

ALGORITMOS DE REGIÃO DE CONFIANÇA PARA PROBLEMAS DE
OTIMIZAÇÃO NÃO LINEAR RESTRITA

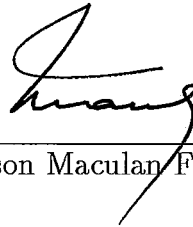
Pedro José Di Novella Cordero

TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS PRO-
GRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE FE-
DERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁ-
RIOS PARA OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS EM ENGE-
NHARIA DE SISTEMAS E COMPUTAÇÃO.

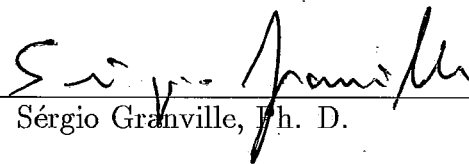
Aprovada por:



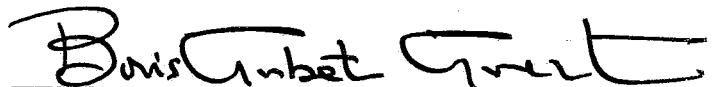
Clóvis Caesar Gonzaga, D. Sc.
(presidente)



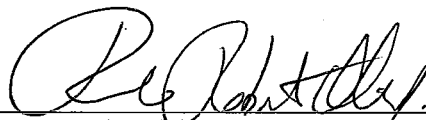
Nelson Maculan Filho, D. Sc.



Sérgio Granville, Ph. D.



Boris Garbati Gorenstein, D. Sc.



Paulo Roberto Oliveira, D. Sc.

RIO DE JANEIRO, RJ – BRASIL
MARÇO DE 1996

DI NOVELLA CORDERO, PEDRO JOSÉ

Algoritmos de Região de Confiança para Problemas de Otimização
não Linear Restrita [Rio de Janeiro] 1996

X, 128 p., 29.7 cm, (COPPE/UFRJ, D. Sc., ENGENHARIA DE SISTEMAS
E COMPUTAÇÃO, 1996)

TESE – Universidade Federal do Rio de Janeiro, COPPE

1 – Otimização não Linear 2 – Algoritmos de Região de Confiança

3 – Programação Quadrática Seqüencial 4 – Estratégia de Pontos Interiores

I. COPPE/UFRJ II. Título(Série).

Aos meus pais
À Ciléa

Agradecimentos

Ao professor Clóvis C. Gonzaga pela orientação, amizade e valiosos ensinamentos sem os quais não teria sido possível a realização desta pesquisa.

Ao professor Nelson Maculan F. pelo apoio e estímulo demonstrados ao longo do desenvolvimento do trabalho.

Ao pesquisador Sérgio Granville, do Cepel, pela sua colaboração e úteis sugestões feitas durante a realização deste trabalho.

À Capes e ao Cepel pelas bolsas de estudos concedidas que viabilizaram a realização do meu doutorado.

Aos colegas e amigos do Cepel e da Coppe pelo apoio e companheirismo mostrado durante o tempo em que nessas instituições permaneci como doutorando.

Aos meus pais, e demais familiares, por terem suportado com paciência minha longa ausência durante o período de realização do meu doutorado.

À Ciléa, cujo constante apoio, carinho e dedicação superaram todas as expectativas e viabilizaram a conclusão com sucesso deste trabalho.

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

Algoritmos de Região de Confiança para Problemas de Otimização não Linear
Restrita

Pedro José Di Novella Cordero

Março de 1996

Orientador: Clóvis Caesar Gonzaga

Programa: Engenharia de Sistemas e Computação

Neste trabalho, apresentamos dois algoritmos de região de confiança para resolver problemas de otimização não linear com restrições. Primeiro desenvolvemos um algoritmo para resolver problemas de minimização de funções escalares não lineares de varias variáveis sujeitas somente a restrições de caixa, logo generalizamos o algoritmo para poder tratar o caso em que restrições não lineares de igualdade são adicionadas ao problema. No desenvolvimento dos algoritmos propostos, empregamos a programação quadrática seqüencial com região de confiança e uma estratégia de pontos interiores para o tratamento das restrições de desigualdade. A escolha da programação quadrática seqüencial com região de confiança se justifica pelas atrativas características de ambos métodos, como a rápida taxa de convergência local, e a convergência global quando combinados os dois métodos, enquanto que a estratégia de pontos interiores evita os problemas associados aos métodos de restrições ativas, no tratamento das restrições de desigualdade. A região de confiança utilizada, tem um formato elíptico, enquanto que os outros métodos utilizam uma região de confiança esférica. Para ambos os algoritmos são apresentados os resultados de convergência global.

O segundo algoritmo apresentado foi implementado e testado para um conjunto de vinte problemas não lineares obtidos nas referências bibliográficas. Os resultados numéricos obtidos nos testes realizados com este algoritmo foram comparados com os obtidos por outros sete códigos de otimização. Este estudo comparativo mostrou a robustez do algoritmo proposto. Também para cada teste é mostrado graficamente o desempenho do algoritmo proposto no decurso das iterações.

Palavras-chave: Otimização não Linear, Algoritmos de Região de Confiança, Programação Quadrática Seqüencial, Estratégia de Pontos Interiores.

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

Trust Region Algorithms for Nonlinearly Constrained Optimization Problems

Pedro José Di Novella Cordero

March 1996

Thesis Supervisor: Clóvis Caesar Gonzaga

Department: Programa de Engenharia de Sistemas e Computação

In this work, we present two algorithms that use the trust region technique to solve nonlinearly constrained optimization problems. First, we develop an algorithm for solving problems of minimizing a nonlinear real-valued function of several variables, subject to box constraints, then we generalize the algorithm to handle the inclusion of nonlinear equality constraints in the problem. In the development of the proposed algorithms, we have used the sequential quadratic programming approach with trust region technique and an interior point method to deal with the inequality constraints. The choice of the sequential quadratic programming approach with trust region technique is based on the attractive features of both method, namely: a fast local convergence property and global convergence when both methods are used together, while the interior point method avoids the problems that appear in active set strategy, when dealing with the inequality constraints. The trust region used, has an ellipsoidal shape, in contrast to the other methods, which use a trust region with spherical shape. Global convergence results are presented for both algorithms.

The generalized algorithm was implemented and tested for a set of twenty nonlinear problems found in the literature. The numerical results obtained were compared with the results of seven well known optimization codes. This compa-

rative numerical study has shown the robustness of the proposed algorithm. The performance of the algorithm is presented graphically, for each test.

Key Words: Nonlinear Optimization, Trust Region Algorithms, Sequential Quadratic Programming, Interior Point Method.

Índice

I	Introdução	1
II	Revisão Conceitual da Programação não Linear	5
II.1	Introdução	5
II.2	Notação	7
II.3	Abordagens na Resolução dos Problemas de PNL	8
II.3.1	Restrições de Igualdade	8
II.3.2	Restrições de Igualdade e Desigualdade	18
III	Um Algoritmo de Região de Confiança para Programação não Linear com Variáveis Canalizadas.	22
III.1	Introdução.	22
III.2	A Região de Confiança.	23
III.3	Geração de um Passo do Algoritmo de Região de Confiança.	24
III.4	Algoritmo.	26
III.5	Convergência Global de Primeira Ordem.	27

IV Um Algoritmo de Região de Confiança para Programação não Linear com Restrições não Lineares de Igualdade e Variáveis Canalizadas.	46
IV.1 Introdução.	46
IV.2 A Região de Confiança.	48
IV.3 Geração de um Passo do Algoritmo de Região de Confiança.	49
IV.4 Estimativas dos multiplicadores de K.K.T.	53
IV.5 Algoritmo.	54
IV.6 Convergência Global de Primeira Ordem.	56
V Testes Computacionais	79
V.1 Introdução	79
V.2 Testes Realizados	79
VI Conclusões	121
Referências Bibliográficas	123

Capítulo I

Introdução

Os problemas de programação não linear aparecem, sob as mais diversas formas, em muitas áreas do saber humano, como nas ciências naturais, na física, na engenharia, na economia, assim como no planejamento governamental, no mundo dos negócios, na indústria e na área militar. Tais problemas podem aparecer por exemplo, no desenho ótimo de um veículo, a melhor utilização de recursos escassos, planejamento de sistemas elétricos, etc. Nas últimas quatro décadas, para atender exigências crescentes nessas áreas, tem se apresentado um grande desenvolvimento nos modelos e métodos de otimização. O aparecimento de computadores cada vez mais velozes, assim como com uma maior capacidade de memória, tem ajudado muito no desenvolvimento e aplicação de novas técnicas de otimização.

Outro aspecto que tem influenciado o desenvolvimento de novos algoritmos de programação não linear é o rápido incremento do tamanho e da complexidade dos problemas a serem resolvidos, como resultado do crescimento tecnológico depois de segunda guerra mundial. Daí a necessidade de algoritmos mais eficientes, mesmo dispondo-se de computadores de grande velocidade.

Esta pesquisa foi motivada pela necessidade de se desenvolver algoritmos robustos para resolver o problema da otimização não linear. Neste trabalho, apresentamos dois novos algoritmos para resolver o problema de programação não linear. O primeiro deles é um algoritmo de região de confiança para otimização de problemas com restrições de caixa, isto é, problemas de forma:

$$\begin{aligned} &\text{minimizar } f(x) \\ &\text{sujeito a } l \leq x \leq u, \end{aligned} \tag{I.1}$$

onde x é um vetor de dimensão n , cuja i -ésima componente x_i tem l_i e u_i como limites inferior e superior respectivamente, podendo ser algumas coordenadas de l iguais a $-\infty$ e algumas coordenadas de u iguais a $+\infty$ e $f : \mathcal{R}^n \rightarrow \mathcal{R}$ é uma função escalar.

O algoritmo desenvolvido para resolver o problema (I.1) é depois estendido para resolver o problema geral da otimização não linear, onde agora temos tanto restrições não lineares de igualdade, como de caixa:

$$\begin{aligned} &\text{minimizar } f(x) \\ &\text{sujeito a } c(x) = 0 \\ & \quad \quad \quad l \leq x \leq u, \end{aligned}$$

onde $f : \mathcal{R}^n \rightarrow \mathcal{R}$ é uma função escalar e $c : \mathcal{R}^n \rightarrow \mathcal{R}^m$ é um vetor de funções não lineares.

No desenvolvimento dos algoritmos propostos, empregamos a programação quadrática seqüencial com região de confiança e uma estratégia de pontos interiores para o tratamento das restrições de desigualdade. A escolha da programação quadrática seqüencial com região de confiança se justifica pelas atrativas características de ambos métodos, como a rápida taxa de convergência local, e a convergência global quando combinados os dois métodos, enquanto que a estratégia de pontos interiores evita os problemas associados aos métodos de restrições ativas, no tratamento das restrições de desigualdade.

Embora vários trabalhos têm sido feitos para desenvolver algoritmos de região de confiança para a otimização não linear, a maioria deles só consideram o caso da otimização sem restrições, problemas só com restrições de caixa ou problemas linearmente restritos, como os trabalhos de Byrd, Schnabel e Shultz [48], Gay [20], Conn, Gould e Toint [10], e Sagara e Fukushima [44]. Os resultados obtidos por esses e outros autores, mostraram que os métodos de região de confiança foram

muito bem sucedidos e robustos, para essa classe de problemas.

No caso da otimização com restrições não lineares, alguns trabalhos onde emprega-se a programação quadrática seqüencial e técnicas de região de confiança têm sido apresentados, mas só com restrições de igualdade. Entre eles podemos mencionar os trabalhos de Byrd, Schnabel e Shultz [3], Celis, Dennis e Tapia [5], Powell e Yuan [43] e Vardi [49].

É portanto o objetivo desta pesquisa, descrever e desenvolver algoritmos robustos para o problema de programação não linear com restrições não lineares de igualdade e caixa, os quais possam ser implementados eficientemente e analisar os resultados obtidos.

O trabalho está organizado em seis capítulos. A seguir apresentamos um sumário dos próximos capítulos da tese.

No capítulo II, descrevemos o problema de programação não linear na sua forma geral e no formato padrão. A seguir é apresentada a notação utilizada na tese. É feita uma revisão das diferentes abordagens empregadas na resolução do problema de programação não linear, fazendo-se especial ênfase na programação quadrática seqüencial, nos métodos de região de confiança e no uso das funções de mérito nesses dois métodos. Além disso, é apresentado o algoritmo afim-escala, no qual baseia-se o tratamento feito das restrições de desigualdade.

No capítulo III, descrevemos e apresentamos um novo algoritmo de região de confiança, para resolver problemas de minimização de funções não lineares sujeitas somente a restrições de caixa. Descrevemos a região de confiança utilizada, a qual tem um formato elíptico, distintamente dos outros métodos que utilizam uma região de confiança esférica. Na parte final do capítulo são apresentados os resultados de convergência global.

No capítulo IV, generalizamos o algoritmo desenvolvido no capítulo III para poder resolver o problema de programação não linear quando temos tanto restrições não lineares de igualdade como restrições de caixa. Mostramos como é feito o cálculo das estimativas dos multiplicadores de Karush-Kuhn-Tucker e a escolha da função de mérito. Finalmente, sob a hipótese de ser a seqüência gerada pelo

algoritmo convergente e uma hipótese sobre a região de confiança que explicitaremos neste capítulo, apresentamos os resultados de convergência global.

O capítulo V detalha os resultados obtidos com a implementação do algoritmo proposto no capítulo IV. Para tal fim, foram realizados vinte testes computacionais. Os resultados obtidos nos primeiros treze testes realizados com o algoritmo proposto são comparados com os obtidos pelos seis códigos de otimização apresentados em [30]. Os resultados dos últimos sete testes são comparados com os obtidos pelo código de otimização MINOS 5.1. Também para cada teste é mostrado o desempenho do algoritmo proposto, no decurso das iterações, através de 4 gráficos.

Por fim, no capítulo VI, apresentamos as conclusões do nosso trabalho e as recomendações para futuros desenvolvimentos.

Capítulo II

Revisão Conceitual da Programação não Linear

II.1 Introdução

O problema de programação não linear, na sua forma mais geral, pode ser enunciado como:

$$\text{minimizar } \bar{f}(\bar{x}) \quad (\text{II.1.1})$$

$$\text{sujeito a } \bar{l}_i \leq \bar{c}_i(\bar{x}) \leq \bar{u}_i, \quad i = 1, \dots, m \quad (\text{II.1.2})$$

$$\bar{l}_{m+i} \leq \bar{x}_i \leq \bar{u}_{m+i}, \quad i = 1, \dots, p, \quad (\text{II.1.3})$$

onde \bar{x} é um vetor de dimensão p , cuja i -ésima componente \bar{x}_i tem \bar{l}_i e \bar{u}_i como limitantes inferior e superior respectivamente, podendo ser algumas coordenadas de \bar{l} iguais a $-\infty$ e algumas coordenadas de \bar{u} iguais a $+\infty$. Se $\bar{l}_i = \bar{u}_i$ então a i -ésima restrição é uma restrição de igualdade. $\bar{f} : \mathcal{R}^p \rightarrow \mathcal{R}$ e $\bar{c}_i : \mathcal{R}^p \rightarrow \mathcal{R}$, $i = 1, \dots, m$ são funções escalares, geralmente suaves, isto é, com derivadas contínuas. As restrições de desigualdade (II.1.3) se denominam de restrições de caixa.

Temos então que um problema de programação não linear (PNL) consiste de uma função escalar, as maioria das vezes de várias variáveis, a qual queremos minimizar, sujeita, quiçá, a uma ou várias outras funções que servem para limitar ou definir os valores das variáveis. A função a ser minimizada é chamada de função objetivo enquanto que as outras são chamadas de restrições.

Devido à dificuldade de se resolverem, os problemas de programação não linear tem sido classificados em diferentes classes. Uma classe particularmente bem estudada é a formada pelos problemas cujas restrições são todas lineares. Estes problemas são conhecidos como problemas linearmente restritos. Nesta classe cabe ressaltar dois casos muito especiais: os problemas cuja função objetivo é quadrática, os quais são conhecidos como problemas de programação quadrática (PQ) e aqueles cuja função objetivo também é linear. Estes são conhecidos como problemas de programação linear (PL). Devido às suas características muito particulares, a programação linear tem se desenvolvido de certa forma independente do resto da PNL, gerando metodologias de estudo específicas para esta área. Um outro caso especial é a chamada otimização não restrita ou desvinculada, na qual os problemas não tem restrições.

Introduzindo-se variáveis de folga adequadamente, as restrições de desigualdade (II.1.2) podem ser transformadas em igualdades. Desta forma o problema (II.1.1 - II.1.3) pode ser re-escrito como:

$$\begin{aligned} & \text{minimizar} && f(x) \\ & \text{sujeito a} && c(x) = 0 \\ & && l \leq x \leq u, \end{aligned} \tag{II.2}$$

onde agora, $x \in \mathcal{R}^n$ contém tanto as variáveis originais como as variáveis de folga, $f : \mathcal{R}^n \rightarrow \mathcal{R}$ é uma função escalar e $c : \mathcal{R}^n \rightarrow \mathcal{R}^m$ é um vetor de funções. Esta formulação se denomina forma padrão e será a utilizada neste trabalho.

Um ponto $\hat{x} \in \mathcal{R}^n$ é viável para o problema (II.2) se $c(\hat{x}) = 0$ e $l \leq \hat{x} \leq u$. Caso contrário denomina-se inviável.

Um ponto viável x^* é um minimizador local (ou solução local) do problema (II.2) se existe uma vizinhança $V(x^*)$ tal que

$$f(x^*) \leq f(x) \text{ para todo ponto viável } x \in V(x^*). \tag{II.3}$$

Se a desigualdade (II.3) é estrita para todos os pontos viáveis $x \in V(x^*)$, $x \neq x^*$, então x^* denomina-se minimizador local estrito, forte ou isolado.

II.2 Notação

Nesta seção vamos apresentar a notação que será utilizada neste trabalho:

escalares: serão denotados usando-se letras minúsculas e as letras gregas (exceto λ , γ , μ , ϕ e ∇).

vetores: serão denotados pelas letras minúsculas e pelas letras gregas λ , γ e μ . Ao longo da tese consideraremos os vetores como vetores coluna. Os vetores com um superíndice t denotarão vetores linhas. As coordenadas do vetor serão indicadas através de subíndices.

matrizes: serão denotadas pelas letras maiúsculas. As matrizes com um superíndice t denotarão a matriz trasposta. O elemento ij de uma matriz será denotado através do subíndice ij .

\mathcal{N} : denota o conjunto dos números naturais.

\mathcal{R} : denota o conjunto dos números reais.

\mathcal{R}^n : denota o espaço real de dimensão n .

$|p|$: denota o valor absoluto do escalar p .

$\|x\|_\infty$: denota a norma infinito do vetor x .

$\|x\|_1$: denota a norma um do vetor x .

$\|x\|$: denota a norma dois do vetor x .

$\|A\|$: denota a norma espectral da matriz A .

$f : \mathcal{R}^n \rightarrow \mathcal{R}$: denota a função objetivo de um problema de (PNL).

$\nabla f(x)$: denota o vetor gradiente da função f no ponto x .

$\nabla^2 f(x)$: denota a matriz hessiana da função f no ponto x .

$c : \mathcal{R}^n \rightarrow \mathcal{R}^m$: denota o vetor das restrições de um problema de (PNL).

$c_i : \mathcal{R}^n \rightarrow \mathcal{R}$: denota a i -ésima restrição de um problema de (PNL).

$\nabla c_i(x)$: denota o vetor gradiente da função c_i no ponto x .

$\nabla^2 c_i(x)$: denota a matriz hessiana da função c_i no ponto x .

A , $A(x)$: denota a matriz jacobiana da função c no ponto x , isto é, a matriz cuja i -ésima coluna é o vetor $\nabla c_i(x)$.

D : denota uma matriz diagonal.

$D(x)$: denota uma matriz diagonal, na qual a diagonal é formada pelo vetor x .

II.3 Abordagens na Resolução dos Problemas de PNL

Muitos algoritmos tem sido propostos para resolver o problema de (PNL), utilizando-se as mais variadas técnicas e metodologias. O caso mais simples é o da otimização sem restrições. Para este tipo de problemas tem-se aplicado vários métodos, os quais podem ser catalogados entre os que são baseados em desenvolvimento em série de Taylor e os que não. A grande maioria dos métodos de otimização para funções diferenciáveis são baseados na expansão em série de Taylor da função objetivo. Eles estão divididos em métodos localmente convergentes, para os quais a convergência a um minimizador local é garantida somente se o ponto inicial esta dentro de uma certa vizinhança do minimizador local, em geral uma vizinhança bastante pequena, e os métodos globalmente convergentes, para os quais a convergência esta garantida, independentemente do ponto selecionado como ponto inicial. Entre os primeiros podemos destacar o método de Newton e os algoritmos quase-Newton. A estratégia básica dos métodos globalmente convergentes é diminuir o valor da função objetivo em cada iteração. Isto é obtido através de uma direção de descida a partir do ponto da iteração presente, uma direção a partir do ponto na qual a função decresce. Entre estes métodos podemos citar o método de máxima descida, os métodos de busca linear e os métodos de região de confiança. Entre os métodos não baseados em desenvolvimento em série de Taylor podemos citar o método das direções conjugadas e como caso particular, os métodos de gradiente conjugados e gradiente conjugados preconditionado. Uma ótima referência para se obter informação sobre otimização não linear sem restrições é [13].

II.3.1 Restrições de Igualdade

Consideremos agora o caso da otimização só com restrições de igualdade, isto é, problemas da forma

$$\begin{aligned} &\text{minimizar } f(x) \\ &\text{sujeito a } c(x) = 0. \end{aligned} \tag{II.4}$$

Uma condição necessária para que um ponto viável x^* seja um minimiza-

dor local de (II.4) é que exista $\lambda^* \in \mathcal{R}^m$ único tal que $\nabla f(x^*) = \sum_{i=1}^m \lambda_i^* \nabla c_i(x^*)$. (λ^* é chamado multiplicador de Lagrange do problema.)

Chamamos lagrangiano do problema (II.4) à função $\ell(x, \lambda) = f(x) - \lambda^t c(x)$. Note-se que se x^* é um minimizador local de (II.4) e λ^* o seu multiplicador de Lagrange associado então $\nabla \ell(x^*, \lambda^*) = 0$. O lagrangiano possui um papel preponderante em muitos dos métodos propostos para a resolução de problemas otimização restrita, como veremos a seguir.

