

UNIVERSIDADE FEDERAL DE SANTA CATARINA
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA E
COMPUTAÇÃO CIENTÍFICA

**Um Algoritmo de Regiões de Confiança
Baseado em Trajetória Central e sua
Utilização em Método de Filtro**

ANA MARIA BASEI
FLORIANÓPOLIS, ABRIL DE 2007

UNIVERSIDADE FEDERAL DE SANTA CATARINA
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA E
COMPUTAÇÃO CIENTÍFICA

Um Algoritmo de Regiões de Confiança Baseado em Trajetória Central e sua Utilização em Método de Filtro

Dissertação apresentada ao Curso de Pós Graduação em Matemática e Computação Científica, do Centro de Ciências Físicas e Matemáticas da Universidade Federal de Santa Catarina, para obtenção do grau de Mestre em Matemática, com Área de Concentração em Matemática Aplicada.

Ana Maria Basei

Florianópolis

2007

**Um algoritmo de regiões de confiança
baseado em trajetória central e sua
utilização em métodos de filtro**

por

Ana Maria Basei

Esta Dissertação foi julgada para a obtenção do Título de “Mestre”,
Área de Concentração em Matemática Aplicada, e aprovada em sua forma
final pelo Curso de Pós-Graduação em Matemática e
Computação Científica.

Clóvis Caesar Gonzaga
Coordenador

Comissão Examinadora

Prof. Dr. Clóvis Caesar Gonzaga (UFSC-Orientador)

Prof. Dr. Carlos Humes Junior (USP)

Prof. Dr. Mário César Zambaldi (UFSC)

Prof. Dr. Juliano de Bem Francisco (UFSC)

Florianópolis, abril de 2007.

*A Deus,
à meus amados pais, Fernando e Zelinda,
à minha querida irmã Fernanda
e ao meu amor Paulo.*

Agradecimentos

A Deus pela vida, saúde, paz, proteção, e principalmente pela família maravilhosa, amigos e pessoas queridas que coloca em minha vida.

Aos meus amados pais pelo amor verdadeiro e incondicional. Agradeço ao carinho, zelo e apoio que os torna meu porto seguro. Pelas longas conversas, conselhos e palavras de confiança e motivação. Ao meu pai agradeço em especial ao maior exemplo que deixou de persistência, otimismo e coragem. A minha mãe agradeço ao exemplo de luta, trabalho, perseverança e coragem.

A minha querida irmã e amiga Fernanda, pelo amor, pelos momentos alegres e divertidos que passamos juntas e ainda pela sua presença nos momentos de dificuldade.

Ao meu amor Paulo H. Bermejo pelo carinho, cuidados, pelos momentos de alegria e apoio nas horas difíceis e principalmente por ser meu maior incentivador, sempre acreditando em mim e fazendo parte da minha vida.

As minhas tias Zaira e Adelaide pela torcida e por se fazerem presentes em momentos importantes e aos demais familiares e amigos queridos.

Aos colegas do mestrado, em especial a amizade de Diane, Fabiana, João, Luciane e Priscila, pelos momentos agradáveis que compartilhamos e pelo companheirismo nos momentos difíceis.

Aos colegas Ismael, Gilberto e Rafael pelo apoio e companheirismo.

Agradeço aos professores membros da banca por dedicarem seu tempo a avaliação do trabalho contribuindo no aprimoramento da versão final.

Agradeço ao professor Igor e demais professores do mestrado pelo conhecimento transmitido. Ao professor Gustavo A.T.F. da Costa pela orientação no estágio de docência.

À Elisa Amaral pela atenção, dedicação e eficiência com que sempre nos atende na secretaria da pós graduação.

Ao CNPQ pelo suporte financeiro.

E finalmente agradeço em especial ao professor Clóvis Gonzaga pela sua orientação e pela forma paciente, sábia e carinhosa como me atendeu durante a realização do trabalho. Acredito que sua inteligência, sua paixão pela matemática, a forma como ensina e conduz suas aulas, além do seu coração enorme e humildade são algumas das suas virtudes que despertam admiração em todos alunos que tem a oportunidade de trabalhar com ele.

Resumo

A utilização do método de região de confiança é muito frequente em problemas de otimização, já que precisamos minimizar modelos de funções. Temos métodos que utilizam regiões com formato de bolas euclidianas e métodos que utilizam regiões com formato de caixas. Na presença de restrições lineares, a utilização de um ou outro pode apresentar desvantagens. Neste trabalho apresentamos um algoritmo de região de confiança baseado em trajetória central. É um método intermediário entre os citados acima. No método apresentado as regiões de confiança são conjuntos de nível da função barreira logarítmica, e evoluem continuamente de pequenas bolas euclidianas tendendo à caixa completa. Este algoritmo é utilizado em seguida em um algoritmo de filtro para resolver um problema de otimização com formato bem geral de restrições de igualdade e variáveis não negativas.

Abstract

Trust region methods are frequently used in Optimization, whenever we need to minimize function models. We have methods based on regions shaped as Euclidean balls and methods based on box shaped regions. In the presence of linear equality constraints each of these shapes has its disadvantages. In this work we present a trust region method based on the central path. This method is in a way intermediate between the other two. Its trust regions are level sets of the logarithmic barrier function, and evolve continuously from small Euclidean balls towards the whole box. This algorithm is then used in a filter method for solving general optimization problems with equality constraints and non-negative variables.

Sumário

Lista de Figuras	x
Introdução	1
1 Introdução:	1
2 Região de Confiança	6
2.1 Algoritmos com Região de Confiança	7
2.2 Região de confiança e a busca de Armijo	10
2.2.1 Busca de Armijo	11
2.3 Convergência global	12
2.4 O problema quadrático e as duas trajetórias: Trajetória de Levenberg- Marquardt e a Trajetória Central	14
2.5 Trajetória Ótima para regiões tipo caixa	19
2.6 Comparação entre os métodos	23
2.7 Generalização	26
2.8 Algoritmos de regiões de confiança definidas pela barreira	27
3 Métodos de Pontos Interiores Primais-Duais	31
3.1 Revisão de conceitos	31
3.1.1 Trajetória Primal-Dual	32
3.1.2 Proximidade e Vizinhanças da Trajetória Central	34
3.2 Algoritmos:	36
3.3 Algoritmo de passo mais longo	38

4	Programação Quadrática Sequencial e Filtros	42
4.1	Programação Não Linear com Restrições de Igualdade e Variáveis não Negativas	42
4.2	Programação Quadrática Sequencial, Restauração Inexata e um Método de Filtro	45
4.2.1	A Fase de Restauração ou Fase de Viabilidade	50
4.2.2	Fase de Otimalidade	51
5	Utilização do Algoritmo de Região de Confiança no Método de Filtro	55
5.1	Fase de otimalidade	56
5.2	O algoritmo	56
5.2.1	O centro analítico do conjunto formado pelas restrições	57
5.2.2	Formato primal de programação quadrática	61
	Considerações Finais	65
	Referências	67

Lista de Figuras

2.1	Ponto de Cauchy.	13
2.2	Ponto de Cauchy	13
2.3	Regiões de confiança e trajetórias.	26
2.4	Regiões de confiança e trajetórias no caso geral.	28
2.5	Uma iteração do algoritmo	29
3.1	Comparação das vizinhanças da trajetória central	36
3.2	Passo mais longo	39
4.1	Construção do par (f^k, h^k)	46
4.2	Uma iteração do método de filtro	47
4.3	Uma iteração f e o filtro para a próxima iteração	49
4.4	Uma iteração h e o filtro para a próxima iteração	50
5.1	Construção da caixa $[l, u]$	60
5.2	Minimização em uma caixa maior	64

Capítulo 1

Introdução:

Considere o seguinte problema de programação não linear:

$$\begin{aligned} & \text{minimizar} && f(x) \\ & \text{sujeito a} && c(x) = 0 \\ & && x \geq 0 \end{aligned} \tag{1.1}$$

em que $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $c : \mathbb{R}^n \rightarrow \mathbb{R}^m$ são funções de classe \mathcal{C}^2 com $m \leq n$. Esse formato é bastante geral, uma vez que restrições de desigualdades podem ser reduzidas a igualdades com a introdução de variáveis de folga.

Na resolução de um problema restrito como o acima dois objetivos devem ser alcançados: otimalidade e viabilidade. Estes dois normalmente são conflitantes e podem ser tratados de forma independente, é o caso dos métodos de restauração inexata. Neste métodos, introduzidos por Martínez e Pilotta [21] uma iteração é composta por duas fases, chamadas fase de restauração e fase de otimalidade. Na fase de restauração, a partir de um determinado ponto, possivelmente não-viável, é determinado um ponto intermediário z tal que a inviabilidade tenha sido “suficientemente” reduzida. Na fase de otimalidade é feita uma minimização aproximada de uma função quadrática (que geralmente é um modelo quadrático da função objetivo ou de um Lagrangiano) sujeita a uma variedade linear e a uma região de confiança de raio Δ centrada em z . Quando for obtido um ponto que reduza $f(\cdot)$, deve-se decidir se o ponto melhora a otimalidade sem piorar muito a viabilidade. Para esta análise pode-se usar uma função de mérito ou filtros.

Neste trabalho vamos supor que o problema (1.1) seja resolvido com o algoritmo de filtro proposto em [13]. É um algoritmo globalmente convergente que usa as idéias de restauração inexata. Cada iteração é composta pelas duas fases mencionadas. Supondo que a fase de restauração esteja resolvida vamos estudar a fase de otimalidade. Portanto, o problema a ser resolvido é o de minimização de uma função quadrática restrita a uma variedade afim, uma região de confiança, ao primeiro ortante e a restrição do filtro. Vamos supor ainda que a função quadrática a ser minimizada é convexa.

A região de confiança pode ter formato de uma bola euclidiana centrada do ponto z (ponto de restauração) ou de uma caixa com centro geométrico z , dependendo da norma utilizada. Um método que usa bolas euclidianas é eficiente e convergente, e a presença de restrições lineares afeta pouco seu custo computacional. Entretanto usa regiões pequenas e é pouco aplicável a problemas com restrições. Já os métodos que usam caixas são mais atraentes, pois elas são bem maiores do que bolas euclidianas e problemas com restrições de caixas são muito comuns. No entanto, a presença de restrições lineares eleva seu custo computacional, porque cada iteração deve resolver um problema de programação quadrática geral.

No capítulo 2, no contexto de otimização irrestrita, apresentamos um método intermediário, com regiões de confiança que evoluem continuamente de pequenas bolas e tendem a uma caixa completa. As regiões são definidas pela função barreira logarítmica e a minimização da quadrática é feita através de uma busca curvilínea tipo Armijo sobre a trajetória central da caixa.

Para fazer esta busca usamos um método de pontos interiores. Um pouco da teoria de métodos de pontos interiores é apresentada no capítulo 3, com destaque para o algoritmo do passo mais longo, que foi o algoritmo escolhido para seguir a trajetória central. Ainda neste capítulo, apresentamos o teorema elaborado na tese de Behling [1] que fornece o primeiro par primal-dual viável para iniciar o algoritmo.

O capítulo 4, é dedicado ao problema (1.1). Apresentamos hipóteses e definições gerais necessárias no estudo de (1.1) além de uma breve descrição do método de filtro. Em seguida formulamos o problema onde aplicaremos o algoritmo de região de confiança proposto no capítulo 2.

Finalmente no capítulo 5 apresentamos o algoritmo utilizado na fase de otimalidade do método de filtro. Ele é composto de duas etapas: a construção da caixa, que deve ser a interseção da região onde o modelo é confiável e o primeiro ortante. A outra etapa corresponde a busca de Armijo que avança na trajetória até que uma das restrições seja violada: a confiabilidade no modelo usado, a positividade ou a restrição do filtro.

Notações

Apresentamos aqui algumas notações usadas no trabalho:

Os vetores colunas e as matrizes serão denotados respectivamente por letras minúsculas e maiúsculas. Os índices inferiores dos vetores representam as componentes do vetor. A letra e denota o vetor coluna cujas componentes são iguais a um com dimensão explicitada no contexto.

Sejam os vetores $x, s \in \mathbb{R}^n$ as notações $x^{-1}, x \cdot s, xs$ e $\frac{x}{s}$ são usadas para representar vetores com componentes $x_i^{-1}, x_i s_i$ e $\frac{x_i}{s_i}, i = 1, \dots, n$.

Quando v e w são vetores, a notação $v \leq w$ significará sempre $v_i \leq w_i$ para todas as suas coordenadas. Quando falamos da caixa $l \leq x \leq u$, entendemos o conjunto dos $x \in \mathbb{R}^n$ tais que $l_i \leq x_i \leq u_i$.

Dado um vetor x , a letra maiúscula X representa uma matriz diagonal formada pelas componentes do vetor x .

Para uma referência futura, aqui está uma lista dos principais símbolos utilizados no texto:

$\|\cdot\|$: denota uma norma qualquer em \mathbb{R}^n . $N(A)$: núcleo de uma matriz A .

Δ : raio da região de confiança

$D_\Delta = \{d \in \mathbb{R}^n \mid \|d\| \leq \Delta\}$: região de confiança

$\mathcal{B}_\Delta = \{d \in \mathbb{R}^n \mid \|d\| \leq \Delta\}$: região de confiança com a norma 2

$q_\mu(\cdot)$: soma de duas funções quadráticas: a função quadrática objetivo e a função $\frac{\|x^2\|}{2}$

$\mathcal{V} = \{x \in \mathbb{R}^n \mid -e \leq x \leq e\}$

$\mathcal{V}^\circ = \{x \in \mathbb{R}^n \mid -e < x < e\}$

$p(\cdot)$: função barreira logarítmica

$p_\mu(\cdot)$: soma de duas funções quadráticas: a função quadrática objetivo e a função

$p(\cdot)$

$\mathcal{V}_{\mathcal{K}} = \{x \in \mathcal{V}^\circ \mid p(x) \leq \mathcal{K}\}$

$\mathcal{V}^v := \{x \in \mathbb{R}^n \mid -v \leq x \leq v\}$

$\overset{\circ}{\mathcal{V}}^v := \{x \in \mathbb{R}^n \mid -v < x < v\}$

$p_v(\cdot)$: função barreira logarítmica definida em $\overset{\circ}{\mathcal{V}}^v$

$\mathcal{B}_\Delta^v := \{x \in \mathbb{R}^n \mid \|V^{-1}x\| \leq \Delta\}$

$$\mathcal{V}_{\mathcal{K}}^v = \{x \in \mathcal{V}^v \mid p_v(x) - p_v(0) \leq \mathcal{K}\}$$

c : gradiente da função f aplicado em um ponto x

$A(x)$: Jacobiana da função $g(\cdot)$

$\Gamma = \{(x, s) \in \mathbb{R}_+^{2n} \mid Ax = b, -Hx + s + A^T y = c, \text{ para algum } y \in \mathbb{R}^m\}$: conjunto dos pares primais duais viáveis

$\Gamma^\circ = \{(x, s) \in \Gamma \mid x, s > 0\}$: o conjunto de pares primais duais interiores

$$[l, u] = \{x \in \mathbb{R}^n \mid l \leq x \leq u\}$$

z : ponto de restauração

Região proibida em \mathbb{R}^2 , $F_k = \{(f, h) \mid f \geq f^i, h \geq h^i\}$

Região proibida em \mathbb{R}^n , $\mathcal{F}_k = \{x \in \mathbb{R}^n \mid f(x) \geq f^i, h(x) \geq h^i\}$

Filtro temporário, \bar{F}_k

Conjunto de pontos proibidos em \mathbb{R}^n associado ao filtro temporário, $\tilde{\mathcal{F}}_k$

Capítulo 2

Região de Confiança

As primeiras seções deste capítulo tratam do método de região de confiança aplicado a um problema de minimização irrestrita, que é a abordagem clássica deste método. Nas seções seguintes introduzimos o problema quadrático que deve ser resolvido a cada iteração do algoritmo de região de confiança. Inicialmente estudamos o caso em que a região de confiança é a bola euclidiana e a partir daí apresentamos a trajetória de Levenberg-Marquardt. Em seguida estudamos o caso da restrição ser uma caixa. Fixamos o tamanho desta caixa e definimos as regiões de confiança como conjuntos de nível da função barreira logarítmica associada a esta caixa. Definimos a trajetória central na caixa como a trajetória ótima dos minimizadores da quadrática restrita aos conjuntos de nível da barreira. Mostramos algumas relações entre as diferentes regiões de confiança e na última seção enunciamos um algoritmo de minimização conceitual para problemas irrestritos. O método resultante baseia-se em algoritmos de trajetória central para programação quadrática, que não serão estudados neste capítulo: estudaremos algoritmos primais-duais, descritos no capítulo 3 em um contexto mais geral.

Na abordagem clássica de região de confiança o problema a ser resolvido é o de minimização irrestrita de uma função , ou seja,

$$\begin{aligned} &\text{minimizar} && f(x) \\ &\text{sujeito a} && x \in \mathbb{R}^n. \end{aligned} \tag{2.1}$$

com $f : \mathbb{R}^n \rightarrow \mathbb{R}$ duplamente continuamente diferenciável.

Um método de região de confiança é um processo iterativo que gera uma sequência $(x^k)_{k \in \mathbb{N}}$ cujos pontos de acumulação devem ser minimizadores de f . Cada iteração k define uma região ao redor do ponto x^k onde o modelo quadrático seja uma representação adequada da função objetivo e em seguida minimiza o modelo nesta região. Se a solução \bar{d} obtida não é aceita por determinado critério, o tamanho da região é reduzido e é feita nova minimização.

A idéia de construir esta região de confiança no modelo quadrático foi proposta por Levenberg em 1944 [18] e Marquardt em 1963 [19] mediante estudos de problemas de mínimos quadrados. Em 1966, a aplicação do método de região de confiança foi sugerida por Goldef, Quandt e Trotter [33]. Powell [27] foi o primeiro a estabelecer a convergência quadrática dos métodos de região de confiança.

