu - 2x^T s$$

$$\text{Fica fácil perceber que } \min_{\mu} \|xs - \mu e\| = \mu(x, s) = \frac{x^T s}{2n}$$

OBS: Lembre que $\dim(x) = 2n$.

(ii) Seja (x, s) um ponto primal-dual viável. Queremos minimizar $\left\| \frac{xs}{\mu} - e \right\|$ em μ .

Temos novamente que $\arg \min_{\mu} \left\{ \left\| \frac{xs}{\mu} - e \right\| \right\} = \arg \min_{\mu} \left\{ \left\| \frac{xs}{\mu} - e \right\|^2 \right\}$. Seguindo a mesma

estratégia de (i), façamos

$$\frac{d \left(\left\| \frac{xs}{\mu} - e \right\|^2 \right)}{d\mu} = 0.$$

Para isto veja que $\left\langle \frac{xs}{\mu} - e, \frac{xs}{\mu} - e \right\rangle' = \left(\frac{(x^2)^T s^2}{\mu^2} - \frac{2x^T s}{\mu} + n \right)' = -2\mu^{-3} \|xs\|^2 + 2\mu^{-2} x^T s$.

Então o ponto de mínimo procurado é dado por $\mu = \frac{\|xs\|^2}{x^T s}$.

Pelo que pesquisamos na literatura o valor mais utilizado, dentre os dois calculados há pouco, é $\mu(x, s) = \frac{x^T s}{\dim(x)}$, o qual iremos utilizar, mesmo caminhando, digamos assim, contra nossa definição de proximidade. O fato é que podemos garantir certa proximidade entre os valores $\frac{x^T s}{\dim(x)}$ e $\frac{\|xs\|^2}{x^T s}$. Entretanto, não iremos nos focar nesse detalhe, mesmo porque ele não tem grande importância, como veremos no sub-capítulo 4.3.

Vamos definir agora algumas vizinhanças da trajetória central primal-dual. O objetivo de introduzir essas vizinhanças está relacionado ao fato de querermos, mediante um algoritmo de pontos interiores, trabalhar dentro delas. Escolheremos por final, utilizar a maior delas, como veremos.

Faremos uma pequena análise das vizinhanças de trajetórias centrais primais-duais mais encontradas na literatura. Estas são todas definidas através da definição de proximidade em (4.5). Assim, uma vizinhança qualquer da trajetória central primal-dual em questão é expressa pelo conjunto $N(\alpha) = \left\{ (x, s) \in \overset{\circ}{V}; \delta(x, s, \mu) \leq \alpha \right\}$ para algum $\mu > 0$ e $\alpha \in (0, 1)$. Vejamos agora algumas delas.

Vizinhança 1:

$$N_2(\alpha) := \left\{ (x, s) \in \overset{\circ}{V}; \left\| \frac{xs}{\mu} - e \right\|_2 \leq \alpha \right\}, \text{ para algum } \mu > 0 \quad (4.6)$$

Vizinhança 2:

$$N_\infty(\alpha) := \left\{ (x, s) \in \overset{\circ}{V}; \left\| \frac{xs}{\mu} - e \right\|_\infty \leq \alpha \right\}, \text{ para algum } \mu > 0 \quad (4.7)$$

Vizinhança 3:

$$N_{-\infty}(\alpha) := \left\{ (x, s) \in \overset{\circ}{V}; \left\| \frac{xs}{\mu} - e \right\|_{-\infty} \leq \alpha \right\}, \text{ para algum } \mu > 0, \quad (4.8)$$

com $\|y\|_{-\infty} = \max\{|y_i|; \text{tal que } y_i < 0\}$ para $\alpha \in (0,1)$, ou seja, nos preocupamos apenas com o tamanho das componentes negativas. Note que em (4.8) $\|\cdot\|_{-\infty}$ não é uma norma. Para valores pequenos de $(1-\alpha)$ (um valor comum é 10^{-3}), $N_{-\infty}$ pode englobar quase todo conjunto $\overset{\circ}{V}$. Desta maneira podemos redefinir $N_{-\infty}$ como

$$N_{-\infty}(\alpha) := \left\{ (x, s) \in \overset{\circ}{V}; \frac{xs}{\mu} \geq (1-\alpha)e \right\}, \text{ para algum } \mu > 0. \quad (4.9)$$

É possível mostrar que $N_{-\infty} \supset N_\infty \supset N_2$, veja [24]. A figura a seguir ilustra este fato

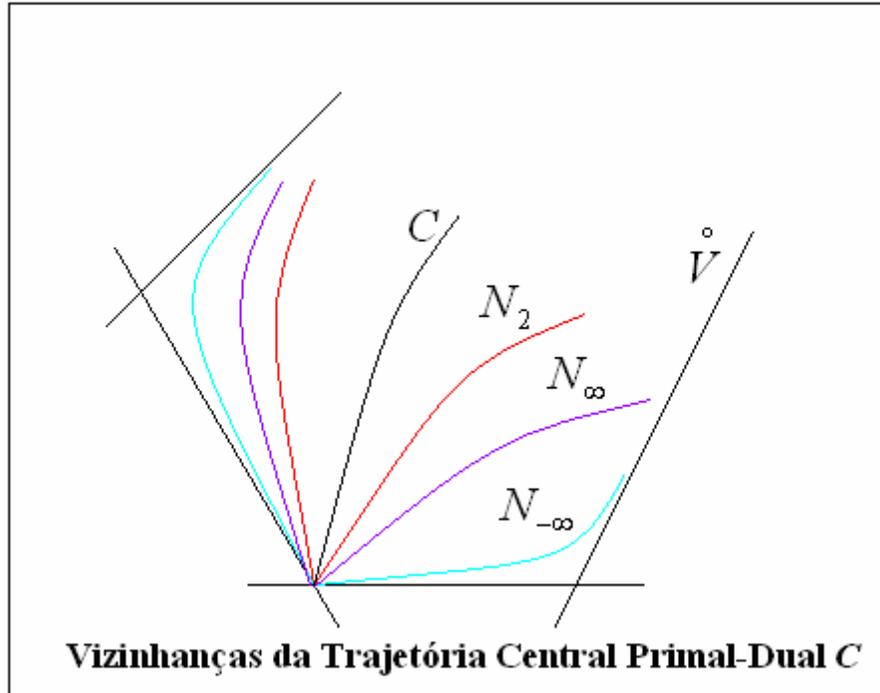


Figura 13

4.2.3 Mudança de Escala

Se e for centro analítico de Ω , já resolvemos o problema de encontrar pelo menos um ponto primal-dual em $\overset{\circ}{V}$, como visto no sub-capítulo 4.2.1. Se e não o for, o que em geral vai acontecer, faremos uma mudança de escala, que será desenvolvida mediante a construção de uma aplicação homeomorfa. Faremos com que no espaço vetorial auxiliar, resultante da aplicação homeomorfa em questão, o ponto e esteja no interior do conjunto $\varphi(\Omega)$ e seja o centro analítico do mesmo.