O princípio básico empregado na resolução de problemas de PNL restritos é o de substituir um problema difícil por outro mais fácil. A aplicação deste princípio conduz a métodos que formulam e resolvem uma seqüência de subproblemas, os quais estão relacionados de uma maneira conhecida com o problema original. Em alguns desses métodos, o subproblema consiste de uma minimização sem restrições de uma função modelo do problema original. Exemplo deste tipo de métodos são os métodos de funções de penalidade e os métodos de multiplicadores ou de lagrangiano aumentado, onde a função objetivo é a função lagrangiana acrescentada com uma função de penalidade. Os métodos de lagrangiano aumentado tiveram uma grande popularidade no início dos anos 70, quando foram propostos independentemente por Hestenes [29] e Powell [41]. Em outros métodos, os subproblemas consistem de um modelo da função objetivo, sujeito a restrições de caixa e/ou restrições lineares obtidas a partir das restrições originais. A forma mais comum de se obter as funções modelo dos subproblemas é a expansão em série de Taylor. Exemplos de este tipo de método são os métodos seqüenciais linearmente restritos, também conhecidos como métodos de lagrangiano reduzido, os métodos de programação quadrática seqüencial (PQS) e métodos relacionados à PQS, incluindo-se aqui o método de região de confiança. Estes últimos são considerados atualmente como os mais eficientes para resolver os problemas de PNL. Vários deles tem sido implementados em pacotes computacionais altamente confiáveis, os quais tem tido um desempenho extremamente bom na prática, inclusive em problemas de teste considerados difíceis. Mais detalhes sobre os métodos aplicados em PNL restrita podem ser achados em [21] e [22].

Nos algoritmos que serão mostrados a seguir, assim como em todos os outros que serão apresentados neste trabalho não será especificada nenhuma regra ou

condição de parada, usando-se o teste “enquanto não convergir”. Em geral as regras de parada baseiam-se nas condições de otimalidade do problema a ser resolvido e em algum critério escolhido para estipular a precisão desejada.

II.3.1.1 Programação Quadrática Seqüencial

Os algoritmos propostos neste trabalho baseiam-se nos métodos de PQS e região de confiança. Os métodos de programação quadrática seqüencial podem ser considerados generalizações do método de Newton para o problema geral de otimização. Neles a solução do problema original é obtida através da resolução de uma seqüência de subproblemas, onde a função objetivo é substituída por uma aproximação quadrática da função lagrangiana e as restrições por equações lineares, obtidas a partir da aproximação linear das restrições originais. Dessa maneira, o subproblema a ser resolvido em cada iteração é um problema de programação quadrática que, em comparação ao problema original, pode ser considerado mais simples. A seguir, no algoritmo II.1, é apresentado o método de PQS, na forma “pura”, para a resolução do problema II.4.

Algoritmo II.1 (PQS):

0. $k = 0$; seja $x^0 \in \mathcal{R}^n$ o ponto inicial.

Enquanto não convergir fazer

1. Calcular $x = x^k$, $g = \nabla f(x)$, $A = \nabla c(x)$ e escolher B , uma matriz simétrica.
2. Calcular o passo d , resolvendo o seguinte problema quadrático:

$$\begin{aligned} &\text{minimizar} && g^t d + \frac{1}{2} d^t B d \\ &\text{sujeito a} && A^t d + c(x) = 0 \end{aligned} \tag{II.5}$$

3. Fazer $x^{k+1} = x + d$; $k = k + 1$.

Fim do enquanto

Um elemento da maior importância no algoritmo apresentado anteriormente é a escolha da matriz B . A melhor escolha dessa matriz ainda hoje é um problema aberto, particularmente nos métodos quase-Newton e é objeto de muita pesquisa atualmente.

Quando dispomos de derivadas segundas tanto da função objetivo, como das restrições, uma escolha boa perto do ótimo é tomar a própria matriz hessiana do lagrangiano no modelo quadrático, i.e, $B^k = \nabla^2 \ell(x^k, \lambda^k)$. De fato, esta foi a escolha feita por Wilson em 1963, quando na sua tese de doutorado, propôs o primeiro método de PQS para otimização com restrições. Como estimativa para os multiplicadores, Wilson utilizou os multiplicadores do subproblema na iteração anterior. Com esta escolha, o método de PQS “puro”, tem convergência quadrática. O problema desta escolha é que a matriz hessiana reduzida pode ser indefinida, o qual levaria a que o subproblema quadrático (II.5) fosse não limitado. Quando derivadas segundas não são disponíveis a alternativa é utilizar aproximações quase-Newton para a hessiana do modelo quadrático e estimativas diferentes para os multiplicadores de Lagrange.

O método de PQS, da forma como foi apresentado no algoritmo II.4, apresenta alguns problemas. O primeiro deles é que assim como acontece com o método de Newton, esta versão do método de PQS não tem boas propriedades de convergência global, isto é, se o ponto inicial x^0 está muito longe da solução x^* , então a convergência do algoritmo não está garantida. Em segundo lugar, como dizemos anteriormente, o subproblema quadrático (II.5) pode ser não limitado se a matriz B for não definida positiva ou mal condicionada. Em terceiro lugar os multiplicadores λ podem não ser limitados se a matriz A estiver perto de ser de posto deficiente. Por último mesmo convergindo, uma boa razão de convergência não é garantida, mesmo perto da solução ótima.

No entanto, se algumas modificações forem introduzidas no algoritmo II.1, os problemas anteriormente relatados podem ser evitados. A convergência global pode ser alcançada através do uso de uma busca linear ou uma estratégia de região de confiança com alguma função de mérito. Os segundo e terceiro problemas também podem ser resolvidos através do uso de uma técnica de região de confiança enquanto

que o último problema pode ser evitado utilizando-se uma correção de segunda ordem na escolha do passo d . A utilização dessa correção de segunda ordem foi introduzida em 1982 por Chamberlain et. al. [6] e Fletcher [18]. Nas seguintes seções vamos descrever os métodos de região de confiança e algumas funções de mérito utilizadas quando o método de região de confiança é empregado na resolução de problemas de PNL com restrições. Mais vejamos primeiramente como é o método de região de confiança no caso da otimização desvinculada.

II.3.1.2 Métodos de Região de Confiança

Uma das razões da grande popularidade dos métodos de região de confiança é o fato de eles ter convergência global e uma rápida taxa de convergência local. Nestes métodos, distintamente da maioria dos métodos de otimização, primeiro é determinado um limite superior para o tamanho do passo a ser dado, e logo usando-se um modelo quadrático é determinado o melhor passo possível, dentre aqueles que tem no máximo o tamanho predeterminado. Na k -ésima iteração de um método de região de confiança, a partir do ponto x^k , o tamanho do passo está restrito por um parâmetro $\Delta_k > 0$. Este passo é calculado resolvendo-se o seguinte problema de PQ:

$$\begin{aligned} &\text{minimizar} && q^k(d) = f(x^k) + \nabla f(x^k)^t d + \frac{1}{2} d^t B_k d \\ &\text{sujeito a} && \|d\| \leq \Delta_k, \end{aligned} \tag{II.6}$$

onde B_k é $\nabla^2 f(x^k)$ ou uma aproximação desta matriz.

O parâmetro Δ_k representa o raio de uma bola ao redor do ponto x^k , onde acredita-se que o modelo quadrático é uma aproximação confiável do problema que queremos resolver. Dai o nome de métodos de região de confiança. O parâmetro Δ_k denomina-se raio da região de confiança e é ajustado no decorrer das iterações, dependendo da adequação da aproximação quadrática no ponto $x^k + d$. Se a aproximação quadrática no ponto $x^k + d$ não for considerada aceitável, o passo calculado na iteração presente é rejeitado o parâmetro Δ_k é diminuído e é calculado um novo passo d . Por outro lado, se a aproximação quadrática for muito boa, além de ser aceito o passo d , o parâmetro Δ_k é aumentado.

O ajuste do parâmetro Δ_k , feito em cada iteração, é fundamental no desempenho de um método de região de confiança. Ele evita que no caso da aproximação quadrática for considerada boa, a vizinhança fique muito pequena, o qual faria que os passos gerados pelo algoritmo sejam muito pequenos, acarretando um aumento do número de iterações. No caso contrário evita que vizinhança fique muito grande, o qual prejudicaria a qualidade do passo gerado na iteração presente.

Seja d^k um passo gerado pela resolução do problema (II.6). A redução real na função objetivo, ao ir de x^k a $x^k + d^k$, a qual denotaremos com $ared_k(d^k)$, está dada por:

$$ared_k(d^k) = f(x^k) - f(x^k + d^k).$$

A variação do modelo quadrático devida ao passo d^k , a qual pode ser vista como uma predição da redução real, de acordo com nosso modelo, a qual denotaremos com $pred_k(d^k)$, está dada por:

$$pred_k(d^k) = q(0) - q(d^k) = -\nabla f(x^k)^t d^k - \frac{1}{2} d^{k^t} B_k d^k.$$

Para que o passo d^k seja considerado aceitável devemos determinar se a adequação da aproximação quadrática no ponto $x^k + d^k$ é boa. Para quantificar a aproximação entre a função objetivo e seu modelo quadrático tomamos a seguinte razão:

$$\frac{ared_k(d^k)}{pred_k(d^k)}.$$

Quanto mais perto esta razão estiver de 1, melhor será a aproximação da função f pela aproximação quadrática no ponto $x^k + d^k$. Se tomamos $\eta \in (0, 1)$ como o grau de precisão desejado, então cada vez que a condição

$$\frac{ared_k(d^k)}{pred_k(d^k)} \geq \eta \tag{II.7}$$

for satisfeita o passo será aceito e $x^{k+1} = x^k + d^k$. Caso contrário, reduzimos o raio Δ_k da região de confiança e calculamos um novo passo provisório d^k , o qual é novamente examinado, usando-se a condição (II.7).

Note-se que se o passo d^k não for nulo, então $pred_k(d^k) > 0$ e portanto, se a condição (II.7) for satisfeita, então um decréscimo na função objetivo é obtido.

A seguir, no algoritmo II.2, é apresentado o método geral de região de confiança no caso da otimização desvinculada.

Algoritmo II.2 (Algoritmo de Região de Confiança para Otimização sem Restrições): Dados $\Delta > 0$, $0 < \eta_1 < \eta_2 < 1$, $\tau_1 \in (0, 1)$ e $\tau_2 > 1$.

0. $k = 0$; seja $x^0 \in \mathcal{R}^n$ o ponto inicial.

Enquanto não convergir fazer

1. Calcular $x = x^k$, $f = f(x)$, $g = \nabla f(x)$ e escolher B , uma matriz simétrica.

2. Calcular o passo d , resolvendo o seguinte problema quadrático:

$$\text{minimizar } q(d) = f + g^t d + \frac{1}{2} d^t B d$$

$$\text{sujeito a } \|d\| \leq \Delta$$

3. Calcular $pred = q(0) - q(d) = -g^t d - \frac{1}{2} d^t B d$.

4. Calcular $ared = f - f(x + d)$.

5. Se $\frac{ared}{pred} \geq \eta_1$ então

$$x^{k+1} = x + d; \quad k = k + 1.$$

caso contrário fazer

$$\Delta = \tau_1 \Delta; \quad \text{ir a 2.}$$

6. Se $\frac{ared}{pred} \geq \eta_2$ fazer $\Delta = \tau_2 \Delta$.

Fim do enquanto

Os valores usuais dos parâmetros dados no algoritmo são $\eta_1 = 0.25$, $\eta_2 = 0.75$, $\tau_1 = 0.25$ e $\tau_2 = 2$. Esses valores são arbitrários, mas a escolha deles não altera significativamente o desempenho do algoritmo.

Na prática é comum escolher como matriz B_k a própria matriz hessiana da função objetivo, ou uma aproximação quase-Newton de ela e tomar o passo de

Newton (ou quase-Newton) se ele satisfaz a restrição de região de confiança. A razão desta escolha é simples, aproveitar a rápida taxa de convergência do método de Newton. Se B_k for definida positiva, o passo de Newton, ou quase-Newton, vem dado por $d_k = -B_k^{-1}\nabla f(x^k)$.

Algoritmos baseados na técnica de região de confiança para problemas de otimização desvinculados tem tido grande sucesso. Devido ao fato de eles ter convergência global não é preciso assumir nenhuma hipótese de proximidade do ponto inicial a solução do problema. Além disso, não é necessário exigir que matriz B seja definida positiva o bem condicionada, pois o uso da restrição de região de confiança $\|d\| \leq \Delta$ garante que o passo d^k é limitado.

O método de região de confiança foi inicialmente utilizado na resolução do problema de mínimos quadrados não linear por Levenberg em 1944 [32] e Marquardt em 1963 [33]. Algoritmos baseados em técnicas de região de confiança para problemas de PNL desvinculados e com restrições de igualdade tem sido desenvolvidos por Goldfeldt, Quandt e Trotter [23], Byrd, Schnabel e Shultz [3], [4] e [48], Fletcher [16], Moré [35] e Vardi [49]. Gay [20] e Schnabel, Koontz e Weiss [47] implementaram códigos baseados em métodos quase-Newton com busca linear e métodos de região de confiança usando uma aproximação quase-Newton da matriz hessiana e os compararam usando um conjunto de problemas de teste proposto por Moré, Garbow e Hillstom [37]. Ambos estudos mostraram que se bem em alguns problemas específicos o desempenho dos métodos apresentavam diferenças consideráveis, nenhum dos métodos se mostrou consistentemente mais eficiente ou robusto que o outro. Na media foi pequena a diferença no desempenho de ambos métodos.

II.3.1.3 Funções de Mérito

Vamos considerar agora as funções de mérito e ver como elas são usadas em métodos de região de confiança para resolver o problema (II.4). Elas também são empregadas nos métodos de PQS para garantir convergência global. No caso desvinculado, ao final de cada iteração de um algoritmo dado, procura-se obter um novo ponto o qual melhora o valor da função objetivo. No caso da otimização com restrições não lineares, salvo alguns poucos casos especiais, resulta impossível gerar uma seqüência

de pontos viáveis com valores decrescentes da função objetivo. Procura-se então que ao final de cada iteração de um algoritmo dado, obter um ponto que além de melhorar o valor da função objetivo também esteja mais perto de satisfazer as restrições do problema. Mais estas duas condições podem ser conflitantes. Para resolver este conflito, é preciso usar uma função que leve em consideração ambas condições, a qual nos permita decidir se o ponto obtido na iteração presente é aceito ou rejeitado. Existem varias funções que podem ser utilizadas com tal intuito, as quais são conhecidas como funções de mérito. A seguir vamos descrever o uso de uma função de mérito num algoritmo de região de confiança para resolver problemas com restrições não lineares.

Seja ϕ uma função de mérito. Consideremos a k -ésima iteração de um método de região de confiança, seja x^k o ponto obtido na iteração anterior e d^k o passo obtido na iteração presente. Nos métodos de região de confiança para o caso restrito, as funções de mérito são usadas para quantificar a adequação do modelo utilizado ao problema original no ponto $x^k + d^k$ e conseqüentemente decidir se o ponto obtido na iteração presente é aceito ou não. Para fazer isto é calculada, da mesma forma que foi obtido o modelo quadrático a partir do problema original, uma aproximação da função de mérito. Se $pred_k(d^k)$ denota a variação da aproximação da função de mérito devida ao passo d^k e $ared_k(d^k) = \phi(x^k) - \phi(x^k + d^k)$ denota a redução real na função de mérito, ao ir de x^k a $x^k + d^k$, então a adequação da aproximação quadrática no ponto $x^k + d^k$ é considerada boa se a condição

$$\frac{ared_k(d^k)}{pred_k(d^k)} \geq \eta$$

for satisfeita, para algum $\eta \in (0, 1)$ dado.

Além de determinar se o ponto gerado em uma iteração do método de região de confiança melhora a função objetivo e a viabilidade, outras características desejáveis que as funções de mérito devem possuir são:

- i. O decréscimo de ϕ de iteração em iteração deve levar a uma solução do problema original.
- ii. Uma solução x^* do problema original deve ser um minimizador local de ϕ .

- iii. Sob hipóteses razoáveis, a função ϕ deve garantir a convergência global do método.
- iv. A taxa de convergência do algoritmo não deve ser prejudicada por ϕ .
- v. A avaliação de ϕ não deve ser dispendiosa desde o ponto de vista computacional.

Estas propriedades desejáveis das funções de mérito, sugerem que tanto a função objetivo como as restrições devem formar parte da função ϕ . Um número considerável de funções de mérito tem sido propostas e utilizadas na otimização com restrições de igualdade.

A função de mérito mais amplamente usada é a função de penalidade ℓ_1 ou função de penalidade de valor absoluto:

$$\phi_\mu(x) = f(x) + \sum_{i=1}^m \mu_i |c_i(x)|, \quad (\text{II.8})$$

onde os μ_i são pesos positivos. Estes pesos são incrementados caso seja necessário, em qualquer iteração, mas eventualmente, eles permanecem fixos. A função de penalidade ℓ_1 tem sido usada durante muitos anos (sob diferentes nomes) em problemas de desenho mecânico e de estruturas.

Esta função de mérito foi proposta por Han em 1977 [28], para assegurar a convergência global de um método de PQS, e tem sido utilizada em algoritmos desenvolvidos por Fletcher [16], Vardi [49], Byrd, Schnabel e Shultz [3], Coleman e Conn [7] e [8] e Powell [42]. Esta também é a função de mérito escolhida para ser usada no algoritmo desenvolvido no capítulo IV deste trabalho. Dois inconvenientes apresentados por esta função de mérito são o fato de ela ser não diferenciável nos pontos viáveis para o problema (II.4) e que $\phi_\mu(x^k + d^k)$ pode ser maior que $\phi_\mu(x^k)$.

Uma função de mérito alternativa, também muito empregada, é o lagrangiano aumentado de Hestenes:

$$\phi_\rho(x, \lambda) = f(x) - \lambda^t c(x) + \frac{1}{2} \rho \|c(x)\|^2, \quad (\text{II.9})$$

onde λ é uma estimativa dos multiplicadores de Lagrange e $\rho > 0$ é um parâmetro de penalidade. O uso de (II.9) como função de mérito foi sugerido por Wright em 1976 [50] e por Schittkowski em 1981 [46].

Ao se escolher (II.9) para ser usada como função de mérito, pontos delicados devem ser considerados cuidadosamente, em particular, um extremo cuidado deve ser tomado na escolha das estimativas dos multiplicadores de Lagrange. No caso da PQS, se λ é tomado como o multiplicador do subproblema quadrático da iteração anterior, a função de mérito sofre mudanças significativas em cada iteração, o qual pode criar dificuldades para provar a convergência global. Fletcher em 1973 [19], propôs o uso da função de penalidade ℓ_1 com a seguinte escolha de λ :

$$\lambda(x) = (\nabla c(x^k)^t \nabla c(x^k))^{-1} \nabla c(x)^t \nabla f(x^k)^t,$$

a qual também foi usada por Powell e Yuan [43]. Uma outra escolha para λ , apresentada em [5] por Celis, Dennis e Tapia e:

$$\lambda(x) = (\nabla c(x^k)^t B^{-1} \nabla c(x^k))^{-1} (c(x) - \nabla c(x^k)^t B^{-1} \nabla f(x^k)),$$

onde B é uma aproximação da matriz $\nabla_{xx}^2 \ell(x, \lambda)$.

A escolha do parâmetro de penalidade ρ é um outro ponto delicado no uso da função (II.9). Este parâmetro deve ser o suficientemente grande para garantir a convergência do método a uma solução de (II.4), mais paralelamente, não pode ser muito grande, já que isto ocasionaria problemas de mal condicionamento. Uma análise minuciosa desta situação pode ser achada em [22].

As implementações mais bem sucedidas de métodos de PQS e de região de confiança quando só derivadas de primeira ordem são disponíveis usam uma aproximação quase-Newton da matriz do termo quadrático e as funções (II.8) ou (II.9) como funções de mérito.

II.3.2 Restrições de Igualdade e Desigualdade

Consideremos inicialmente o caso da otimização onde somente restrições de caixa estão presentes, isto é, problemas da forma

$$\begin{aligned} &\text{minimizar} && f(x) \\ &\text{sujeito a} && l \leq x \leq u. \end{aligned} \tag{II.10}$$

Este tipo de problemas é provavelmente o mais comum dos problemas de otimização restrita que aparecem como conseqüência de aplicações práticas. Mesmo

no caso desvinculado, em [21] Gill, Murray e Wright propõem a introdução de limites nas variáveis de decisão, como uma maneira de aferir a formulação do problema. Mesmo que, devido a sua estrutura, o problema (II.10) seja o mais simples dos problemas de otimização com restrições de desigualdade, ele é mais complexo do que muitos problemas que só contêm restrições de igualdade, pelo fato de envolver um problema combinatório, o qual consiste em determinar o conjunto das restrições ativas em uma solução, isto é, o conjunto das restrições que são satisfeitas como igualdade pela solução. Os algoritmos baseados em conjuntos ativos, em cada iteração, fazem uma predição do conjunto ativo no ótimo, e resolvem então uma seqüência de subproblemas, os quais estão relacionados de uma maneira conhecida com o problema original, onde agora só existem restrições de igualdade. O ponto negativo destes algoritmos é que a predição do conjunto ativo no ótimo pode mudar muito de uma iteração para outra, o qual dificulta a convergência do método.

A maioria dos algoritmos de região de confiança existente na literatura, só considera o problema de PNL com restrições de igualdade, ou então como o problema (II.10). A razão disso, é a dificuldade de demonstrar convergência para o caso em que ambos os tipos de restrições aparecem. Algoritmos de região de confiança para resolver o problema (II.10) foram apresentados por Conn, Gould e Toint [10], Coleman e Li [9], Dennis e Vicente [14] e Sagara e Fukushima [44]. Para problemas onde também aparecem restrições de igualdade, Bonnans e Bouhtou [2] apresentam um algoritmo de região de confiança para o problema de PQ e Dennis, Heinkenschloss e Vicente [15] um algoritmo de região de confiança para o problema de controle ótimo. Em [24], Gonzaga propõe um algoritmo de região de confiança para o problema de PNL, com restrições lineares de desigualdade.