2.1 Algoritmos com Região de Confiança

Considere $k \in \mathbb{N}$ e $x^k \in \mathbb{R}^n$ o ponto corrente. O modelo quadrático da variação da função objetivo em torno de x^k é dado por:

$$d \in \mathbb{R}^n \mapsto m_k(d) = \nabla f(x^k)^T d + \frac{1}{2} d^T B_k d,$$

em que B_k uma matriz que pode ser uma aproximação da Hessiana $\nabla^2 f(x^k)$ ou qualquer outra matriz simétrica que satisfaça

$$\|B_k\| \leq \beta$$

para algum $\beta > 0$ independente de k . Como

$$f(x^k + d) = f(x^k) + \nabla f(x^k)^T d + \frac{1}{2} d^T \nabla^2 f(x^k + td) d,$$

para algum $t \in [0, 1]$, e como $m_k(d) = \nabla f(x^k)^T d + O(\|d\|^2)$, a diferença entre $m_k(d)$ e $f(x^k + d) - f(x^k)$ é limitada por $O(\|d\|^2)$, e então o erro da aproximação é pequeno quando d é pequeno. Na prática, a abordagem de região confiança é muito geral pois precisamos exigir muito pouco de B_k , apenas simetria e limitação uniforme no

índice k .

No caso em que $B_k = \nabla^2 f(x^k)$ o algoritmo obtido é chamado método de Newton para região de confiança.

Como o modelo quadrático só aproxima bem a função em uma vizinhança do ponto x^k define-se o conjunto D_Δ chamado região de confiança:

$$D_\Delta = \{d \in \mathbb{R}^n \mid \|d\| \leq \Delta\},$$

com Δ o raio da bola e $\|\cdot\|$ uma norma qualquer em \mathbb{R}^n .

Desta forma a cada iteração k do método de região de confiança formulamos o sub-problema:

$$\begin{aligned} & \text{minimizar } m_k(d) \\ & \text{sujeito a } \|d\| \leq \Delta. \end{aligned} \tag{2.2}$$

Este problema tem solução única, sempre que não ocorra o “caso difícil”, descrito em Nocedal [25] e comentado a seguir. Supondo que a solução de (2.2) seja única, esta é o passo de região de confiança e será denotada por $d(x^k, \Delta)$, com dependência de Δ já que ao alterarmos o raio Δ a direção do passo é alterada. O ponto $\bar{x} = x^k + d(x^k, \Delta)$ seria uma boa aproximação do minimizador de f na região D_Δ .

Além do seu tamanho outra característica importante da região é seu formato que depende da norma utilizada. Ao usarmos a norma ℓ_∞ , D_Δ é uma caixa de arestas de comprimento 2Δ . O formato de caixa se destaca em relação à bola euclidiana pelo seu tamanho em relação a ela. Além disso, em problemas restritos o formato da região de confiança pode ser definido de maneira a adaptar-se às restrições do problema. Assim em problema com restrições do tipo caixa pode-se usar a norma ℓ_∞ .

No entanto a minimização na caixa é mais difícil do que a minimização na bola. Métodos de região de confiança do tipo caixa foram estudados e implementados pela equipe de Otimização da UNICAMP no Brasil, coordenada por J. M. Martínez [23].

Neste trabalho pretendemos estudar a minimização em caixas sujeita ainda a uma variedade linear, o que torna o problema mais complexo; um problema geral de programação quadrática com restrições não lineares.

Para simplificar, iniciamos o estudo com regiões esféricas e cúbicas, que serão posteriormente generalizadas para regiões elipsoidais e caixas quaisquer.

O algoritmo

Considere x^k o ponto corrente tal que $\nabla f(x^k) \neq 0$ e $d(x^k, \Delta)$ a solução do sub-problema (2.2) formulado na k -ésima iteração. O critério usualmente estabelecido para avaliar a solução $d(x^k, \Delta)$ é baseado nas definições a seguir,

- Redução prevista pelo modelo:

$$pred(x^k, \Delta) = -m_k(d(x^k, \Delta)) > 0.$$

- Redução efetiva da função:

$$ared(x^k, \Delta) = f(x^k) - f(x^k + d(x^k, \Delta)).$$

- A relação entre as duas reduções, na k -ésima iteração é dada por

$$\rho = \frac{ared}{pred}.$$

O valor de ρ informa se o modelo quadrático representa bem a função na região D_Δ . A condição de aceitação do passo pode ser escrita como

$$ared \geq \alpha \, pred \text{ ou seja } \rho \geq \alpha.$$

Em implementações é usual estabelecer $\alpha = 10^{-4}$. Para análise pode-se fazer $\alpha = 0.25$, por exemplo [25].

Podemos observar que como $d(x^k, \Delta)$ é obtido minimizando o modelo m_k em uma região que inclui a direção $d = 0$, a redução prevista sempre será positiva. Desta forma se ρ é negativo, $f(x + d(x^k, \Delta))$ é maior do que o valor corrente $f(x^k)$ logo o passo $d(x^k, \Delta)$ deve ser rejeitado.

Apresentamos a seguir o algoritmo de região de confiança.

Algoritmo 2.1.1 *Região de confiança*

Dados: $x^0 \in \mathbb{R}^n$, $\beta, \alpha \in (0, 1)$ fixos e $\Delta > 0$. (Por exemplo, $\alpha = 0.25$, $\beta = 0.25$)

$k = 0$

ENQUANTO $\|\nabla f(x^k)\| \neq 0$

Construa o modelo $d \mapsto m_k(d) = \nabla f(x^k)^T d + \frac{1}{2}d^T B_k d$.

REPITA

Calcule $d = d(x^k, \Delta)$, solução de:

minimizar $m_k(h)$

sujeito a $\|h\| \leq \Delta$.

$ared = f(x^k) - f(x^k + d)$

$pred = -m_k(d)$

$\rho = ared/pred$

Se $\rho < \alpha$, faça $\Delta = \beta\Delta$.

Até que $\rho \geq \alpha$

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

$k = k + 1$

Fim.

A definição do raio Δ inicial em cada iteração pode ser feita da seguinte forma: para evitar raios demasiadamente pequenos “herdados” da iteração anterior, Δ é tomado maior que um valor mínimo fixado Δ_{min} . Além disso, pode-se considerar o sucesso da iteração anterior. Se ρ for próximo de 1 ($\rho > 0.75$), o modelo é bom e pode-se expandir a região de confiança. Se ρ for próximo de zero ($\rho < 0.25$) então a região deve ser reduzida na próxima iteração. Se ρ for positivo mas não próximo de zero ($0.25 < \rho < 0.75$) o raio permanece inalterado.

2.2 Região de confiança e a busca de Armijo

A seguir descreveremos brevemente a popular busca unidirecional de Armijo para em seguida apresentarmos a relação entre este método e o método de região de

confiança.

2.2.1 Busca de Armijo

Considere o problema de minimização de uma função $f : \mathbb{R}^n \rightarrow \mathbb{R}$ de classe \mathcal{C}^2 a partir de um ponto $x^k \in \mathbb{R}^n$ e uma direção $d \in \mathbb{R}^n$. O método de busca de Armijo visa uma boa redução da função na direção d , não tenta resolver o problema proposto.

Dados x^k , d e $\bar{\Delta}$ o algoritmo busca $\lambda \in [0, \bar{\Delta}]$ tal que $f(x^k + \lambda d) < f(x^k)$. A condição que determina se a redução é suficiente é a seguinte desigualdade

$$f(x^k + \lambda d) \leq f(x^k) + \eta \lambda \nabla f(x^k)^T d, \quad (2.3)$$

para alguma constante $\eta \in (0, 1)$.

A desigualdade (2.3) é chamada a “condição de Armijo”.

O algoritmo que determina λ é simples. Inicia com $\lambda = \bar{\Delta}$ e reduz λ até encontrar um ponto aceitável.

Algoritmo 2.2.1 Busca de Armijo

Dados: $x, d, \bar{\Delta}$ e $f'(x, d) = \nabla f(x)^T d$, $\eta, \beta \in (0, 1)$, por exemplo $\eta = \beta = 0,5$.

$\lambda = \bar{\Delta}$

ENQUANTO $f(x + \lambda d) > f(x) + \eta \lambda f'(x, d)$

$\lambda = \beta \lambda$

FIM

É possível melhorar o algoritmo permitindo que λ aumente quando $\lambda = \bar{\Delta}$ é aceitável, mas isso raramente é feito.

Região de Confiança com modelo linear e Armijo

Vamos supor que em vez do modelo quadrático no método de região de confiança usemos o modelo linear, ou seja, $B_k = 0$.

Assim o modelo m_k em torno do ponto x^k é dado por

$$m_k(d) = \nabla f(x^k)^T d.$$

Como o modelo é linear o ponto que minimiza o modelo na região de raio λ é o ponto $x^k + \lambda h$ onde $h = -\frac{\nabla f(x^k)}{\|\nabla f(x^k)\|}$, ou equivalentemente, $d(x^k, \Delta) = \lambda h$. O método de região de confiança exige que seja satisfeita a condição $\text{ared } d(x^k, \lambda) \geq \alpha \text{ pred } d(x^k, \lambda)$, que pode ser reescrita como

$$f(x^k + \lambda d) \leq f(x^k) + \alpha \lambda \nabla f(x^k)^T d,$$

que é exatamente a condição de Armijo (2.3). Ou seja, para o caso do modelo linear o método de região de confiança coincide com a busca de Armijo.

2.3 Convergência global

Nos métodos de região de confiança, apesar de estarmos procurando a solução ótima do sub-problema (2.2), é suficiente para se ter convergência global que a cada iteração k , o ponto $x^{k+1} = x^k + d(x^k, \Delta_k)$ esteja dentro da região de confiança e que forneça uma *redução suficiente* no modelo. Esta redução pode ser quantificada em termos do ponto de Cauchy denotado na k -ésima iteração por d_k^c e definido a seguir.

O ponto de Cauchy é definido como minimizador de m_k ao longo da direção oposta ao gradiente $\nabla f(x^k)$ na região de confiança D_Δ . Dependendo da curvatura do modelo pode estar no interior ou na fronteira da região de confiança, como mostra figura 2.1.

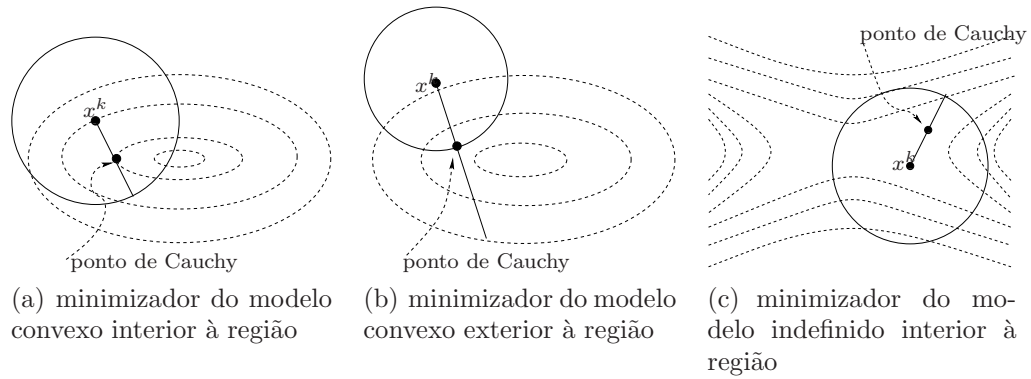


Figura 2.1: Ponto de Cauchy.

O ponto de Cauchy é expresso por $d_k^c = -\alpha^c \Delta \frac{\nabla f(x^k)}{\|\nabla f(x^k)\|}$ com $\alpha^c = 1$ quando o minimizador esta na fronteira e

$$\alpha^c = \frac{\|\nabla f(x^k)\|^3}{\nabla f(x^k)^T B_k \nabla f(x^k)}$$

quando o minimizador for interior a região, veja [25].

A figura 2.2 mostra o ponto de Cauchy numa iteração k . As elipses representam as curvas de nível do modelo $m_k(\cdot)$. A área hachurada corresponde ao conjunto de pontos que satisfazem $-m_k(d(x^k)) \geq -m_k(d_k^c)$.

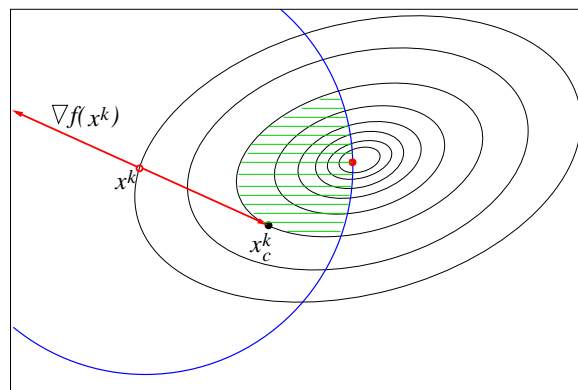


Figura 2.2: Ponto de Cauchy

Uma estimativa para a redução obtida no modelo m com uma solução aproximada $d(x^k, \Delta)$ é:

$$-m_k(d(x^k, \Delta)) \geq c_1 \|\nabla f(x^k)\| \min \left\{ \Delta, \frac{\|\nabla f(x^k)\|}{\|B_k\|} \right\} \quad (2.4)$$

onde $c_1 \in [0, 1)$ é uma constante. Esta estimativa considera que a redução obtida no modelo é proporcional ao tamanho do passo e ao gradiente.

O ponto de Cauchy satisfaz (2.4) com $c_1 = 1/2$. Este resultado está demonstrado em [25].

Toda solução aproximada $d(x^k, \Delta_k)$ que proporcione uma redução no modelo igual ou superior à redução obtida com o ponto de Cauchy, $m_k(d(x^k, \Delta_k)) \leq m_k(d_c^k)$ satisfaz (2.4) com $c_1 = 1/2$.

2.4 O problema quadrático e as duas trajetórias: Trajetória de Levenberg-Marquardt e a Trajetória Central

Esta é a seção principal deste capítulo, onde definimos as duas trajetórias: a trajetória de Levenberg-Marquardt e a trajetória central. Isto será feito a partir do estudo do sub-problema quadrático a ser resolvido na k -ésima iteração do método de região de confiança:

$$\begin{aligned} \text{minimizar} \quad & m_k(d) = \nabla f(x^k)^T d + \frac{1}{2} d^T B_k d \\ \text{sujeito a} \quad & \|d\| \leq \Delta. \end{aligned} \quad (2.5)$$

Como a função $m_k(\cdot)$ é contínua e $D_\Delta = \{d \in \mathbb{R}^n \mid \|d\| \leq \Delta\}$ é um conjunto fechado e limitado garantimos que este problema sempre tem solução.

O estudo é feito de forma independente das demais seções deste capítulo usando a notação comum em problemas de otimização com função objetivo quadrática

$$x \in \mathbb{R}^n \mapsto q(x) = c^T x + \frac{1}{2} x^T H x.$$

Temos como hipótese H simétrica semidefinida positiva.

No problema (2.5) q representa a variação do modelo quadrático de f em um ponto x_k .

O problema a ser estudado é

$$\begin{aligned} & \text{minimizar} && c^T x + \frac{1}{2} x^T H x \\ & \text{sujeito a} && \|x\| \leq \Delta \end{aligned} \tag{2.6}$$

Trajectoria de Levenberg-Marquardt

A trajetória de Levenberg-Marquardt surge do estudo do problema (2.6) utilizando a restrição de bola euclidiana, que denotamos por

$$\mathcal{B}_\Delta = \{x \in \mathbb{R}^n \mid \|x\|_2 \leq \Delta\}.$$

Consideremos o seguinte problema irrestrito:

Consideremos os seguintes problemas restrito e irrestrito:

$$(P_\Delta) \quad \begin{aligned} & \text{minimizar} && q(x) \\ & \text{sujeito a} && \|x\| \leq \Delta \end{aligned}$$

$$(PQ_\mu) \quad \begin{aligned} & \text{minimizar} && q_\mu(x) = q(x) + \mu \frac{\|x\|^2}{2}, \quad \mu \geq 0. \\ & && x \in \mathbb{R}^n \end{aligned}$$

O teorema a seguir relaciona (P_Δ) e (PQ_μ) . Este teorema é demonstrado em várias referências, por exemplo [25]. Para o nosso tratamento ficar completo, fazemos a demonstração para o caso convexo, que é mais simples.

Teorema 2.4.1 \bar{x} é solução de (P_Δ) se e só se existe $\mu \geq 0$ tal que \bar{x} é única solução de (PQ_μ) , com $\mu(\Delta - \|\bar{x}\|) = 0$.

Prova: (\Rightarrow) Suponhamos $\Delta > 0$ e \bar{x} solução de (P_Δ) . Logo existe $\mu \geq 0$ tal que

$$(H + \mu I)\bar{x} = -c \quad (2.7)$$

$$\mu(\Delta - \|\bar{x}\|) = 0 \quad (2.8)$$

$$\|\bar{x}\| \leq \Delta \quad (2.9)$$

$$\mu \geq 0. \quad (2.10)$$

A condição (2.8) significa ou \bar{x} é minimizador global de $q(\cdot)$ e $\|\bar{x}\| \leq \Delta$ (caso $\mu = 0$) ou $\|\bar{x}\| = \Delta$.

O segundo caso ocorre quando a restrição $\|\bar{x}\| \leq \Delta$ é ativa e $\mu > 0$.

Como para $\mu > 0$ \bar{x} satisfaz (2.7), que é a condição necessária de otimalidade de (PQ_μ) e $(H + \mu I)$ é definida positiva, segue que \bar{x} é solução de (PQ_μ) com $\|\bar{x}\| = \Delta$.