Suponha então que nosso ponto inicial primal seja o ponto dado $x^0 \in \overset{\circ}{\Omega}$, tal que $x^0 \neq e$ e que x^0 seja centro analítico de Ω . Faremos agora uma mudança de escala dada pela aplicação a seguir

$$\begin{aligned} \varphi: R^{2n} &\rightarrow R^{2n} \\ x &\mapsto \varphi(x) = \bar{x} = X^{-1}x \end{aligned} \tag{4.10}$$

Em que $X = \text{diag}(x^0)$. Observe que X^{-1} está bem definida já que $x^0 > 0$. Temos então que $x = X \bar{x}$. A figura a seguir ilustra a mudança de escala em R^2 que estamos propondo.

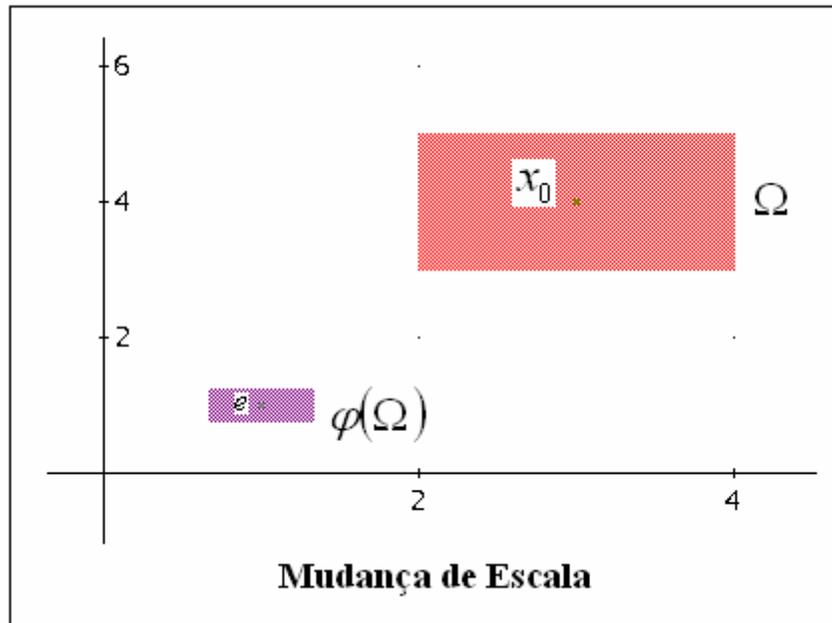


Figura 14

Vamos caminhar agora em direção a construção de uma folga dual s , para isto considere a função $\hat{f} = f \circ \varphi^{-1}$.

Temos então que $\nabla_x f(x) = \nabla_x \hat{f}(\bar{x}) = \nabla_x \hat{f}(X^{-1}x) = X^{-1} \nabla_{\bar{x}} \hat{f}(\bar{x})$.

Logo, $-\nabla_{\bar{x}} \hat{f}(\bar{x}) = -X \nabla f(x)$ (direção de máximo declive de \hat{f} a partir de \bar{x}).

Lembre que estamos sujeitos a restrição $Ax = b$, ou seja, $AX \bar{x} = b$. Isto sugere a projeção de $-\nabla_{\bar{x}} \hat{f}(\bar{x}) = -X \nabla f(x)$ sobre $N(AX)$. Chamaremos esta projeção de \bar{d} . Assim sendo, temos

$$\bar{d} = -P_{AX}(X \nabla f(x)) \quad (4.11)$$

Essa direção é de suma importância no nosso estudo porque nos proporcionará uma folga dual viável s como veremos em seguida. A direção \bar{d} também define a direção afim-escala $d = X \bar{d}$ apresentada por Dikin na década de 80. Considere a figura 14 que ilustra a construção que queremos fazer.

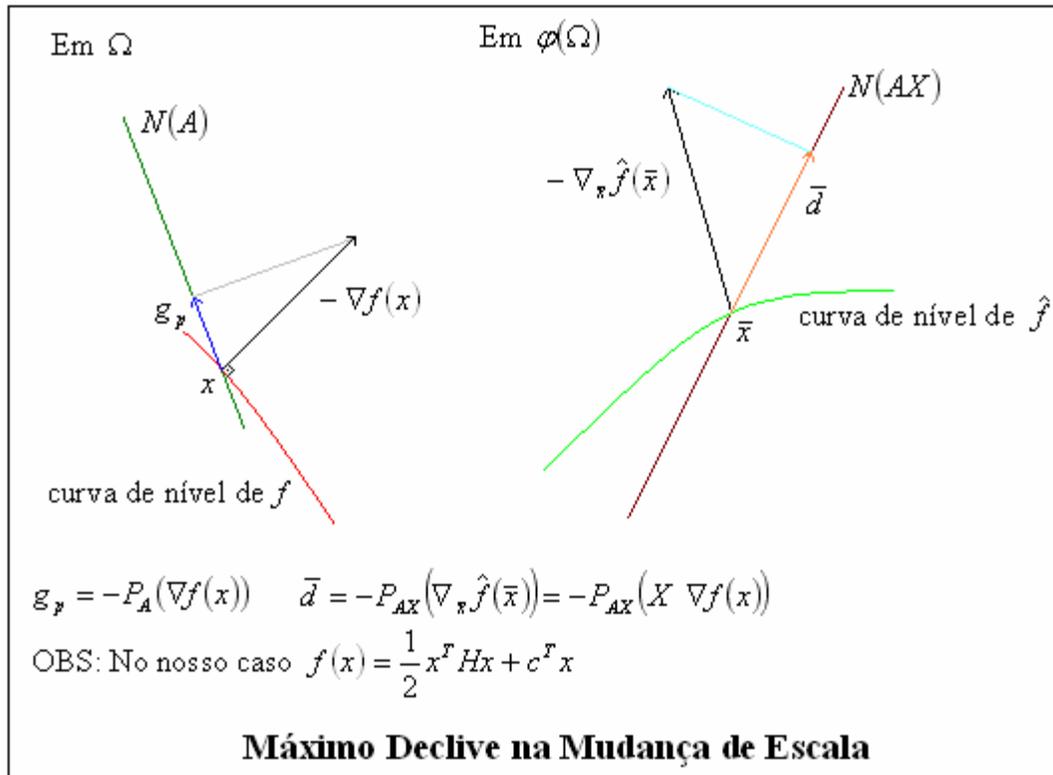


Figura 15

LEMA 4.5: Seja $\bar{s} = -\bar{d} + \mu e$, então $P_{AX}(\nabla_{\bar{x}} \hat{f}(\bar{x})) = P_{AX}(\bar{s})$.