II.3.2.1 Algoritmo Afim-Escala

A abordagem empregada na resolução de problemas de PNL com restrições de desigualdade, nos algoritmos propostos nos capítulos III e IV deste trabalho, está baseada no algoritmo afim-escala para o caso linearmente restrito.

Convencionou-se chamar de algoritmos afins aos algoritmos que geram pontos apenas no conjunto viável de problemas linearmente restritos, em contraste

com outros algoritmos que geram pontos num conjunto maior, como por exemplo o algoritmo de Karmarkar, o qual gera pontos num cone gerado pelo conjunto viável.

Consideremos o problema de PNL, linearmente restrito:

$$\begin{aligned} & \text{minimizar } f(x) \\ & \text{sujeito a } A^t x - b = 0 \\ & x \geq 0, \end{aligned} \tag{II.11}$$

onde $f : \mathcal{R}^n \rightarrow \mathcal{R}$ é uma função diferenciável, A é uma matriz $n \times m$ dada e b um vetor de \mathcal{R}^m dado. Vamos supor que o conjunto viável é limitado.

A seguir apresentamos o algoritmo afim-escala. A diferença para os algoritmos dados anteriormente, é que neste vamos supor que o ponto inicial é viável para o problema (II.11). Mais ainda, o ponto inicial deve estar no interior do hiper-cubo determinado pelas restrições de desigualdade, que no caso do problema (II.11) é o ortante não negativo de \mathcal{R}^n . Seja $e^t = (1, \dots, 1) \in \mathcal{R}^n$.

Algoritmo II.3 (Algoritmo Afim-escala): Dado $\rho \in (0, 1)$.

0. $k = 0$; seja $x^0 \in \mathcal{R}^n$, $x^0 > 0$ viável, o ponto inicial.

Enquanto não convergir fazer

1. Calcular $x = x^k$, $g = \nabla f(x)$.
2. Calcular D , matriz diagonal, onde $D_{ii} = x_i$, $\forall i = 1, \dots, n$.
3. Fazer mudança de escala: $\bar{A} = DA$, $\bar{g} = Dg$.
4. Calcular P , a matriz de projeção: $P = I - \bar{A}(\bar{A}^t \bar{A})^{-1} \bar{A}^t$.
5. Calcular o passo d : $d = -P\bar{g}$.
6. Calcular o tamanho do passo:

$$\Delta = \min_{1 \leq i \leq n} \left\{ \frac{1}{-d_i} \mid d_i < 0 \right\}.$$

7. Fazer $x^{k+1} = D(e + \rho \Delta d)$; $k = k + 1$.

Fim do enquanto

No algoritmo (II.3), a mudança de escala, no passo 3, leva o ponto x^k ao ponto e . O tamanho do passo está sujeito à variável Δ e ao parâmetro ρ . O valor da variável Δ mede o máximo comprimento do passo a partir do ponto e até alguma componente anular-se. O teste de razão feito no passo 6 do algoritmo sempre tem solução para um problema com conjunto viável limitado, já que neste caso pelo menos uma componente de d deve ser negativa. Mas, para poder repetir o procedimento na próxima iteração, devemos manter o ponto x^{k+1} no interior do ortante não negativo. Por isso o método afim-escala encontra-se na classe dos métodos de pontos interiores. No algoritmo (II.11) tomamos um passo um pouco menor que Δ , reduzido por um fator heurístico ρ . O valor usual do parâmetro ρ é 0.95.

Informações mais detalhadas sobre os algoritmos afins, assim como outros métodos de pontos interiores podem ser achadas em [25] e [26].

Uma das vantagens de usar uma técnica baseada no algoritmo afim-escala, para o tratamento das restrições de desigualdade, nos algoritmos propostos neste trabalho, é que a seqüência de pontos gerada pelo algoritmo está contida no interior do hipercubo determinado pelos limites inferior e superior das variáveis, eliminando-se desta forma o problema combinatório de determinar o conjunto das restrições ativas na solução. Outra vantagem é que na prática, implementações do algoritmo afim-escala aplicadas na programação linear, tem dado resultados espetaculares, resolvendo problemas de grande porte, com milhares de restrições e variáveis, até uma precisão de 10 casas decimais em cerca de 30 iterações.

Capítulo III

Um Algoritmo de Região de Confiança para Programação não Linear com Variáveis Canalizadas.

III.1 Introdução.

Neste capítulo apresentamos um algoritmo, baseado na estratégia de região de confiança e em técnicas de pontos interiores, para resolver o problema de minimizar uma função não linear onde as variáveis estão limitadas tanto inferior, como superiormente. Os algoritmos de região de confiança mais difundidos até hoje, usualmente trabalham em uma região esférica. Em nosso trabalho, determinamos a forma e o tamanho da região de confiança para gerar um elipsóide adaptado ao hipercubo determinado pelos limites inferior e superior das variáveis. O algoritmo proposto gera uma seqüência de pontos dentro do hipercubo nos quais se obtém uma melhoria na função objetivo em cada iteração. Demonstramos que qualquer ponto de acumulação desta seqüência satisfaz as condições necessárias de otimalidade de Karush-Kuhn-Tucker.

Consideremos o seguinte problema de programação não linear:

$$\begin{aligned} \min \quad & f(x) \\ \text{s. a.} \quad & l \leq x \leq u, \end{aligned} \tag{III.1}$$

onde $f : \mathcal{R}^n \rightarrow \mathcal{R}$ é uma função não linear, $f \in C^2$ e os vetores l e u são conhecidos, sendo $l < u$, podendo ser algumas coordenadas de l iguais a $-\infty$ e algumas coordenadas de u iguais a $+\infty$.

As condições necessárias para que x^* seja um mínimo local de (III.1) são:
Existem $\lambda^* \in \mathcal{R}^n$, $\mu^* \in \mathcal{R}^n$ tais que

$$l \leq x^* \leq u$$

$$\nabla f(x^*) = \lambda^* - \mu^*$$

$$\lambda^* \geq 0, \mu^* \geq 0$$

$$\forall i = 1, \dots, n \quad \lambda_i^*(-x_i^* + l_i) = 0$$

$$\forall i = 1, \dots, n \quad \mu_i^*(x_i^* - u_i) = 0.$$

Estas condições são conhecidas como condições de Karush-Kuhn-Tucker.
Em nosso caso, elas podem ser reescritas como:

$$l \leq x^* \leq u \quad (\text{III.2.1})$$

$$\text{se } x_i^* = l_i \text{ então } \frac{\partial f}{\partial x_i}(x^*) \geq 0 \quad (\text{III.2.2})$$

$$\text{se } l_i < x_i^* < u_i \text{ então } \frac{\partial f}{\partial x_i}(x^*) = 0 \quad (\text{III.2.3})$$

$$\text{se } x_i^* = u_i \text{ então } \frac{\partial f}{\partial x_i}(x^*) \leq 0. \quad (\text{III.2.4})$$

No caso de ser f convexa, as condições de Karush-Kuhn-Tucker também são condições suficientes de otimalidade.

III.2 A Região de Confiança.

Consideremos os seguintes conjuntos:

$$H_1 = \{x \in \mathcal{R}^n / l \leq x \leq u\} \text{ e}$$

$$H_1^0 = \{x \in \mathcal{R}^n / l < x < u\}.$$

Note-se que devido ao fato de ser $l < u$ temos que $H_1^0 \neq \emptyset$. Dados um ponto $x^k \in H_1^0$ e números $\Delta_k \in \mathcal{R}_+$ e $\rho \in (0, 1)$, nossa região de confiança está dada pelo elipsóide simples centrado em x^k , o qual é representado pela equação:

$$d^t D^{-2} d \leq \rho^2$$

onde D é uma matriz diagonal cujos elementos são

$$d_{ii} = \min\{\Delta_k, x_i^k - l_i, u_i - x_i^k\}, \quad \forall i = 1, \dots, n.$$

Esta escolha da matriz D nos permite alterar tanto o tamanho quanto o formato da região de confiança. Os algoritmos de região de confiança só alteram o tamanho da região de confiança e não o seu formato. No nosso caso ao mudar o tamanho da região de confiança também modificamos o seu formato. Se o valor do raio Δ_k for grande, nossa região de confiança será um elipsóide, o qual se adapta melhor ao hipercubo definido pelas restrições de desigualdade, permitindo um passo maior que se tomássemos uma bola. Entretanto, se Δ_k for suficientemente pequeno então a região de confiança será uma bola. Neste último caso temos que as restrições de desigualdade estão longe de ser ativas, e em conseqüência nosso problema equivale a um problema sem restrições e portanto o melhor formato para a região de confiança é o de bola, já que este não favorece nenhuma direção em particular.

III.3 Geração de um Passo do Algoritmo de Região de Confiança.

O algoritmo a ser desenvolvido neste capítulo, para resolver o problema (III.1), está baseado na programação quadrática seqüencial, no método de região de confiança e no algoritmo afim-escala, os quais foram apresentados no capítulo II. Na k -ésima iteração do algoritmo que vamos propor, dados x^k , o ponto gerado na iteração $k-1$ do algoritmo, o qual não satisfaz as condições necessárias de primeira ordem para uma solução de (III.1), $g^k = \nabla f(x^k)$, $B_k = \nabla^2 f(x^k)$, o raio Δ_k da região de confiança e a matriz D , é resolvido o seguinte subproblema quadrático:

$$\begin{aligned} \min \quad & q_k(d) = g^{k^t} d + \frac{1}{2} d^t B_k d \\ \text{s. a.} \quad & \| D^{-1} d \| \leq \rho. \end{aligned} \tag{III.3}$$

O parâmetro ρ do problema (III.3) é utilizado para garantir que o ponto resultante na k -ésima iteração do algoritmo pertença a H_1^0 . O valor usual para ρ é 0.95. Na prática permite-se ao parâmetro ρ assumir valores maiores que 1, sempre que o ponto resultante na k -ésima iteração do algoritmo estiver no interior de H_1 .

Note-se que se d é solução de (III.3) então

$$\|d\| \leq \rho \Delta_k. \quad (\text{III.4})$$

As condições necessárias para que d^k seja um mínimo local de (III.3) são:

Existe $\psi \in \mathcal{R}$ tal que

$$\|D^{-1}d^k\| \leq \rho \quad (\text{III.5.1})$$

$$g^k + B_k d^k = -\psi D^{-2} d^k \quad (\text{III.5.2})$$

$$\psi(\|D^{-1}d^k\| - \rho) = 0 \quad (\text{III.5.3})$$

$$\psi \geq 0. \quad (\text{III.5.4})$$

A condição (III.5.2) pode ser reescrita como $(B_k + \psi D^{-2})d^k = -g^k$, sendo a matriz $B_k + \psi D^{-2}$ semidefinida positiva. A demonstração deste fato poder ser achada em [12] e em [17].

Seja d^k solução do problema (III.3). Agora, devemos determinar se o ponto $x^k + d^k$ melhora suficientemente a função objetivo.

A redução real na função, ao ir de x^k a $x^k + d^k$, a qual denotaremos com $ared_k(d^k)$, está dada por:

$$ared_k(d^k) = f(x^k) - f(x^k + d^k).$$

A predição da redução real, de acordo com nosso modelo, a qual denotaremos com $pred_k(d^k)$, está dada por:

$$pred_k(d^k) = -g^{k^t} d^k - \frac{1}{2} d^{k^t} B_k d^k.$$

Note-se que no caso no qual $g^k \neq 0$ temos que $pred_k(d^k) > 0$, já que $H_1^0 \neq \emptyset$. Mais adiante demonstraremos que qualquer ponto gerado pelo nosso algoritmo é sempre viável para o problema (III.1) e portanto temos que x^k não satisfaz as condições necessárias de primeira ordem para uma solução de (III.1) somente se $g^k \neq 0$.

Se a melhora em f é uma proporção suficiente da predição de nosso modelo, isto é, se $\frac{ared_k(d^k)}{pred_k(d^k)} \geq \eta$, onde $\eta \in (0, 1)$ é uma constante fixa, então o passo será aceito e $x^{k+1} = x^k + d^k$. Caso contrário, reduzimos o raio Δ_k da região de confiança e calculamos um novo passo provisório d^k , o qual é novamente examinado.

III.4 Algoritmo.

Algoritmo III.1: Dados $0 < \eta < 1$, $0 < \omega < 1$, $0 < \tau < 1$ e $0 < \rho < 1$.

0. $k = 0$; seja $x^0 \in H_1^0$ o ponto inicial.

Enquanto não convergir fazer

1. Calcular

$$x = x^k; f = f(x); g = \nabla f(x); B = \nabla^2 f(x).$$

2. Calcular

$$\Delta = \max_{1 \leq i \leq n} \{m_i\} + \omega,$$

onde:

$$m_i = \begin{cases} 1 & \text{se } l_i = -\infty \text{ e } u_i = +\infty \\ \min\{x_i - l_i, u_i - x_i\} & \text{caso contrário.} \end{cases}$$

3. Calcule D, matriz diagonal, onde

$$d_{ii} = \min\{\Delta, x_i - l_i, u_i - x_i\}, \quad \forall i = 1, \dots, n.$$

4. Calcule o passo d , resolvendo o problema (III.3).

5. Calcule

$$pred = -g^t d - \frac{1}{2} d^t B d.$$

6. Calcule

$$f_+ = f(x + d).$$

7. Calcule

$$ared = f - f_+.$$

8. Se $\frac{ared}{pred} \geq \eta$ então

$$x^{k+1} = x + d; \quad k = k + 1.$$

caso contrário fazer

$$\Delta = \tau \Delta; \quad \text{ir a 3.}$$

Fim do enquanto

Os valores usuais dos parâmetros dados no algoritmo são $\eta = 0.25$, $\omega = 0.1$, $\tau = 0.25$ e $\rho = 0.95$. Estes valores são arbitrários, mas a escolha dos mesmos não altera significativamente o desempenho do algoritmo.

III.5 Convergência Global de Primeira Ordem.

Para os resultados de convergência de global que serão apresentados a seguir assumiremos certas hipóteses, que denominaremos hipóteses padrões.

Hipóteses padrões.

Dado $x^0 \in H_1^0$ definimos os seguintes conjuntos:

$$\bar{H} = \{x \in H_1 / f(x) \leq f(x^0)\} \text{ e}$$

$$\bar{H}^0 = \{x \in \bar{H} / l < x < u\}.$$

Seja $\{x^k\}$ uma seqüência gerada pelo algoritmo, com um ponto inicial $x^0 \in H_1^0$ arbitrário, e suponhamos que $\{x^k\}$ está contida em algum subconjunto aberto S de \mathcal{R}^n , tal que $\bar{H}^0 \subset S$. Suponhamos que $f(x)$ é duas vezes continuamente diferenciável em S e que $g(x) = \nabla f(x)$ e $B(x) = \nabla^2 f(x)$ são limitados em S . Suporemos também que \bar{H} é limitado.

No primeiro lema que apresentamos, demonstramos que se tomamos como ponto inicial para nosso algoritmo $x^0 \in H_1^0$ então a viabilidade é mantida em cada iteração do algoritmo.

LEMA III.1. Seja x^k um ponto gerado pelo algoritmo e d^k um passo gerado pela solução de (III.3) na iteração k . Se $x^k \in H_1^0$ então $x^{k+1} = x^k + d^k$ também pertence a H_1^0 .

Demonstração.

Seja x^k um ponto gerado pelo algoritmo, o qual pertence a H_1^0 e d^k um passo gerado

pela solução de (III.3) na iteração k . Devemos ver que $l < x^{k+1} = x^k + d^k < u$. Suponhamos que $x_i^{k+1} \leq l_i$ para algum i . Então, pela definição de d_{ii} temos que

$$0 < d_{ii} \leq x_i^k - l_i = x_i^{k+1} - d_i^k - l_i \leq -d_i^k$$

e portanto

$$\frac{-d_i^k}{d_{ii}} \geq 1 > \rho$$

e em consequência

$$\|D^{-1}d\| > \rho$$

o qual contradiz o fato de ser d solução de (III.3). Temos por conseguinte que $x_i^{k+1} > l_i \forall i = 1, \dots, n$. De maneira análoga demonstra-se que $\forall i = 1, \dots, n$ $x_i^{k+1} < u_i$. •

LEMA III.2. Suponhamos que as hipóteses padrões se cumprem e seja x^k um ponto gerado pelo algoritmo, o qual não satisfaz as condições necessárias de primeira ordem para uma solução de (III.1) e d^k um passo gerado pela solução de (III.3) na iteração k . Então

$$pred_k(d^k) \geq \frac{1}{2} \frac{g^{k^t} Dg^k}{\|g^k\|} \min\{\rho, \frac{g^{k^t} Dg^k}{\|g^k\| \|DB_k D\|}\}.$$

Demonstração.

Consideremos um ponto qualquer x^k (fixo) gerado pelo algoritmo, o qual não satisfaz as condições necessárias de primeira ordem para uma solução de (III.1). Então existe um índice i tal que $|g_i^k| > 0$ e portanto, $\|g^k\| > 0$.

Consideremos a função $\phi : \mathcal{R} \rightarrow \mathcal{R}$ definida por:

$$\phi(\delta) = q_k(-\delta D(\frac{g^k}{\|g^k\|})) = -\frac{\delta}{\|g^k\|} g^{k^t} Dg^k + \frac{1}{2} \frac{\delta^2}{\|g^k\|^2} g^{k^t} DB_k Dg^k.$$

Se $g^{k^t} DB_k Dg^k \leq 0$ então $\phi(\delta) \leq -\frac{\delta}{\|g^k\|} g^{k^t} Dg^k$. Em particular, para $\delta = \frac{\rho}{2}$ temos $\phi(\frac{\rho}{2}) \leq -\frac{\rho}{2\|g^k\|} g^{k^t} Dg^k$. Como d^k é solução de (III.3) e $-\frac{\rho}{2} D(\frac{g^k}{\|g^k\|})$ é viável para (III.3) então

$$-pred_k(d^k) = q_k(d^k) \leq q_k(-\frac{\rho}{2} D(\frac{g^k}{\|g^k\|})) = \phi(\frac{\rho}{2}) \leq -\frac{\rho}{2\|g^k\|} g^{k^t} Dg^k. \quad (III.6)$$

Suponhamos agora que $g^{kt} DB_k Dg^k > 0$. Então

$$\begin{aligned} \phi(\delta) &= -\frac{\delta}{\|g^k\|} g^{kt} Dg^k + \frac{1}{2} \frac{\delta^2}{\|g^k\|^2} g^{kt} DB_k Dg^k = \\ &= \frac{1}{2} \left(\frac{(g^{kt} Dg^k)^2}{g^{kt} DB_k Dg^k} - \frac{2\delta}{\|g^k\|} g^{kt} Dg^k + \frac{\delta^2}{\|g^k\|^2} g^{kt} DB_k Dg^k \right) - \frac{1}{2} \frac{(g^{kt} Dg^k)^2}{g^{kt} DB_k Dg^k} = \\ &= \frac{1}{2} \left(\frac{\delta}{\|g^k\|} \sqrt{g^{kt} DB_k Dg^k} - \frac{g^{kt} Dg^k}{\sqrt{g^{kt} DB_k Dg^k}} \right)^2 - \frac{1}{2} \frac{(g^{kt} Dg^k)^2}{g^{kt} DB_k Dg^k}. \end{aligned}$$

Seja δ^* o mínimo de ϕ no intervalo $[0, \rho]$. Então

$$-pred_k(d^k) = q_k(d^k) \leq q_k(-\delta^* D(\frac{g^k}{\|g^k\|})) = \phi(\delta^*). \quad (\text{III.7})$$

Se $\delta^* < \rho$ então

$$\delta^* = \frac{\|g^k\| g^{kt} Dg^k}{g^{kt} DB_k Dg^k}$$

e

$$\phi(\delta^*) = -\frac{1}{2} \frac{(g^{kt} Dg^k)^2}{g^{kt} DB_k Dg^k} \leq -\frac{1}{2} \frac{(g^{kt} Dg^k)^2}{\|g^k\|^2 \|DB_k D\|}. \quad (\text{III.8})$$

Se $\delta^* = \rho$ então

$$\rho \leq \frac{\|g^k\| g^{kt} Dg^k}{g^{kt} DB_k Dg^k}$$

e

$$\phi(\delta^*) = \phi(\rho) \leq -\frac{\rho}{\|g^k\|} g^{kt} Dg^k + \frac{\rho}{2 \|g^k\|} g^{kt} Dg^k = -\frac{\rho}{2 \|g^k\|} g^{kt} Dg^k. \quad (\text{III.9})$$

Portanto, por (III.6), (III.7), (III.8) e (III.9) temos que

$$\begin{aligned} pred_k(d^k) &\geq \min\left\{ \frac{\rho}{2 \|g^k\|} g^{kt} Dg^k, \frac{1}{2} \frac{(g^{kt} Dg^k)^2}{\|g^k\|^2 \|DB_k D\|} \right\} = \\ &= \frac{1}{2} \frac{g^{kt} Dg^k}{\|g^k\|} \min\left\{ \rho, \frac{g^{kt} Dg^k}{\|g^k\| \|DB_k D\|} \right\}. \quad \bullet \end{aligned}$$

COROLÁRIO III.1. Suponhamos que as hipóteses padrões se cumprem. Seja x^k um ponto gerado pelo algoritmo, o qual não satisfaz as condições necessárias de primeira ordem para uma solução de (III.1) e d^k um passo gerado pela solução de (III.3) na iteração k . Se

$$\Delta_k \leq \min\{x_i^k - l_i, u_i - x_i^k\}, \quad \forall i = 1, \dots, n$$

então existe uma constante $\kappa_1 > 0$ tal que

$$\text{pred}_k(d^k) \geq \frac{1}{2} \|g^k\| \min\{\rho \Delta_k, \kappa_1 \|g^k\|\}.$$

Demonstração.