(\Leftarrow) Por outro lado, dado $\mu > 0$ definimos x_μ o (único) minimizador de $q_\mu(\cdot)$ e $\Delta_\mu = \|x_\mu\|$, satisfazendo

$$\forall x \in \mathbb{R}^n, q(x) + \frac{\mu}{2}\|x\|^2 \leq q(x) + \frac{\mu}{2}\|x_\mu\|^2.$$

A desigualdade vale em particular para $\|x\| \leq \Delta_\mu$, $x \neq x_\mu$:

$$q(x) > q(x_\mu) + \frac{\mu}{2}(\|x_\mu\|^2 - \|x\|^2) = q(x_\mu) + \frac{\mu}{2}(\Delta_\mu^2 - \|x\|^2),$$

ou seja ,

$$q(x) > q(x_\mu).$$

Portanto x_μ é solução única de (P_Δ) com $\Delta = \Delta_\mu$ □

De acordo com o teorema para resolver (P_Δ) quando o ponto de Newton não é viável ou não existe pode-se usar (PQ_μ) : devemos encontrar $\mu > 0$ tal que a solução x_μ encontrada para (PQ_μ) satisfaça $\|x_\mu\| = \Delta$. Isso é feito por tentativas: trata-se de um problema unidirecional em μ . A relação entre valores diferentes de

x_μ e $\|x_\mu\|$ é dada pelo seguinte teorema:

Teorema 2.4.2 *Consideremos os pontos x_{μ_1} , x_{μ_2} para $0 < \mu_1 < \mu_2$. Temos $\|x_{\mu_2}\| \leq \|x_{\mu_1}\|$.*

Prova: Como x_{μ_i} minimiza q_{μ_i} , $i = 1, 2$

$$q(x_{\mu_1}) + \mu_1 \|x_{\mu_1}\|^2 \leq q(x_{\mu_2}) + \mu_1 \|x_{\mu_2}\|^2$$

e

$$q(x_{\mu_2}) + \mu_2 \|x_{\mu_2}\|^2 \leq q(x_{\mu_1}) + \mu_2 \|x_{\mu_1}\|^2.$$

Subtraindo

$$(\mu_2 - \mu_1) \|x_{\mu_1}\|^2 \geq (\mu_2 - \mu_1) \|x_{\mu_2}\|^2$$

e como $(\mu_2 - \mu_1) > 0$

$$\|x_{\mu_1}\|^2 \geq \|x_{\mu_2}\|^2.$$

□

Portanto quando μ decresce os pontos x_μ correspondem a soluções de (P_Δ) para raios Δ crescentes. Para problemas de grande porte existem métodos muito eficientes computacionalmente para resolver (P_Δ) de forma aproximada, como Dog-Leg e o método de Steinhaug, [25].

Notemos que para cada $\mu \in (0, +\infty)$ existe uma única solução x_μ . Os pontos x_μ descrevem a curva

$$\mu \in (0, +\infty) \mapsto \chi(\mu) := \operatorname{argmin} \left\{ q(x) + \frac{\mu}{2} \|x_\mu\|^2 \right\},$$

chamada caminho de Levenberg-Marquardt.

Por construção, esta trajetória coincide com a trajetória $\Delta \in [0, +\infty) \rightarrow d_\Delta$ das soluções do problema de Região de Confiança (2.6) (ignorando o caso difícil).

Busca curvilínea. Da discussão acima concluímos que o passo de região de confiança é o resultado de uma busca sobre a trajetória de Levenberg-Marquardt, com a mesma formulação do método de Armijo, substituindo o modelo quadrático pelo linear .

Caso difícil. Definimos $\lambda_1, \lambda_2, \dots, \lambda_n$ autovalores de H em ordem crescente e v_1, v_2, \dots, v_n autovetores correspondentes. O caso difícil ocorre quando $-\lambda_1 = \mu$, ou seja, $(H + \mu I)$ é semidefinida positiva, e o vetor c é ortogonal ao subespaço gerado pelos autovetores associados a $\lambda_1 = -\mu$. Neste caso o problema (PQ_μ) tem uma variedade linear de minimizadores que são soluções de $(H + \mu I)x = -c$. Se $\mu > -\lambda_1$, então a solução é única e podemos definir o caminho de Levenberg-Marquardt para $\mu > -\lambda_1$. Neste trabalho estamos supondo $\lambda_1 > 0$.

O lema a seguir mostra que para $\mu \rightarrow +\infty$, x_μ tende a um passo na direção de Cauchy.

Lema 1 : Para $\mu \rightarrow +\infty$, temos que $\Delta \rightarrow 0$ e $\frac{x_\mu}{\|x_\mu\|} \rightarrow \frac{-c}{\|c\|}$ (direção de Cauchy).

Prova: Para qualquer $\mu > 0$, temos que x_μ satisfaz

$$(H + \mu I)x_\mu = -c$$

ou

$$\mu x_\mu = -c - Hx_\mu.$$

Pelo teorema anterior, x_μ é limitado. O lado direito é limitado. Então μx_μ é limitado. Concluímos que $\lim_{\mu \rightarrow +\infty} x_\mu = 0$, $\lim_{\mu \rightarrow +\infty} Hx_\mu = 0$ e $\lim_{\mu \rightarrow +\infty} \mu x_\mu = -c$. Dividindo por $\mu \|x_\mu\|$ e tomando limite para

$$\frac{x_\mu}{\|x_\mu\|} = -\frac{c}{\mu \|x_\mu\|} - \frac{Hx_\mu}{\mu \|x_\mu\|}, \text{ temos}$$

$$\lim_{\mu \rightarrow +\infty} \frac{x_\mu}{\|x_\mu\|} = -\frac{c}{\|c\|}.$$

□

2.5 Trajetória Ótima para regiões tipo caixa

A trajetória central será definida a partir do problema (2.6) com a norma ℓ_∞ ,

$$(P_B) \quad \begin{array}{ll} \text{minimizar} & c^T x + \frac{1}{2} x^T H x \\ \text{sujeito a} & \|x\|_\infty \leq \Delta \end{array} \quad (2.11)$$

Como já comentamos o interesse neste problema esta no fato de que $D_\Delta = \{x \in \mathbb{R}^n \mid -\Delta e \leq x \leq \Delta e\}$ tem o formato que se adapta a problemas com restrições do tipo caixa, muito comuns em problemas práticos, e além disso, tem tamanho maior do que a bola euclidiana.

No algoritmo de região de confiança clássico cada iteração determina um minimizador do modelo em uma região onde se supõe que este é confiável; caso não o seja confiável deve-se reduzir o tamanho da região e determinar novo minimizador. No caso de caixas, o problema de minimização a ser resolvido é um problema de programação quadrática geral.

Temos então dois métodos: um com bolas euclidianas que é eficiente e convergente, mas usa regiões pequenas e é pouco aplicável a problemas com restrições. Um algoritmo ideal é uma busca curvilínea ao longo da Trajetória de Levenberg. A presença de restrições lineares afeta pouco a sua implementação. Temos um método baseado em caixas, com alto custo computacional porque cada iteração deve resolver um problema de programação quadrática geral. Vamos desenvolver um método intermediário, com regiões de confiança que evoluem continuamente de pequenas bolas e tendem à caixa completa: as regiões são definidas pela função barreira logarítmica.

Função Barreira Logarítmica

Considere uma caixa fixa dada por $\|x\|_\infty \leq \Delta$. Nesta seção vamos usar $\Delta = 1$ e as seguintes notações:

$$\mathcal{V} = \{x \in \mathbb{R}^n \mid -e \leq x \leq e\}, \quad \mathcal{V}^\circ = \{x \in \mathbb{R}^n \mid -e < x < e\},$$

em que $e = [1, 1, \dots, 1]$.

A função barreira logarítmica associada a caixa \mathcal{V} é definida por

$$x \in \mathcal{V}^\circ \mapsto p(x) = -\frac{1}{2} \sum_{i=1}^n \log(x_i + 1) + \log(1 - x_i).$$

Por outro lado $p(x) = -\frac{1}{2} \log \prod_{i=1}^n (x_i + 1)(1 - x_i)$. Daí conclui-se que se o valor de alguma das restrições estiver próximo de um, ou seja, x próximo da fronteira da caixa \mathcal{V} , o produtório fica próximo de zero e conseqüentemente $p(x)$ cresce indefinidamente, fato que motiva o seu nome de barreira. Foi usada pela primeira vez em otimização por Frisch [10] em 1955.

Centro Analítico. Vimos que ao nos aproximarmos internamente da fronteira de \mathcal{V} , p cresce indefinidamente. Como \mathcal{V} é limitado, concluímos que a função p possui minimizador no interior da caixa. De fato, as derivadas de $p(x)$ são:

$$[\nabla p(x)]_i = -\frac{1}{x_i + 1} + \frac{1}{1 - x_i}, \quad \nabla^2 p(x) = \text{diag} \left(\frac{1}{(x_i + 1)^2} + \frac{1}{(1 - x_i)^2} \right) > I,$$

e concluímos que $\nabla^2 p(x)$ é definida positiva e tem como único minimizador o ponto $x = 0$, uma vez que $\nabla p(0) = 0$.

O ponto que minimiza p foi definido por Sonnevend [31] como o centro analítico de \mathcal{V} .

Trajectoria Central

Dado $\mu > 0$, define-se a função penalizada associada a $(P_{\mathcal{B}})$ por

$$x \in \mathcal{V}^\circ \rightarrow p_\mu(x) = q(x) + \mu p(x)$$

e o problema penalizado como

$$(PP_\mu) \begin{array}{ll} \text{minimizar} & p_\mu(x) \\ \text{sujeito a} & x \in \mathcal{V}^\circ. \end{array}$$

Para cada $\mu \in (0, +\infty)$ temos que p_μ cresce indefinidamente perto da

fronteira do conjunto compacto \mathcal{V} e $\nabla^2 p_\mu(x) = H + \mu \nabla^2 p(x) > H + \mu I > 0$. Portanto existe único ponto x_μ que resolve (2.5) e x_μ é chamado ponto central. O mapeamento ponto conjunto

$$\mu \in \mathbb{R}_{++} \mapsto \chi(\mu) := \operatorname{argmin} \{p_\mu(x) \mid x \in \mathcal{V}^\circ\}$$

define a curva denominada de *trajetória central*.

Vamos mostrar que o problema (PQ_μ) está relacionado com o seguinte problema restrito:

$$(P_{\mathcal{K}}) \quad \begin{array}{ll} \text{minimizar} & q(x) \\ \text{sujeito a} & p(x) \leq \mathcal{K}, \end{array}$$

$\mathcal{K} \in \mathbb{R}$.

Teorema 2.5.1 *Suponha que $q(\cdot)$ é convexa e $x_{\mathcal{K}}$ solução de $(P_{\mathcal{K}})$. Então $x_{\mathcal{K}}$ é solução de $(P_{\mathcal{K}})$ se e só se existe $\mu \geq 0$ tal que $x_{\mathcal{K}}$ é solução de (2.5) com $p(x) = \mathcal{K}$.*

Prova: (\Rightarrow) Suponha que $x_{\mathcal{K}}$ é solução de $(P_{\mathcal{K}})$ então existe $\mu \geq 0$ tal que

$$\begin{array}{ll} \nabla q(x_{\mathcal{K}}) + \mu \nabla p(x_{\mathcal{K}}) & = 0 \\ \mu(\mathcal{K} - p(x_{\mathcal{K}})) & = 0 \\ p(x_{\mathcal{K}}) & \leq \mathcal{K} \\ \mu & \geq 0 \end{array}$$

A condição $\mu(\mathcal{K} - p(x_{\mathcal{K}})) = 0$ significa ou $x_{\mathcal{K}}$ é minimizador global de q e $p(x_{\mathcal{K}}) \leq \mathcal{K}$ (caso $\mu = 0$) ou $p(x_{\mathcal{K}}) = \mathcal{K}$. O segundo caso ocorre quando a restrição $p(x_{\mathcal{K}}) \leq \mathcal{K}$ é ativa e $\mu > 0$.

Observemos que para $\mu > 0$, $x_{\mathcal{K}}$ satisfaz a primeira condição de KKT, que coincide com a condição necessária de otimalidade para (PP_μ) . Como $H + \mu \nabla^2 p(x_{\mathcal{K}})$ é definida positiva segue que $x_{\mathcal{K}}$ é solução única de (PP_μ) .

(\Leftarrow) Por outro lado, considere $\mu > 0$ e x_μ o (único) minimizador de $p_\mu(\cdot)$, satisfazendo

$$\forall x \in \mathcal{V}^\circ \quad q(x) + \mu p(x) \leq q(x_\mu) + \mu p(x_\mu).$$

A desigualdade vale em particular para x tal que $p(x) \leq p(x_\mu) = \mathcal{K}$, $x \neq x_\mu$:

$$q(x) > q(x_\mu) + \mu(p(x_\mu) - p(x)) = q(x_\mu) + \frac{\mu}{2}(\mathcal{K} - p(x))$$

$$q(x) > q(x_\mu).$$

Portanto x_μ é minimizador único de q no conjunto $\{x \in \mathcal{V}^\circ \mid p(x) \leq \mathcal{K}\}$. \square

De acordo com o teorema, se um ponto x_μ pertence a trajetória central, x_μ é solução do problema quadrático restrito ao conjunto $\mathcal{V}_{\mathcal{K}} = \{x \in \mathcal{V}^\circ \mid p(x) \leq \mathcal{K}\}$ para algum $\mathcal{K} \in \mathbb{R}$. Portanto ao percorrermos a trajetória central estamos resolvendo problemas restritos aos conjuntos $\mathcal{V}_{\mathcal{K}}$ para valores diferentes de \mathcal{K} .

O teorema a seguir determina como ocorre a variação de $p(x_\mu) = \mathcal{K}$ ao variarmos μ .

Teorema 2.5.2 *Consideremos os pontos x_{μ_1} , x_{μ_2} para $\mu_1 < \mu_2 < +\infty$ soluções dos problemas penalizados (PP_μ) . Temos que $p(x_{\mu_1}) = \mathcal{K}_1 < \mathcal{K}_2 = p(x_{\mu_2})$.*

Prova: Como x_{μ_1} é solução de (PP_{μ_1})

$$q(x_{\mu_1}) + \mu_1 p(x_{\mu_1}) < q(x_{\mu_2}) + \mu_1 p(x_{\mu_2}).$$

Da mesma forma, como x_{μ_2} é solução de (PP_{μ_2})

$$q(x_{\mu_1}) + \mu_2 p(x_{\mu_1}) > q(x_{\mu_2}) + \mu_2 p(x_{\mu_2}).$$

Subtraindo as duas desigualdades obtemos

$$(\mu_1 - \mu_2)p(x_{\mu_1}) < (\mu_1 - \mu_2)p(x_{\mu_2}).$$

Como $(\mu_1 - \mu_2) > 0$ segue que

$$p(x_{\mu_1}) < p(x_{\mu_2}).$$

Ou $\mathcal{K}_1 < \mathcal{K}_2$. □

2.6 Comparação entre os métodos

No caso de regiões de confiança do tipo \mathcal{B}_Δ uma iteração do algoritmo faz uma busca na trajetória de Levenberg-Marquardt: se o modelo não é confiável reduz Δ (ou aumenta-se μ) até o modelo ser confiável. No caso de regiões tipo caixa, vamos seguir a trajetória central a partir do centro analítico que coincide com o ponto a partir do qual o modelo foi construído e percorrer a trajetória (aumentando e reduzindo μ) até o modelo deixar de ser confiável. Desta forma as regiões de confiança são os conjuntos $\mathcal{V}_{\mathcal{K}}$. Na seção 2.8 escrevemos um algoritmo completo, incluindo a escolha do tamanho das caixas.

Os conjuntos \mathcal{B}_Δ e $\mathcal{V}_{\mathcal{K}}$

Nesta seção apresentamos dois lemas importantes que relacionam os conjuntos \mathcal{B}_Δ e $\mathcal{V}_{\mathcal{K}}$ e caracterizam o comportamento entre as duas trajetórias.

Esses dois lemas foram formulados a partir do estudo da relação entre as funções: $p(x)$ e $\frac{\|x\|^2}{2}$.

Considere $x \in \mathcal{V}$. Temos que $p(x) = -\frac{1}{2} \sum_{i=1}^n \log(x_i + 1) + \log(1 - x_i)$. Vamos expandir cada parcela em série de Taylor: definimos para $i = 1, \dots, n$

$$p_i(x_i) = -\frac{1}{2} \log(x_i + 1) + \log(1 - x_i).$$

Sabemos que para $|x_i| < 1$,

$$\log(1 + x_i) = x_i - \frac{x_i^2}{2} + \frac{x_i^3}{3} - \dots$$

Portanto a expansão de p em torno de 0 em cada componente é dada por :

$$p_i(x_i) = \frac{x_i^2}{2} + \frac{x_i^4}{4} + \frac{x_i^6}{6} + \dots \geq \frac{x_i^2}{2} \tag{2.12}$$

e

$$p(x) = \frac{\|x\|_2^2}{2} + \frac{\|x\|_4^4}{4} + \frac{\|x\|_6^6}{6} + \dots \geq \frac{\|x\|_2^2}{2} \quad (2.13)$$

Podemos obter a expressão para a derivada $p'_i(x_i)$,

$$p'_i(x_i) = x_i + x_i^3 + x_i^5 + \dots,$$

$$\nabla p(x) = x + x^2 + x^3 + \dots;$$

e a segunda derivada para cada i ,

$$p''_i(x_i) = 1 + 3x_i^2 + 5x_i^4 + \dots > 1 + 3x_i^2 > 1 \text{ para todo } x_i \neq 0. \quad (2.14)$$

Ao definirmos $X = \text{diag}(x)$ obtemos

$$\nabla^2 p(x) = I + 3X^2 + 5X^4 + \dots > I + 3X^2 > 1 \text{ para } x \neq 0.$$

Concluimos de (2.13) que $p(x) \geq \frac{\|x\|_2^2}{2}$, para todo $x \in \mathcal{V}^\circ$.

Por (2.12) p cresce mais do que $\frac{\|x\|_2^2}{2}$ perto de zero.

De (2.14) concluimos que próximo de zero a hessiana é próxima da identidade.

Lema 2 Consideremos $\Delta > 0$. Para $\mathcal{K} \leq \frac{\Delta^2}{2}$, $\mathcal{V} \subset \mathcal{B}_\Delta$.