Demonstração:

Sabemos que $P_{AX}(\mu e) = \mu P_{AX}(e) = 0$ já que e é centro analítico de $\varphi(\Omega)$. Sabemos também que para qualquer projeção P , $P(a) = P(P(a))$. Com isso temos que

$$\begin{aligned} P_{AX}(\nabla_{\bar{x}} \hat{f}(\bar{x})) &= P_{AX}(X \nabla f(x)) = P_{AX}(P_{AX}(X \nabla f(x)) + P_{AX}(\mu e)) = P_{AX}(P_{AX}(X \nabla f(x)) + \mu e) = \\ &P_{AX}(-\bar{d} + \mu e) = P_{AX}(\bar{s}) \end{aligned}$$

■

LEMA 4.6: Seja $s = X^{-1}(-\bar{d} + \mu e)$ então $P_A(\nabla f(x)) = P_A(s)$, ou seja, (x, s) forma um par primal-dual.

Demonstração:

Do Lema anterior temos que $P_{AX}\bar{s} = P_{AX}(P_{AX}(X \nabla f(x)) + \mu e)$, ou, de forma equivalente $\bar{s} = X \nabla f(x) + (AX)^T \bar{y} = X \nabla f(x) + XA^T \bar{y}$. Multiplicando por X^{-1} obtemos $s = X^{-1}\bar{s} = \nabla f(x) + A^T \bar{y}$. Pela definição de projeção conclui-se que $P_A(\nabla f(x)) = P_A(s)$.

■

Um par primal-dual: Dado o centro analítico x , mostramos agora como obter $s \in R^{2n}$ e $\mu > 0$ tal que $(x, s) \in \overset{\circ}{V}$ e $\delta(x, s, \mu) \leq \alpha$, para $\alpha \in (0, 1)$ dado.

Pela construção acima, fazemos

$$s = X^{-1}(-\bar{d} + \mu e),$$

e impomos as condições $s > 0$ e $\delta(x, s, \mu) \leq \alpha$, ou seja

$$\alpha > \left\| \frac{xs}{\mu} - e \right\| = \left\| \frac{-\bar{d} + \mu e}{\mu} - e \right\| = \left\| \frac{\bar{d}}{\mu} \right\|,$$

Em que $\|\cdot\|$ pode ser $\|\cdot\|_2$, $\|\cdot\|_\infty$ ou $\|\cdot\|_{-\infty}$.

Basta portanto escolher $\mu > 0$ tal que $(-\bar{d} + \mu e) > 0$ e $\|\bar{d}\| \leq \mu\alpha$. Por exemplo, usando a proximidade δ_∞ , basta fazer $\mu > \frac{\|\bar{d}\|_\infty}{\alpha}$. Esta construção é resumida no seguinte lema.

LEMA 4.7: Sejam x o centro analítico de Ω e $\bar{d} = -P_{AX} \nabla f(e)$. Então o par (x, s) com $s = X^{-1}(-\bar{d} + \mu e)$ é viável e satisfaz $\delta_\infty(x, s, \mu) \leq \alpha$ para qualquer $\mu > \frac{\|\bar{d}\|_\infty}{\alpha}$.

Temos portanto um procedimento para obter um par (x^0, s^0) para iniciar um método de trajetória central. A construção acima é descrita em Gonzaga e Cardia [11] para problemas de programação linear. Nesse artigo mostra-se também o seguinte:

Suponha que o centro analítico é o vetor e (sem perda de generalidade pois o caso geral é obtido com uma mudança de escala). Com a construção descrita acima obtemos

$$s = -\bar{d} + \mu e, \quad \delta(e, s, \mu) = \frac{\|\bar{d}\|}{\mu}.$$

Se tomarmos $\tilde{x} = e + \frac{1}{\mu} \bar{d}$, vemos que \tilde{x} é primal viável interior se $\mu > \|\bar{d}\|_\infty$. Em programação linear, o par (\tilde{x}, s) será primal-dual viável interior, e obtemos

$$\delta(\tilde{x}, s, \mu) = \left\| \frac{xs}{\mu} - e \right\| = \left\| \left(e + \frac{1}{\mu} \bar{d} \right) \left(e - \frac{1}{\mu} \bar{d} \right) - e \right\| = \frac{\|\bar{d}\|^2}{\mu^2}.$$

É fácil ver que $\|\bar{d}^2\| \leq \|\bar{d}\|^2$ (a igualdade vale para $\|\cdot\|_\infty$), e então

$$\delta(\tilde{x}, s, \mu) \leq (\delta(e, s, \mu))^2,$$

o que é excelente: o par (\tilde{x}, s, μ) está mais próximo de (x_μ, s_μ) do que (e, s) e portanto é um ponto melhor para iniciar um algoritmo de trajetória central. Por exemplo, tomando $\mu = \sqrt{2}\|\bar{d}\|_\infty$, temos $\delta_\infty(e, s, \mu) = \frac{1}{\sqrt{2}}$ e $\delta_\infty(\tilde{x}, s, \mu) = \frac{1}{2}$.

Tentamos a mesma construção para o problema quadrático, mas agora o par (\tilde{x}, s) não satisfaz em geral a primeira condição de KKT. Isto porque a primeira equação

$$-Hx + A^T y + s = c$$

envolve x e s (no caso linear, $H = 0$). O fato do par (\tilde{x}, s) não ser primal-dual viável implicou numa frustração na nossa dissertação.

Poderíamos ainda, provar que a direção afim-escala a partir do centro analítico é tangente à trajetória central primal, mas nesta dissertação, devido ao fato do ponto (x, s) , da proposição anterior, não ser primal-dual viável não iremos provar este fato, ficando ele apenas ilustrado na figura a seguir. Dessa maneira o melhor ponto inicial primal-dual que temos até agora é o ponto $(x, s) = \left(Xe, X^{-1} \left(e - \frac{\bar{d}}{\mu} \right) \right)$, com $\mu > \|\bar{d}\|_\infty$. Mesmo assim ainda iremos continuar na busca de um par primal-dual viável ainda melhor como foi feito acima para o caso linear.