Já que as hipóteses padrões se cumprem, existe uma constante $\beta > 0$ tal que $\|B_k\| \leq \beta$ para todo k . Seja $\kappa_1 = \frac{1}{\beta}$. Como

$$\Delta_k \leq \min\{x_i^k - l_i, u_i - x_i^k\}, \quad \forall i = 1, \dots, n$$

temos que $D = \Delta_k I$. Então, pelo lema III.2 temos que

$$\begin{aligned} \text{pred}_k(d^k) &\geq \frac{1}{2} \frac{g^{k^t} D g^k}{\|g^k\|} \min\left\{\rho, \frac{g^{k^t} D g^k}{\|g^k\| \|DB_k D\|}\right\} = \\ &\frac{1}{2} \Delta_k \|g^k\| \min\left\{\rho, \frac{\|g^k\|}{\Delta_k \|B_k\|}\right\} \geq \\ &\frac{1}{2} \Delta_k \|g^k\| \min\left\{\rho, \frac{\|g^k\|}{\Delta_k \beta}\right\} = \\ &\frac{1}{2} \|g^k\| \min\{\rho \Delta_k, \kappa_1 \|g^k\|\}. \quad \bullet \end{aligned}$$

COROLÁRIO III.2. Suponhamos que as hipóteses padrões se cumprem e seja x^k um ponto gerado pelo algoritmo, o qual não satisfaz as condições necessárias de primeira ordem para uma solução de (III.1) e d^k um passo gerado pela solução de (III.3) na iteração k . Dados $\epsilon_1 > 0$ e $\epsilon_2 > 0$, se existe uma componente x_i^k de x^k tal que

$$l_i + \epsilon_1 \leq x_i^k \leq u_i - \epsilon_1 \quad \text{e} \quad |g_i^k| = \left| \frac{\partial f}{\partial x_i}(x^k) \right| \geq \epsilon_2$$

então existe uma constante positiva κ_1 independente de x^k e de d^k tal que

$$\text{pred}_k(d^k) \geq \frac{1}{2} \epsilon_2 \min\{\rho \Delta_k, \rho \epsilon_1, \epsilon_2 \kappa_1\}.$$

Demonstração.

Já que as hipóteses padrões se cumprem, existe uma constante $\beta > 0$ tal que $|b_{ii}^k| \leq \|B_k\| \leq \beta$ para todo k . Seja $\kappa_1 = \frac{1}{\beta}$.

Consideremos um ponto qualquer x^k o qual satisfaz as hipóteses do corolário, isto é, existe uma componente x_i^k de x^k tal que

$$l_i + \epsilon_1 \leq x_i^k \leq u_i - \epsilon_1 \quad \text{e} \quad |g_i^k| = \left| \frac{\partial f}{\partial x_i}(x^k) \right| \geq \epsilon_2$$

para $\epsilon_1 > 0$ e $\epsilon_2 > 0$ dados. Temos portanto que $x_i^k - l_i \geq \epsilon_1$ e $u_i - x_i^k \geq \epsilon_1$ e em conseqüência temos que $d_{ii} \geq \min\{\Delta_k, \epsilon_1\}$. Consideremos a função $\phi : \mathcal{R} \rightarrow \mathcal{R}$ definida por:

$$\phi(\delta) = q_k(\delta D e_i) = \delta g_i^k d_{ii} + \frac{1}{2} \delta^2 d_{ii}^2 b_{ii}^k,$$

onde e_i é o i -ésimo vetor unitário.

$\forall \delta \in [-\rho, \rho]$ temos que $\delta D e_i$ é viável para o problema (III.3).

Como d^k é solução de (III.3) então

$$\text{pred}_k(d^k) = -q_k(d^k) \geq -q_k(\delta D e_i) = -\phi(\delta) \quad \forall \delta \in [-\rho, \rho]. \quad (\text{III.10})$$

Consideremos agora o problema:

$$\min \quad q_k(\delta d_{ii}) = g_i^k \delta d_{ii} + \frac{1}{2} \delta d_{ii} b_{ii}^k d_{ii} \delta \quad (\text{III.11})$$

$$\text{s. a.} \quad |d_{ii}^{-1} \delta d_{ii}| \leq \rho.$$

Seja δ^* solução de (III.11). Então, por (III.10) e pelo lema III.2 temos que:

$$\begin{aligned} \text{pred}_k(d^k) &\geq -q_k(\delta^* d_{ii}) \geq \frac{1}{2} \frac{g_i^k d_{ii} g_i^k}{|g_i^k|} \min\left\{\rho, \frac{g_i^k d_{ii} g_i^k}{|g_i^k| |d_{ii} b_{ii}^k d_{ii}|\right\} = \\ &\frac{1}{2} |g_i^k| d_{ii} \min\left\{\rho, \frac{|g_i^k|}{d_{ii} |b_{ii}^k|}\right\} = \frac{1}{2} |g_i^k| \min\left\{\rho d_{ii}, \frac{|g_i^k|}{|b_{ii}^k|}\right\} \geq \\ &\frac{1}{2} \epsilon_2 \min\{\rho \min\{\Delta_k, \epsilon_1\}, \epsilon_2 \kappa_1\} = \frac{1}{2} \epsilon_2 \min\{\rho \Delta_k, \rho \epsilon_1, \epsilon_2 \kappa_1\}. \quad \bullet \end{aligned}$$

O seguinte lema mostra que a redução prevista da função objetivo fornece uma aproximação da redução real da mesma função, cuja precisão é proporcional ao quadrado do comprimento do passo.

LEMA III.3. Suponhamos que as hipóteses padrões se cumprem. Então existe uma constante $\kappa_2 > 0$ tal que para qualquer $x \in S$ e qualquer $d \in \mathcal{R}^n$ tal que o segmento de reta que une x com $x + d$ está contido em S ,

$$|ared(d) - pred(d)| \leq \kappa_2 \|d\|^2,$$

onde

$$pred(d) = -g(x)^t d - \frac{1}{2} d^t B d$$

e

$$ared(d) = f(x) - f(x + d).$$

Demonstração. Por ser f duas vezes continuamente diferenciável temos que

$$f(x + d) = f(x) + g(x)^t d + \frac{1}{2} d^t B d + O(\|d\|^2),$$

o qual implica que

$$f(x) - f(x + d) - (-g(x)^t d - \frac{1}{2} d^t B d) = O(\|d\|^2),$$

e portanto

$$ared(d) - pred(d) = O(\|d\|^2).$$

Por definição de $O(\cdot)$ temos que existe uma constante $\kappa_2 > 0$ tal que

$$|ared(d) - pred(d)| \leq \kappa_2 \|d\|^2. \quad \bullet$$

O seguinte teorema demonstra que o algoritmo está bem definido, no sentido que cada iteração interna terminará com um passo aceitável depois de um número finito de iterações.

TEOREMA III.1. Suponhamos que as hipóteses padrões se cumprem. Então, a menos que algum ponto x^k satisfaça as condições necessárias de primeira ordem para uma solução de (III.1), cada iteração interna do algoritmo terminará depois de um número finito de iterações.

Demonstração.

Consideremos um ponto qualquer x^k (fixo) gerado pelo algoritmo, o qual não satisfaz as condições necessárias de primeira ordem para uma solução de (III.1). Então $\|g^k\| > 0$. Consideremos primeiro o teste

$$\frac{ared_k(d^k)}{pred_k(d^k)} \geq \eta,$$

o qual é feito em cada iteração interna, com valores decrescentes do raio de região de confiança Δ_k (para o mesmo x^k temos $\Delta_k \downarrow 0$). Para Δ_k suficientemente pequeno, pelo corolário III.1, temos

$$pred_k(d^k) \geq \frac{1}{2}\|g^k\| \min\{\rho\Delta_k, \kappa_1\|g^k\|\},$$

para alguma constante $\kappa_1 > 0$.

Como $\|g^k\| > 0$ então para Δ_k suficientemente pequeno temos que

$$pred_k(d^k) \geq \frac{1}{2}\rho\|g^k\|\Delta_k, \text{ (pois } \kappa_1\|g^k\| \text{ permanece fixo e } \Delta_k \downarrow 0).$$

Por outra parte, pelo lema III.3 e por (III.4) temos que

$$|ared_k(d^k) - pred_k(d^k)| \leq \kappa_2\|d\|^2 \leq \kappa_2\rho^2\Delta_k^2,$$

para alguma constante $\kappa_2 > 0$. Logo,

$$\left| \frac{ared_k(d^k)}{pred_k(d^k)} - 1 \right| \leq \frac{\kappa_2\rho^2\Delta_k^2}{pred_k(d^k)} \leq \frac{2\kappa_2\rho}{\|g^k\|}\Delta_k.$$

Então, para Δ_k suficientemente pequeno temos que

$$\left| \frac{ared_k(d^k)}{pred_k(d^k)} - 1 \right| \leq (1 - \eta), \text{ o qual implica que } \frac{ared_k(d^k)}{pred_k(d^k)} \geq \eta.$$

Portanto, depois de um número finito de iterações internas, o passo será aceito. •

No que segue vamos dar o resultado de convergência global de primeira ordem. Demonstraremos, através de um conjunto de lemas e teoremas, que qualquer ponto de acumulação da seqüência de pontos gerada pelo algoritmo satisfaz as condições necessárias de primeira ordem para uma solução de (III.1).

LEMA III.4. Suponhamos que as hipóteses padrões se cumprem. Então

- i. Qualquer ponto de acumulação \hat{x} de $\{x^k\}$ é viável para o problema (III.1).
- ii. $\{f(x^k)\}$ é estritamente decrescente.
- iii. $f(\hat{x}^1) = f(\hat{x}^2)$ para qualquer par de pontos de acumulação \hat{x}^1, \hat{x}^2 de $\{x^k\}$.

Demonstração.

- i. Pelo lema III.1, a seqüência $\{x^k\}$ está contida em H_1 o qual é um conjunto fechado e em conseqüência contém todos seus pontos de acumulação.
- ii. Pelo teorema III.1, no final da k -ésima iteração do algoritmo, depois de um número finito de iterações internas, o teste

$$\frac{ared_k(d^k)}{pred_k(d^k)} \geq \eta,$$

é satisfeito. Temos portanto que

$$f(x^k) - f(x^{k+1}) = f(x^k) - f(x^k + d^k) = ared_k(d^k) \geq \eta pred_k(d^k) > 0.$$

Portanto, $f(x^k) > f(x^{k+1}) \forall k$.

- iii. Como conseqüência do lema III.1 e de ii. deste lema, temos que a seqüência $\{x^k\}$ está contida em \bar{H}^0 . Portanto, todos seus pontos de acumulação estão contidos em \bar{H} . Por ser f contínua e o conjunto \bar{H} compacto, então f é limitada em \bar{H} . Temos portanto que $\{f(x^k)\}$ é limitada inferiormente em \bar{H} . Sejam \hat{x}^1 e \hat{x}^2 dois pontos de acumulação de $\{x^k\}$. Então existem subseqüências $\{x^{k_i}\}$ e $\{x^{k_j}\}$ tais que

$$\{x^{k_i}\} \longrightarrow \hat{x}^1 \quad \text{e} \quad \{x^{k_j}\} \longrightarrow \hat{x}^2.$$

Pela continuidade de f temos que

$$\{f(x^{k_i})\} \longrightarrow f(\hat{x}^1) \quad \text{e} \quad \{f(x^{k_j})\} \longrightarrow f(\hat{x}^2).$$

Temos então que $f(\hat{x}^1)$ e $f(\hat{x}^2)$ são pontos de acumulação de $\{f(x^k)\}$. Por outra parte, a seqüência $\{f(x^k)\}$ é limitada inferiormente e estritamente decrescente e portanto é convergente. Em conseqüência, $f(\hat{x}^1) = f(\hat{x}^2)$. •

Note-se que se o ponto inicial x^0 da seqüência $\{x^k\}$ gerada pelo algoritmo, está em H_1^0 então, pelos lemas III.1 e III.4, temos que a seqüência inteira $\{x^k\}$ está contida em \bar{H} . Como por hipóteses, o conjunto \bar{H} é compacto, temos que o conjunto de pontos de acumulação de $\{x^k\}$ é não vazio e está contido em \bar{H} . Temos portanto que todo ponto de acumulação de $\{x^k\}$ satisfaz a condição (III.2.1).

No teorema seguinte demonstramos que todo ponto de acumulação da seqüência $\{x^k\}$ satisfaz a condição (III.2.3).

TEOREMA III.2. Suponhamos que as hipóteses padrões se cumprem. Então, para qualquer ponto de acumulação \hat{x} de $\{x^k\}$ temos que

$$\text{se } l_i < \hat{x}_i < u_i \text{ então } \frac{\partial f}{\partial x_i}(\hat{x}) = 0.$$

Demonstração.

Seja \hat{x} um ponto de acumulação de $\{x^k\}$. Então existe uma subsequência $\{x^{k_j}\}$ de $\{x^k\}$ tal que

$$\lim_{j \rightarrow +\infty} x^{k_j} = \hat{x}. \quad (\text{III.12})$$

A prova será feita pelo absurdo. Suponhamos que existe uma componente \hat{x}_i de \hat{x} tal que

$$l_i < \hat{x}_i < u_i \text{ e } \frac{\partial f}{\partial x_i}(\hat{x}) \neq 0.$$

Então existe $\epsilon > 0$ tal que $l_i + \epsilon \leq \hat{x}_i \leq u_i - \epsilon$ e $|\frac{\partial f}{\partial x_i}(\hat{x})| \geq \epsilon$. Pela continuidade de g e por (III.12), temos $\exists \bar{N} > 0$ tal que $\forall j > \bar{N}$

$$|x_i^{k_j} - \hat{x}_i| \leq \frac{\epsilon}{2} \text{ e } |g_i^{k_j}| \geq \frac{\epsilon}{2}.$$

Temos portanto, que para $j > \bar{N}$ $l_i + \frac{\epsilon}{2} \leq x_i^{k_j} \leq u_i - \frac{\epsilon}{2}$. Pelo corolário III.2, existe uma constante positiva κ_1 tal que

$$\text{pred}_{k_j}(d^{k_j}) \geq \frac{\epsilon}{4} \min\{\rho \frac{\epsilon}{2}, \frac{\epsilon}{2} \kappa_1, \rho \Delta_{k_j}\} = w > 0. \quad (\text{III.13})$$

Pelo lema III.3, por (III.13) e por (III.4) temos que existe uma constante positiva κ_2 tal que

$$\left| \frac{\text{ared}_{k_j}(d^{k_j})}{\text{pred}_{k_j}(d^{k_j})} - 1 \right| \leq \frac{\kappa_2 \rho^2 \Delta_{k_j}^2}{w},$$

o qual implica que

$$\frac{ared_{k_j}(d^{k_j})}{pred_{k_j}(d^{k_j})} \geq 1 - \frac{\kappa_2 \rho^2 \Delta_{k_j}^2}{w} \geq \eta$$

se $\Delta_{k_j} \leq \min \left\{ \sqrt{\frac{(1-\eta)\epsilon^2}{8\rho\kappa_2}}, \sqrt{\frac{(1-\eta)\epsilon^2\kappa_1}{8\rho^2\kappa_2}}, \frac{(1-\eta)\epsilon}{4\rho\kappa_2} \right\} = \bar{\Delta}$.

Assim, o passo de região de confiança d^{k_j} será aceito sempre que $\Delta_{k_j} \leq \bar{\Delta}$, e portanto $\Delta_{k_j} \geq \tau_1 \bar{\Delta}$, onde τ_1 é o fator de redução da bola de região de confiança.

Logo, temos que

$$\frac{ared_{k_j}(d^{k_j})}{pred_{k_j}(d^{k_j})} \geq \eta \text{ se } \Delta_{k_j} \in [\tau_1 \bar{\Delta}, \bar{\Delta}].$$

Então, $\forall j > \bar{N}$ temos que

$$ared_{k_j}(d^{k_j}) \geq \eta pred_{k_j}(d^{k_j}) \geq \eta \frac{\epsilon}{4} \min \left\{ \rho \frac{\epsilon}{2}, \frac{\epsilon}{2} \kappa_1, \rho \Delta_{k_j} \right\} \geq \eta \frac{\epsilon}{4} \min \left\{ \rho \frac{\epsilon}{2}, \frac{\epsilon}{2} \kappa_1, \tau_1 \rho \bar{\Delta} \right\} > 0.$$

Se a seqüência inteira converge a \hat{x} então podemos tomar $\{x^{k_j}\} = \{x^k\}$. Neste caso temos uma contradição com o fato de ser $\{f(x^k)\}$ limitada inferiormente em \bar{H} , já que para qualquer $j > \bar{N}$ temos

$$f(x^j) - f(x^{j+1}) = ared_j(d^j) \geq \eta \frac{\epsilon}{4} \min \left\{ \rho \frac{\epsilon}{2}, \frac{\epsilon}{2} \kappa_1, \tau_1 \rho \bar{\Delta} \right\} > 0.$$

Suponhamos agora que toda a seqüência não converge a \hat{x} . Então existe um $r_1 > 0$ tal que existem infinitos x^k que ficam fora da bola centrada em \hat{x} de raio r_1 , a qual denotaremos por $B(\hat{x}, r_1)$. Novamente, pela continuidade de g e por (III.12), temos que $\exists r_2 > 0$ tal que

$$B(\hat{x}, r_2) \cap \{x^{k_j}\}_{j=1}^{+\infty} \subset \{x^{k_j} / \|x^{k_j} - \hat{x}\| \leq \frac{\epsilon}{2} \text{ e } |g_i^{k_j}| \geq \frac{\epsilon}{2}\}.$$

Tomemos $r = \min\{r_1, r_2\}$. Seja $\bar{N}_0 > 0$ tal que $\forall j > \bar{N}_0$ $x^{k_j} \in B(\hat{x}, \frac{r}{2})$ e $k > k_j$ o menor índice tal que $x^k \notin B(\hat{x}, r)$. Neste caso temos que

$$x^k = x^{k_j} + \sum_{i=k_j}^{k-1} d^i,$$

o qual implica que $\forall j > \bar{N}_0$

$$\frac{r}{2} < \|x^k - x^{k_j}\| = \left\| \sum_{i=k_j}^{k-1} d^i \right\| \leq \sum_{i=k_j}^{k-1} \|d^i\| \leq \sum_{i=k_j}^{k-1} \rho \Delta_i.$$

Portanto,

$$f(x^{k_j}) - f(x^k) = \sum_{i=k_j}^{k-1} \{f(x^i) - f(x^{i+1})\} = \sum_{i=k_j}^{k-1} \text{ared}_i(d^i) \geq$$

$$\sum_{i=k_j}^{k-1} \eta \text{pred}_i(d^i) \geq \eta \frac{\epsilon}{4} \sum_{i=k_j}^{k-1} \min\{\rho \frac{\epsilon}{2}, \frac{\epsilon}{2} \kappa_1, \rho \Delta_i\} \geq$$

$$\eta \frac{\epsilon}{4} \min\{\rho \frac{\epsilon}{2}, \frac{\epsilon}{2} \kappa_1, \sum_{i=k_j}^{k-1} \rho \Delta_i\} \geq \eta \frac{\epsilon}{8} \min\{\rho \epsilon, \epsilon \kappa_1, r\}.$$

Temos novamente uma contradição com o fato de ser $\{f(x^k)\}$ limitada inferiormente em \bar{H} , pois existem infinitos x^k que produzem um decréscimo pelo menos constante em f . Em conseqüência cumpre-se

$$\frac{\partial f}{\partial x_i}(\hat{x}) = 0 \text{ se } l_i < \hat{x}_i < u_i. \bullet$$

COROLÁRIO III.3. Suponhamos que as hipóteses padrões se cumprem. Então, para qualquer ponto de acumulação \hat{x} de $\{x^k\}$ temos que

$$\min\{\hat{x}_i - l_i, u_i - \hat{x}_i\} \frac{\partial f}{\partial x_i}(\hat{x}) = 0 \quad \forall i = 1, \dots, n.$$

Demonstração.

Pelo teorema III.2 temos que

$$\text{se } l_i < \hat{x}_i < u_i \text{ então } \frac{\partial f}{\partial x_i}(\hat{x}) = 0.$$

Por outro lado, pelo contrário do mesmo teorema e pelo lema III.4 temos que

$$\text{se } \frac{\partial f}{\partial x_i}(\hat{x}) \neq 0 \text{ então } \hat{x}_i = l_i \text{ ou } \hat{x}_i = u_i. \bullet$$

LEMA III.5. Suponhamos que as hipóteses padrões se cumprem. Então

$$pred_k(d^k) \longrightarrow 0.$$

Demonstração.

Pelo teorema III.1, no final da k -ésima iteração do algoritmo, depois de um número finito de iterações internas, o teste

$$\frac{ared_k(d^k)}{pred_k(d^k)} \geq \eta,$$

é satisfeito. Temos portanto que para todo k se cumpre:

$$ared_k(d^k) \geq \eta pred_k(d^k) > 0,$$

o qual implica que

$$f(x^k) - f(x^{k+1}) \geq \eta pred_k(d^k) > 0.$$

Como a seqüência $\{f(x^k)\}$ é convergente, então $f(x^k) - f(x^{k+1}) \longrightarrow 0$ e em conseqüência tem-se que $pred_k(d^k) \longrightarrow 0$. •

LEMA III.6. Suponhamos que as hipóteses padrões se cumprem. Seja \hat{x} um ponto de acumulação de $\{x^k\}$ e $\{x^{k_j}\}$ uma subseqüência de $\{x^k\}$ tal que $\lim_{j \rightarrow +\infty} x^{k_j} = \hat{x}$. Se $g(\hat{x}) \neq 0$ então $\inf \Delta_{k_j} > 0$.

Demonstração.

Devemos provar que $\Delta_{k_j} \not\rightarrow 0$. Suponhamos o contrário. Então $D_{k_j} \longrightarrow 0$.