Prova: Consideremos $x \in \mathbb{R}^n$ tal que $x \in \mathcal{V}_\mathcal{K}$. Temos que $p(x) \leq \mathcal{K} \leq \frac{\Delta^2}{2}$. Como $\frac{\|x\|_2^2}{2} \leq p(x)$ segue que $\|x\| \leq \Delta$. Logo $x \in \mathcal{B}_\Delta$. \square

Lema 3 Consideremos $x \in \mathbb{R}^n$ tal que $\|x\| \leq \frac{1}{2}$. Então

$$\frac{\|x\|_2^2}{2} \leq p(x) \leq \frac{7}{6} \frac{\|x\|_2^2}{2}$$

Se $\Delta \leq \frac{1}{2}$, então para $\mathcal{K} = \frac{\Delta^2}{2}$ e $\Delta' = \sqrt{\frac{6}{7}}\Delta$

$$\mathcal{B}_\Delta \supset \mathcal{V}_\mathcal{K} \supset \mathcal{B}_{\Delta'}$$

Prova: Ao expandirmos a função p em série de Taylor obtemos que a i -ésima componente de $p(x)$ é dada por:

$$p_i(x_i) = \frac{x_i^2}{2} + \frac{x_i^4}{4} + \frac{x_i^6}{6} + \dots \leq \frac{x_i^2}{2} + \frac{x_i^4}{4}(1 + x_i^2 + x_i^4 + \dots).$$

Como $|x_i| < 1$,

$$p_i(x_i) \leq \frac{x_i^2}{2} + \frac{x_i^4}{4} \frac{1}{(1 - x_i^2)} \leq \frac{x_i^2}{2} + \frac{x_i^4}{4} \left(\frac{1}{1 - \|x\|_\infty^2} \right).$$

Logo

$$p(x) \leq \frac{\|x\|_2^2}{2} + \frac{\|x\|_4^4}{4} \left(\frac{1}{1 - \|x\|_\infty^2} \right)$$

e como $\|x\|_4^4 \leq \|x\|_2^2$ para x tal que $\|x\|_\infty < 1$,

$$p(x) \leq \frac{\|x\|_2^2}{2} \left(1 + \frac{\|x\|_2^2}{2(1 - \|x\|_\infty^2)} \right).$$

Para $\|x\| \leq \frac{1}{2}$ temos

$$p(x) \leq \frac{7}{6} \frac{\|x\|_2^2}{2},$$

por outro lado, de acordo com (2.13)

$$p(x) \geq \frac{\|x\|_2^2}{2}.$$

A segunda afirmação do lema segue diretamente da primeira, pelo raciocínio usado no lema 2. □

Concluimos que quando Δ é pequeno e $\mathcal{K} = \frac{\Delta^2}{2}$ as regiões \mathcal{B}_Δ e $\mathcal{V}_\mathcal{K}$ são muito semelhantes. À medida que Δ cresce elas se afastam, com $\mathcal{V}_\mathcal{K}$ tendendo à caixa completa. Como mostra figura 2.3

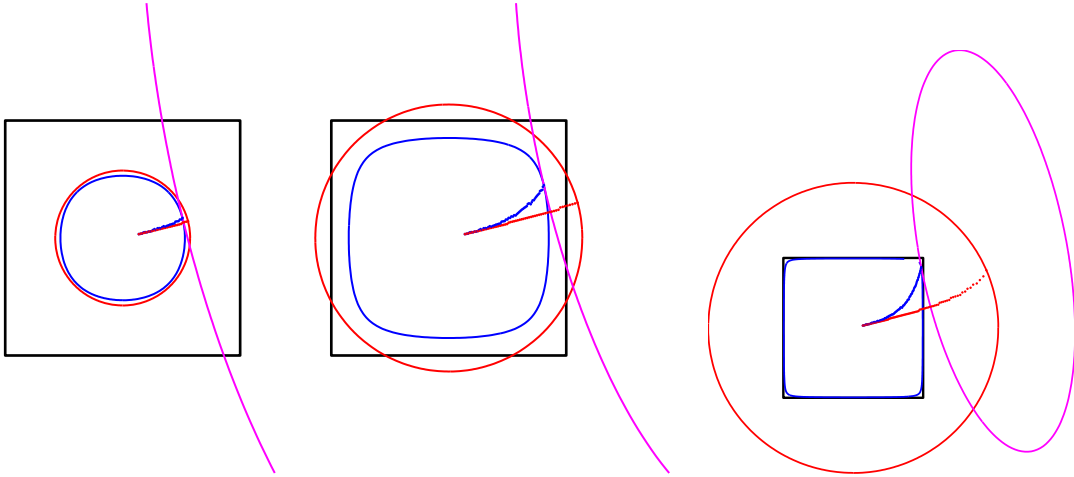


Figura 2.3: Regiões de confiança e trajetórias.

2.7 Generalização

Vamos considerar o problema de região de confiança em caixas gerais:

$$\mathcal{V}^v = \{x \in \mathbb{R}^n \mid -v \leq x \leq v\}, \text{ com } v \in \mathbb{R}_{++}^n, \text{ e interior } \overset{\circ}{\mathcal{V}}^v.$$

Definimos $V = \text{diag}(v)$ e a função barreira para \mathcal{V}^v ,

$$-v < x < v \mapsto p_v(x) = -\frac{1}{2} \sum_{i=1}^n \log(v_i - x_i) + \log(v_i + x_i).$$

Vamos definir também as regiões elipsoidais (elipsóides de Dikin) para $\Delta > 0$.

$$\mathcal{B}_\Delta^v = \{x \in \mathbb{R}^n \mid \|V^{-1}x\| \leq \Delta\}$$

e as regiões para $\mathcal{K} > 0$

$$\mathcal{V}_{\mathcal{K}}^v = \{x \in \mathcal{V}^v \mid p_v(x) - p_v(0) \leq \mathcal{K}\}.$$

Note que $\mathcal{B}_\Delta = \mathcal{B}_\Delta^e$ e $\mathcal{V}_{\mathcal{K}} = \mathcal{V}_{\mathcal{K}}^e$.

Considere a mudança de variáveis (mudança de escala) $x = V\bar{x}$.

Lema 4 *Nas condições acima, temos:*

- (i) $x \in \mathcal{V}^v$ se e só se $\|\bar{x}\|_\infty \leq 1$
- (ii) $x \in \mathcal{B}_\Delta^v$ se e só se $\bar{x} \in \mathcal{B}_\Delta$
- (iii) $(\forall x \in \overset{\circ}{\mathcal{V}}^v) p_v(x) - p_v(0) = p(\bar{x})$.
- (iv) $x \in \mathcal{V}_{\mathcal{X}}^v$ se e só se $\bar{x} \in \mathcal{V}_{\mathcal{X}}$
- (v) x^* minimiza $q(x)$ em \mathcal{B}_Δ^v se e só se \bar{x}^* minimiza $q(V\bar{x})$ em \mathcal{B}_Δ
- (vi) x^* minimiza $q(x) + \mu p_v(x)$ em \mathcal{V}^v se e só se \bar{x}^* minimiza $q(V\bar{x}) + \mu p(x)$ em \mathcal{V} .

Prova: (i) e (ii) são triviais.

(iii) Para $x \in \overset{\circ}{\mathcal{V}}^v$, $\log(v_i - x_i) = \log v_i + \log\left(1 - \frac{x_i}{v_i}\right)$.

Somando, $p_v(x) = -\frac{1}{2} \left[2 \sum_{i=1}^n \log v_i + \sum_{i=1}^n \log(1 - \bar{x}_i) + \sum_{i=1}^n \log(1 + \bar{x}_i) \right]$.

O resultado segue de $p_v(x)(0) = -\sum_{i=1}^n \log v_i$.

(iv), (v) e (vi) seguem trivialmente de (i) a (iii) por mudança de variáveis. \square

Conclusão

Para o método de regiões de confiança tradicional, as regiões \mathcal{B}_Δ^v são os elipsóides de Dikin associados ao centro da caixa. Para o método de região de confiança definidas por p_v , basta fazer uma busca curvilínea sobre a trajetória central, o que é equivalente a fazer uma mudança de escala e seguir a trajetória definida por $\|x\|_\infty \leq 1$.

2.8 Algoritmos de regiões de confiança definidas pela barreira

Vamos descrever uma iteração do algoritmo, usando caixas definidas por $\|d\|_\infty \leq \Delta$. A iteração faz uma busca do tipo Armijo sobre a trajetória central

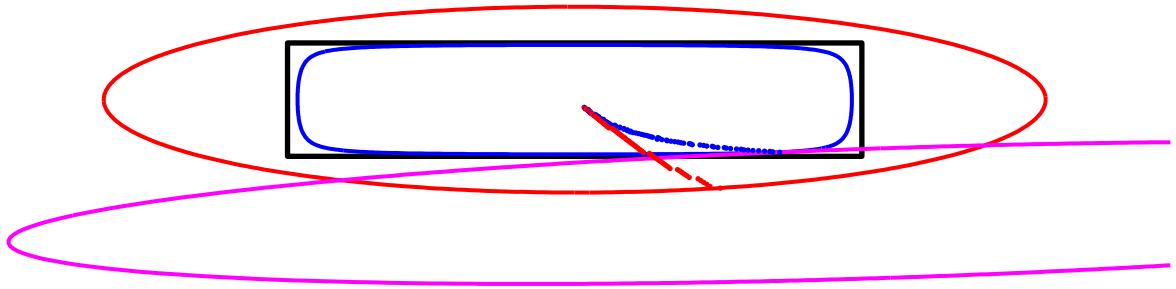


Figura 2.4: Regiões de confiança e trajetórias no caso geral.

na caixa, reduzindo o parâmetro μ enquanto $\mu > \varepsilon$ e $ared(d_\mu) > \alpha pred(d_\mu)$, com $\alpha = 0.5$, (por exemplo) e ε uma precisão para o algoritmo de trajetória.

É necessário discutir dois pontos:

- (i) O algoritmo de trajetória central: usaremos métodos primais-duais, discutidos no capítulo 3.
- (ii) A inicialização do algoritmo: este é um ponto muito delicado, que foi resolvido na tese de R. Behling [1] e será descrito no capítulo seguinte. Vimos que os conjuntos \mathcal{B}_Δ e \mathcal{V}_x são muito semelhantes para um raio Δ menor ou igual a $\frac{\Delta}{2}$. Desta forma é natural que um ponto inicial para seguir a trajetória central possa ser dado pela solução d_μ do problema de região de confiança usando \mathcal{B}_Δ . Resultado surpreendente é obtido na tese de R. Behling. Prova-se que o μ associado a d_μ na trajetória de Levenberg- Marquardt é o μ associado ao ponto da trajetória central x_μ próximo de d_μ .

Notemos ainda que como d_μ é obtido de um problema de região de confiança pode ocorrer que $ared(d_\mu) < \alpha pred(d_\mu)$. Neste caso Δ é reduzido e faz-se o método tradicional de região de confiança.

Apresentamos a seguir uma iteração do algoritmo de região de confiança.

Algoritmo 2.8.1 *Região de confiança baseado na função barreira*

Dados: $x^0 \in \mathbb{R}^n$, $\alpha \in (0, 1)$ fixo, ε e $\Delta > 0$ (Por exemplo, $\alpha = 0.25, \Delta = 1.$)

$k = 0$

ENQUANTO $\|\nabla f(x^k)\| \neq 0$

Construa o modelo $d \mapsto m_k(d) = \nabla f(x^k)^T d + \frac{1}{2}d^T H_k d$.

Calcule $d_\mu = \operatorname{argmin} \left\{ m_k(d) \mid \|d\| \leq \frac{\Delta}{2} \right\}$ e obtenha μ^0 multiplicador de

Lagrange

se $\operatorname{ared}(d_\mu) \leq \alpha \operatorname{pred}(d_\mu)$

 siga a trajetória de Levenberg-Marquardt para $\mu \in (\mu^0, +\infty)$

senão

 siga a trajetória central com $\mu < \mu^0$ até encontrar um ponto d_μ

 com $\operatorname{ared}(d_\mu) \leq \alpha \operatorname{pred}(d_\mu)$ ou $\mu \leq \varepsilon$

$k = k + 1$

Fim.

A figura 2.5 ilustra uma iteração do algoritmo. A trajetória de Levenberg-Marquardt está em vermelho e a trajetória central está em azul. Para um raio menor ou igual a 1/2 as trajetórias são muito parecidas. Determinamos um ponto inicial na trajetória de Levenberg-Marquardt, se o modelo for confiável neste ponto abandonamos esta e seguimos a central. Podemos dizer que no algoritmo de região de confiança proposto seguimos uma trajetória híbrida.

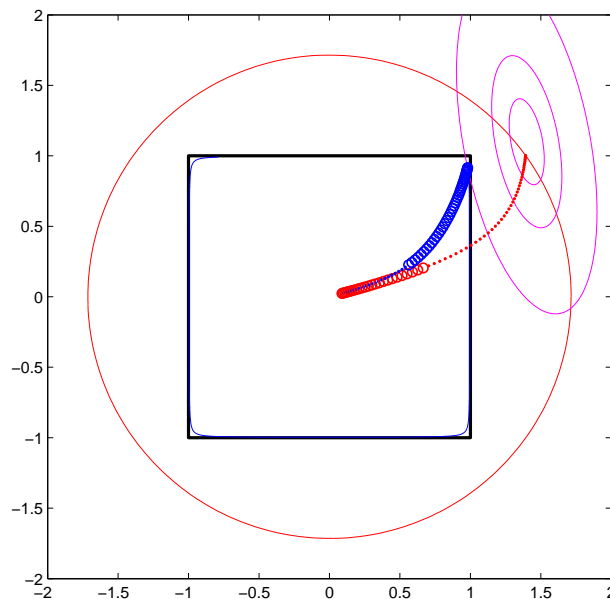


Figura 2.5: Uma iteração do algoritmo

Determinação de Δ para a iteração seguinte: seja d aceito.

- Se $ared(d) > \beta pred(d)$, com $\beta \in (\alpha, 1)$, (por exemplo, $\beta = 0.75$) aumentar Δ com $\Delta = 2\|d\|_\infty$
- Senão, fazer $\Delta = \min\{\Delta, \gamma\|d\|_\infty\}$, com $\gamma > 1$ (por exemplo, $\gamma = 1.5$.)

No caso de uma caixa qualquer $[l, u] = \{x \in \mathbb{R}^n \mid l \leq x \leq u\}$, definimos o vetor $v = \frac{u-l}{2}$ e $V = \text{diag}(v)$. Com a mudança de escala $x = V\bar{z}$ aplicamos o lema anterior. Deve-se minimizar $m(V\bar{z})$ na bola $\{\bar{z} \in \mathbb{R}^n \mid \bar{z} \leq \Delta/2\}$ ou seguir a trajetória (Levenberg-Marquardt ou central) na caixa $[-e, e]$.

Restrições lineares. Suponhamos que restrições de igualdade $Ax = b$, $A \in \mathbb{R}^{m \times n}$ de posto completo com $m < n$ sejam adicionadas ao problema irrestrito (2.1). Neste caso, o problema de região de confiança em torno de um ponto x^k tal que $Ax^k = b$ e a trajetória central passam a ser restritos a $\mathcal{N}(A)$.

Capítulo 3

Métodos de Pontos Interiores Primais-Duais

Como o algoritmo de região de confiança proposto nesta dissertação deve seguir a trajetória central, o objetivo neste capítulo é apresentar um algoritmo primal-dual que gera pontos em uma vizinhança da trajetória, possibilitando a busca a ser feita no algoritmo 2.8.1.

Além do algoritmo, enunciamos o resultado obtido na dissertação de Behling [1], já introduzido na capítulo 2.

Antes de apresentarmos o algoritmo vamos revisar na primeira seção alguns conceitos sobre a teoria primal dual de pontos interiores, com base em Gonzaga [14] e nos livros de S.Wright [32], de Bonnans e outros [4], Nocedal e Wright [25].

3.1 Revisão de conceitos

O formato padrão para o problema a ser resolvido neste capítulo é o problema geral de programação quadrática convexa com restrições lineares:

$$\begin{aligned} \text{minimizar} \quad & c^T x + \frac{1}{2} x^T H x \\ \text{sujeito a} \quad & Ax = b \\ & x \geq 0 \end{aligned} \tag{3.1}$$

com $x, c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times n}$ e $H \in \mathbb{R}^{n \times n}$ uma matriz simétrica semi-definida positiva. Vamos supor que A tem posto completo.

3.1.1 Trajetória Primal-Dual

A condição necessária e suficiente de otimalidade (KKT) para o problema é:

$$-Hx + A^T y + s = c \quad (3.2)$$

$$Ax = b \quad (3.3)$$

$$x s = 0 \quad (3.4)$$

$$x, s \geq 0, \quad (3.5)$$

em que $x s$ denota o produto componente a componente dos vetores.

Para resolver (3.1) devemos encontrar (x, s, y) que resolvem as condições de KKT. A dificuldade está em resolver (3.4), já que este é um problema combinatório. Obteremos um problema mais simples se relaxarmos a complementariedade fazendo para $\mu > 0$

$$x_i s_i = \mu, \quad i = 1, \dots, n.$$

$$-Hx + A^T y + s = c \quad (3.6)$$

$$Ax = b \quad (3.7)$$

$$x s = \mu e \quad (3.8)$$

$$x, s > 0. \quad (3.9)$$

O método primal-dual resolve aproximadamente as condições de KKT relaxadas. Com isso elimina-se o problema combinatório e as desigualdades passam a ser $x, s > 0$, definindo um conjunto aberto.

Para cada $\mu > 0$ este sistema tem solução única (x_μ, s_μ, y_μ) chamado ponto

central associado ao parâmetro μ . Variando μ define-se a curva

$$\mathcal{C} = \{(x_\mu, s_\mu, y_\mu) \mid \mu > 0\},$$

chamada trajetória central.

Para caracterizar um ponto central primal-dual é suficiente especificar (x_μ, s_μ) , uma vez que y é univocamente determinado por (3.6) com A de posto completo. Daqui em diante falaremos em soluções primais-duais (x, s) . O ponto x_μ é chamado parte primal, s_μ é chamada parte dual, e o par (x_μ, s_μ) é chamado par primal dual.

Quando μ tende a zero a solução do sistema de KKT relaxado tende a uma solução do sistema de KKT exato, satisfazendo (3.4).

O lema a seguir relaciona a trajetória central primal dual com a trajetória primal definida no capítulo 2.

Lema 5 *Dado $\mu > 0$, x_μ é um ponto central primal de (3.1) associado a μ se e só se (x_μ, s_μ) é ponto central primal-dual associado a μ .*

Ou seja, um ponto x_μ da trajetória central primal associado a um determinado $\mu > 0$ é a parte primal x_μ do ponto central associado ao mesmo μ .

Vamos denotar por Γ o conjunto dos pares primais duais viáveis

$$\Gamma = \{(x, s) \in \mathbb{R}_+^{2n} \mid Ax = b, -Hx + s + A^T y = c, \text{ para algum } y \in \mathbb{R}^m\} \quad (3.10)$$

e por Γ° o conjunto de pares primais duais interiores

$$\Gamma^\circ = \{(x, s) \in \Gamma \mid x, s > 0\}. \quad (3.11)$$

3.1.2 Proximidade e

Vizinhanças da Trajetória Central

Os algoritmos de pontos interiores seguem a trajetória central ou alguma vizinhança dela. Dado $\mu > 0$, queremos encontrar (x, s) tal que $x s = \mu e$. Podemos avaliar o desvio desta condição por uma medida de proximidade.