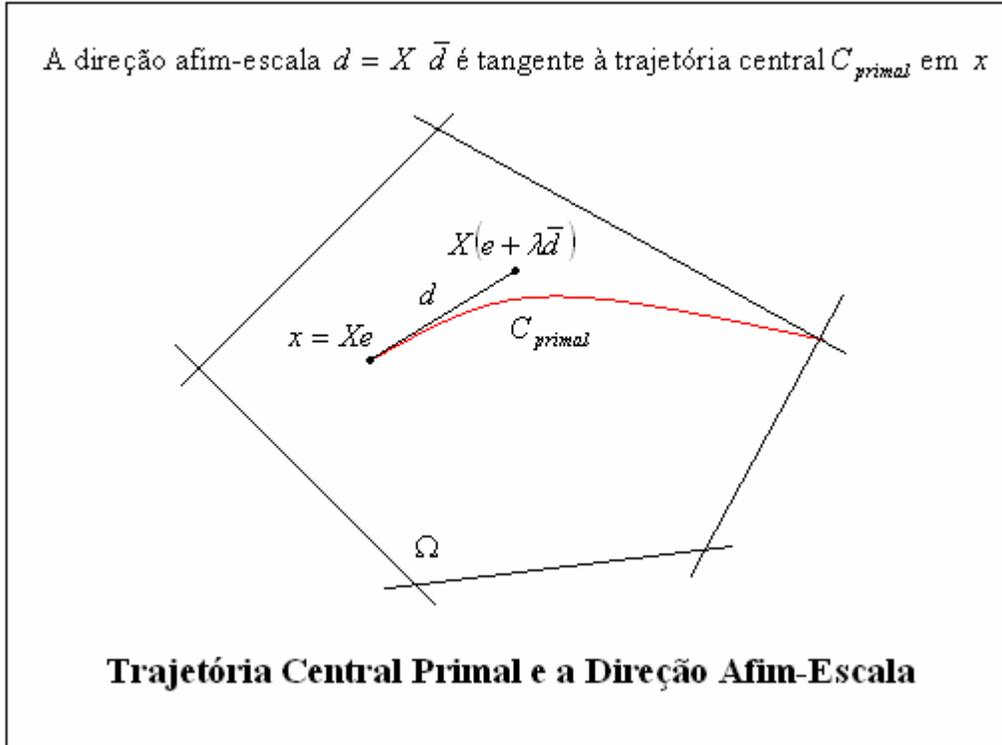


Figura 16

4.2.4 Um novo ponto primal-dual viável

Na seção 4.2.3 vimos como lidar em relação a uma mudança de escala. Encontramos também um par primal-dual viável, mas como já frisamos queremos um par primal-dual um pouco melhor.

Considere então o problema de minimizar uma função convexa quadrática $f : R^{2n} \rightarrow R$ restrito a $Ax = b$. Dados $x \in R^{2n}$ e $\Delta > 0$, fixos, definimos o problema de região de confiança associado a f numa bola $B_\Delta(x)$ por

$$\begin{aligned} &\text{Minimizar } f(x+h) && (4.12) \\ &\text{s.a. } Ah = 0 \\ &\quad \|h\|_2^2 \leq \Delta^2. \end{aligned}$$

Já tínhamos definido Trajetória ótima de região de confiança no capítulo 1 como sendo o conjunto $A = \{x + h \in \mathbb{R}^{2n}; h \text{ é solução de (4.12)}\}$, para algum $\Delta \geq 0$.

$$f(x+h) = f(x) + \nabla^T f(x)h + \frac{1}{2}h^T Hh.$$

Suponha agora que estamos interessados em abordar o problema de região de confiança em torno do ponto $x = e$ para um raio $\Delta > 0$ em termos de suas condições de KKT, as quais são

$$\begin{aligned} \nabla f(e+h) + A^T y + \mu h &= 0 \\ Ah &= 0 \\ \mu(\|h\|^2 - \Delta^2) &= 0. \end{aligned}$$

Ou equivalentemente,

$$\begin{aligned} (H + \mu I)h + A^T y &= -\nabla f(e) \\ Ah &= 0 \\ \mu(\|h\| - \Delta) &= 0. \end{aligned}$$

Veja que a primeira condição significa que $\mu h = -P(\nabla f(e+h))$. Observe então que se um par $(\bar{\mu}, \bar{h})$ satisfaz $\bar{\mu}\bar{h} = -P(\nabla f(e+h))$ e $A\bar{h} = 0$, \bar{h} é minimizador de (4.12) na bola $B_{\|\bar{h}\|}(e)$, ou seja, o ponto $w = e + \bar{h}$ está sobre a trajetória ótima de região de confiança, veja a figura.

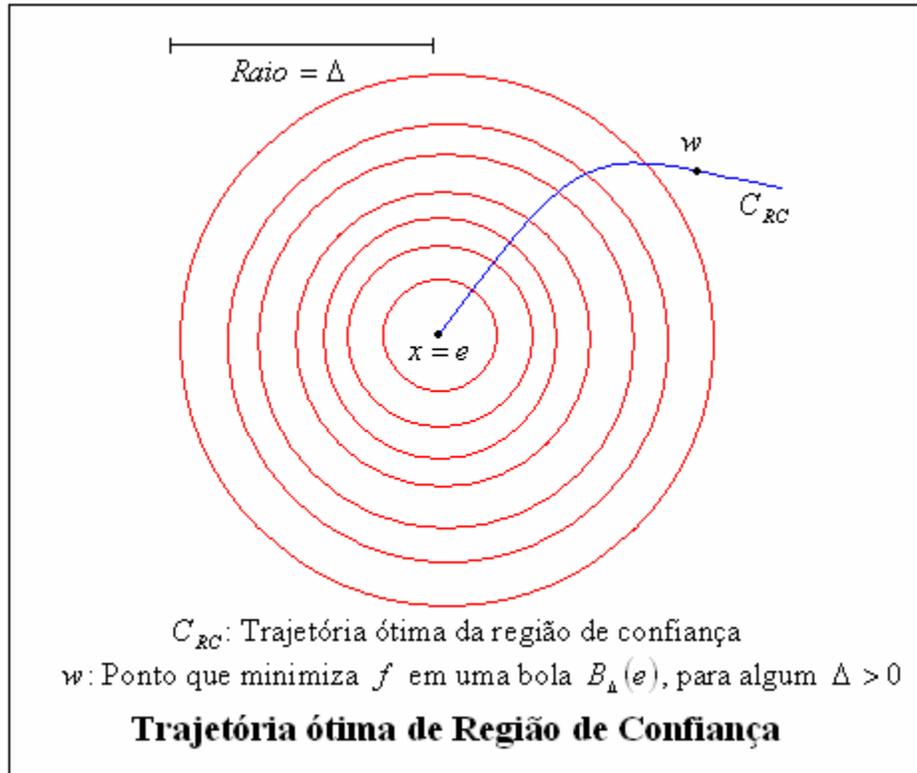


Figura 17

Temos então o seguinte:

- (i) $(e + h, s)$ é um par primal-dual se $P(s) = P(\nabla f(e + h))$ e $Ah = 0$
- (ii) $e + h_{\mu}$ pertence a trajetória ótima do problema de região de confiança se

$$\mu h_{\mu} = -P(\nabla f(e + h_{\mu}))$$

É importante salientar que o resultado a seguir é um resultado novo, desenvolvido por nós nesta dissertação.

TEOREMA 4.1: Seja h_{μ} uma direção tal que $Ah_{\mu} = 0$ e $\mu h_{\mu} = -P(\nabla f(e + h_{\mu}))$. Então $s = \mu(e - h_{\mu})$ é folga dual.

Demonstração:

$$\text{Temos } P(s) = \mu P(e) - \mu P(h_{\mu}) = -P(h_{\mu}) = P(\nabla f(e + h_{\mu})).$$

■

Para μ suficientemente grande, teremos $\|h_\mu\|_\infty < 1$ e, portanto, $s > 0$.

Agora podemos examinar o seguinte par primal-dual:

$$\begin{aligned}x &= e + h_\mu \\s &= \mu(e - h_\mu).\end{aligned}$$

$$\text{Obtemos } \delta(x, s, \mu) = \|(e + h_\mu)(e - h_\mu) - e\| = \|h_\mu^2\| \leq \|h_\mu\|^2.$$

Para que tenhamos então uma proximidade de, por exemplo $\frac{1}{4}$, basta encontrar h_μ tal que $\|h_\mu\|$ seja da ordem de $\frac{1}{2}$. Para isto vamos fazer um método de busca, como descrevemos a seguir, que já está adaptado para quando nosso ponto inicial não for $x = e$.

Algoritmo 4.1 Ponto primal-dual inicial

Dado o problema (4.3) com x centro analítico de Ω

Faça $X = \text{diag}(x)$ e $\mu_0 = 2\|\nabla f(e)\|$

Para $k = 0, 1, 2, 3, \dots$

Calcule com $\mu_k = 2^k \mu_0$ o ponto $h_k = h_{\mu_k}$ resolvendo o sistema

$$(H + \mu_k I)h_{\mu_k} + XA^T y = -X\nabla f(e)$$

$$AXh_{\mu_k} = 0$$

Pare quando $\|h_k\| < \frac{1}{2}$

Fim

Resultado $h_K = h_\mu$

Deste modo nosso ponto inicial primal-dual definitivo é $(X(e + h_\mu), X^{-1}\mu(e - h_\mu))$, com h_μ resultado do Algoritmo 4.1. A Figura 18 ilustra nossa escolha quanto ao ponto inicial primal.

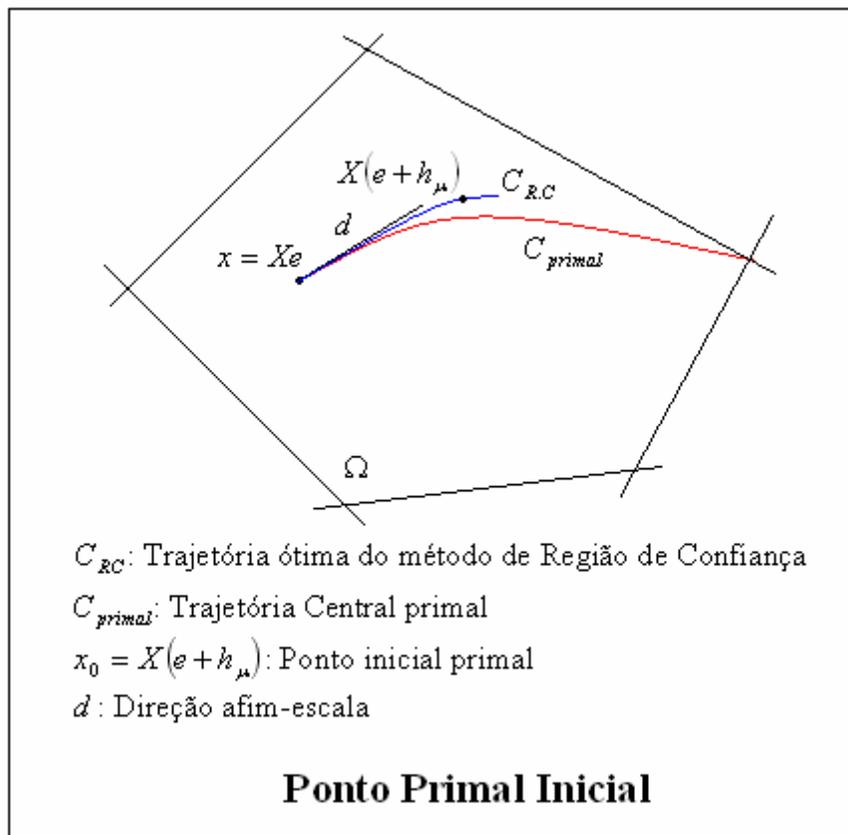


Figura 18

4.3 Método de Pontos Interiores de Trajetória

O método de pontos interiores que iremos utilizar será inicializado a partir de um ponto primal-dual $(x^k, s^k) \in \overset{\circ}{V}$. Faremos então o cálculo de um ponto de Newton em relação ao sistema de equações determinadas por KKT que veremos logo em frente. Esse ponto e o ponto (x^k, s^k) determinarão a direção a qual chamaremos de direção de Newton d^k . Caminharemos sobre esta direção de modo a não ultrapassar a fronteira da vizinhança $N_{-\infty}$ fazendo uma busca unidirecional do tipo Armijo, achando assim o ponto (x^{k+1}, s^{k+1}) . Um método tal qual é chamado de método de pontos interiores de trajetória central.