Pelo lema III.2, temos

$$pred_{k_j}(d^{k_j}) \geq \frac{1}{2} \frac{g^{k_j t} D_{k_j} g^{k_j}}{\|g^{k_j}\|} \min\left\{\rho, \frac{g^{k_j t} D_{k_j} g^{k_j}}{\|g^{k_j}\| \|D_{k_j} B_{k_j} D_{k_j}\|}\right\}.$$

Para j suficientemente grande temos que $\|g^{k_j}\| > \frac{\|g(\hat{x})\|}{2}$ e como $D_{k_j} \longrightarrow 0$

e $B(x)$ é limitada em S então temos que

$$pred_{k_j}(d^{k_j}) \geq \frac{1}{2} \rho \frac{g^{k_j t} D_{k_j} g^{k_j}}{\|g^{k_j}\|}.$$

Por outra parte, pelo lema III.3 e por (III.4) temos que

$$|\text{ared}_{k_j}(d^{k_j}) - \text{pred}_{k_j}(d^{k_j})| \leq \kappa_2 \|d^{k_j}\|^2 \leq \kappa_2 \rho^2 \Delta_{k_j}^2,$$

para alguma constante $\kappa_2 > 0$. Logo,

$$\left| \frac{\text{ared}_{k_j}(d^{k_j})}{\text{pred}_{k_j}(d^{k_j})} - 1 \right| \leq \frac{\kappa_2 \rho^2 \Delta_{k_j}^2}{\text{pred}_{k_j}(d^{k_j})} \leq \frac{2 \kappa_2 \rho \|g^{k_j}\|}{g^{k_j t} D_{k_j} g^{k_j}} \Delta_{k_j}^2.$$

Em conseqüência, se $\Delta_{k_j} \rightarrow 0$ então $\frac{\text{ared}_{k_j}(d^{k_j})}{\text{pred}_{k_j}(d^{k_j})} \rightarrow 1$. Isto leva a uma contradição com a regra de atualização de Δ , já que Δ não pode decrescer quando $\frac{\text{ared}_{k_j}(d^{k_j})}{\text{pred}_{k_j}(d^{k_j})} \geq \eta$. Portanto, $\Delta_{k_j} \not\rightarrow 0$. •

Seja \hat{x} um ponto de acumulação de $\{x^k\}$ e $\hat{g} = \nabla f(\hat{x})$. Se $\hat{g} = 0$ então as condições necessárias de otimalidade de primeira ordem são satisfeitas em \hat{x} . Suponhamos agora que $\hat{g} \neq 0$. Definimos os seguintes conjuntos:

$$N = \{i = 1, \dots, n / \hat{g}_i \neq 0\} \text{ e } \bar{N} = \{i = 1, \dots, n\} - N,$$

$$\text{e a constante } \epsilon = \frac{1}{2} \min_{i \in N} |\hat{g}_i|. \quad (\text{III.14})$$

Pelo corolário III.3 temos que para todo $i \in N$ $\hat{x}_i = l_i$ ou $\hat{x}_i = u_i$.

$$\text{Seja } N_l = \{i \in N / \hat{x}_i = l_i\} \text{ e } N_u = \{i \in N / \hat{x}_i = u_i\}.$$

Para demonstrar que \hat{x} satisfaz as condições necessárias de otimalidade de primeira ordem só falta demonstrar que os sinais do gradiente no ponto \hat{x} são corretos, isto é,

$$\hat{g}_i > 0 \quad \forall i \in N_l \quad \text{e} \quad \hat{g}_i < 0 \quad \forall i \in N_u. \quad (\text{III.15})$$

Nos resultados obtidos até agora não foi preciso colocar a hipótese de convexidade na função objetivo, mas para poder demonstrar (III.15) vamos ter que exigir que a função f seja convexa. Portanto, no que segue, suporemos que a função f é convexa.

O CONJUNTO FINAL.

O ponto \hat{x} pertence à face do conjunto viável

$$F = \{x \in H_1 / x_i = l_i \quad \forall i \in N_l \text{ e } x_i = u_i \quad \forall i \in N_u\}.$$

Nesta face vamos definir o seguinte conjunto final:

$$\Omega = \{ x \in F / f(x) = f(\hat{x}) \}.$$

LEMA III.7. O conjunto Ω é convexo.

Demonstração.

Seja $x \in CO(\Omega)$. Devemos provar que $x \in \Omega$. Já que $x \in CO(\Omega)$ temos que $x_i = l_i \forall i \in N_l$ e $x_i = u_i \forall i \in N_u$, e pela convexidade de f , $f(x) \leq f(\hat{x})$. Devemos ver agora que $f(x) \geq f(\hat{x})$. Seja $h = x - \hat{x}$. Note-se que $h_N = 0$. Pela convexidade de f , $f(x) \geq f(\hat{x}) + \nabla f(\hat{x})^t h = f(\hat{x})$, já que $(\hat{g})^t h = 0$. Isto prova que $f(x) \geq f(\hat{x})$, completando a demonstração. •

O seguinte lema é um resultado conhecido da análise convexa.

LEMA III.8. Seja $f : \mathcal{R}^n \rightarrow \mathcal{R}$ uma função continuamente diferenciável. Se f é constante em um conjunto convexo $\Omega \subset \mathcal{R}^n$ então ∇f é constante em Ω .

Demonstração.

A demonstração pode ser achada em [27]. •

LEMA III.9. Para qualquer $x \in \Omega$, $\nabla f(x) = \nabla f(\hat{x})$.

Demonstração.

Pelo lema III.7, Ω é um conjunto convexo, no qual f é constante. Portanto, pelo lema III.8, ∇f é constante em Ω . Segue-se disto que $\nabla f(x) = \nabla f(\hat{x}) \forall x \in \Omega$. •

Consideremos agora o conjunto final Ω . Dado um número $\tau > 0$, definimos uma vizinhança Ω_τ de Ω por

$$\Omega_\tau = \{ x \in \bar{H} / \|x - y\|_\infty \leq \tau \text{ para algum } y \in \Omega \}.$$

LEMA III.10. Seja $\epsilon > 0$ definido como em (III.14). Então existe um $\bar{\delta} > 0$ tal que $\forall x \in \Omega_{\bar{\delta}}$ temos:

- i. $\|\nabla f(x) - \nabla f(\hat{x})\| \leq \epsilon$,
- ii. $\forall i \in N$ temos que $g_i(x)\hat{g}_i > 0$.

Demonstração.

Por ser ∇f contínuo em \bar{H} e ser este conjunto compacto, temos que ∇f é uniformemente contínuo em \bar{H} . Portanto, $\exists \bar{\delta} > 0$ tal que

$$\forall x, y \in \bar{H} \quad \|x - y\|_{\infty} \leq \bar{\delta} \implies \|\nabla f(x) - \nabla f(y)\| \leq \epsilon.$$

Da definição de $\Omega_{\bar{\delta}}$, para cada $x \in \Omega_{\bar{\delta}} \exists y \in \Omega$ tal que $\|x - y\|_{\infty} \leq \bar{\delta}$. Pela continuidade uniforme de ∇f e pelo lema III.9 temos

$$\|\nabla f(x) - \nabla f(y)\| = \|\nabla f(x) - \nabla f(\hat{x})\| \leq \epsilon,$$

provando i.

Prova de ii.: Pela definição de ϵ , para cada $i \in N$ temos que

$$|g_i(x) - \hat{g}_i| \leq \|g(x) - \hat{g}\| \leq \epsilon \leq \frac{1}{2}|\hat{g}_i|,$$

e deste modo, $g_i(x)$ e \hat{g}_i tem o mesmo sinal, portanto $g_i(x)\hat{g}_i > 0 \quad \forall i \in N$. •

Por ser \bar{H} compacto, existem vetores \hat{l} e \hat{u} tais que todas suas coordenadas são finitas, $l \leq \hat{l} < \hat{u} \leq u$ e $\bar{H} \subset \{x \in \mathcal{R}^n / \hat{l} \leq x \leq \hat{u}\}$.

$$\text{Seja } \delta = \frac{1}{2} \min \left\{ \bar{\delta}, \frac{1}{2} \min_{1 \leq i \leq n} \{\hat{u}_i - \hat{l}_i\} \right\}.$$

O seguinte lema relaciona outros pontos de acumulação de $\{x^k\}$ ao conjunto Ω associado a \hat{x} .

LEMA III.11. Seja \tilde{x} um ponto de acumulação de $\{x^k\}$. Então, $\tilde{x} \in \Omega$ ou $\tilde{x} \notin \Omega_{\delta}$.

Demonstração.

Ab absurdo. Suponhamos que \tilde{x} é um ponto de acumulação de $\{x^k\}$ tal que $\tilde{x} \in \Omega_\delta$ e $\tilde{x} \notin \Omega$. Pelo lema III.4 temos que $f(\tilde{x}) = f(\hat{x})$, assim nossa hipótese requer que $\tilde{x}_i \neq l_i$ para algum $i \in N_l$ ou $\tilde{x}_i \neq u_i$ para algum $i \in N_u$. Suponhamos que $\tilde{x}_i \neq l_i$ para algum $i \in N_l$. Como $\tilde{x} \in \Omega_\delta$ então $\tilde{x}_i \neq u_i$. Pelo lema III.10 temos que $g_i(\tilde{x}) \neq 0$ já que $\tilde{x} \in \Omega_\delta$ e $i \in N$. Isto contradiz o corolário III.3, completando a prova. •

LEMA III.12. Seja $\{x^k\}$ uma seqüência gerada pelo algoritmo. Se f é convexa então, $\forall k \ g^{k^t} d^k < 0$.

Demonstração.

Pelo teorema de Taylor temos que:

$$f(x^k + d^k) = f(x^k) + g^{k^t} d^k + \frac{1}{2} d^{k^t} B(x^k + \theta d^k) d^k,$$

para algum $0 \leq \theta \leq 1$. Como f é convexa temos que $d^{k^t} B(x^k + \theta d^k) d^k \geq 0$.

Então, pelo lema III.4 tem-se:

$$g^{k^t} d^k \leq f(x^k + d^k) - f(x^k) < 0. \quad \bullet$$

LEMA III.13. As seqüências geradas pelo algoritmo satisfazem $g^{k^t} d^k \rightarrow 0$.

Demonstração.

Por ser f convexa temos que $\forall k \ d^{k^t} B_k d^k \geq 0$. Temos então que

$$\forall k \ pred_k(d^k) = d^{k^t} \left(\frac{1}{2} B_k + \psi D^{-2} \right) d^k \geq \frac{1}{2} d^{k^t} B_k d^k \geq 0.$$

Pelo lema III.5 temos que $pred_k(d^k) \rightarrow 0$. Portanto, $d^{k^t} B_k d^k \rightarrow 0$.

Pela definição de $pred_k(d^k)$ temos que:

$$g^{k^t} d^k = -pred_k(d^k) - \frac{1}{2} d^{k^t} B_k d^k.$$

Como os dois termos do lado direito da última equação tendem a zero, então temos que $g^{k^t} d^k \rightarrow 0$. •

LEMA III.14. Existe $\tilde{k} > 0$ tal que $\forall k \geq \tilde{k}$ se $x^k \in \Omega_\delta$ então $\forall i \in N \hat{g}_i d_i^k < 0$.

Demonstração.

Seja $K^0 \subset \mathcal{N}$ tal que $x^k \xrightarrow{K^0} \hat{x}$ com $\hat{x} \in \Omega$. Por (III.5.2) e (III.5.4) $\forall k \in \mathcal{N}$ existe $\psi_k \in \mathcal{R}$ tal que

$$g^k + B_k d^k = -\psi_k D^{-2} d^k$$

$$\psi_k \geq 0.$$

Se $d^k \xrightarrow{K^0} 0$ então, por ser $B(x)$ limitada em S , $B_k d^k \xrightarrow{K^0} 0$ e portanto $\psi_k D^{-2} d^k \xrightarrow{K^0} -\hat{g} \neq 0$. Portanto existe $\tilde{k} > 0$ tal que $\forall k \in K^0$ se $k \geq \tilde{k}$ então $x^k \in \Omega_\delta$ e $\forall i \in N \hat{g}_i d_i^k < 0$.

Suponhamos agora que $d^k \not\xrightarrow{K^0} 0$. Como a seqüência $\{x^k\}$ é limitada, devem existir $K \subset K^0$ e $\tilde{x} \in H_1$ tal que $x^k \xrightarrow{K} \hat{x}$, $x^{k+1} \rightarrow \tilde{x}$ e $d^k \xrightarrow{K} \hat{d} = \tilde{x} - \hat{x} \neq 0$. Demonstraremos que também neste caso $B_k d^k \xrightarrow{K} 0$.

Seja $I = \{x = \hat{x} + \lambda \hat{d} / \lambda \in [0, 1]\} \subset H_1$. Vejamos que f é constante no conjunto I . Pelo lema anterior temos que $\hat{g}^t \hat{d} = 0$. Em conseqüência para qualquer $x \in I$ temos que

$$f(x) = f(\hat{x} + \lambda \hat{d}) = f(\hat{x}) + \lambda \hat{g}^t \hat{d} + O(\hat{x}, \lambda \hat{d}) = f(\hat{x}) + O(\hat{x}, \lambda \hat{d}),$$

sendo $O(\hat{x}, \lambda \hat{d})$ crescente em λ , devido ao fato de ser f convexa. Tomando $\lambda = 1$ temos que $x = \tilde{x}$ e como $f(\hat{x}) = f(\tilde{x})$ então $O(\hat{x}, 0\hat{d}) = O(\hat{x}, 1\hat{d}) = 0$, o qual implica que f é constante em I e conseqüentemente, pelo lema III.9, temos $g(x) = \nabla f(x)$ é constante em I .

Pela continuidade de B temos que $B_k d^k \xrightarrow{K} B(\hat{x})\hat{d}$. Vamos demonstrar que $B(\hat{x})\hat{d} = 0$.

$$\forall x \in I \text{ temos que } g(x) = g(\hat{x} + \lambda \hat{d}) = g(\hat{x}) + \lambda B(\hat{x})\hat{d} + O(\hat{x}, \lambda \hat{d}).$$

Portanto, $\lambda B(\hat{x})\hat{d} + O(\hat{x}, \lambda \hat{d}) = 0$, ou seja

$$B(\hat{x})\hat{d} = \frac{O(\hat{x}, \lambda \hat{d})}{\lambda}.$$

Tomando limite quando $\lambda \rightarrow 0$, por ser g diferenciável temos que $B(\hat{x})\hat{d} = 0$.

Novamente, como $B_k d^k \xrightarrow{K} 0$ temos que existe $\tilde{k} > 0$ tal que $\forall k \in K$ se $k \geq \tilde{k}$ então $x^k \in \Omega_\delta$ e $\forall i \in N \hat{g}_i d_i^k < 0$. •

LEMA III.15. Sejam \hat{x} , $\hat{g} = \nabla f(\hat{x})$ e δ como nos lemas anteriores. Se $\hat{g}_i < 0$ para algum $i \in N_l$ ou $\hat{g}_i > 0$ para algum $i \in N_u$ então $\{x^k\}$ tem um ponto de acumulação \tilde{x} tal que $\tilde{x} \notin \Omega_\delta$.

Demonstração.

Suponhamos que $\hat{g}_i < 0$ para algum $i \in N_l$ e suponhamos pelo absurdo que existe um $\bar{k} > 0$ tal que $\forall k \geq \bar{k} x^k \in \Omega_\delta$. Seja $\hat{k} = \max\{\bar{k}, \tilde{k}\}$, onde \tilde{k} é como no lema anterior. Pelo lema III.14 temos que $\forall k \geq \hat{k} \hat{g}_i d_i^k < 0$ e portanto $d_i^k > 0$. Segue-se que $\forall k \geq \hat{k} x_i^{k+1} = x_i^k + d_i^k > x_i^k$, e a seqüência $\{x_i^k\}$ cresce monotonamente. Isto contradiz o fato de que $\liminf x_i^k = l_i$. De maneira análoga demonstra-se o lema quando $\hat{g}_i > 0$ para algum $i \in N_u$. •

O último teorema completa a prova que os sinais de \hat{g}_N são corretos e portanto \hat{x} satisfaz as condições necessárias de otimalidade de Karush-Kuhn-Tucker.

TEOREMA III.3. Sejam \hat{x} e $\hat{g} = \nabla f(\hat{x})$ como acima. Então $\hat{g}_{N_l} > 0$ e $\hat{g}_{N_u} < 0$.

Demonstração.

Ab absurdo. Suponhamos primeiro que existe um $i \in N_l$ tal que $\hat{g}_i < 0$. Vejamos em primeiro lugar que

$$\exists K \subset \mathcal{N} \text{ t. q. } x^k \xrightarrow{K} \bar{x}^1, x^{k+1} \xrightarrow{K} \bar{x}^2, \text{ com } \bar{x}^1 \in \Omega \text{ e } \bar{x}^2 \notin \Omega_\delta. \quad (\text{III.16})$$

Pelo lema III.15 existem pontos de acumulação em Ω e fora de Ω_δ e em conseqüência existe $K^0 \subset \mathcal{N}$ t. q. $\forall k \in K^0 x^k \in \Omega_\delta$ e $x^{k+1} \notin \Omega_\delta$. Como a seqüência $\{x^k\}$ é limitada, devem existir $K \subset K^0$, \bar{x}^1 e \bar{x}^2 t. q. $x^k \xrightarrow{K} \bar{x}^1$, e $x^{k+1} \xrightarrow{K} \bar{x}^2$. Devemos provar que $\bar{x}^1 \in \Omega$ e $\bar{x}^2 \notin \Omega_\delta$. Já que Ω_δ é fechado, $\bar{x}^1 \in \Omega_\delta$, e pelo lema III.11 $\bar{x}^1 \in \Omega$.

Suponhamos pelo absurdo que $\bar{x}^2 \in \Omega_\delta$. Então pelo lema III.11 devemos ter que

$x^2 \in \Omega$. Então para k suficientemente grande, $k \in K$, $\|x^{k+1} - x^2\|_\infty < \delta$, o qual implica que $x^{k+1} \in \Omega_\delta$. Isto contradiz a definição de K^0 e prova que $x^2 \notin \Omega_\delta$. Com isto fica demonstrado (III.16).

Demonstraremos agora que

$$\exists j \in N_1 \text{ t. q. } \hat{g}_j < 0 \text{ e } x_2^j > x_1^j = l_j \text{ ou}$$

$$\exists j \in N_u \text{ t. q. } \hat{g}_j > 0 \text{ e } x_2^j < x_1^j = n_j. \quad (\text{III.17})$$

Como $f(x_1) = f(x_2) \in \Omega$ então $x_2^{N_u} \neq l_{N_u}$ ou $x_2^{N_u} \neq n_{N_u}$. Suponhamos que para algum $j \in N_1$ $x_2^j > x_1^j = l_j$. Já que $x^{j+1} \xrightarrow{K} x_2^j - x_1^j$, para k suficientemente grande temos que $x^k \in \Omega_\delta$ e $d_k^j = x_{k+1}^j - x_k^j > 0$. Agora, pelo lema III.14, $\hat{g}_j > 0$. Analogamente prova-se que se $\exists j \in N_u$ t. q. $x_2^j < x_1^j = n_j$ então $\hat{g}_j < 0$.

Temos demonstrado assim (III.17).

Consideremos agora a sequência de passos $d^k = x^{k+1} - x^k$ e estudemos a sequência de valores $g^{k_t} d^k$. Pelo lema III.13, temos que $g^{k_t} d^k \rightarrow 0$. Tomando limites no conjunto K obtemos:

$$g^{k_t} d^k \xrightarrow{K} \hat{g}_t(x_2 - x_1) = g_t^N(x_2^N - x_1^N)$$

e em consequência

$$g_t^N(x_2^N - x_1^N) = 0. \quad (\text{III.18})$$

Por outra parte, para $k \in K$ suficientemente grande $x^k \in \Omega_\delta$ e pelo lema III.14, $\hat{g}_t d^k > 0$. Tomando limites nesses produtos para $k \in K$ concluímos que

$$\forall i \in N \quad \hat{g}_i(x_2^i - x_1^i) \leq 0.$$

Mas por (III.17), uma dessas desigualdades deve ser estrita, o qual nos leva a

$$\hat{g}_t^N(x_2^N - x_1^N) > 0,$$

o qual contradiz (III.18). Portanto, $\forall i \in N_1$ tem-se que $\hat{g}_i > 0$. Analogamente demonstra-se que $\forall i \in N_u$ $\hat{g}_i > 0$.

Capítulo IV

Um Algoritmo de Região de Confiança para Programação não Linear com Restrições não Lineares de Igualdade e Variáveis Canalizadas.

IV.1 Introdução.

Neste capítulo apresentamos um novo algoritmo, baseado na estratégia de região de confiança e técnicas de pontos interiores, para resolver o problema de minimizar uma função não linear sujeita a restrições não lineares de igualdade e onde as variáveis estão limitadas tanto inferior, como superiormente. Os algoritmos baseados na estratégia de região de confiança, usualmente, trabalham na interseção de uma região esférica com o conjunto viável. Em nosso trabalho, determinamos a forma e o tamanho da região de confiança para gerar um elipsóide adaptado ao hiper-cubo determinado pelos limites inferior e superior das variáveis. O algoritmo resultante gera uma seqüência de pontos dentro do hiper-cubo nos quais se obtém uma melhoria tanto no valor da função objetivo como na viabilidade. Esta melhoria é medida por uma função de mérito que leva em consideração ambas condições, a qual nos permite decidir se o ponto obtido na iteração presente é aceito e como atualizar a região de confiança. Na prática é necessário utilizar uma técnica de conjuntos ativos no tratamento das restrições que definem o hiper-cubo, assim como em algumas iterações é preciso fazer uma correção de segunda ordem para que o ponto resultante da mesma

fique mais perto da região viável e evitar a possível ocorrência do efeito Maratos, ocasionado pela utilização de uma função de mérito não diferenciável.

Consideremos o seguinte problema de programação não linear:

$$\begin{aligned} \min \quad & f(x) \\ \text{s. a.} \quad & c(x) = 0 \\ & l \leq x \leq u, \end{aligned} \tag{IV.1}$$

onde $f : \mathcal{R}^n \rightarrow \mathcal{R}$ é uma função não linear, $f \in C^2$, $c : \mathcal{R}^n \rightarrow \mathcal{R}^m$, $0 < m < n$, é um vetor de funções não lineares, $c \in C^2$, e os vetores l e u são conhecidos, sendo $l < u$, podendo ser algumas coordenadas de l iguais a $-\infty$ e algumas coordenadas de u iguais a $+\infty$.