Definição 3.1.1 *Dados um ponto $(x, s) \in \Gamma^\circ$ e $\mu > 0$, a proximidade de (x, s) ao ponto (x_μ, s_μ) é definida por*

$$\delta(x, s, \mu) = \left\| \frac{x s}{\mu} - e \right\|.$$

Uma proximidade $\delta(x, s, \mu) < \alpha$, com $\alpha \in (0, 0.5)$ indica que (x, s) está próximo ao ponto central associado a μ .

Frequentemente são dados x e s e procura-se um valor de μ tal que $\delta(x, s, \mu)$ seja pequeno. Assim, dado $(x, s) \in \Gamma^\circ$ define-se a seguinte medida de proximidade

$$\delta(x, s) = \min_{\mu > 0} \left\| \frac{x s}{\mu} - e \right\|. \quad (3.12)$$

A solução deste problema resulta em

$$\mu^*(x, s) = \frac{\|x s\|^2}{x^T s}. \quad (3.13)$$

Porém, o valor de μ mais utilizado, muito próximo deste, é dado pelo minimizador de

$$\delta(x, s) := \min_{\mu > 0} \|x s - \mu e\|,$$

dado por

$$\mu(x, s) = \frac{x^T s}{n}. \quad (3.14)$$

Em programação linear e quadrática $n\mu(x, s)$ é o valor do gap de dualidade e em particular, em um ponto central, $x s = \mu e$, $\mu = \frac{x^T s}{n}$.

Isto justifica os algoritmos de trajetória central, que geram seqüências (x^k, s^k, μ^k) com $\mu^k \rightarrow 0$.

Vizinhanças da Trajetória Central

Dado $\alpha \in (0, 1)$, a vizinhança $N(\alpha)$ da trajetória central é definida pelo conjunto:

$$N(\alpha) = \{(x, s) \in \Gamma^\circ \mid \delta(x, s, \mu) \leq \alpha, \text{ para algum } \mu > 0\}.$$

Vamos descrever um algoritmo que segue a trajetória central gerando pares primais-duais em $N(\alpha)$ para um dado α , por exemplo $\alpha = 0.5$. Existem outras medidas de proximidade que geram vizinhanças maiores:

$$\delta(x, s, \mu)_\infty = \left\| \frac{x s}{\mu} - e \right\|_\infty$$

e

$$\delta(x, s, \mu)_{-\infty} = \left\| \frac{x s}{\mu} - e \right\|_{-\infty},$$

com $\|y\|_{-\infty} = \max\{|y_i| : y_i \leq 0\}$. Apesar da notação, $\|\cdot\|_{-\infty}$ não é uma norma.

A vizinhança “menos infinito” é definida por

$$N_{-\infty}(\alpha) = \{(x, s) \in \Gamma^\circ \mid \delta(x, s, \mu)_{-\infty} \leq \alpha \text{ para algum } \mu > 0\} \quad (3.15)$$

Para valores pequenos de $(1 - \alpha)$ (um valor comum é 10^{-3}), $N_{-\infty}$ pode englobar quase toda região Γ° . Desta maneira podemos redefinir N como

$$N_{-\infty}(\alpha) = \{(x, s) \in \Gamma^\circ \mid \frac{x s}{\mu} \geq (1 - \alpha)e, \mu > 0\}.$$

Assim $N_{-\infty}$ é um conjunto muito maior do que N_2 . Mas em geral a teoria de convergência com $N_{-\infty}$ é muito mais complicada. A figura 3.1 ilustra as três vizinhanças.

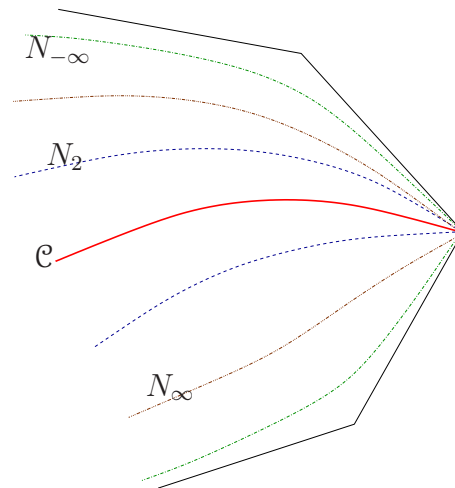


Figura 3.1: Comparação das vizinhanças da trajetória central

3.2 Algoritmos:

Cada iteração de um algoritmo viável primal-dual da trajetória central parte de um ponto $(x^k, s^k, y^k) \in \Gamma^\circ$ e de um parâmetro $\mu > 0$ satisfazendo uma medida de proximidade $\delta(x^k, s^k, \mu^k) \leq \alpha$ para algum α fixo. Dado um parâmetro $\sigma \in (0, 1)$ o algoritmo faz iterações de Newton para encontrar aproximadamente o ponto da trajetória associado a $\sigma\mu$.

Passo de Newton

Vamos definir $F(\cdot)$ uma função vetorial definida pelas condições de KKT relaxadas

$$F(x, s, y) = \begin{bmatrix} -Hx + A^T y + s - c \\ Ax - b \\ XSe - \sigma\mu e \end{bmatrix} \quad (3.16)$$

para $y \in \mathbb{R}^m$, $x, s \in \mathbb{R}^n$, S e X denotam as matrizes cujas diagonais são dadas pelos vetores s e x , respectivamente.

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 $(x, s) \in V^\circ$. Para isso

precisamos resolver a equação

$$F'(x, s, y)(d_x, d_s, d_y) = -F(x, s, y). \quad (3.17)$$

Temos que

$$-F(x, s, y) = \begin{bmatrix} 0 \\ 0 \\ \sigma\mu e - XSe \end{bmatrix},$$

pois temos como hipótese $(x, s, y) \in \Gamma^\circ$. Substituindo em (3.17) obtemos em formato matricial

$$\begin{bmatrix} -H & I & A^T \\ A & 0 & 0 \\ S & X & 0 \end{bmatrix} \begin{bmatrix} d_x \\ d_s \\ d_y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \sigma\mu e - XSe \end{bmatrix}. \quad (3.18)$$

A solução $(\Delta x, \Delta s, \Delta y)$ do sistema é o passo de Newton.

A iteração termina com uma busca ao longo da direção de Newton tal que o novo ponto $(x_\gamma, s_\gamma, y_\gamma) := (x + \gamma\Delta x, s + \gamma\Delta s, y + \gamma\Delta y)$ pertença a uma determinada vizinhança da trajetória central, ou seja, satisfaça $\delta_\gamma(x_\gamma, s_\gamma) \leq \alpha$, com

$$\delta_\gamma(x_\gamma, s_\gamma) = \left\| \frac{x_\gamma s_\gamma}{\mu_\gamma} - e \right\|, \text{ com } \mu_\gamma = \frac{x_\gamma^T s_\gamma}{n} \quad (3.19)$$

Os diversos algoritmos primais-duais de trajetória central diferem no cálculo do passo γ , na escolha de σ a cada iteração, na forma de obter a direção (d_x, d_s, d_y) e na vizinhança $N(\alpha)$ na qual os pontos devem permanecer.

A eficiência dos algoritmos está baseada no seguinte teorema.

Teorema 3.2.1 *Dados $(x, s) \in \Gamma^\circ$ e $\mu^+ > 0$ tais que $\delta(x, s, \mu^+) = \delta < 1$, seja (x^+, s^+) o resultado de um passo de Newton para 3.18 a partir de (x, s) . Então $\delta(x^+, s^+, \mu^+) \leq \frac{1}{\sqrt{8}} \frac{\delta^2}{1 - \delta}$. Em particular, se $\delta \leq 0.5$, então $\delta(x^+, s^+, \mu^+) \leq \delta^2$.*

A demonstração está feita em detalhes em [4]. Isto mostra que o passo de Newton é muito eficiente em uma ampla região.

3.3 Algoritmo de passo mais longo

Existe uma vasta literatura sobre o passo primal-dual. A questão central é a escolha de σ : σ pequeno força um passo longo, mas o passo de Newton pode não ser aceitável; σ próximo de 1 faz uma “centralização”, aproximando o ponto da trajetória central sem avançar.

Sabemos que o passo de Newton é eficiente se todos os pontos estiverem em uma região “próxima” da trajetória central. O método exposto adiante calcula a cada iteração o passo de Newton mais longo possível, tal que o novo ponto permaneça em uma vizinhança da trajetória central.

No algoritmo de passo mais longo são calculadas separadamente, a partir de (x, s, y) , duas direções extremas $d^c, d^a \in \mathbb{R}^{2n+m}$: d^c é a solução do sistema (3.18) quando $\sigma = 1$, chamada direção de *centralização*; d^a é a solução de (3.18) quando $\sigma = 0$, chamada solução *afim-escala*, d^a ;

$$\sigma = 0 : (x^a, s^a, y^a) = (x^k, s^k, y^k) + d^a \quad (3.20)$$

e

$$\sigma = 1 : (x^a, s^a, y^a) = (x^k, s^k, y^k) + d^c \quad (3.21)$$

A direção de Newton que seria obtida resolvendo o passo de Newton para $\sigma \in [0, 1]$ é uma combinação convexa destas direções, dada por

$$d = (d_x, d_s, d_y) = \sigma d^c + (1 - \sigma) d^a. \quad (3.22)$$

O valor de σ deve ser tal que o novo ponto (x^+, s^+, y^+) permaneça em uma vizinhança da trajetória central, mas determine o maior passo possível nesta região.

Por exemplo, ao escolhermos a vizinhança $N_2(0.5)$, σ deve satisfazer

$$\left\| \frac{(\sigma x^c + (1 - \sigma)x^a)(\sigma s^c + (1 - \sigma)s^a)}{\sigma \mu} - e \right\|_2 = 0.5. \quad (3.23)$$

Desenvolvendo a equação acima, obtém-se uma equação de quarto grau. Ao

resolvê-la escolhemos $\sigma \in [0, 1)$ real: caso tenha mais de uma raiz em $(0, 1)$ tomamos a menor. Se não existir raiz real, o passo de Newton completo com $\sigma = 0$ é aceito e o problema é fácil.

Definimos $(x^{k+1}, s^{k+1}, y^{k+1}) := (x, s, y) + \sigma d^c + (1 - \sigma) d^a$ e $\mu^{k+1} := \sigma \mu$. E é feita nova iteração.

A figura 3.2 ilustra uma iteração do método do passo mais longo.

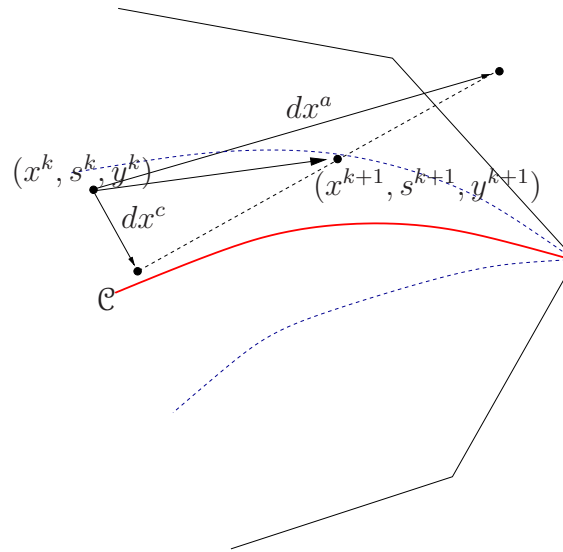


Figura 3.2: Passo mais longo

O algoritmo possui a estrutura que se segue:

Algoritmo 3.3.1 *Algoritmo do Passo mais longo*

Dados: Dados: $(x^0, s^0, y^0) \in N_2(0.5)$ e $\mu^0 > 0$ tais que $\delta(x^0, s^0, \mu^0) \leq 0.5$.

$k = 0$

REPITA

$x := x^k; s := s^k; y := y^k; \mu = \mu^k$.

Calcule a direção afim-escala d^a resolvendo o sistema (3.18) para $\sigma = 0$.

Calcule a direção afim-escala d^c resolvendo o sistema (3.18) para $\sigma = 1$.

Encontre σ solução de (3.23).

Defina o novo ponto: $(x^{k+1}, s^{k+1}, y^{k+1}) := (x, s, y) + \sigma d^c + (1 - \sigma) d^a$.

$\mu^{k+1} := \sigma \mu$

$k := k + 1$.

Até a convergência

O algoritmo descrito nesta seção foi desenvolvido por Mcshane [24] e estudado por Gonzaga [14] e Gonzaga e Bonnans [15]. Para detalhes e teoria, ver [14].

Par viável inicial. Para utilizar o algoritmo é necessário conhecer um par primal-dual (x^0, s^0) viável inicial, próximo à trajetória central. Se o centro analítico existe e é conhecido, este problema foi resolvido na tese de mestrado de Behling [1].

Supondo inicialmente que o centro analítico do conjunto

$$\{x \in \mathbb{R}^n \mid Ax = b, x \geq 0\}$$

é o ponto e , Behling considera o problema de região de confiança em torno de e :

$$\begin{aligned} & \text{minimizar} && q(e + d) \\ & \text{sujeito a} && Ad = 0 \\ & && \|d\|_2 \leq \Delta. \end{aligned} \tag{3.24}$$

Considere d_μ solução de (3.24) com $\Delta = \|d_\mu\|$ e $\mu > 0$ o multiplicador de Lagrange associado a restrição da bola. Behling mostra que $s = \mu(e - d_\mu)$ é folga dual viável para $x = e + d_\mu$, ou seja, (x, s) é par dual viável.

O teorema formulado é o seguinte:

Teorema 3.3.1 : *Seja d_μ uma direção tal que $Ad_\mu = 0$ e $d_\mu\mu = -P(\nabla f(e + d_\mu))$. Então $s = \mu(e - d_\mu)$ é folga dual viável.*

Mostra-se ainda que se $\|d_\mu\|_2 \leq \frac{1}{\sqrt{2}}$ então $\delta(x, s, \mu) \leq \frac{1}{2}$.

Ou seja, para μ suficientemente grande tal que $\|d_\mu\|_2 \leq \frac{1}{\sqrt{2}}$ o ponto $x = e + d_\mu$ na trajetória de Levenberg- Marquardt de (3.24) determina o par primal-dual viável $x = e + d_\mu, s = \mu(e - d_\mu)$ tal que $(x, s) \in N_2(0.5)$.

O resultado para centro analítico qualquer segue por uma mudança de escala. Suponha que o centro analítico primal é $\theta \in \mathbb{R}^n$. Fazendo a mudança de escala $\bar{x} = x/\theta$ e resolvendo o problema de região de confiança com raio $1/2$, o teorema

acima fornece uma direção \bar{d} . O par primal-dual associado será $(\theta(e + \bar{d}), \theta^{-1}(e - \bar{d}))$. Note que a direção $\theta\bar{d}$ é o resultado do problema de região de confiança no elipsóide definido por $\|d/\theta\| \leq 1/2$.

Um método de busca de (x, s) , já adaptado para um centro analítico $x = \theta$ qualquer esta descrito em [1]. Ele parte de um valor de μ inicial positivo, e segue a trajetória de Levenberg-Marquardt aumentando μ até que o ponto esteja na região de raio $\Delta/2$, isto é, na região correspondente ao elipsóide $\{x \in \mathbb{R}^n \mid \|(x - \theta)/\theta\| \leq \Delta/2\}$.

Em nosso algoritmo no capítulo anterior, supusemos que está disponível um algoritmo para encontrar um minimizador do modelo quadrático em uma bola de raio menor ou igual a $1/2$ (com a devida mudança de escala). Essa minimização, pelo teorema acima, fornece um par primal-dual viável associado ao valor de μ obtido no problema de região de confiança, que será usado como ponto inicial para o algoritmo de trajetória central.

Capítulo 4

Programação Quadrática

Seqüencial e Filtros

Os dois capítulos anteriores descrevem o algoritmo de região de confiança proposto nesta dissertação. Já este capítulo e o próximo são dedicados exclusivamente a utilização do algoritmo de região de confiança no método de filtro proposto por Gonzaga, Karas e Vanti em [13] para a resolução de um problema de programação não linear com com formato geral.

Neste capítulo apresentamos o problema de programação não linear a ser resolvido, algumas definições e hipóteses gerais e finalmente descrevemos brevemente o método de filtro.

A aplicação do algoritmo de região de confiança no método de filtro está descrita no capítulo 5.

4.1 Programação Não Linear com Restrições de Igualdade e Variáveis não Negativas

O problema de programação não linear estudado possui restrições de igualdades e variáveis não negativas:

$$\begin{aligned} & \text{minimizar } f(x) \\ (P) \quad & \text{sujeito a } g(x) = 0 \\ & x \geq 0. \end{aligned}$$

com $f : \mathbb{R}^n \rightarrow \mathbb{R}$ e $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$. Vamos supor ambas de classe \mathcal{C}^2 e $m \leq n$. Como já comentamos o interesse neste formato está no fato de que problemas com restrições de desigualdades podem ser transformadas em restrições de igualdade com inclusão de variáveis de folga.

Apresentamos a seguir as principais definições e hipóteses.