4.3.1 Cálculo do Passo de Newton

Considere o sistema a seguir em que $\sigma \in (0,1)$ é um valor fixo e $\mu > 0$ é dado.

$$\begin{aligned} -Hx + s + A^T y &= c \\ Ax &= b \\ xs &= \sigma\mu e \end{aligned} \tag{4.13}$$

Seja $F(\cdot)$ uma função tal que $F(x, s, y) = \begin{bmatrix} -Hx + s + A^T y - c \\ Ax - b \\ xs - \sigma\mu e \end{bmatrix}$, com $x, s \in \mathbb{R}^{2n}$ e $y \in \mathbb{R}^{2m}$.

Vamos dar um passo de Newton para a resolução de $F(x, s, y) = 0$, supondo que estamos saindo de um ponto (x, s, y) tal que $(x, s) \in \overset{\circ}{V}$ e tal que (x, s, y) satisfaz as primeiras duas condições de (4.13). Para isso precisamos resolver a equação

$$F'(x, s, y)(\Delta x, \Delta s, \Delta y) = -F(x, s, y). \tag{4.14}$$

Note que $F(x, s, y) = \begin{bmatrix} 0 \\ 0 \\ xs - \sigma\mu e \end{bmatrix}$, já que supomos (x, s, y) viável.

Temos então que

$$\begin{bmatrix} 0 \\ 0 \\ xs - \sigma\mu e \end{bmatrix} = \begin{bmatrix} -H\Delta x + \Delta s + A^T \Delta y \\ A\Delta x \\ \Delta x \quad s + x \quad \Delta s \end{bmatrix}.$$

Escrevendo de forma matricial obtemos

$$\begin{bmatrix} -H & I & A^T \\ A & 0 & 0 \\ S & X & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta s \\ \Delta y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ xs - \sigma\mu e \end{bmatrix}. \quad (4.15)$$

Falaremos em pares primais-duais (x, s) , sem mencionar as variáveis y . Para obter y basta resolver

$$-Hx + s + A^T y = c \quad (4.16)$$

que tem solução única porque $\text{rank}(A) = 2m$.

Concluimos que a trinca (x, s, y) primal-dual é perfeitamente caracterizada pelo par (x, s) , e podemos omitir a menção de y sempre que isto for conveniente.

Dado um par $(x^k, s^k) \in \overset{\circ}{V}$, associamos a ele o valor $\mu_k = \frac{(x^k)^T s^k}{2n}$, como vimos na seção 4.2.2. Um passo de Newton a partir de (x^k, s^k) como descrevemos acima visa aproximar-se do ponto central $(x_{\sigma\mu}, s_{\sigma\mu_k})$ associado a $\mu = \sigma\mu_k < \mu_k$. Com $\sigma = 1$, o passo é chamado de passo de centralização. Com $\sigma < 1$, procura-se evoluir ao longo da trajetória central.

A figura a seguir ilustra o que estamos desenvolvendo.

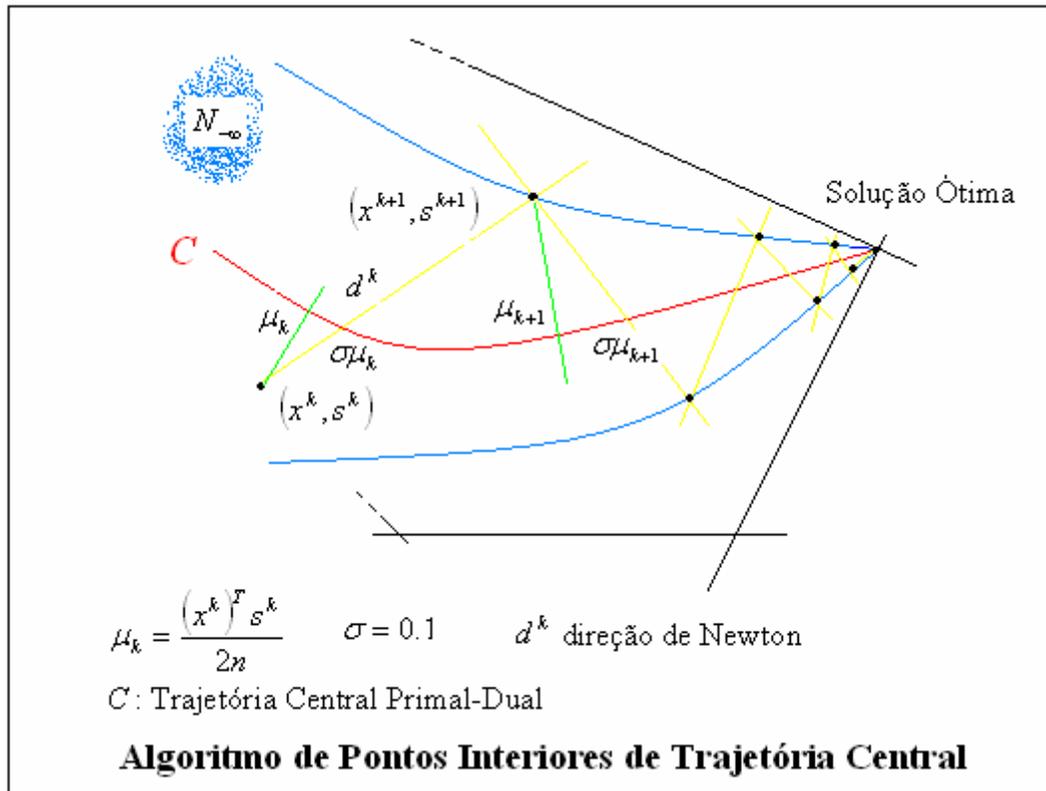


Figura 19

Em seguida expomos o algoritmo de pontos interiores de trajetória que utilizaremos para resolver o problema foco desta dissertação. Para um estudo extenso de algoritmos primais-duais de trajetória central, veja [25].

Algoritmo 4.2 Pontos Interiores

Dados: O problema (4.1), $\sigma = 0.1$ e $\alpha \in (0,1)$, $1-10^{-3}$ por exemplo

Reformule (4.1) para (4.2)

Ponto Inicial:

Dado o problema (4.2) calcule um ponto inicial (x^0, s^0) viável pelo algoritmo 4.1.

Calcule y^0 por (4.15).

Algoritmo Preditor Corretor:

Para $k = 0, 1, \dots, k \max$

Resolva o sistema (4.14) com $(x, s, y) = (x^k, s^k, y^k)$ para encontrar $(\Delta x^k, \Delta s^k, \Delta y^k)$.