Dado um ponto viável x^* para o problema (IV.1) consideremos o conjunto $N = \{ i = 1, \dots, n \mid x_i^* = l_i \text{ ou } x_i^* = u_i \}$, o qual representa o conjunto das restrições ativas de desigualdade. Em 1951 Kuhn e Tucker, em [31], demonstraram que se uma condição, chamada de qualificação de vínculos de primeira ordem, era satisfeita pelo ponto x^* então as seguintes condições são necessárias para x^* ser um mínimo local do problema (IV.1):

Existem $\lambda^* \in \mathcal{R}^m$, $\gamma^* \in \mathcal{R}^n$ tais que

$$c(x^*) = 0 \tag{IV.2.1}$$

$$l \leq x^* \leq u \tag{IV.2.2}$$

$$\nabla f(x^*) = \sum_{i=1}^m \lambda_i^* \nabla c_i(x^*) + \gamma^* \tag{IV.2.3}$$

$$\forall i = 1, \dots, n \quad \min\{x_i^* - l_i, u_i - x_i^*\} \gamma_i^* = 0 \tag{IV.2.4}$$

$$\text{se } x_i^* = l_i \text{ então } \gamma_i^* \geq 0 \tag{IV.2.5}$$

$$\text{se } x_i^* = u_i \text{ então } \gamma_i^* \leq 0 \tag{IV.2.6}$$

Estas condições são conhecidas como condições necessárias de otimalidade de Karush-Kuhn-Tucker de primeira ordem. No caso de restrições não lineares não é fácil determinar se a condição de qualificação de vínculos é satisfeita por um ponto dado. Não obstante, uma condição suficiente para que ela seja satisfeita no ponto

x^* é que os vetores $\nabla c_i(x^*)$, para $i = 1, \dots, m$, I_i para $i \in N$ sejam linearmente independentes, onde I_i é a i -ésima coluna da matriz identidade. A demonstração deste fato pode ser achada em [34]. Um ponto viável para o problema (IV.1) que satisfaz essa relação denomina-se ponto regular. Note-se que todo ponto regular pode ter no máximo $n - m$ componentes na fronteira do hipercubo determinado pelas restrições de desigualdade.

IV.2 A Região de Confiança.

Consideremos os seguintes conjuntos:

$$H_1 = \{x \in \mathcal{R}^n / l \leq x \leq u\} \quad e$$

$$H_1^0 = \{x \in \mathcal{R}^n / l < x < u\}.$$

Note-se que devido ao fato de ser $l < u$ temos que $H_1^0 \neq \emptyset$. Dados um ponto $x^k \in H_1^0$ e números $\Delta_k \in \mathcal{R}_+$ e $\rho \in (0, 1)$, nossa região de confiança está dada pelo elipsóide simples centrado em x^k , o qual é representado pela equação:

$$d^t D^{-2} d \leq \rho^2$$

onde D é uma matriz diagonal cujos elementos são

$$d_{ii} = \min\{\Delta_k, x_i^k - l_i, u_i - x_i^k\}, \quad \forall i = 1, \dots, n.$$

Esta escolha da matriz D nos permite alterar tanto o tamanho quanto o formato da região de confiança. Os algoritmos de região de confiança só alteram o tamanho da região de confiança e não o seu formato. No nosso caso ao mudar o tamanho da região de confiança também modificamos o seu formato. Se o valor do raio Δ_k for grande, nossa região de confiança será um elipsóide, o qual se adapta melhor ao hipercubo definido pelas restrições de desigualdade, permitindo um passo maior que se tomássemos uma bola. Entretanto, se Δ_k for suficientemente pequeno então a região de confiança será uma bola. Neste último caso temos que as restrições de desigualdade estão longe de ser ativas, e em conseqüência nosso problema equivale a um problema sem restrições e portanto o melhor formato para a região de confiança é o de bola, já que este não favorece nenhuma direção em particular.

IV.3 Geração de um Passo do Algoritmo de Região de Confiança.

O algoritmo a ser desenvolvido neste capítulo, para resolver o problema (IV.1), a exemplo do realizado no capítulo anterior, também está baseado na programação quadrática seqüencial, no método de região de confiança e no algoritmo afim-escala, os quais foram apresentados no capítulo II. No início da k -ésima iteração do algoritmo que vamos propor dispomos do ponto x^k , o qual foi gerado na iteração anterior. A partir dele são calculados $g^k = \nabla f(x^k)$, a matriz $A_k = A(x^k)$ que é a matriz jacobiana de c no ponto x^k , o vetor $\lambda(x^k)$ que representa o multiplicador de Lagrange estimado no ponto x^k , cujo cálculo será apresentado na seção IV.4 deste capítulo, a matriz $B_k = \nabla^2 f(x^k) + \sum_{i=1}^m \lambda_i(x^k) \nabla^2 c_i(x^k)$ e $\alpha_k \in (0, 1]$ que é um fator de relaxação da linearização das restrições de igualdade, cujo cálculo será explicitado nesta seção. A seguir é resolvido o seguinte subproblema quadrático:

$$\min q_k(d) = g^{kt} d + \frac{1}{2} d^t B_k d \quad (\text{IV.3.1})$$

$$\text{s. a. } A_k^t d = -\alpha_k c(x^k) \quad (\text{IV.3.2})$$

$$\| D^{-1} d \| \leq \rho. \quad (\text{IV.3.3})$$

O fator de relaxação α_k pode ser no máximo 1 e deve ser escolhido de maneira tal que a interseção dos conjuntos dados por (IV.3.2) e (IV.3.3) seja não vazia e contenha mais de um ponto.

Note-se que se d satisfaz (IV.3.3) então

$$\| \bar{d} \| \leq \rho \Delta_k. \quad (\text{IV.3.4})$$

Fazendo $d = D\bar{d}$, e tomando $\bar{g}^k = Dg^k$, $\bar{A}_k = DA_k$, $\bar{B}_k = DB_k D$ o problema (IV.3.1)-(IV.3.3) pode ser escrito como:

$$\min \bar{q}_k(\bar{d}) = \bar{g}^{kt} \bar{d} + \frac{1}{2} \bar{d}^t \bar{B}_k \bar{d} \quad (\text{IV.4.1})$$

$$\text{s. a. } \bar{A}_k^t \bar{d} = -\alpha_k c(x^k) \quad (\text{IV.4.2})$$

$$\| \bar{d} \| \leq \rho. \quad (\text{IV.4.3})$$

Seja \bar{Z}_k uma matriz $n \times (n-m)$ cujas colunas formam uma base ortonormal para o espaço nulo de \bar{A}_k^t , isto é, $\bar{A}_k^t \bar{Z}_k = 0$ e $\bar{Z}_k^t \bar{Z}_k = I$. As colunas da matriz $Z_k = D \bar{Z}_k$, formam uma base para o espaço nulo de A_k^t .

Qualquer solução da equação (IV.4.2) deve ter a forma

$$\bar{d} = \alpha_k \bar{v}^k + \bar{Z}_k \bar{u}, \quad (\text{IV.5})$$

onde \bar{v}^k é um vetor no espaço imagem de \bar{A}_k e $\bar{u} \in \mathcal{R}^{n-m}$. Em particular, tomaremos $\bar{v}^k = -\bar{A}_k [\bar{A}_k^t \bar{A}_k]^{-1} c(x^k)$, o qual corresponde a tomar a solução mínimos quadrados da equação $\bar{A}_k^t \bar{d} = -c(x^k)$.

Nossa restrição de região de confiança requer $\|\alpha_k \bar{v}^k\| \leq \rho$, o que implica que $\alpha_k \leq \frac{\rho}{\|\bar{v}^k\|}$. Adicionalmente, requeremos que quando $\alpha_k < 1$, para que o passo no espaço imagem não seja muito pequeno em relação à região de confiança, tomamos $\alpha_k \|\bar{v}^k\| \geq \theta \rho$, onde θ é uma constante, $\theta \in (0, 1]$. Portanto, lembrando que $\alpha_k \leq 1$, temos que:

$$\alpha_k \in [\min\{1, \theta \frac{\rho}{\|\bar{v}^k\|}\}, \min\{1, \frac{\rho}{\|\bar{v}^k\|}\}].$$

Voltando ao espaço original, temos:

$$d = D \bar{d} = \alpha_k D \bar{v}^k + D \bar{Z}_k \bar{u} = \alpha_k v^k + D \bar{Z}_k \bar{u}, \quad \text{onde}$$

$$v^k = -D^2 A_k [A_k^t D^2 A_k]^{-1} c(x^k).$$

Nossa restrição de região de confiança é $\|D^{-1} d\| \leq \rho$. Em conseqüência, o fator de relaxação α_k deve ser escolhido no intervalo

$$[\min\{1, \theta \frac{\rho}{\|D^{-1} v^k\|}\}, \min\{1, \frac{\rho}{\|D^{-1} v^k\|}\}].$$

Em particular, no algoritmo a ser apresentado na seção IV.5, vamos tomar

$$\alpha_k = \min\{1, \theta \frac{\rho}{\|D^{-1} v^k\|}\}.$$

Exemplo.

Consideremos o seguinte subproblema quadrático, onde a primeira restrição representa a linearização da restrição de igualdade e a segunda a restrição de região de confiança.

$$\begin{aligned} \min \quad & q_k(d) = g^{kt}d + \frac{1}{2}d^t B_k d \\ \text{s. a.} \quad & d_1 + d_2 = -2 \\ & \|D^{-1}d\| \leq 0.95, \end{aligned}$$

onde

$$D = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{pmatrix}.$$

Neste caso, $A^t = (1, 1)$, $c = 2$ e $\rho = 0.95$.

Temos que $v = -D^2 A[A^t D^2 A]^{-1} c$. Logo,

$$v = - \begin{pmatrix} \frac{2}{5} \\ \frac{8}{5} \end{pmatrix}.$$

Pela restrição de região de confiança, $\alpha \|D^{-1}v\| \leq 0.95$. Então,

$$\alpha \leq \frac{0.95}{\|D^{-1}v\|} = 0.95 \frac{\sqrt{5}}{4} < 1.$$

Tomando $\alpha = \frac{2.85\sqrt{5}}{16}$ (75 % do valor máximo possível para α) o problema relaxado é:

$$\begin{aligned} \min \quad & q_k(d) = g^{kt}d + \frac{1}{2}d^t B_k d \\ \text{s. a.} \quad & d_1 + d_2 = -\frac{2.85\sqrt{5}}{8} \\ & \|D^{-1}d\| \leq 0.95 \end{aligned}$$

Na figura IV.1 a região de confiança é representada pela elipse simples centrada na origem. Note-se que quando $\alpha = 1$, o qual equivale a não relaxar a linearização da restrição de igualdade, a interseção das restrições é vazia. Se tomamos $\alpha = \frac{0.95\sqrt{5}}{4}$ a interseção é um ponto, enquanto que se tomamos $\alpha = \frac{2.85\sqrt{5}}{16}$, o conjunto resultante da interseção das restrições é não vazio e contém mais de um ponto. Neste conjunto é onde vamos resolver o subproblema quadrático.

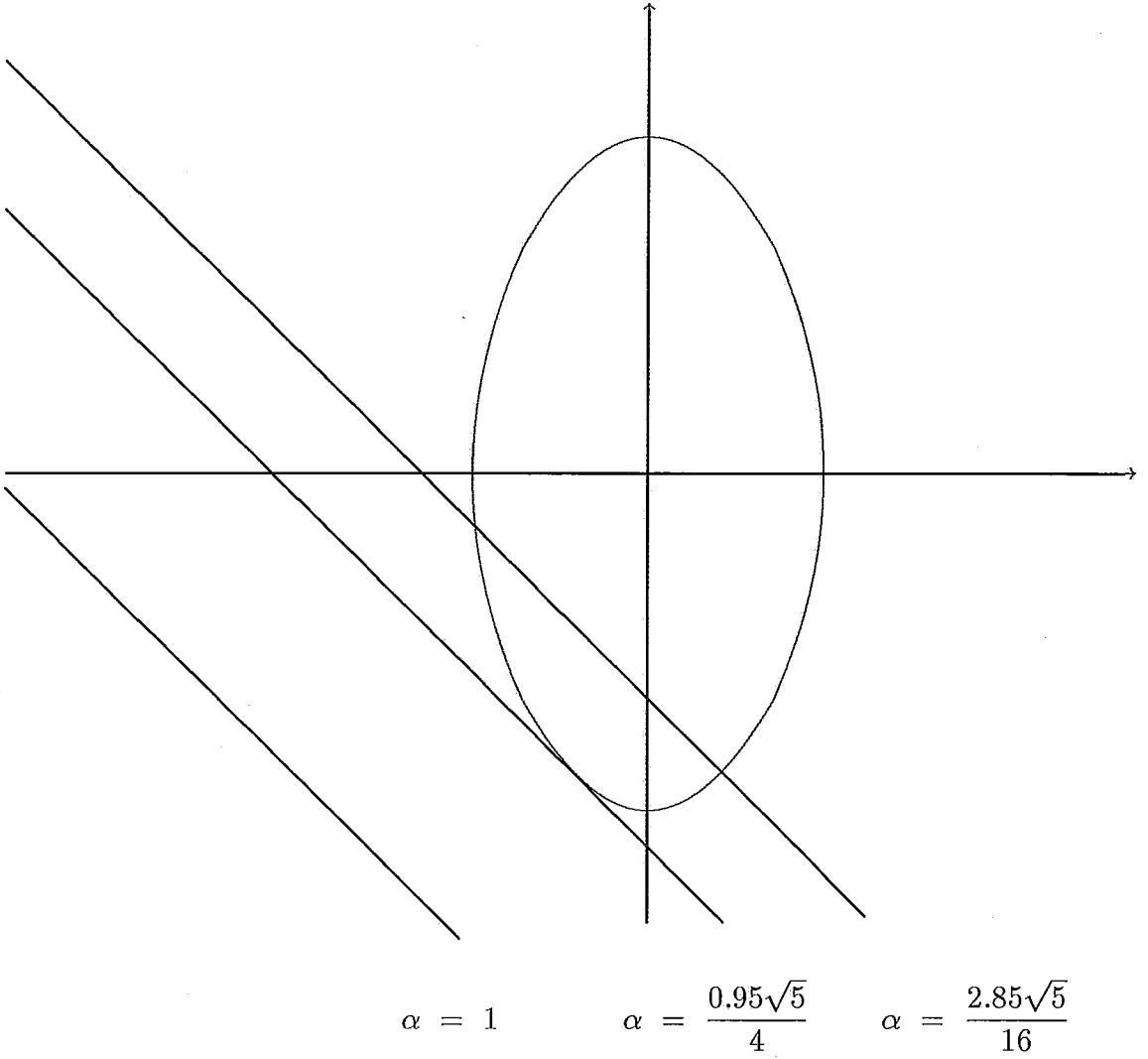


Figura IV.1

Consideremos novamente o problema (IV.4.1)-(IV.4.3) e seja \bar{d} uma solução a (IV.4.2) e (IV.4.3). Então, $\bar{d} = \alpha_k \bar{v}^k + \bar{Z}_k \bar{u}$ e temos portanto $\|\bar{d}\|^2 = \alpha_k^2 \|\bar{v}^k\|^2 + \|\bar{u}\|^2$. Em conseqüência, para que a restrição (IV.4.3) seja satisfeita deve cumprir-se:

$$\|\bar{u}\| \leq \sqrt{\rho^2 - \alpha_k^2 \|\bar{v}^k\|^2}.$$

Em termos da decomposição (IV.5) o problema (IV.4.1)-(IV.4.3) torna-se:

$$\begin{aligned} \min \quad & (\bar{g}^k + \alpha_k \bar{B}_k \bar{v}^k)^t \bar{Z}_k \bar{u} + \frac{1}{2} \bar{u}^t \bar{Z}_k^t \bar{B}_k \bar{Z}_k \bar{u} \\ \text{s. a.} \quad & \|\bar{u}\|^2 \leq \rho^2 - \alpha_k^2 \|\bar{v}^k\|^2. \end{aligned} \tag{IV.6}$$

Uma vez achado \bar{u}^k , solução de (IV.6), fazemos $\bar{d} = \alpha_k \bar{v}^k + \bar{Z}_k \bar{u}^k$ e $d^k = D\bar{d}$.

Desta forma, a resolução do problema (IV.4.1)-(IV.4.3) ficou reduzida a calcular $\bar{v}^k = -\bar{A}_k [\bar{A}_k^t \bar{A}_k]^{-1} c(x^k)$ e a resolver o problema (IV.6) para calcular \bar{u}^k ,

a componente no espaço nulo de \bar{A}_k^t de \bar{d} . Note-se que este problema se reduz a minimizar uma função quadrática sujeita a uma restrição de região de confiança, o qual é equivalente a aplicar o método de região de confiança para resolver um problema sem restrições.

Dado um passo d^k gerado pelo procedimento anteriormente descrito, devemos determinar se o ponto $x^k + d^k$ melhora a função objetivo e a viabilidade. Mediremos esta melhoria com a função de mérito seguinte:

$$\phi(x) = f(x) + \sum_{i=1}^m \mu_i |c_i(x)|,$$

onde os μ_i são pesos positivos. Estes pesos serão incrementados caso seja necessário, em qualquer iteração, mas eventualmente, eles permanecerão fixos.

A redução real na função, ao ir de x^k a $x^k + d^k$, a qual denotaremos com $ared_k(d^k)$, está dada por:

$$\begin{aligned} ared_k(d^k) &= \phi(x^k) - \phi(x^k + d^k) = \\ &= f(x^k) - f(x^k + d^k) + \sum_{i=1}^m \mu_i \{|c_i(x^k)| - |c_i(x^k + d^k)|\}. \end{aligned}$$

A predição da redução real, de acordo com nosso modelo, a qual denotaremos com $pred_k(d^k)$, está dada por:

$$pred_k(d^k) = -g^{k^t} d^k - \frac{1}{2} d^{k^t} B_k d^k + \sum_{i=1}^m \mu_i \{|c_i(x^k)| - |c_i(x^k) + a^i(x^k)^t d^k|\}.$$

Se a melhora em ϕ é uma proporção suficiente da predição de nosso modelo, isto é, se

$$\frac{ared_k(d^k)}{pred_k(d^k)} \geq \eta,$$

onde $\eta \in (0, 1)$ é uma constante fixa, então o passo será aceito e $x^{k+1} = x^k + d^k$. Caso contrário, reduzimos o raio Δ_k da região de confiança e calculamos um novo passo provisório d^k , o qual é novamente examinado, usando a função de mérito ϕ .

IV.4 Estimativas dos multiplicadores de K.K.T.

Olhando as condições de K.K.T. para o problema (IV.1) temos que se x^* é uma solução ótima, então existem $\lambda^* \in \mathcal{R}^m$ e $\gamma^* \in \mathcal{R}^n$ tais que

$$\nabla f(x^*) = A_* \lambda^* + \gamma^*,$$

onde $A_* = A(x^*)$.

Projetando esta equação em $N(A_*^t)$ temos: $P_{A_*^t} \nabla f(x^*) = P_{A_*^t} \gamma^*$.

No que segue, consideraremos um multiplicador estimado em um ponto $x^k \in H_1$, qualquer vetor $\gamma^k \in \mathcal{R}^n$ tal que:

$$P_{A_k^t} \gamma^k = P_{A_k^t} \nabla f(x^k).$$

Vejam como escolher um multiplicador estimado entre todos os vetores que satisfazem a equação anterior. Consideremos $\bar{\gamma}^k = P_{A_k^t D} D \nabla f(x^k)$. Mostraremos a seguir que $D^{-1} \bar{\gamma}^k$ é um multiplicador estimado. Em efeito,

$$D \nabla f(x^k) - \bar{\gamma}^k = D \nabla f(x^k) - P_{A_k^t D} D \nabla f(x^k) = (I - P_{A_k^t D}) D \nabla f(x^k) = D A_k w,$$

para algum $w \in \mathcal{R}^m$. Portanto, $D \nabla f(x^k) = D A_k w + \bar{\gamma}^k$. Multiplicando esta equação por D^{-1} , temos $\nabla f(x^k) = A_k w + D^{-1} \bar{\gamma}^k$. Logo, $\gamma^k = D^{-1} \bar{\gamma}^k$ é um multiplicador estimado. Temos então $\gamma^k = D^{-1} P_{A_k^t D} D \nabla f(x^k)$, e substituindo $P_{A_k^t D} = I - D A_k (A_k^t D^2 A_k)^{-1} A_k^t D$, temos:

$$\gamma^k = (I - A_k (A_k^t D^2 A_k)^{-1} A_k^t D^2) \nabla f(x^k).$$

A partir desta última equação, devido ao fato de serem os espaços $\text{imagem}(A^t)$ e $\text{nulo}(A)$ espaços complementares, obtemos os multiplicadores estimados para as restrições de igualdade, os quais vêm dados pela seguinte expressão:

$$\lambda^k = (A_k^t D^2 A_k)^{-1} A_k^t D^2 \nabla f(x^k).$$

Note que esta escolha é equivalente a resolver o seguinte sistema utilizando mínimos quadrados: $A_k \lambda^k = \nabla f(x^k) - \gamma^k$.

IV.5 Algoritmo.

Algoritmo IV.1: Dados $\eta \in (0, 1)$, $\omega \in (0, 1)$, $\theta \in (0, 1]$, $\tau \in (0, 1)$, $\mu \in \mathcal{R}_+^m$, $\pi > 0$ e $\rho \in (0, 1)$.

0. $k = 0$; seja $x^0 \in H_1^0$ o ponto inicial.

Enquanto não convergir fazer

1. Calcular $x = x^k$, $f = f(x)$, $g = g(x)$, $c = c(x)$, $A = A(x)$, $\bar{Z} = \bar{Z}(x)$.

2. Calcular

$$\Delta = \max_{1 \leq i \leq n} \{m_i\} + \omega,$$

onde:

$$m_i = \begin{cases} 1 & \text{se } l_i = -\infty \text{ e } u_i = +\infty \\ \min\{x_i - l_i, u_i - x_i\} & \text{caso contrário.} \end{cases}$$

3. Calcule D, matriz diagonal, onde

$$[d]_{ii} = \min\{\Delta, x_i - l_i, u_i - x_i\}, \forall i = 1, \dots, n.$$

4. Calcular

$$\lambda = (A^t D^2 A)^{-1} A^t D^2 g, \quad B = \nabla^2 f(x) + \sum_{i=1}^m \lambda_i(x) \nabla^2 c_i(x).$$

5. Calcule $\bar{v} = -DA[A^t D^2 A]^{-1}c$.

Se $\bar{v} = 0$ então $\alpha = 1$; ir a 7.