Denotamos a matriz Jacobiana da função g aplicada ao ponto $x \in \mathbb{R}^n$ por $A(x)$ com

$$A^T(x) = [\nabla g_1(x), \dots, \nabla g_m(x)].$$

Função Lagrangiano. A função Lagrangiano associado ao problema (P) é definida por

$$(x, \lambda, s) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}_+^n \rightarrow L(x, \lambda, s) = f(x) + \lambda^T g(x) - s^T x.$$

Sua derivada é dada por

$$\nabla_x L(x, \lambda, s) = \nabla f(x) + A^T(x)\lambda - s.$$

Para estabelecermos as condições de Karush-Kuhn-Tucker vamos assumir que vale a condição de qualificação de Mangassarian-Fromovitz (CQMF):

Um ponto $x \in \mathbb{R}^n$ viável para (P) satisfaz CQMF se e só se

- (i) *A matriz $A(x)$ tem posto completo*
- (ii) *Existe uma direção $d \in N(A(x))$ tal que $d_{I(x)} > 0$, em que $I(x) = \{i = 1, \dots, n \mid x_i = 0\}$ é o conjunto ativo em x .*

Condições Necessárias de Otimalidade (KKT). Se $\bar{x} \in \mathbb{R}^n$ é solução ótima de

(P) e satisfaz a CQMF, então existem $\bar{\lambda} \in \mathbb{R}^m$ e $\bar{s} \in \mathbb{R}^n$ tais que

$$\nabla f(\bar{x}) + \mathbf{A}(\bar{x})^T \bar{\lambda} - \bar{s} = 0 \quad (4.1)$$

$$c(\bar{x}) = 0 \quad (4.2)$$

$$\bar{x} \bar{s} = 0 \quad (4.3)$$

$$\bar{x}, \bar{s} \geq 0 \quad (4.4)$$

Além disso para toda direção $d \in \mathbb{R}^n$ tal que $A(\bar{x})^T d = 0$ e $d_{I(x)} = 0$,

$$d^T \nabla_{xx} L(\bar{x}, \bar{\lambda}, \bar{s}) d \geq 0.$$

Condição Suficiente de Otimalidade. Suponha que em um ponto $\bar{x} \in \mathbb{R}^n$ são satisfeitas as condições de KKT com multiplicadores $\bar{\lambda}, \bar{s}$. Suponha também que \bar{x} satisfaz CQMF e além disso

$$\forall d \in N(A(\bar{x})) \text{ com } d_{I(x)} = 0, \quad d^T \nabla_{xx} L(\bar{x}, \bar{\lambda}, \bar{s}) d > 0.$$

Então \bar{x} é solução ótima de (P).

Pontos interiores. Em nosso algoritmo vamos exigir que

$$x^k > 0, \quad k = 0, 1, \dots$$

Vamos ainda definir a função inviabilidade:

Medida de inviabilidade. A inviabilidade é medida por

$$x \in \mathbb{R}_{++}^n \rightarrow h(x) := \|g(x)\|,$$

em que $\|\cdot\|$ é uma norma qualquer em \mathbb{R}^n . Usaremos a norma euclidiana.

Na próxima seção descrevemos o método de Programação Quadrática Sequencial (PQS) e o algoritmo de filtro.

4.2 Programação Quadrática Sequencial, Restauração Inexata e um Método de Filtro

A família dos métodos de Programação Quadrática Sequencial originou-se da tese de Wilson [34] em 1963 e a partir de então, vários pesquisadores tem se interessado pelo tema. Na década de 70 popularizou-se com os trabalhos de Garcia-Palomares e Mangasarian [11], Han [16, 17] e Powell [28, 29, 30]. Um estudo bastante detalhado sobre o assunto é feito por Boggs e Tolle em [3]. Uma referência básica em português é [20].

Num problema geral de otimização temos uma função objetivo e restrições geralmente não lineares. A estratégia básica de PQS consiste, em cada passo, em fazer uma aproximação quadrática da função objetivo e uma aproximação linear das restrições.

A convergência local rápida (superlinear ou quadrática) deste método é estabelecida devido a equivalência com o método de Newton aplicado às condições de KKT do problema, veja Bonnans e outros [4]. Para convergência global pode-se usar busca linear ou região de confiança. A prova da convergência do uso de busca linear e região de confiança em PQS pode ser encontrada em [4] e [6] respectivamente.

Um algoritmo de Programação Não Linear deve trabalhar com dois objetivos normalmente conflitantes, reduzir a função objetivo e obter viabilidade. Tradicionalmente otimalidade e viabilidade são combinadas em uma única função como no caso dos métodos que usam funções de mérito, nos métodos de penalidade, Lagrangiano aumentado e nos métodos de restauração inexata [21, 22], entre outros.

Os Métodos de Filtro. Como alternativa aos métodos que combinam viabilidade e otimalidade em uma única função, Fletcher e Leyffer [8] introduziram os chamados Métodos de Filtro para globalização em PNL. Nestes métodos inviabilidade e otimalidade são trabalhados separadamente, priorizando a redução da inviabilidade.

São baseados em PQS e caracterizados pelo uso do conceito de dominância - conceito da otimização multi-critério :

$$x \text{ domina } y \text{ se, e somente se, } f(y) \geq f(x) \text{ e } h(y) \geq h(x)$$

A idéia do algoritmo de filtro é interferir o mínimo possível nas iterações de PQS, mas induzir a convergência.

O filtro F é uma lista de pares (f^i, h^i) , em que cada par (f^i, h^i) define uma região proibida em \mathbb{R}^2 , $\{(f, h) \mid f \geq f^i, h \geq h^i\}$, que é composta pelos pontos de \mathbb{R}^2 dominados por (f^i, h^i) , e uma região proibida em \mathbb{R}^n , $\{x \in \mathbb{R}^n \mid f(x) \geq f^i, h(x) \geq h^i\}$.

O algoritmo trabalha com o filtro F_k em \mathbb{R}^2 e com o conjunto de pontos proibidos associados, \mathcal{F}_k em \mathbb{R}^n . O conjunto não é construído. Dado um ponto $x \in \mathbb{R}^n$, é fácil verificar se ele pertence à região proibida, fazendo apenas comparações em \mathbb{R}^2 .

A estrutura do filtro é simples e facilmente implementável, visto que apenas um número de pares são armazenados.

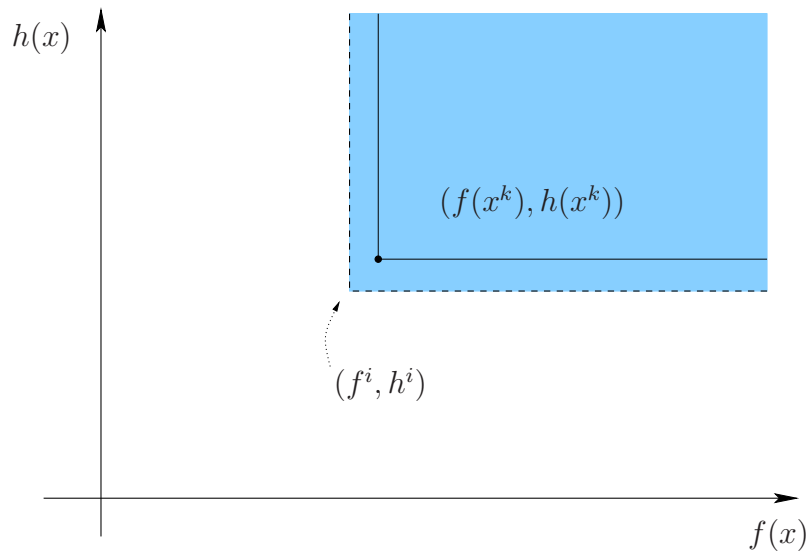


Figura 4.1: Construção do par (f^k, h^k)

Cada elemento (f^i, h^i) do filtro é construído a partir de um iterando x^k fazendo

$$f^i = f(x^k) - \alpha h(x^k), \quad h^i = h(x^k) - \alpha h(x^k)$$

em que $\alpha \in (0, 1)$ é uma constante fixa, por exemplo, $\alpha = 0.2$, como mostramos na figura 4.1.

Vamos descrever uma iteração do algoritmo de filtro:

Consideremos um ponto x^k , um filtro F_k . No início de cada iteração, o par

$$(f(x^k) - \alpha h(x^k), (1 - \alpha)h(x^k)) \text{ com } \alpha \in (0, 1),$$

é temporariamente introduzido no filtro, definindo o chamado *filtro temporário*, denotado por \bar{F}_k , e definindo o conjunto de pontos proibidos em \mathbb{R}^n associado ao filtro temporário, denotado por $\bar{\mathcal{F}}_k$. Após a iteração ser completada, este par será adicionado ao filtro permanente F_k apenas se a iteração não produzir um decréscimo em f . Sempre que um novo par é introduzido no filtro, os pares dominados podem ser eliminados do filtro.

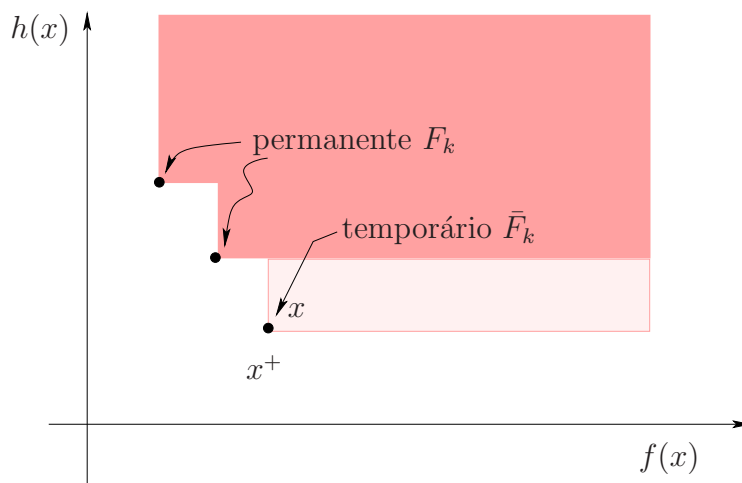


Figura 4.2: Uma iteração do método de filtro

Na figura 4.2 por abuso de notação indicamos $x = (f(x^k), h(x^k))$ e $x^+ = (f(x^{k+1}), h(x^{k+1}))$. Os círculos pretos representam elementos do filtro.

O algoritmo de filtro que vamos estudar foi proposto por Gonzaga, Karas e Vanti [13]. É um algoritmo globalmente convergente para programação não linear e segue as idéias de restauração inexata de Martínez [22] e Martínez e Pilotta [21].

A principal idéia dos métodos de restauração inexata é que a viabilidade, por ser um aspecto muito importante do problema, pode ser tratada independentemente da otimalidade. Se diferenciam dos demais métodos de restauração (métodos viáveis e quase viáveis) pela metodologia de relaxarem as restrições quando longe da solução

e gradualmente diminuir a inviabilidade quando o método avança, o que é útil no caso de restrições fortemente não lineares.

Desta forma, seguindo as idéias de restauração inexata, o algoritmo que vamos estudar gera uma seqüência $(x^k)_{k \in \mathbb{N}}$, sendo que cada iteração inicia em um ponto x^k e tem duas fases :

- (i) Fase de viabilidade: gera-se um ponto z^k em que a medida de inviabilidade h é reduzida;
- (ii) Fase de otimalidade: a partir de z^k , gera-se x^{k+1} com $f(x^{k+1}) < f(z^k)$. Estas duas fases são totalmente independentes, o único acoplamento entre elas é estabelecido pelo filtro. Elas são descritas a seguir.

Antes de apresentarmos o algoritmo precisamos da definição de Linearização do conjunto viável que é comentada na seção fase de otimalidade.

Linearização do conjunto viável. A cada ponto z viável $\in \mathbb{R}_+^n$ associamos o seguinte poliedro:

$$L(z) = \{x \in \mathbb{R}^n \mid A(z)^T(x - z) = 0, x \geq 0\}.$$

Algoritmo 4.2.1 *Algoritmo de filtro.*

Dados: $x^0 \in \mathbb{R}^n$, $F_0 = \emptyset$, $\mathcal{F}_0 = \emptyset$, $\alpha \in (0, 1)$.

$k = 0$

REPITA

$$(\tilde{f}, \tilde{h}) = (f(x^k) - \alpha h(x^k), (1 - \alpha)h(x^k)).$$

Construa o conjunto $\bar{F}_k = F_k \cup \{(\tilde{f}, \tilde{h})\}$.

Defina o conjunto $\bar{\mathcal{F}}_k = \mathcal{F}_k \cup \{x \in \mathbb{R}^n; \mid f(x) \geq \tilde{f}, h(x) \geq \tilde{h}\}$.

Fase de viabilidade:

se $h(x^k) = 0$, então faça $z^k = x^k$

senão, calcule $z^k \notin \bar{\mathcal{F}}_k$ tal que $h(z^k) < (1 - \alpha) h(x^k)$.

se impossível, então pare com insucesso.

Fase de otimalidade:

se z^k é estacionário, então pare com sucesso

senão, calcule $x^{k+1} \notin \bar{\mathcal{F}}_k$ tal que $x^{k+1} \in L(z^k)$ e $f(x^{k+1}) \leq f(z^k)$.

Atualização do filtro:

se $f(x^{k+1}) < f(x^k)$, então

$$F_{k+1} = F_k, \quad \mathcal{F}_{k+1} = \mathcal{F}_k \quad (\text{iteração-}f)$$

senão,

$$F_{k+1} = \bar{F}_k, \quad \mathcal{F}_{k+1} = \bar{\mathcal{F}}_k \quad (\text{iteração-}h)$$

$$k = k + 1.$$

As figuras 4.3 e 4.4 ilustram uma iteração completa do algoritmo, assim como o filtro para a próxima iteração, onde, por abuso de notação, indicamos $x = (f(x^k), h(x^k))$, $z = (f(z^k), h(z^k))$ e $x^+ = (f(x^{k+1}), h(x^{k+1}))$.

Na figura 4.3 a iteração é do tipo f e conseqüentemente o par (\tilde{f}, \tilde{h}) não é incluído no filtro para a próxima iteração. A região proibida é toda a parte hachurada, sendo que a parte vermelha é temporária e deixa de ser proibida para a iteração seguinte.

Já na figura 4.4 a iteração é do tipo h , o par (\tilde{f}, \tilde{h}) é incluído no filtro e a região temporária passa a ser permanente.

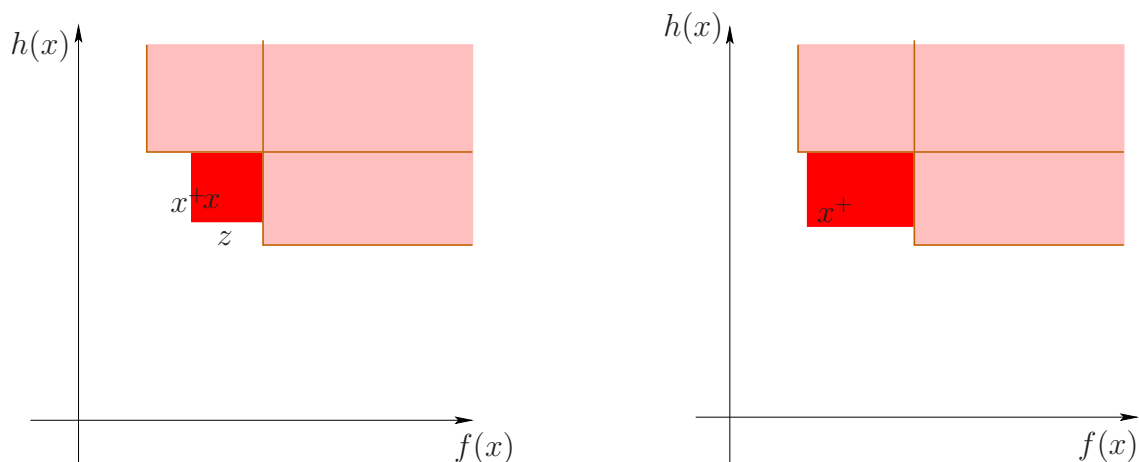


Figura 4.3: Uma iteração f e o filtro para a próxima iteração

A convergência global deste algoritmo é estudada em [13]. Mostra-se que sob condições razoáveis sobre os passos de restauração e otimalidade e usando as CQMF, qualquer seqüência (x^k) gerada pelo algoritmo tem ponto de acumulação estacionário (isto é, que satisfaz as condições de KKT.)

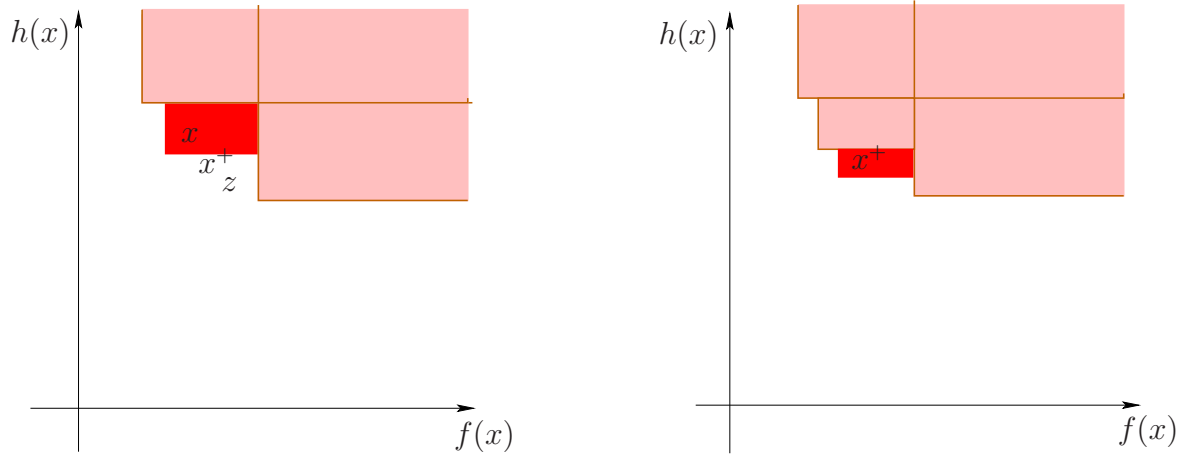


Figura 4.4: Uma iteração h e o filtro para a próxima iteração

4.2.1 A Fase de Restauração ou Fase de Viabilidade

Na fase de restauração partimos de um ponto $x^k > 0$, $x^k \in \bar{\mathcal{F}}_k$, e procuramos um ponto $z^k > 0$ tal que $z^k \notin \bar{\mathcal{F}}_k$ e $h(z^k) \leq \beta h(x^k)$, com $\beta \in (0, 1)$ fixo. O problema é resolvido usando uma ou mais iterações de um algoritmo para minimizar h no conjunto \mathbb{R}_+^n , parando quando as duas condições acima são satisfeitas.

Passo de Newton. Desprezando as restrições $x \geq 0$, o passo de Newton para a solução de $g(x) = 0$ é uma solução do sistema indeterminado

$$A(x)d = -g(x).$$

A direção de Newton normal é obtida forçando $d \perp N(A(x))$, ou seja, fazendo $d_N = A(x)^T y$ e resolvendo o sistema

$$A(x)A(x)^T y = -g(x).$$

Se $A(x)$ tem posto completo, obtem-se

$$d_N = -A(x)^T (A(x)A(x)^T)^{-1} g(x).$$

Se $x^k + d_N > 0$, então o passo d_N (com ou sem busca unidirecional) pode

ser usado na restauração. Mas em geral isto não é garantido.