Faça uma busca unidirecional tipo Armijo sobre $d^k = (\Delta x^k, \Delta s^k)$ a partir de (x^k, s^k) para caminhar no máximo até a fronteira $N_{-\infty}$, achando λ . Comece com $\lambda = 1$. Se $x^k + \lambda d^k$ estiver em $N_{-\infty}$, ótimo, senão faça $\lambda = \frac{1}{2}$ e assim sucessivamente até $x^k + \lambda d^k$ pertencer a $N_{-\infty}$. faça então

Faça $(x^{k+1}, s^{k+1}) = (x^k, s^k) + \lambda d^k$ e calcule y^{k+1} por $-H x^{k+1} + s^{k+1} + A^T y^{k+1} = c$.

Fim

Uma outra alternativa na construção do nosso algoritmo, proposta por Gonzaga [10], seria de resolver o sistema de Newton (4.14), a cada passo k , para $\mu = 0$ e para

$\mu_k = \frac{(x^k)^T s^k}{2n}$, sempre com $\sigma = 1$. Os pontos determinados por esta resolução são

chamados respectivamente de ponto de Newton completo P_{Newton} e passo de centralização

$P_{Centralização}$. A proposta se resume a fazer uma busca em θ (um parâmetro real) para que

encontremos um ponto P_θ sobre o segmento que une P_{Newton} com $P_{Centralização}$ de tal maneira

que P_θ esteja a uma certa proximidade da trajetória central primal-dual, definida por $N_{-\infty}$ por exemplo. Desse modo teríamos então que P_θ seria o ponto (x^{k+1}, s^{k+1}) . Permanece então essa alternativa que está ilustrada na figura a seguir

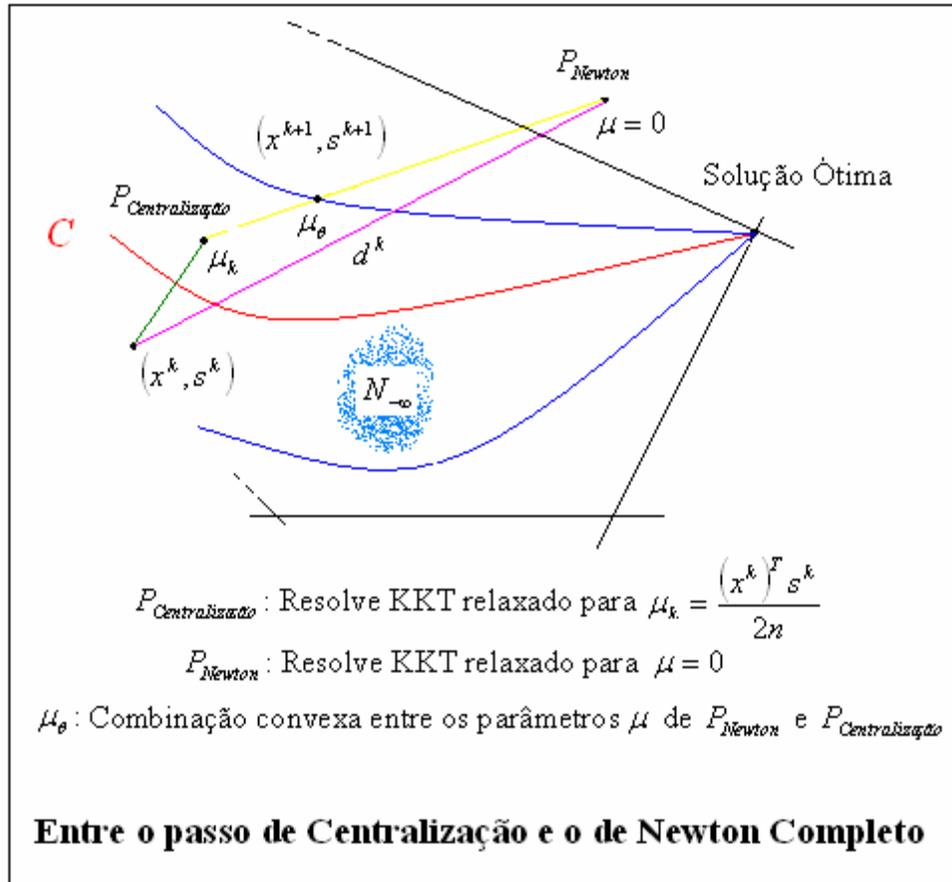


Figura 20

4.3.2 Passo Tangente do PQS numa caixa

O maior objetivo desta dissertao j foi alcanado, que consistia em dar base para a resoluo de (4.1). Comentamos que uma das motivaes que nos levou abordar o problema (4.1)  o fato dele ser um subproblema de PQS. Visaremos agora relacionar os

problemas (2.11) (passo tangente do PQS) numa caixa, e (4.1) para podermos utilizar o algoritmo 4.1 na resolução de (2.11) numa caixa.

Em (2.11) queríamos

$$\begin{aligned} &\text{minimizar } \nabla^T f_0(\bar{x})h + \frac{1}{2}h^T W(\bar{x}, \bar{\lambda})h \\ &\text{s.a. } A(\bar{x})h = 0. \end{aligned}$$

Mas isto numa bola $\|h\|_\infty \leq \Delta$. Ou seja, temos um problema do tipo

$$\begin{aligned} &\text{minimizar } \frac{1}{2}h^T Hh + c^T h && (4.17) \\ &\text{s.a. } Ah = 0 \\ &\quad \|h\|_\infty \leq \Delta. \end{aligned}$$

Veja que $\|h\|_\infty \leq \Delta$ é equivalente a $-\Delta e \leq h \leq \Delta e$. Faça então $x = h + \delta$ em que $\delta = \Delta e = [\Delta \ \Delta \ \dots \ \Delta]^T \in R^n$. Logo (4.17) é equivalente a

$$\begin{aligned} &\text{minimizar } \frac{1}{2}(x - \delta)^T H(x - \delta) + c^T (x - \delta) \\ &\text{s.a. } Ax = A\delta \\ &\quad 0 \leq x \leq 2\delta. \end{aligned}$$

Que é equivalente a

$$\begin{aligned} &\text{minimizar } \frac{1}{2}x^T Hx + c^T x && (4.18) \\ &\text{s.a. } Ax = A\delta \\ &\quad 0 \leq x \leq 2\delta. \end{aligned}$$

Veja que com isso temos um problema como em (4.1), já que temos também que $x = \delta$ é centro analítico da caixa $[0, 2\delta]$ e satisfaz $Ax = A\delta$. Para resolver (4.18) basta então aplicar

o algoritmo de pontos interiores 4.2 (O Algoritmo 4.2 utiliza o Algoritmo 4.1. Neste, utilize $x = \delta$ para encontrar um par primal-dual viável).