6. Calcular

$$\alpha = \min\left\{1, \theta \frac{\rho}{\|\bar{v}\|}\right\}.$$

7. Calcule o passo \bar{u} no espaço nulo, resolvendo o problema:

$$\min \quad (Dg + \alpha DBD\bar{v})^t \bar{Z}\bar{u} + \frac{1}{2} \bar{u}^t \bar{Z}^t DBD\bar{Z}\bar{u}$$

$$\text{s. a.} \quad \|\bar{u}\|^2 \leq \rho^2 - \alpha^2 \|\bar{v}\|^2.$$

Seja $v = D\bar{v}$ e $d = \alpha v + D\bar{Z}\bar{u}$.

8. Calcule $\tilde{\mu}^k = -(A^t D^2 A)^{-1} A^t D^2 (g + \frac{1}{2} \alpha Bv)$, fator de atualização dos pesos.

Para cada i , se $\mu_i < |\tilde{\mu}_i^k| + \pi$ então $\mu_i = |\tilde{\mu}_i^k| + 2\pi$.

9. Calcule

$$\text{pred} = -g^t d - \frac{1}{2} d^t B d + \sum_{i=1}^m \mu_i \{|c_i| - |c_i + a^{i^t} d|\}.$$

10. Calcule $f_+ = f(x + d)$ e $c^+ = c(x + d)$.

11. Calcule

$$\text{ared} = f - f_+ + \sum_{i=1}^m \mu_i \{|c_i| - |c_i^+|\}.$$

12. Se $\frac{ared}{pred} \geq \eta$ então

$$x^{k+1} = x + d; \quad k = k + 1.$$

caso contrário fazer

$$\Delta = \tau\Delta; \quad \text{ir a 3.}$$

Fim do enquanto

Os valores usuais dos parâmetros dados no algoritmo são $\eta = 0.25$, $\omega = 0.1$, $\theta = 0.75$, $\tau = 0.25$ e $\rho = 0.95$. Esses valores são arbitrários, mas a escolha deles não altera significativamente o desempenho do algoritmo.

IV.6 Convergência Global de Primeira Ordem.

Nesta seção vamos dar o resultado de convergência global de primeira ordem, obtido assumindo certas hipóteses, que denominaremos hipóteses padrões.

Hipóteses padrões.

Dado $x^0 \in H_1^0$ definimos os seguintes conjuntos:

$$\bar{H} = \{x \in H_1 / \phi(x) \leq \phi(x^0)\} \text{ e}$$

$$\bar{H}^0 = \{x \in \bar{H} / l < x < u\}.$$

Seja $\{x^k\}$ uma seqüência gerada pelo algoritmo, com um ponto inicial $x^0 \in H_1^0$ arbitrário, e suponhamos que $\{x^k\}$ está contida em algum subconjunto aberto S de \mathcal{R}^n . Vamos supor que \bar{H} é limitado, que todos os pontos viáveis de \bar{H} são regulares, que $f(x)$ e $c(x)$ são duas vezes continuamente diferenciáveis em S , que a matriz jacobiana $A(x)$ tem posto m para todo $x \in \bar{H}$, e que $A(x)$, $(A(x)^t A(x))^{-1}$, $g(x)$, $\nabla^2 f(x^k)$ e cada $\nabla^2 c_i(x)$, $\forall i = 1, \dots, m$ são limitados em S . Suporemos também que existe um $\pi > 0$ tal que $\mu_i \geq |\tilde{\mu}_i^k| + \pi$ para todo k e que para algum $\beta > 0$, $\|B_k\| \leq \beta$ para todo k .

Seja \hat{x} um ponto de acumulação de $\{x^k\}$ e $K_1 \subset \mathcal{N}$ tal que $x^k \xrightarrow{K_1} \hat{x}$. Por construção temos que a seqüência $\{\Delta_k\}$ é limitada e $\Delta_k > 0 \forall k \in \mathcal{N}$. Seja $\hat{\Delta} = \inf_{k \in K_1} \{\Delta_k\}$ e $K \subset K_1$ tal que $\Delta_k \xrightarrow{K} \hat{\Delta}$. Vamos estabelecer também os conjuntos $N = \{i = 1, \dots, n / \hat{x}_i = l_i \text{ ou } \hat{x}_i = u_i\}$ e $\bar{N} = \{i = 1, \dots, n\} - N$. No que se segue, vamos considerar \hat{x} , $\hat{\Delta}$, K , N e \bar{N} como descritos neste parágrafo.

Além das hipóteses padrões, para poder demonstrar que o ponto \hat{x} satisfaz as condições (IV.2.1) e (IV.2.4), vamos precisar da hipótese de que $\hat{\Delta} > 0$ e para demonstrar que satisfaz as condições (IV.2.5) e (IV.2.6), vamos exigir também que a seqüência gerada pelo algoritmo seja convergente. Estas duas hipóteses são facilmente verificáveis, embora muito fortes. No capítulo III, foi preciso usar a hipótese da convexidade da função objetivo para poder demonstrar que os sinais do gradiente nas restrições ativas são corretos. Sem esta hipótese, mesmo no caso da otimização só com restrições de caixa, não tem sido possível até o momento, demonstrar convergência global. Neste capítulo, não vamos exigir que a função objetivo seja convexa, mais por ser o problema estudado neste capítulo muito mais difícil, por causa da presença das restrições não lineares de igualdade, que o estudado no capítulo anterior, foi necessário impor essas duas hipóteses adicionais. O seguinte exemplo ilustra o que acontece quando a hipótese de ser $\hat{\Delta} > 0$ não é satisfeita. Consideremos o seguinte problema de programação não linear:

$$\begin{aligned} &\text{minimizar} && -x_1 \\ &\text{sujeito a} && c(x_1, x_2) = -x_1^2 + x_2 = 0 \\ &&& -\sqrt{2}/2 \leq x_1 \leq 1/2 \\ &&& 1/3 \leq x_2 \leq 1 \end{aligned}$$

A solução deste problema é atingida no ponto $(-\sqrt{3}/3, 1/3)$ e o valor da função objetivo no ótimo é $\sqrt{3}/3$. Se usamos o ponto $(0.49, 0.5)$ como ponto inicial, o algoritmo tende ao ponto $(1/2, 1/3)$, o qual é inviável para o problema dado, pois $c(1/2, 1/3) = 1/12$. Neste caso, o raio da região de confiança tende a zero, o que inviabiliza a convergência ao ponto $(-\sqrt{3}/3, 1/3)$. Note-se que no ponto $(1/2, 1/3)$ temos as duas restrições de caixa ativas, quando no máximo só poderíamos ter uma restrição de desigualdade ativa. Cabe ser ressaltado que, usando os pontos $(0.49, 0.5)$

e $(0, 0)$ como pontos iniciais, o programa MINOS 5.1 também convergiu ao ponto $(1/2, 1/3)$.

O uso da hipótese da seqüência $\{x^k\}$ ser convergente, é equivalente à utilização da hipótese de convexidade no capítulo III, na demonstração da correção dos sinais dos multiplicadores de Karush-Kuhn-Tucker. O uso desta hipótese, assim como o uso da hipótese de ser $\hat{\Delta} > 0$, ficara claramente explicitado, nos resultados em que sejam empregadas.

No primeiro lema que apresentamos, demonstramos que se tomamos como ponto inicial para nosso algoritmo $x^0 \in H_1^0$ então a restrição $l \leq x \leq u$ é satisfeita em cada iteração do algoritmo.

LEMA IV.1. Seja x^k um ponto gerado pelo algoritmo e d^k um passo gerado pela solução de (IV.3.1) - (IV.3.3) na iteração k. Se $x^k \in H_1^0$ então $x^{k+1} = x^k + d^k$ também pertence a H_1^0 .

Demonstração.

Seja x^k um ponto gerado pelo algoritmo, o qual pertence a H_1^0 e d^k um passo gerado pela solução de (IV.3.1) - (IV.3.3) na iteração k. Devemos ver que $l < x^{k+1} = x^k + d^k < u$. Suponhamos que $x_i^{k+1} \leq l_i$ para algum i . Então, pela definição de d_{ii} temos que

$$0 < d_{ii} \leq x_i^k - l_i = x_i^{k+1} - d_i^k - l_i \leq -d_i^k$$

e portanto

$$\frac{-d_i^k}{d_{ii}} \geq 1 > \rho$$

e em conseqüência

$$\| D^{-1}d \| > \rho$$

o qual contradiz o fato de d satisfazer (IV.3.3). Temos por conseguinte que $x_i^{k+1} > l_i \forall i = 1, \dots, n$. De maneira análoga demonstra-se que $\forall i = 1, \dots, n \ x_i^{k+1} < u_i$. •

COROLÁRIO IV.1. Suponhamos que as hipóteses padrões se cumprem. Então qualquer ponto de acumulação \hat{x} de $\{x^k\}$ satisfaz a condição (IV.2.2).

Demonstração.

Pelo lema IV.1, a seqüência $\{x^k\}$ está contida em H_1 o qual é um conjunto fechado e em consequência contém todos seus pontos de acumulação. •

Devido às hipóteses padrões e ao lema IV.1, $\forall k \in \mathcal{N}$ a matriz jacobiana $A(x^k)$ tem posto m e a matriz $D(\Delta_k, x^k)$ tem posto n para $\Delta_k > 0$. Portanto, a matriz $(A(x^k)^t D(\Delta_k, x^k)^2 A(x^k))^{-1}$ existe e é limitada para todo $\forall k \in \mathcal{N}$. Mais adiante, no teorema IV.1, demonstraremos que $\forall k \in \mathcal{N}$, $\Delta_k > 0$.

Vejamos agora que se $\hat{\Delta} > 0$ então a matriz $A(\hat{x})^t \hat{D}(\hat{\Delta}, \hat{x})^2 \hat{A}(\hat{x})$ é não singular.

LEMA IV.2. Suponhamos que as hipóteses padrões se cumprem e seja \hat{x} um ponto de acumulação de $\{x^k\}$ tal que vetores $\nabla c_i(\hat{x})$, para $i = 1, \dots, m$, I_i para $i \in N$ sejam linearmente independentes, onde I_i é a i -ésima coluna da matriz identidade.

i. Se $\hat{\Delta} > 0$ então $\hat{A}^t \hat{D}$ tem posto m , onde $\hat{A} = A(\hat{x})$ e $\hat{D} = D(\hat{\Delta}, \hat{x})$.

ii. Se $\hat{\Delta} = 0$ então $\hat{A}^t \tilde{D}$ tem posto m , onde $\hat{A} = A(\hat{x})$ e \tilde{D} é uma matriz diagonal cujos elementos são $\tilde{D}_{ii} = \min \left\{ 1, \lim_{\substack{k \rightarrow +\infty \\ k \in K}} \frac{x_i^k - l_i}{\Delta_k}, \lim_{\substack{k \rightarrow +\infty \\ k \in K}} \frac{u_i - x_i^k}{\Delta_k} \right\}$.

Demonstração.

Devido à hipótese da independência linear dos gradientes das restrições ativas no ponto \hat{x} temos que $p = \text{card}(N) \leq n - m$ e a matriz (\hat{A}, I_N) tem posto $m + p$, onde I_N é formada pelas colunas da matriz identidade I tais que $i \in N$.

Seja (\hat{A}_N, I) a submatriz de (\hat{A}, I_N) formada pelas linhas de (\hat{A}, I_N) tais que $i \in N$. Então (\hat{A}_N, I) tem posto p e como o posto linha é igual ao posto coluna, temos que $(\hat{A}_{\bar{N}}, 0)$ tem posto m . Logo $\hat{A}_{\bar{N}}$ tem posto m .

Vamos demonstrar primeiramente [i.]. Rearranjando as matrizes \hat{A} e \hat{D} de modo que

$$\bar{A} = \begin{pmatrix} \hat{A}_N \\ \hat{A}_{\bar{N}} \end{pmatrix} \quad \text{e} \quad \bar{D} = \begin{pmatrix} \hat{D}_N & 0 \\ 0 & \hat{D}_{\bar{N}} \end{pmatrix},$$

onde $\hat{D}_N = 0$ e $\hat{D}_{\bar{N}} > 0$, já que $\hat{\Delta} > 0$. Então temos que

$$\bar{A}^t \bar{D} = (\hat{A}_N \hat{D}_N, \hat{A}_{\bar{N}} \hat{D}_{\bar{N}}) = (0, \hat{A}_{\bar{N}} \hat{D}_{\bar{N}}).$$

Portanto, o posto de $\bar{A}^t \bar{D}$ é igual ao posto de $\hat{A}_{\bar{N}}$. Agora, como o posto de $\hat{A}^t \hat{D}$ é igual ao posto de $\bar{A}^t \bar{D}$ então temos que $\hat{A}^t \hat{D}$ tem posto m .

Vamos agora demonstrar [ii.]. Vejamos como é a matriz \tilde{D} . $\forall k \in K$ e $\forall i \in N$ temos que $0 < \min\left\{1, \frac{x_i^k - l_i}{\Delta_k}, \frac{u_i - x_i^k}{\Delta_k}\right\} \leq 1$. Logo, $\forall i \in N$

$$0 \leq \tilde{D}_{ii} = \min\left\{1, \lim_{\substack{k \rightarrow +\infty \\ k \in K}} \frac{x_i^k - l_i}{\Delta_k}, \lim_{\substack{k \rightarrow +\infty \\ k \in K}} \frac{u_i - x_i^k}{\Delta_k}\right\} \leq 1.$$

Por outro lado, $\forall i \in \bar{N}$ $\tilde{D}_{ii} = 1$ já que k suficientemente grande e $i \in \bar{N}$ temos que $\min\left\{1, \frac{x_i^k - l_i}{\Delta_k}, \frac{u_i - x_i^k}{\Delta_k}\right\} = 1$.

Rearranjando as matrizes \hat{A} e \tilde{D} de modo que

$$\bar{A} = \begin{pmatrix} \hat{A}_N \\ \hat{A}_{\bar{N}} \end{pmatrix} \quad \text{e} \quad \bar{D} = \begin{pmatrix} \tilde{D}_N & 0 \\ 0 & \tilde{D}_{\bar{N}} \end{pmatrix},$$

onde $0_N \leq \tilde{D}_N \leq I_N$ e $\tilde{D}_{\bar{N}} = I_{\bar{N}}$. Então temos que

$$\bar{A}^t \bar{D} = (\hat{A}_N \tilde{D}_N, \hat{A}_{\bar{N}} \tilde{D}_{\bar{N}}) = (\hat{A}_N \tilde{D}_N, \hat{A}_{\bar{N}}).$$

Portanto, o posto de $\bar{A}^t \bar{D}$ é igual ao posto de $\hat{A}_{\bar{N}}$. Agora, como o posto de $\hat{A}^t \hat{D}$ é igual ao posto de $\bar{A}^t \bar{D}$ então temos que $\hat{A}^t \hat{D}$ tem posto m . •

COROLÁRIO IV.2. Suponhamos que as hipóteses padrões se cumprem e seja \hat{x} um ponto de acumulação de $\{x^k\}$ tal que vetores $\nabla c_i(\hat{x})$, para $i = 1, \dots, m$, I_i para $i \in N$ sejam linearmente independentes, onde I_i é a i -ésima coluna da matriz identidade.

i. Se $\hat{\Delta} > 0$ então a matriz $\hat{A}^t \hat{D}^2 \hat{A}$ é não singular, sendo $\hat{A} = A(\hat{x})$ e $\hat{D} = D(\hat{\Delta}, \hat{x})$.

ii. Se $\hat{\Delta} = 0$ então a matriz $\hat{A}^t \tilde{D}^2 \hat{A}$ é não singular, sendo $\hat{A} = A(\hat{x})$ e \tilde{D} uma matriz diagonal cujos elementos são $\tilde{D}_{ii} = \min\left\{1, \lim_{\substack{k \rightarrow +\infty \\ k \in K}} \frac{x_i^k - l_i}{\Delta_k}, \lim_{\substack{k \rightarrow +\infty \\ k \in K}} \frac{u_i - x_i^k}{\Delta_k}\right\}$.

Demonstração.

A matriz $\hat{A}^t \hat{D}^2 \hat{A} = (\hat{A}^t \hat{D})(\hat{A}^t \hat{D})^t$ é não singular se e somente se $\hat{A}^t \hat{D}$ tem posto m . Mas, pelo lema anterior, esta última condição é satisfeita, já que por hipótese temos que os gradientes das restrições ativas no ponto \hat{x} são linearmente independentes e $\hat{\Delta} > 0$. De maneira análoga demonstra-se [ii]. •

Vamos demonstrar agora que, como conseqüência das hipóteses padrões assim como dos resultados anteriores, os vetores $\tilde{\mu}^k$ são limitados em S e a norma dos vetores $v(x)$ e $c(x)$ satisfazem uma útil desigualdade.

LEMA IV.3. Suponhamos que as hipóteses padrões se cumprem. Então existe uma constante $\gamma_1 > 0$ tal que

$$\|v(x)\| \leq \gamma_1 \|c(x)\| \leq \gamma_1 \|c(x)\|_1, \quad (\text{IV.7})$$

e os vetores $\tilde{\mu}^k$ são limitados em S .

Demonstração.

Pelas hipóteses padrões temos que as matrizes $A(x)$ e $(A(x)^t A(x))^{-1}$ são limitadas em S e por construção a matriz D também é limitada em S . Seja \hat{x} um ponto de acumulação de $\{x^k\}$ e $K \subset \mathcal{N}$ tal que $x^k \xrightarrow{K} \hat{x}$ e $\Delta^k \xrightarrow{K} \hat{\Delta}$.

Suponhamos primeiro que $\hat{\Delta} > 0$. Por [i.] do corolário IV.2 temos que a matriz $\hat{A}^t \hat{D}^2 \hat{A}$ é não singular. Portanto, existe uma constante γ_1 tal que $0 < \|D^2 A_k [A_k^t D^2 A_k]^{-1}\| < \gamma_1, \forall k \in K$. Logo,

$$\begin{aligned} \|v^k\| &= \|D^2 A_k [A_k^t D^2 A_k]^{-1} c(x^k)\| \leq \|D^2 A_k [A_k^t D^2 A_k]^{-1}\| \|c(x^k)\| \leq \\ &\gamma_1 \|c(x^k)\| \leq \gamma_1 \|c(x)\|_1. \end{aligned}$$

Consideremos agora o caso quando $\hat{\Delta} = 0$. Por [ii.] do corolário IV.2 temos que a matriz $\hat{A}^t \tilde{D}^2 \hat{A}$ é não singular. Dado que $\hat{D}^2 \hat{A} [\hat{A}^t \hat{D}^2 \hat{A}]^{-1} = \tilde{D}^2 \hat{A} [\hat{A}^t \tilde{D}^2 \hat{A}]^{-1}$ temos que existe uma constante γ_1 tal que $0 < \|D^2 A_k [A_k^t D^2 A_k]^{-1}\| < \gamma_1, \forall k \in K$. Novamente tem-se

$$\|v^k\| = \|D^2 A_k [A_k^t D^2 A_k]^{-1} c(x^k)\| \leq \|D^2 A_k [A_k^t D^2 A_k]^{-1}\| \|c(x^k)\| \leq$$

$$\gamma_1 \|c(x^k)\| \leq \gamma_1 \|c(x)\|_1.$$

Demonstraremos agora a segunda parte do lema. Segundo o visto na demonstração da primeira parte do lema, independentemente do fato de $\hat{\Delta}$ tender a zero ou não, temos que a matriz $[A_k^t D^2 A_k]^{-1} A_k^t D^2$ é limitada. Portanto, devido às hipóteses padrões para ver que os vetores $\tilde{\mu}^k = -[A_k^t D^2 A_k]^{-1} A_k^t D^2 (g^k + \frac{1}{2} \alpha_k B_k v^k)$ são limitados em S , basta provar que os vetores $\alpha_k v^k$ são limitados já que

$$\|\tilde{\mu}^k\| \leq \|[A_k^t D^2 A_k]^{-1} A_k^t D^2\| (\|g^k\| + \frac{1}{2} \|B_k\| \|\alpha_k v^k\|),$$

∇f é limitado em \bar{H}^0 e $\|B\| \leq \beta$. Em efeito, $\|\bar{d}^k\|^2 = \alpha_k^2 \|\bar{v}^k\|^2 + \|\bar{u}^k\|^2 \geq \alpha_k^2 \|\bar{v}^k\|^2$. Logo, $\alpha_k \|\bar{v}^k\| \leq \|\bar{d}^k\| \leq \rho$. Temos portanto, $\|\alpha_k v^k\| = \alpha_k \|D \bar{v}^k\| \leq \rho \|D\| \leq \rho \Delta_k$. Como por construção Δ_k é limitado $\forall k \in \mathcal{N}$ então temos o resultado desejado. •

Como consequência do lema anterior temos que, para k suficientemente grande, os vetores de pesos $\tilde{\mu}^k$ permanecerão constantes, já que eles são limitados e pelo algoritmo, quando eles são atualizados é acrescentando uma quantidade fixa $2\pi > 0$.

O seguinte lema mostra que a redução prevista da função de mérito fornece uma aproximação da redução real da mesma função, cuja precisão é proporcional ao quadrado do comprimento do passo.

LEMA IV.4. Suponhamos que as hipóteses padrões se cumprem. Então existe uma constante $\kappa > 0$ tal que para qualquer $x \in \bar{H}^0$, $B \in \mathcal{R}^{n \times n}$ com $\|B\| \leq \beta$ e qualquer $d \in \mathcal{R}^n$ tal que o segmento de reta que une x com $x + d$ esta contido em \bar{H}^0 ,

$$|ared(d) - pred(d)| \leq \kappa \|d\|^2,$$

onde

$$pred(d) = -g(x)^t d - \frac{1}{2} d^t B d + \sum_{i=1}^m \mu_i \{|c_i(x)| - |c_i(x) + a^i(x)^t d|\}$$

e

$$ared(d) = f(x) - f(x + d) + \sum_{i=1}^m \mu_i \{|c_i(x)| - |c_i(x + d)|\}.$$

Demonstração.