Regiões de Confiança. Se os passos de Newton puros não são eficientes, deve-se usar um algoritmo de região de confiança para resolver o problema

$$\begin{aligned} &\text{minimizar} && \frac{1}{2}\|g(x)\|^2 \\ &\text{sujeito a} && x \geq 0. \end{aligned}$$

A resolução é feita usando o modelo quadrático de Gauss-Newton, ver [6]

$$\frac{1}{2}\|g(x+d)\|^2 \approx \frac{1}{2}\|g(x)\|^2 + g(x)^T A(x)^T d + \frac{1}{2}d^T A(x)^T A(x)d.$$

A matriz $A(x)^T A(x)$ é semi definida positiva e não é definida positiva para $m < n$. O modelo quadrático

$$q(d) = g(x)^T A(x)d + \frac{1}{2}d^T A(x)^T A(x)d$$

pode ser usado em algoritmos de região de confiança com o seguinte cuidado: se d_N pertence à região de confiança, então deve-se usar o passo d_N .

Concluimos que a fase de restauração é um caso especial (sem restrições de igualdade) do problema P_O que está detalhado no capítulo 5.

A resolução deste problema, usando um método de pontos interiores baseado no algoritmo de Coleman e Li [5], foi desenvolvida na tese de doutorado de Juliano B. Francisco [9].

4.2.2 Fase de Otimalidade

Na fase de restauração obtemos $z > 0$ tal que $\|g(z)\| < \|g(x)\|$ e $z \notin \bar{\mathcal{F}}$. No fase de otimalidade pretende-se encontrar $w \in \mathbb{R}^n$ tal que $f(w) < f(z)$ e ainda sem perder muito da viabilidade adquirida com o passo de restauração.

A fase de otimalidade utiliza um algoritmo de regiões de confiança aplicado

ao problema de minimização de $f(\cdot)$ em $L(z)$:

$$(P_O) \quad \begin{array}{ll} \text{minimizar} & f(x) \\ \text{sujeito a} & A(z)(x - z) = 0 \\ & x \geq 0 \\ & x \notin \bar{\mathcal{F}}. \end{array}$$

A resolução é feita executando passos de região de confiança para o problema acima sem a restrição $x \notin \bar{\mathcal{F}}$, mas reduzindo a região de modo a satisfazer essa restrição. Vamos utilizar o algoritmo de região de confiança proposto nesta dissertação para determinar o passo de otimalidade detalhado no capítulo 5.

Aqui vamos somente comentar a construção do modelo quadrático de $f(\cdot)$ em torno do ponto z .

Problemas com restrição de igualdade. Como fizemos no problema de viabilização vamos inicialmente desprezar as restrições $x \geq 0$ e $x \notin \bar{\mathcal{F}}$ e estudar um passo ao longo de $N(A(z))$.

No passo de otimalidade devido à condição (4.4) e ao fato de o Lagrangiano incorporar informações sobre a curvatura das restrições, faz-se o seguinte: em vez de minimizar f , calcula-se um estimador λ de multiplicador ótimo de Lagrange e minimiza-se o modelo quadrático do Lagrangiano associado a este λ restrito ao conjunto $N(A(z))$.

Multiplicador de Lagrange. Caso $\bar{\lambda}$ seja desconhecido, podemos calcular um estimador para os multiplicadores de Lagrange utilizando a primeira condição de KKT (4.1). A solução do problema de mínimos quadráticos para esta igualdade é dada por

$$\tilde{\lambda} = -(A(x)A(x)^T)^{-1}A(x)\nabla f(x).$$

Redução do Modelo Quadrático do Lagrangiano

Considere $z \in \mathbb{R}^n$ e suponha $\lambda \in \mathbb{R}^m$ um estimador para o multiplicador ótimo $\bar{\lambda}$. Já comentamos que se x é ótimo local de (P) então existe $\bar{\lambda} \in \mathbb{R}^m$ que satisfaz as condições de KKT e como $\text{posto}(A(x)) = m$ ele é único.

O modelo quadrático da variação do Lagrangiano em z e λ é dado por

$$m(d) = \nabla f(z)^T d + \lambda^T A(z)d + \frac{1}{2} d^T H d,$$

com $H = \nabla_{xx} L(z, \lambda)$. Usando $A(z)d = 0$ na expressão para $m(\cdot)$, o problema fica

$$\begin{aligned} \text{minimizar} \quad & \nabla f(z)^T d + \frac{1}{2} d^T H d \\ \text{sujeito a} \quad & A(z)d = 0 \end{aligned} \tag{4.5}$$

Aplicando as condições de KKT temos:

$$\nabla f(z) + H d + A(z)^T \mu = 0 \tag{4.6}$$

$$A(z)^T h = 0 \tag{4.7}$$

Seja (d_N, μ) a solução obtida. Se o passo d_N satisfaz as restrições de (P_O) ele pode ser utilizado. Neste caso, μ é um estimador de multiplicador de Lagrange para a próxima iteração.

Concluimos que no passo tangente irrestrito procuramos minimizar ao longo de $N(A(z))$ um modelo quadrático de $f(\cdot)$ dado por

$$d \in \mathbb{R}^n \rightarrow q(d) = \nabla f(z)^T d + \frac{1}{2} d^T \nabla_{xx} L(z, \lambda) d,$$

isto é, um modelo em que a Hessiana de $f(\cdot)$ é substituída pela Hessiana do Lagrangiano com um estimador de multiplicadores λ . Para este problema o passo de Newton é dado por (4.6) e (4.7).

Se o passo de Newton não for aceitável é necessário usar regiões de confiança para restringir o passo. Faremos isto para o problema completo (P_O) .

Passo tangente para o problema restrito.

A construção do modelo quadrático para o problema restrito é mais complexa. A estimativa do multiplicador ótimo é mais difícil e não será estudada nesta dissertação. Ver Bonnans e outros em [4] e Conn e outros em [6]. Em [4] mostra-se que se a minimização de um modelo quadrático é feita no conjunto $L(z)$, então (como ocorre no caso irrestrito acima) os multiplicadores (λ, s) obtidos nessa minimização fornecem um estimador para a iteração seguinte.

Suponha que (λ, s) é um estimador de multiplicadores para (P_O) . A discussão do problema irrestrito motiva o seguinte: o modelo quadrático de $f(\cdot)$ a utilizar em (P_O) é o mesmo modelo usado no caso irrestrito, isto é,

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

em que $H = \nabla_{xx}^2 \left(\sum_{i=1}^n \lambda_i c_i(z) \right)$. As restrições $x \geq 0$ são lineares e não influenciam H .

Capítulo 5

Utilização do Algoritmo de Região de Confiança no Método de Filtro

O algoritmo de região de confiança que apresentamos foi desenvolvido para ser utilizado na fase de otimalidade do algoritmo de filtro 4.2.1, aplicado na resolução de um problema de programação não linear com formato geral. Das duas fases do método de filtro escolhemos a de otimalidade, já que na de restauração não temos restrições lineares. Como comentamos, o algoritmo proposto pode ser aplicado também na obtenção do passo de restauração.

Lembramos que o problema de programação não linear a ser resolvido com o algoritmo de filtro é o seguinte:

$$\begin{aligned} & \text{minimizar} && f(x) \\ & \text{sujeito a} && g(x) = 0 \\ & && x \geq 0 \end{aligned} \tag{5.1}$$

com $f : \mathbb{R}^n \rightarrow \mathbb{R}$ e $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$. Vamos supor ambas de classe \mathcal{C}^2 e $m \leq n$.

Hipótese importante. Neste capítulo temos como hipótese que passo de restauração z é conhecido e $z \geq 0$.

5.1 Fase de otimalidade

A fase de otimalidade utiliza um algoritmo de região de confiança aplicado ao problema de minimização de $f(\cdot)$ no conjunto linearizado $L(z) = \{x \in \mathbb{R}^n \mid A(z)(x - z) = 0, x \geq 0\}$:

$$(P_O) \quad \begin{array}{ll} \text{minimizar} & f(x) \\ \text{sujeito a} & A(z)(x - z) = 0 \\ & x \geq 0 \\ & x \notin \bar{\mathcal{F}}. \end{array}$$

Conforme discutido no final da seção 4.2.2, o modelo quadrático da variação de f que vamos utilizar em (P_O) é

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

em que

$$H = \nabla_{xx}^2 \left(\sum_{i=1}^n \lambda_i c_i(z) \right).$$

O problema a ser resolvido é:

$$\begin{array}{ll} \text{minimizar} & c^T d + \frac{1}{2} d^T H d \\ \text{sujeito a} & A(z)d = 0 \\ & \|d\|_\infty \leq \Delta \\ & z + d \geq 0 \\ & z + d \notin \bar{\mathcal{F}}, \end{array} \tag{5.2}$$

com $c = \nabla f(z)$.

5.2 O algoritmo

No capítulo 2 apresentamos o algoritmo de região de confiança baseado em trajetória central. Para seguir a trajetória central escolhemos um algoritmo de pontos interiores, apresentado no capítulo 3.

A teoria primal-dual é feita para o formato primal de programação quadrática. O conteúdo desta seção tem como objetivo reescrever o problema (5.2) de forma a possibilitar que apliquemos a teoria apresentada no capítulo 3. Fazendo $x = z + d$ em (5.2) obtemos

$$\begin{aligned}
&\text{minimizar} && c^T(x - z) + \frac{1}{2}(x - z)^T H(x - z) \\
&\text{sujeito a} && A(z)x = A(z)z \\
&&& \|x - z\|_\infty \leq \Delta \\
&&& x \geq 0 \\
&&& x \notin \bar{\mathcal{F}},
\end{aligned} \tag{5.3}$$

que é equivalente a

$$\begin{aligned}
&\text{minimizar} && c^T x + \frac{1}{2}x^T Hx \\
&\text{sujeito a} && Ax = b \\
&&& \|x - z\|_\infty \leq \Delta \\
&&& x \geq 0 \\
&&& x \notin \bar{\mathcal{F}},
\end{aligned} \tag{5.4}$$

com $A = A(z)$, $A(z)z = b$.

5.2.1 O centro analítico do conjunto formado pelas restrições

Vimos que o problema de inicialização de um algoritmo primal-dual para programação quadrática foi resolvido em [1] quando o centro analítico do conjunto das restrições for conhecido. Nesta seção ajustamos as restrições do problema (5.4) de modo a satisfazer a hipótese do teorema apresentado em [1].

Restrição da caixa .

Consideremos inicialmente a restrição da caixa, $\|x - z\|_\infty \leq \Delta$ determinada pela precisão do modelo. Veja que $\|x - z\|_\infty \leq \Delta$ é equivalente a $-\delta + z \leq x \leq z + \delta$, $\delta = \Delta e$. Vamos denotar esta caixa por \mathcal{V} .

O lema seguinte determina uma importante propriedade do ponto z .

Lema 6 *O centro analítico da caixa \mathcal{V} é $\theta = z$.*

Prova: A função barreira logarítmica associada a \mathcal{V} é definida por:

$$x \in \mathcal{V}^\circ \mapsto p(x) = -\frac{1}{2} \sum_{i=1}^n \log(x_i - z_i + \Delta) + \log(z_i + \Delta - x_i)$$

Devemos mostrar que o minimizador da função p é $x = z$.

Na forma de produto p é escrita como

$$p(x) = -\frac{1}{2} \log \prod_{i=1}^n (x_i - z_i + \Delta)(z_i + \Delta - x_i).$$

Como a função logarítmica é uma função crescente, é fácil ver que o que procuramos é na verdade um maximizador para o produto. Cada fator do produto em questão é independente, logo basta determinarmos o maximizador de cada um, que é trivialmente $x_i = z_i$. Ou seja, o centro analítico de \mathcal{V} é seu centro geométrico $x = z$.

□

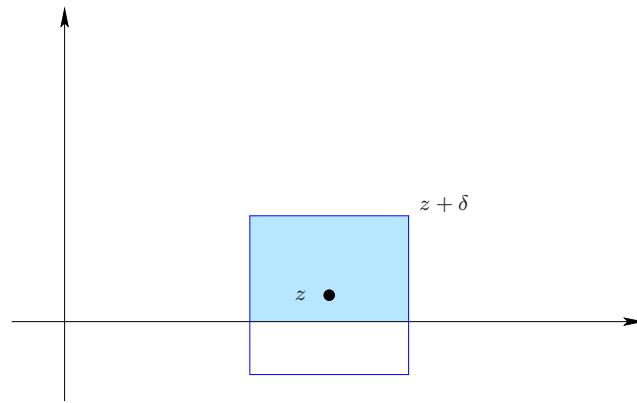
De forma similar mostra-se que o centro analítico de uma caixa qualquer no \mathbb{R}^n , $[l, u] = \{x \in \mathbb{R}^n \mid l \leq x \leq u\}$ é o ponto $\theta = \frac{l + u}{2}$.

No problema a ser resolvido temos ainda além da restrição da caixa três restrições: a de positividade, a da variedade linear e a de filtro. Nesta seção não vamos considerar a restrição do filtro.

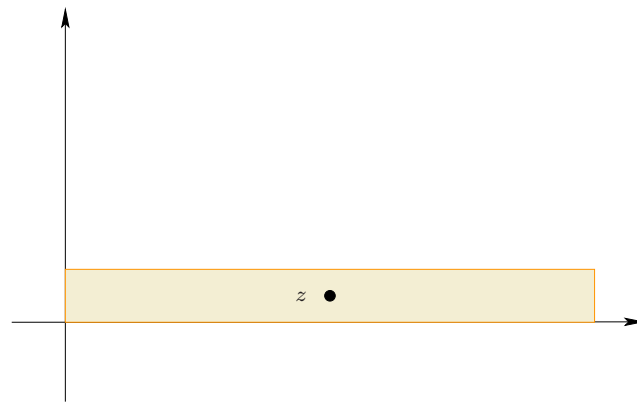
Restrição de Positividade.

O centro geométrico da caixa $[-\delta + z, \delta + z]$ é o ponto z . Ao restringirmos a caixa ao primeiro ortante obtemos uma nova caixa $[0, z + \delta]$ e perdemos a propriedade do ponto z .

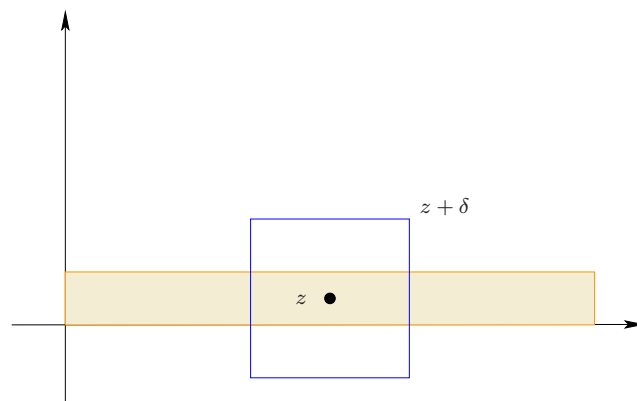
Pretendemos construir uma caixa $[l, u]$, $l \geq 0$ com centro geométrico $z = \frac{l + u}{2}$. A construção proposta é a seguinte: considere a maior caixa no ortante positivo com centro geométrico z , $C = [0, 2z]$. A interseção entre as caixas $[-\delta + z, \delta + z]$ e $[0, 2z]$ é a caixa $[l, u]$ com $l = \max\{0, z - \delta\}$, $u = \min\{2z, z + \delta\}$ e com centro geométrico $z = \frac{l + u}{2}$. Esta construção está ilustrada na figura 5.1. A caixa resultante está contida na região de raio Δ onde estamos supondo o modelo confiável, e também contida no primeiro ortante.



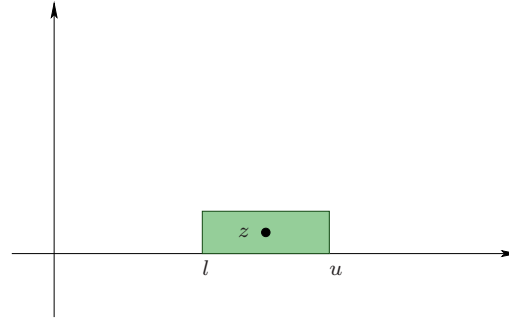
(a) Caixa de tamanho Δ



(b) Caixa C



(c) Interseção entre as duas caixas



(d) Caixa final

Figura 5.1: Construção da caixa $[l, u]$.

Retrições lineares. Como z minimiza a função barreira logarítmica associada a caixa $[l, u]$ e satisfaz $Ax = b$, segue que z é centro analítico do conjunto $\Omega = \{x \in [l, u] \mid Ax = b\}$.

Note que pela construção feita temos que a hipótese do centro analítico de Ω ser conhecido sempre será satisfeita no nosso problema possibilitando a construção de um par primal dual viável inicial para seguir a trajetória em Ω .

O problema a ser resolvido é dado por:

$$\begin{aligned}
 \text{minimizar} \quad & m(x) = c^T x + \frac{1}{2} x^T H x \\
 \text{sujeito a} \quad & Ax = b \\
 & l \leq x \leq u \\
 & x \notin \bar{\mathcal{F}}.
 \end{aligned} \tag{5.5}$$

Uma iteração do algoritmo de região de confiança baseado em trajetória central para resolver (5.5) determina o par inicial primal dual viável (x^0, s^0) , x^0 na trajetória de Levenberg Marquardt. Se x^0 for aceitável reduz $m(\cdot)$ seguindo a trajetória central em Ω até encontrar um ponto não aceitável; caso contrário segue a trajetória de Levenberg-Marquardt entre z e o ponto x^0 até encontrar um ponto aceitável.

Um ponto $x = z + d$ obtido é aceitável se $ared(d) \geq \alpha pred(d)$, $x \notin \bar{\mathcal{F}}$ e $x \geq 0$.

Ajuste de Δ . Para a iteração seguinte, faz-se:

Se $ared(d) > 0.75 pred(d)$, $\Delta = \max\{\Delta/2, 2\|d\|_\infty\}$.

Senão, $\Delta = \min\{\Delta, 1.5\|d\|_\infty\}$.

5.2.2 Formato primal de programação quadrática

A teoria de algoritmos primais-duais é feita para o formato primal de programação quadrática. Faremos uma translação e introdução de variáveis de folga de forma a deixar o problema (5.5) neste formato.

Transformação do problema.