Conclusão

Esta dissertação garantiu um contato positivo com várias ferramentas de Programação não Linear restrita e irrestrita, tendo firmado vários conceitos supostamente já assimilados. O presente trabalho também conciliou uma interação de assuntos, que até pouco tempo atrás me pareciam desconexos. Neste sentido vejo que o estudo feito nos quatro capítulos desta dissertação deixa essa interação bem clara.

Lembre que um dos principais motivos que nos levou a abordar o problema de minimização em caixas foi o artigo de Ferris, Wathen e Armand [5]. Neste contexto creio termos conseguido suprir um dos nossos objetivos que se resumia a propor a obtenção de um ponto inicial com propriedades interessantes dentro da problemática. Nesse sentido os resultados teóricos obtidos foram melhores do que o esperado, já que estes culminaram com uma descoberta do professor Dr. Clóvis Caesar Gonzaga em relação à procura de um par primal-dual viável inicial para o problema de minimização de uma quadrática convexa numa caixa sobre uma variedade afim. Isto gerou uma motivação que sem dúvida me instiga a querer crescer na área de otimização através de um doutorado.

Já o que se pode dizer em relação a conclusões mediante testes numéricos é que seria necessário um pouco mais de tempo na sua abordagem já que somente obtivemos resultados teóricos que nos interessavam no começo de maio de 2006. Entretanto, foi feita uma programação em Matlab que, no entanto, não fez parte desta dissertação. Mas esperamos dar continuidade ao estudo da problemática proposta mediante testes numéricos e, se conseguirmos, teoricamente, por meio do enfraquecimento de algumas hipóteses que havíamos assumido.

Referências Bibliográficas

[1] BENTERKI, D.; CROUZEIX, J. P. and MERIKHI, B. A numerical feasible interior point method for linear semidefinite programs, 2005.

[2] BERTSEKAS, D.P. *Nonlinear Programming*. Athena Scientific, Belmont, Massachusetts, 1995.

[3] BONNANS, J. F.; GILBERT, J. C.; LEMARÉCHAL, C.; SAGATIZÁBAL, C. A. *Numerical Optimization: Theoretical and Practical Aspects*. Springer Verlag, Berlin, 2002.

[4] CONN, A. R.; GOULD N. I. M.; TOINT P. L. *Trust-Region Methods*. MPS-SIAM Series on Optimization, SIAM, Philadelphia, 2000.

[5] FERRIS, M. C.; WATHEN A. J. and ARMAND P. *Limited Memory Solution of Complementarity Problems arising in Video Games*, Oxford University, 2005.

[6] FLETCHER, R. and LEYFFER, S. Nonlinear programming without a penalty function. *Mathematical Programming - Ser. A*, 91(2):239-269, 2002.

[7] FRANCISCO, J. B. *Viabilidade em Programação não Linear*. Tese (Doutorado), UNICAMP, 2005.

[8] GONZAGA, C. C. Notas de Aula de PNL. Universidade Federal de Santa Catarina, 2005.

[9] GONZAGA, C. C. *Path-Following Methods for Linear Programming*, COPPE, Federal University of Rio de Janeiro, 1991.

- [10] GONZAGA, C. C. *The largest step path following algorithm for monotone linear complementarity problems*. *Mathematical Programming* 76(1997) 309-332.
- [11] GONZAGA, C.C.; CARDIA, M. *Properties of the Central Point in Linear Programming problems*. *Numerical Algorithms*, 35 p. 185-204, 2004.
- [12] GONZAGA, C. C.; KARAS, E. W. and VANTI M. A globally convergent filter method for nonlinear programming. *SIAM J. Optimization*, 14(3):646-669, 2003.
- [13] KARAS, E. *Exemplos de Trajetória Central mal comportada em Otimização Convexa e um algoritmo de Filtros para Programação não Linear*. Universidade Federal de Santa Catarina e Universidade de Paris I-Panthéon-Sorbonne. Tese (Doutorado), 2002.
- [14] MARTINEZ, J. M. Box-quacan and the implementation of augmented lagrangian algorithms for minimization with inequality constrains. *Computational & Applied Mathematics*, v. 19, p. 31-56, 2000.
- [15] MARTÍNEZ, J. M. Inexact-restoration method with Lagrangian tangent decrease and a new merit function for nonlinear programming. *Journal of Optimization Theory and Applications*, 111: 39-58, 2001.
- [16] MARTÍNEZ, J. M. and PILOTTA, E. A. Inexact restoration algorithm for constrained optimization. *Journal of Optimization Theory and Applications*, 104:135-163, 2000.
- [17] MARTÍNEZ, J. M.; SANTOS, S. A. *Métodos Computacionais de Otimização*, IMPA, 1995.
- [18] MEHROTA, S.; SUN, J. On the implementation of a (primal-dual) interior point method. *SIAM Journal on Optimization*, v. 2, n. 4, p. 575-601, 1992.

- [19] MIZUNO, S.; TODD, M. J.; YE, Y. On adaptive step primal-dual interior-point algorithms for linear programming. *Mathematics of Operations Research*, v. 18, n. 4, p. 964-981, 1993.
- [20] MORÉ, J. J.; SORESENSEN, D. C. Computing a trust region step. *SIAM Journal on Sci. Statist. Comput.*, v.4, p. 553-572, 1983.
- [21] NOCEDAL, J. and WRIGHT, S. J. *Numerical Optimization*. Springer Series in Operations Research. Springer – Verlag, 1999.
- [22] OMOJOKUN, E. *Trust Region Algorithms for Optimization with Nonlinear Equality and Inequality Constraints*. Tese (Doutorado) – Dept. of Computer Science, University of Colorado, 1991.
- [23] VANDERBEI, R. J. *Linear Programming, Foundations and Extensions*, Kluwer Academic Publishers, 1996.
- [24] VANTI, M. R. *Melhoria da Segurança Dinâmica por Otimização e Algoritmos de Programação Não-Linear*, Universidade Federal de Santa Catarina, Programa de Pós-Graduação em Engenharia Elétrica, Florianópolis, 2003.
- [25] WRIGHT, S. J. *Primal-Dual Interior-Point Methods*. Philadelphia: SIAM Publication, 1997.