Definimos $x^{(1)} := x - l$ (translação) e introduzimos variáveis de folga $x^{(2)} \in \mathbb{R}^n$. Definindo

$$b^{(1)} := b - Al, \quad b^{(2)} := u - l,$$

obtemos

$$\begin{aligned} \text{minimizar} \quad & c^T(x^{(1)} + l) + \frac{1}{2}(x^{(1)} + l)^T H(x^{(1)} + l) \\ \text{sujeito a} \quad & Ax^{(1)} = b^{(1)} \\ & x^{(1)} + x^{(2)} = b^{(2)} \\ & x^{(2)} \geq 0. \end{aligned} \tag{5.6}$$

Levando em consideração que $c^T l$ e $l^T H l$ são constantes resolver (5.6) é equivalente a resolver

$$\begin{aligned} \text{minimizar} \quad & [c + Hl]^T x^{(1)} + \frac{1}{2}x^{(1)T} H x^{(1)} \\ \text{sujeito a} \quad & Ax^{(1)} = b^{(1)} \\ & x^{(1)} + x^{(2)} = b^{(2)} \\ & x^{(2)} \geq 0. \end{aligned} \tag{5.7}$$

Vamos definir $c^{(1)} = c + (Hl)^T$ e escrever na forma matricial o problema anterior:

$$\text{Fazendo } \bar{H} = \begin{bmatrix} H & 0 \\ 0 & 0 \end{bmatrix}, \bar{A} = \begin{bmatrix} A & 0 \\ I & I \end{bmatrix}, \bar{x} = \begin{bmatrix} x^{(1)} \\ x^{(2)} \end{bmatrix}, \bar{c} = \begin{bmatrix} c^{(1)} \\ 0 \end{bmatrix} \text{ e}$$

$$\bar{b} = \begin{bmatrix} b^{(1)} \\ b^{(2)} \end{bmatrix} \text{ obtemos}$$

$$\begin{aligned} & \text{minimizar} && \bar{c}^T \bar{x} + \frac{1}{2} \bar{x}^T \bar{H} \bar{x} \\ & \text{sujeito a} && \bar{A} \bar{x} = \bar{b} \\ & && \bar{x} \geq 0 \end{aligned} \tag{5.8}$$

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

Com a introdução de variáveis de folga $x^{(2)}$, o tamanho do problema dobrou, já que $\bar{x} \in \mathbb{R}^{2n}$ e no formato original $x \in \mathbb{R}^n$. Por outro lado \bar{H} continua simétrica semidefinida positiva e $\text{rank}(\bar{A}) = m + n$.

As vantagens do novo formato do problema (5.5) são os fatos de que ele simplifica a exposição das condições de KKT, e os algoritmos de pontos interiores para Programação Quadrática se baseiam neste formato.

Ponto inicial do algoritmo.

Temos que o centro analítico de $\Omega = \{x \in [l, u] \mid Ax = b\}$ é o ponto $\theta = z$. Após a transformação do problema o centro analítico de $\bar{\Omega} = \{\bar{x} \in \mathbb{R}^{2n} \mid \bar{A}\bar{x} = \bar{b}, \bar{x} \geq 0\}$ é dado por $\bar{\theta} = \begin{bmatrix} \frac{u-l}{2} \\ \frac{u-l}{2} \end{bmatrix}$.

Para determinar o par inicial viável usamos o método de busca proposto na dissertação de Behling [1]. A busca é feita na trajetória de Levenberg-Marquardt determinada pelo subproblema quadrático de região de confiança aplicado em (5.8) no ponto $\bar{\theta}$.

Aplicação do algoritmo de região de confiança na fase de otimalidade do método de filtro.

Consideremos $z \geq 0$ ponto obtido com o passo de restauração.

Algoritmo 5.2.1 *Fase de otimalidade usando algoritmo de região de confiança baseado na função barreira*

Dados: $z \in \mathbb{R}_+^n$, $\alpha \in (0, 1)$ fixo, ε e $\Delta > 0$ (Por exemplo, $\alpha = 0.25$, $\Delta = 1$.)

Construa o modelo $x \mapsto m(x) = \nabla f(z)^T x + \frac{1}{2} x^T H x$.

Construa a caixa $[l, u]$ proposta na seção 5.2.1

Transforme o formato de (5.4) para o formato (5.8)

Obtenha o o par viável inicial (\bar{x}^0, \bar{s}^0) e μ^0 multiplicador de Lagrange a partir de $\bar{\theta}$

Obtenha o ponto inicial x_μ na trajetória de Levenberg-Marquardt em Ω dado pelas n primeiras componentes do ponto \bar{x}^0 adicionadas ao ponto l .

Calcule $ared(d_\mu) = f(z) - f(x_\mu)$, $pred = m(z) - m(x_\mu)$

se $ared(d_\mu) \leq \alpha pred(d_\mu)$, ou $x_\mu \in \bar{\mathcal{F}}$, ou $x_\mu < 0$

 siga a trajetória de Levenberg-Marquardt para $\mu \in (\mu^0, +\infty)$ até encontrar um ponto x_μ aceitável: $ared(d_\mu) \geq \alpha pred(d_\mu)$, $x_\mu \notin \bar{\mathcal{F}}$ e $x_\mu \geq 0$

senão

 siga a trajetória central com $\mu < \mu^0$ até encontrar um ponto x_μ com $ared(d_\mu) \leq \alpha pred(d_\mu)$ ou $x_\mu \in \bar{\mathcal{F}}$ ou $x_\mu < 0$ ou $\mu \leq \varepsilon$

Fim.

Minimização em uma caixa maior. Na construção da caixa podemos utilizar algum método de identificação de conjunto ativo para eliminar algumas das restrições de não negatividade.

O seguinte procedimento pode ser usado: avaliamos o gradiente de f em z , projetado sobre núcleo de $A(z)$, Pc . Calculamos o ângulo entre Pc e cada um dos vetores e_i da base canônica, dado por $\phi_i = [Pc]_i / \|Pc\|$. Se $\phi_i < \varepsilon$, com por exemplo $\varepsilon = -0.2$, a restrição $x_i \geq 0$ pode ser ignorada e com isso obtemos uma caixa maior $[l, u]$, fazendo para essas componentes $l_i = z_i - \Delta$, $u_i = z_i + \Delta$. Veja figura 5.2 a seguir.

Entretanto como $[l, u] \not\subset \mathbb{R}_+^n$ os pontos x_μ da trajetória podem não pertencer ao primeiro ortante e devemos incluir a restrição de $x_\mu > 0$ para caracterizar um ponto aceitável no algoritmo, como fizemos em 5.2.1.

Aplicamos o algoritmo de região de confiança proposto no trabalho na fase de otimalidade do algoritmo de filtro para o problema (5.1). Mas como já comen-

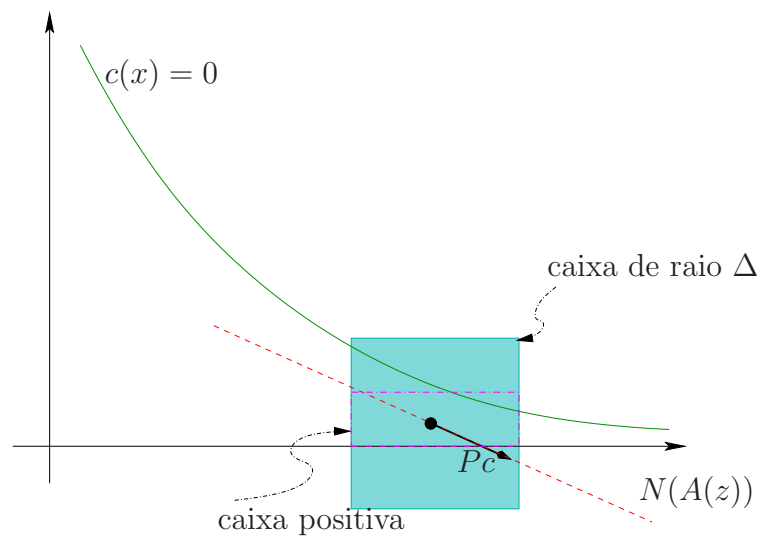


Figura 5.2: Minimização em uma caixa maior

tamos poderíamos aplicá-lo também na fase de viabilidade, minimizando o modelo quadrático da função $\frac{1}{2}\|g(\cdot)\|^2$ sem a restrição da variedade afim.

Considerações finais

Estudamos ao longo deste trabalho um algoritmo de região de confiança e sua aplicação em um método de filtro.

Algoritmos de região de confiança que utilizam regiões do tipo caixa são atraentes devido ao tamanho da caixa em relação a bola euclidiana. No entanto, a presença de restrições lineares eleva seu custo computacional. Diante deste inconveniente, apresentamos no capítulo 2 um método de região de confiança alternativo, baseado em trajetória central. Fixamos o tamanho de uma caixa e com um algoritmo de pontos interiores avançamos ou recuamos na trajetória central, até encontrarmos um ponto que reduza o modelo e que pertença a uma região onde este seja confiável. Temos que neste algoritmo as regiões de confiança são os conjuntos de nível da função barreira logarítmica definida na caixa.

Já era conhecido de [1] que a trajetória central está muito próxima da trajetória de Levenberg-Marquardt para um raio $\Delta \leq 1/2$. Resultado que foi obtido no estudo da inicialização do mesmo algoritmo em meio ao contexto primal-dual. O lema 3, obtido durante o estudo do algoritmo de região de confiança confirmou o resultado. O lema afirma que para o mesmo raio Δ os conjuntos de nível da função barreira e as bolas euclidianas são muito parecidas e conseqüentemente temos que as trajetórias de Levenberg-Marquardt e a central são muito próximas.

Com este lema conseguimos caracterizar melhor o algoritmo de região de confiança proposto. Ele corresponde a uma busca curvilínea tipo Armijo sobre uma trajetória híbrida: para $\Delta \leq 1/2$ a trajetória é a de Levenberg-Marquardt, para $\Delta > 1/2$ é a trajetória central. E as regiões de confiança são conjuntos que evoluem de bolas euclidianas e tendem a caixa completa (curvas de nível da função barreira).

Utilização do algoritmo de região de confiança proposto.

Problemas de minimização de um modelo quadrático sujeito a uma região de confiança e restrições lineares devem ser resolvidos frequentemente em programação não linear. O algoritmo de região de confiança proposto foi desenvolvido para resolver a fase de otimalidade do método de filtro, proposto em [13], aplicado na resolução de um problema de PNL com formato bem geral de restrições de igualdade e variáveis não negativas. Nesta fase deve-se reduzir a função objetivo, restrita a uma variedade linear e as restrições do filtro e positividade.

No último capítulo apresentamos um algoritmo que resolve esta fase. O algoritmo é composto por duas etapas. A construção de uma caixa diferenciada, que é construída de forma que esteja totalmente contida no primeiro ortante e preserve a propriedade do ponto sobre qual o modelo é construído. A outra etapa é a busca sobre a trajetória híbrida desta caixa. É nesta busca que o algoritmo mostra-se adequado para resolução do passo tangente. A cada ponto da trajetória obtido verifica-se se cada restrição é satisfeita. Se um ponto não satisfaz uma das restrições, por exemplo, não pertencer ao filtro ou à região onde o modelo é confiável, o algoritmo pára e utiliza o ponto anterior para a próxima iteração. O tamanho da caixa é modificado, se necessário, somente na próxima iteração.

Ainda neste capítulo dedicamos uma seção para comentar a possibilidade de trabalhar com uma caixa maior, utilizando um método de identificação do conjunto de ativo. O algoritmo é praticamente o mesmo. A busca é feita em um conjunto maior, que pode conter pontos que não pertençam ao ortante positivo e conseqüentemente deve-se incluir a positividade para testar em cada ponto da trajetória obtido.

Implementações. Neste trabalho não apresentamos exemplos de implementação do método. Ele foi testado em exemplos simples no trabalho de Bleyer [2].

Trabalhos futuros. Como trabalhos futuros sugerimos além de implementações, o estudo do enfraquecimento da hipótese de convexidade e estimativa dos multiplicadores de Lagrange.

Referências Bibliográficas

- [1] BEHLING, R. *Minimização de quadráticas em caixas sobre variedades afins, um subproblema de PQS*. Dissertação de mestrado. Universidade Federal de Santa Catarina, Santa Catarina, 2006.
- [2] BLEYER, I. *Um algoritmo de pontos interiores para o sub-problema tangencial em métodos de filtro para programação não linear*. Trabalho de conclusão de Curso. Universidade Federal de Santa Catarina, 2007.
- [3] BOGGS P. T. and TOLLE J. W. Sequential quadratic programming. *Acta Numer.*, v. 4 p. 1-51, 1996.
- [4] BONNANS, J. F. ; LEMARÉCHAL, C. ; GILBERT, J. C. e SAGASTIZÁBAL, C. A. *Numerical Optimization: Theoretical and Practical Aspects* Springer Series in Operations Research. Springer - ,2006.
- [5] COLEMAN, T. F. and LI, Y. An interior trust region approach for nonlinear minimization subject to bounds. *SIAM Journal of Optimization* , 6: 418 -445, 1996.
- [6] CONN, A. R . ; GOULD, N. I. M. e TOINT, P. L. *Trust-Region Methods*. Philadelphia, MPS-SIAM Series on Optimization , pp. 751 , 2000
- [7] FERRIS, M. ; WATHEN , A. and ARMAND P. *Limited memory solution of complementarity problems arising in video games*. Technical report, Oxford University, 2006.
- [8] FLETCHER, R. and LEYFFER, S. Nonlinear programming without a penalty function. *Mathematical Programming-Ser.A*. 91(2):239-269, 2002.

- [9] FRANCISCO, J. B. *Viabilidade em Programação Não Linear* Tese (Doutorado), UNICAMP, 2005.
- [10] FRISCH, K. R. The Logarithmic Potential Method of convex Programming, University Institute of Economics, memorandum, Oslo, Norway, 1955
- [11] GARCIA-PALOMARES U. M. and Mangasarian O. L. Superlinearly convergent quasi-newton methods for nonlinearly constrained optimization problems. *Mathematical Programming*, v 11 p. 1-13, 1976.
- [12] GONZAGA, C. C. *Path following methods for linear programming*. SIAM Review, 34(2); 167-227, 1992.
- [13] GONZAGA, C. C.; KARAS, E. and VANTI, M. *A globally convergent filter method for nonlinear programming*. SIAM Journal on Optimization, 14(3): 646-669, 2003.
- [14] GONZAGA, C. C. The largest step path following algorithm for monotone linear complementarity problems. *Mathematical Programming*, v.76, p. 309-332, 1997.
- [15] GONZAGA, C. C. ; BONNANS, F. Fast convergence of the simplified largest step path following algorithm. *Mathematical Programming*, v.76, p. 95-116, 1997.
- [16] HAN S - P. Superlinearly convergent variable metri algorithms for general nonlinear programming problems. *Mathematical Programming*, v.11, p. 263-282, 1976.
- [17] HAN S - P. A globally convergent method for nonlinear programming. *Journal of Optimization Theory and Applications*, v.22, p. 297-309, 1977.
- [18] LEVENBERG, K. A method for the solution of certain nonlinear problems in least squares. *Quart. Appl. Math.*, 2:164-168, 1944.
- [19] MARQUARDT, D. W. An algorithm for least squares estimation of nonlinear parameters. *SIAM J. Appl. Math.*, 11:431-441, 1963.

- [20] MARTÍNEZ, J. M. and SANTOS, S. A. Métodos computacionais de otimização. 20 Colóquio Brasileiro de Matemática - IMPA , July 1955 .
- [21] MARTÍNEZ, J. M. and PILOTTA, E. A. Inexact restoration algorithms for constrained optimization. *Journal of Optimization Theory and Applications*, 104: 135-163, 2000.
- [22] MARTÍNEZ, J. M. Inexact restoration method with lagrangian tangent decrease and new merit function for nonlinear programming. *Journal of Optimization Theory and Applications*, 111:39-58, 2001.
- [23] MARTÍNEZ, J. M. Box-quacan and implementation of augmented lagrangian algortithms for minimization with inequality constrains. *Computacional E Applied Mathematics*, 19: 31-56, 2000.
- [24] MCSHANE, K. A. A superlinearly convergent $O(\sqrt{n}L)$ iteration primal-dual linear programming algorithm. *SIAM Journal on Optimization*, v.4, p. 247-261, 1994.
- [25] NOCEDAL, J. e WRIGHT, S. J. *Numerical Optimization*. Springer Series in Operations Research. Springer - Verlag, 1999.
- [26] OMOJOKUN, E. Trust Region Algorithms for Optimization with Nonlinear Equality and Inequality Constrains. Tese Doutorado, Dept. of Computer Science, University of Colorado, 1991.
- [27] POWELL, M. J. D. A hybrid method for nonlinear equations. In P. Ravinowitz, editor, *Numerical Methods for Nonlinear Algebraic Equations*, Gordon and Breach, London, 1970.
- [28] POWELL, M. J. D. A fast algorithm for nonlinearly constrained optimization calculations. In G. A. Watson, editor, *Numerical Analysis Dundee*, pages 144-157. Springer -Verlag, Berlin, 1977.
- [29] POWELL, M. J. D. Algorithms for nonlinear constraints that use Lagrangian functions. *Mathematical Programming*, v.14, p. 224-248, 1978.

- [30] POWELL, M. J. D. The convergence of variable metric methods for nonlinearly constrained optimization calculations. In R. R. Meyer O. L. Mangasarian and S. M. Robinson, editors, *Nonlinear Programming 3*, pages 27-64. Academic Press, New York, 1978.
- [31] SONNEVEND, G. An Analytical Centre Polyhedrons and New Classes of Global Algorithms for Linear (Smooth, Convex) Programming, in Lecture Notes Control Inform. Sci. 84, Springer-Verlag, New York, NY, 1985, p p.866-876.
- [32] WRIGHT, S. J. *Primal-Dual Interior-Point Methodos*. Philadelphia, SIAM: 1997.
- [33] QUANDT, S. M. R. E. and LEVENBERG, H. F. TROTTER (1966) Maximization by quadratic hill-climbing. *Econometrica* , 34, 541-551.
- [34] WILSON, R. B. (1963) *A Simplicial Algorithm for Concave Programming*, PH.D. Dissertation, Harvard Universtity, Graduate School of Business Administration.
- [35] VANTI, M. R. *Melhoria da segurança dinâmica por otimização e algoritmos de programação não linear*. Tese de doutorado. Universidade Federal de Santa Catarina, Santa Catarina, Brazil, 2003.