Pelo teorema do valor médio existem pontos ψ e $\bar{\psi}$ no segmento de reta que une x com $x + d$, tais que

$$f(x + d) - f(x) = g(\psi)^t d \quad \text{e} \quad g(x) - g(\psi) = \nabla^2 f(\bar{\psi})(x - \psi).$$

Pelo mesmo teorema, $\forall i = 1, \dots, m$ existem pontos ψ^i e $\bar{\psi}^i$ no segmento de reta que une x com $x + d$, tais que

$$c_i(x + d) - c_i(x) = a^i(\psi^i)^t d \quad \text{e} \quad a^i(x) - a^i(\psi^i) = \nabla^2 c_i(\bar{\psi}^i)(x - \psi^i).$$

Temos portanto que

$$\begin{aligned} |ared(d) - pred(d)| &= \left| f(x) - f(x + d) + \sum_{i=1}^m \mu_i \{ |c_i(x)| - |c_i(x + d)| \} \right. \\ &\quad \left. + g(x)^t d + \frac{1}{2} d^t B d - \sum_{i=1}^m \mu_i \{ |c_i(x)| - |c_i(x) + a^i(x)^t d| \} \right| \\ &\leq \left| f(x) - f(x + d) + g(x)^t d + \frac{1}{2} d^t B d \right| + \left| \sum_{i=1}^m \mu_i \{ -|c_i(x + d)| + |c_i(x) + a^i(x)^t d| \} \right| \\ &= \left| -g(\psi)^t d + g(x)^t d + \frac{1}{2} d^t B d \right| + \left| \sum_{i=1}^m \mu_i \{ -|c_i(x) + a^i(\psi^i)^t d| + |c_i(x) + a^i(x)^t d| \} \right| \\ &\leq \|d\| \|g(x) - g(\psi)\| + \frac{1}{2} \|B\| \|d\|^2 + \sum_{i=1}^m \mu_i \|d\| \|a^i(x) - a^i(\psi^i)\| \\ &\leq \|d\| \|\nabla^2 f(\bar{\psi})\| \|x - \psi\| + \frac{1}{2} \|B\| \|d\|^2 + \sum_{i=1}^m \mu_i \|d\| \|\nabla^2 c_i(\bar{\psi}^i)\| \|(x - \psi^i)\| \\ &\leq \left(\|\nabla^2 f(\bar{\psi})\| + \frac{1}{2} \|B\| + \sum_{i=1}^m \mu_i \|\nabla^2 c_i(\bar{\psi}^i)\| \right) \|d\|^2 \end{aligned}$$

Como $\nabla^2 f$ e $\nabla^2 c^i$ para $i = 1, \dots, m$ são limitadas em \bar{H}^0 e $\|B\| \leq \beta$ então obtemos o resultado desejado. •

Consideremos novamente o problema (IV.6).

$$\text{Seja } \tilde{g}^k = \bar{g}^k + \alpha_k \bar{B}_k \bar{v}^k \text{ e } \tilde{\Delta}_k = \sqrt{\rho^2 - \alpha_k^2 \|\bar{v}^k\|^2}.$$

A predição da variação do lagrangiano, devida ao passo $\bar{Z}_k \bar{u}^k$, a qual denotaremos com $hpred_k(\bar{u}^k)$, está dada por:

$$hpred_k(\bar{u}^k) = -\tilde{q}_k(\bar{u}^k) = -\tilde{g}^{kt} \bar{Z}_k \bar{u}^k - \frac{1}{2} \bar{u}^{kt} \bar{Z}_k^t \bar{B}_k \bar{Z}_k \bar{u}^k.$$

A direção de Cauchy, ou direção de máximo declive desde o centro da região de confiança de raio $\tilde{\Delta}_k$, a qual denotaremos por \bar{u}^c , é $\bar{u}^c = -t_c \bar{Z}_k^t \tilde{g}^k$, onde

$$t_c = \begin{cases} \frac{\|\bar{Z}_k^t \tilde{g}^k\|^2}{\tilde{g}^{kt} \bar{Z}_k \bar{Z}_k^t \bar{B}_k \bar{Z}_k \bar{Z}_k^t \tilde{g}^k} & \text{se } \frac{\|\bar{Z}_k^t \tilde{g}^k\|^3}{\tilde{g}^{kt} \bar{Z}_k \bar{Z}_k^t \bar{B}_k \bar{Z}_k \bar{Z}_k^t \tilde{g}^k} \leq \tilde{\Delta}_k \\ \frac{\tilde{\Delta}_k}{\|\bar{Z}_k^t \tilde{g}^k\|} & \text{caso contrário.} \end{cases}$$

Dizemos que \bar{u}^k , a solução do problema (IV.6), satisfaz a condição do decréscimo da fração de Cauchy, se existe uma constante $\sigma > 0$ tal que

$$hpred_k(\bar{u}^k) \geq -\sigma \tilde{q}_k(\bar{u}^c), \quad \forall k = 1, \dots \quad (\text{IV.8})$$

Para poder assegurar a convergência global do nosso algoritmo é necessário que a solução do problema (IV.6), satisfaça a condição do decréscimo da fração de Cauchy. No seguinte lema demonstramos que \bar{u}^k satisfaz a condição do decréscimo da fração de Cauchy assim como também uma útil desigualdade, a qual será de grande importância para os resultados a serem obtidos mais adiante, e que é satisfeita sempre que (IV.8) se cumpre.

LEMA IV.5. Suponhamos que as hipóteses padrões se cumprem e seja x^k um ponto gerado pelo algoritmo, o qual não satisfaz as condições necessárias de primeira ordem para uma solução de (IV.1) e \bar{u}^k um passo gerado pela solução de (IV.6) na iteração k . Então \bar{u}^k satisfaz a condição do decréscimo da fração de Cauchy e

$$hpred_k(\bar{u}^k) \geq \frac{1}{2} \|\bar{Z}_k^t \tilde{g}^k\| \min \left\{ \tilde{\Delta}_k, \frac{\|\bar{Z}_k^t \tilde{g}^k\|}{\|\bar{Z}_k^t \bar{B}_k \bar{Z}_k\|} \right\}.$$

Demonstração.

Consideremos um ponto qualquer x^k (fixo) gerado pelo algoritmo, o qual não satisfaz as condições necessárias de primeira ordem para uma solução de (IV.1).

Suponhamos primeiro que $\bar{Z}_k^t \tilde{g}^k = 0$. Então o problema (IV.6) se reduz a:

$$\begin{aligned} \min \quad & \frac{1}{2} \bar{u}^t \bar{Z}_k^t \bar{B}_k \bar{Z}_k \bar{u} \\ \text{s. a.} \quad & \|\bar{u}\| \leq \tilde{\Delta}_k. \end{aligned}$$

Se $\bar{Z}_k^t \bar{B}_k \bar{Z}_k$ é semi-definida positiva então $hpred_k(\bar{u}^k) = 0$.

Suponhamos então que $\bar{Z}_k^t \bar{B}_k \bar{Z}_k$ é indefinida. Seja m_k o autovalor mais negativo de $\bar{Z}_k^t \bar{B}_k \bar{Z}_k$ e r^k um autovetor associado a esse autovalor. Então $\bar{u}^k = \tilde{\Delta}_k \frac{r^k}{\|r^k\|}$ e $hpred_k(\bar{u}^k) = -m_k \tilde{\Delta}_k^2 > 0$.

Em ambos os casos temos que se satisfaz a relação

$$hpred_k(\bar{u}^k) \geq \frac{1}{2} \|\bar{Z}_k^t \tilde{g}^k\| \min \left\{ \tilde{\Delta}_k, \frac{\|\bar{Z}_k^t \tilde{g}^k\|}{\|\bar{Z}_k^t \bar{B}_k \bar{Z}_k\|} \right\} = 0.$$

Vamos considerar agora o caso $\bar{Z}_k^t \tilde{g}^k \neq 0$ e portanto $\|\bar{Z}_k^t \tilde{g}^k\| > 0$.

Consideremos a função $\phi : \mathcal{R} \rightarrow \mathcal{R}$ definida por:

$$\begin{aligned} \phi(\delta) = \tilde{q}_k \left(-\delta \frac{\bar{Z}_k^t \tilde{g}^k}{\|\bar{Z}_k^t \tilde{g}^k\|} \right) &= -\delta \tilde{g}^{kt} \bar{Z}_k \frac{\bar{Z}_k^t \tilde{g}^k}{\|\bar{Z}_k^t \tilde{g}^k\|} + \frac{1}{2} \frac{\delta^2}{\|\bar{Z}_k^t \tilde{g}^k\|^2} \tilde{g}^{kt} \bar{Z}_k \bar{Z}_k^t \bar{B}_k \bar{Z}_k \bar{Z}_k^t \tilde{g}^k = \\ &= -\delta \|\bar{Z}_k^t \tilde{g}^k\| + \frac{1}{2} \frac{\delta^2}{\|\bar{Z}_k^t \tilde{g}^k\|^2} \tilde{g}^{kt} \bar{Z}_k \bar{Z}_k^t \bar{B}_k \bar{Z}_k \bar{Z}_k^t \tilde{g}^k. \end{aligned}$$

Se $\tilde{g}^{kt} \bar{Z}_k \bar{Z}_k^t \bar{B}_k \bar{Z}_k \bar{Z}_k^t \tilde{g}^k \leq 0$ então $\phi(\delta) \leq -\delta \|\bar{Z}_k^t \tilde{g}^k\|$. Em particular, para $\delta = \tilde{\Delta}_k$ temos $\phi(\tilde{\Delta}_k) \leq -\tilde{\Delta}_k \|\bar{Z}_k^t \tilde{g}^k\|$. Como \bar{u}^k é solução de (IV.6) e $-\tilde{\Delta}_k \frac{\bar{Z}_k^t \tilde{g}^k}{\|\bar{Z}_k^t \tilde{g}^k\|}$ é viável para (IV.6) então

$$-hpred_k(\bar{u}^k) = \tilde{q}_k(\bar{u}^k) \leq \tilde{q}_k \left(-\tilde{\Delta}_k \frac{\bar{Z}_k^t \tilde{g}^k}{\|\bar{Z}_k^t \tilde{g}^k\|} \right) = \phi(\tilde{\Delta}_k) \leq -\tilde{\Delta}_k \|\bar{Z}_k^t \tilde{g}^k\|. \quad (\text{IV.9})$$

Suponhamos agora que $\tilde{g}^{kt} \bar{Z}_k \bar{Z}_k^t \bar{B}_k \bar{Z}_k \bar{Z}_k^t \tilde{g}^k > 0$. Então

$$\begin{aligned} \phi(\delta) &= -\delta \|\bar{Z}_k^t \tilde{g}^k\| + \frac{1}{2} \frac{\delta^2}{\|\bar{Z}_k^t \tilde{g}^k\|^2} \tilde{g}^{kt} \bar{Z}_k \bar{Z}_k^t \bar{B}_k \bar{Z}_k \bar{Z}_k^t \tilde{g}^k = \\ &= \frac{1}{2} \left(\frac{\|\bar{Z}_k^t \tilde{g}^k\|^4}{\tilde{g}^{kt} \bar{Z}_k \bar{Z}_k^t \bar{B}_k \bar{Z}_k \bar{Z}_k^t \tilde{g}^k} - 2\delta \|\bar{Z}_k^t \tilde{g}^k\| + \frac{\delta^2}{\|\bar{Z}_k^t \tilde{g}^k\|^2} \tilde{g}^{kt} \bar{Z}_k \bar{Z}_k^t \bar{B}_k \bar{Z}_k \bar{Z}_k^t \tilde{g}^k \right) - \end{aligned}$$

$$\frac{1}{2} \left(\frac{\|\bar{Z}_k^t \tilde{g}^k\|^4}{\tilde{g}^{k^t} \bar{Z}_k \bar{Z}_k^t \bar{B}_k \bar{Z}_k \bar{Z}_k^t \tilde{g}^k} \right) =$$

$$\frac{1}{2} \left(-\frac{\|\bar{Z}_k^t \tilde{g}^k\|^2}{\sqrt{\tilde{g}^{k^t} \bar{Z}_k \bar{Z}_k^t \bar{B}_k \bar{Z}_k \bar{Z}_k^t \tilde{g}^k}} + \frac{\delta}{\|\bar{Z}_k^t \tilde{g}^k\|} \sqrt{\tilde{g}^{k^t} \bar{Z}_k \bar{Z}_k^t \bar{B}_k \bar{Z}_k \bar{Z}_k^t \tilde{g}^k} \right)^2 - \frac{1}{2} \left(\frac{\|\bar{Z}_k^t \tilde{g}^k\|^4}{\tilde{g}^{k^t} \bar{Z}_k \bar{Z}_k^t \bar{B}_k \bar{Z}_k \bar{Z}_k^t \tilde{g}^k} \right).$$

Seja δ^* o mínimo de ϕ no intervalo $[0, \tilde{\Delta}_k]$. Então

$$-hpred_k(\bar{u}^k) = \tilde{q}_k(\bar{u}^k) \leq \tilde{q}_k(-\delta^* \frac{\bar{Z}_k^t \tilde{g}^k}{\|\bar{Z}_k^t \tilde{g}^k\|}) = \phi(\delta^*). \quad (IV.10)$$

Se $\delta^* < \tilde{\Delta}_k$ então

$$\delta^* = \frac{\|\bar{Z}_k^t \tilde{g}^k\|^3}{\tilde{g}^{k^t} \bar{Z}_k \bar{Z}_k^t \bar{B}_k \bar{Z}_k \bar{Z}_k^t \tilde{g}^k}$$

e

$$\phi(\delta^*) = -\frac{1}{2} \frac{\|\bar{Z}_k^t \tilde{g}^k\|^4}{\tilde{g}^{k^t} \bar{Z}_k \bar{Z}_k^t \bar{B}_k \bar{Z}_k \bar{Z}_k^t \tilde{g}^k} \leq -\frac{1}{2} \frac{\|\bar{Z}_k^t \tilde{g}^k\|^2}{\|\bar{Z}_k^t \bar{B}_k \bar{Z}_k\|}. \quad (IV.11)$$

Se $\delta^* = \tilde{\Delta}_k$ então

$$\tilde{\Delta}_k \leq \frac{\|\bar{Z}_k^t \tilde{g}^k\|^3}{\tilde{g}^{k^t} \bar{Z}_k \bar{Z}_k^t \bar{B}_k \bar{Z}_k \bar{Z}_k^t \tilde{g}^k}$$

e

$$\phi(\delta^*) = \phi(\tilde{\Delta}_k) = -\tilde{\Delta}_k \|\bar{Z}_k^t \tilde{g}^k\| + \frac{1}{2} \frac{\tilde{\Delta}_k^2}{\|\bar{Z}_k^t \tilde{g}^k\|^2} \tilde{g}^{k^t} \bar{Z}_k \bar{Z}_k^t \bar{B}_k \bar{Z}_k \bar{Z}_k^t \tilde{g}^k \leq$$

$$-\tilde{\Delta}_k \|\bar{Z}_k^t \tilde{g}^k\| + \frac{1}{2} \tilde{\Delta}_k \|\bar{Z}_k^t \tilde{g}^k\| = -\frac{1}{2} \tilde{\Delta}_k \|\bar{Z}_k^t \tilde{g}^k\|. \quad (IV.12)$$

Portanto, por (IV.9), (IV.10), (IV.11) e (IV.12) temos que

$$hpred_k(\bar{u}^k) \geq \tilde{q}_k(\bar{u}^c) \geq \min \left\{ \tilde{\Delta}_k \|\bar{Z}_k^t \tilde{g}^k\|, \frac{1}{2} \frac{\|\bar{Z}_k^t \tilde{g}^k\|^2}{\|\bar{Z}_k^t \bar{B}_k \bar{Z}_k\|}, \frac{1}{2} \tilde{\Delta}_k \|\bar{Z}_k^t \tilde{g}^k\| \right\} =$$

$$\frac{1}{2} \|\bar{Z}_k^t \tilde{g}^k\| \min \left\{ \tilde{\Delta}_k, \frac{\|\bar{Z}_k^t \tilde{g}^k\|}{\|\bar{Z}_k^t \bar{B}_k \bar{Z}_k\|} \right\}.$$

É importante ressaltar que para demonstrar a convergência global de um algoritmo de região de confiança é necessário que o passo \bar{u}^k satisfaça a condição do decréscimo da fração de Cauchy ou a condição do lema IV.5. Mais detalhes acerca do papel dessas condições na teoria de convergência global dos algoritmos de região de confiança podem ser achados em [3], [11], [35], [40] e [48].

LEMA IV.6. Suponhamos que as hipóteses padrões se cumprem e seja x^k um ponto gerado pelo algoritmo, o qual não satisfaz as condições necessárias de primeira ordem para uma solução de (IV.1) e \bar{u}^k um passo gerado pela solução de (IV.6) na iteração k . Então

- i. $hpred_k(\bar{u}^k) = \bar{q}_k(\alpha_k \bar{v}^k) - q_k(d^k)$.
- ii. $\bar{q}_k(\alpha_k \bar{v}^k) = \frac{1}{2} \alpha_k \bar{v}^{k^t} (\bar{g}^k - \tilde{g}^k) = \alpha_k c(x^k)^t \tilde{\mu}^k$.

Demonstração.

Prova de i. Por definição de d^k temos que:

$$\begin{aligned}
 q_k(d^k) &= q_k(\alpha_k D \bar{v}^k + D \bar{Z}_k \bar{u}^k) = \\
 &g^{k^t} (\alpha_k D \bar{v}^k + D \bar{Z}_k \bar{u}^k) + \frac{1}{2} (\alpha_k D \bar{v}^k + D \bar{Z}_k \bar{u}^k)^t B_k (\alpha_k D \bar{v}^k + D \bar{Z}_k \bar{u}^k) = \\
 &g^{k^t} (\alpha_k D \bar{v}^k) + \frac{1}{2} (\alpha_k D \bar{v}^k)^t B_k (\alpha_k D \bar{v}^k) + \\
 &g^{k^t} D \bar{Z}_k \bar{u}^k + (\alpha_k D \bar{v}^k)^t B_k (D \bar{Z}_k \bar{u}^k) + \frac{1}{2} (D \bar{Z}_k \bar{u}^k)^t B_k (D \bar{Z}_k \bar{u}^k) = \\
 &\bar{g}^{k^t} (\alpha_k \bar{v}^k) + \frac{1}{2} (\alpha_k \bar{v}^k)^t \bar{B}_k (\alpha_k \bar{v}^k) + \\
 &\bar{g}^{k^t} \bar{Z}_k \bar{u}^k + (\alpha_k \bar{v}^k)^t \bar{B}_k (\bar{Z}_k \bar{u}^k) + \frac{1}{2} \bar{u}^{k^t} \bar{Z}_k^t \bar{B}_k \bar{Z}_k \bar{u}^k = \\
 &\bar{g}^{k^t} (\alpha_k \bar{v}^k) + \frac{1}{2} (\alpha_k \bar{v}^k)^t \bar{B}_k (\alpha_k \bar{v}^k) + (\bar{g}^k + \alpha_k \bar{B}_k \bar{v}^k)^t \bar{Z}_k \bar{u}^k + \frac{1}{2} \bar{u}^{k^t} \bar{Z}_k^t \bar{B}_k \bar{Z}_k \bar{u}^k = \\
 &\bar{q}_k(\alpha_k \bar{v}^k) - hpred_k(\bar{u}^k).
 \end{aligned}$$

Temos portanto que $hpred_k(\bar{u}^k) = \bar{q}_k(\alpha_k \bar{v}^k) - q_k(d^k)$, com o qual fica demonstrado i.

Vamos agora demonstrar ii.

$$\begin{aligned}
 \bar{q}_k(\alpha_k \bar{v}^k) &= \bar{g}^{k^t} (\alpha_k \bar{v}^k) + \frac{1}{2} (\alpha_k \bar{v}^k)^t \bar{B}_k (\alpha_k \bar{v}^k) = \\
 &\frac{1}{2} \alpha_k \bar{v}^{k^t} (\bar{g}^k + \bar{g}^k + \bar{B}_k (\alpha_k \bar{v}^k)) = \frac{1}{2} \alpha_k \bar{v}^{k^t} (\bar{g}^k + \tilde{g}^k) = \\
 &-\frac{1}{2} \alpha_k c(x^k)^t [\bar{A}_k^t \bar{A}_k]^{-1} \bar{A}_k^t (\bar{g}^k + \tilde{g}^k) = -\alpha_k c(x^k)^t [\bar{A}_k^t \bar{A}_k]^{-1} \bar{A}_k^t \left(\frac{1}{2} \bar{g}^k + \frac{1}{2} \tilde{g}^k \right) = \\
 &-\alpha_k c(x^k)^t [\bar{A}_k^t \bar{A}_k]^{-1} \bar{A}_k^t \left(\frac{1}{2} \bar{g}^k + \frac{1}{2} \tilde{g}^k + \frac{1}{2} \bar{B}_k (\alpha_k \bar{v}^k) \right) =
 \end{aligned}$$

$$\begin{aligned}
& -\alpha_k c(x^k)^t [\bar{A}_k^t \bar{A}_k]^{-1} \bar{A}_k^t (\bar{g}^k + \frac{1}{2} \bar{B}_k (\alpha_k \bar{v}^k)) = \\
& -\alpha_k c(x^k)^t [A_k^t D^2 A_k]^{-1} A_k^t D^2 (g^k + \frac{1}{2} B_k (\alpha_k v^k)) = \alpha_k c(x^k)^t \tilde{\mu}^k. \quad \bullet
\end{aligned}$$

LEMA IV.7. Suponhamos que as hipóteses padrões se cumprem. Seja x^k um ponto gerado pelo algoritmo e \bar{u}^k um passo gerado pela solução de (IV.6) na iteração k . Então

$$pred_k(d^k) \geq hpred_k(\bar{u}^k) + \pi \alpha_k \|c(x^k)\|_1.$$

Demonstração.

Por definição temos que:

$$\begin{aligned}
pred_k(d^k) &= -q_k(d^k) + \sum_{i=1}^m \mu_i \{|c_i(x^k)| - |c_i(x^k) + a^i(x^k)^t d^k|\} = \\
& -q_k(d^k) + \sum_{i=1}^m \mu_i \{|c_i(x^k)| - |c_i(x^k) - \alpha_k c_i(x^k)|\} = \\
& -q_k(d^k) + \sum_{i=1}^m \mu_i \alpha_k |c_i(x^k)|,
\end{aligned}$$

já que $a^i(x^k)^t d^k = -\alpha_k c_i(x^k)$ e $0 < \alpha_k \leq 1$. Agora, pelo lema IV.6, temos:

$$\begin{aligned}
pred_k(d^k) &= hpred_k(\bar{u}^k) - \bar{q}_k(\alpha_k \bar{v}^k) + \sum_{i=1}^m \mu_i \alpha_k |c_i(x^k)| = \\
& hpred_k(\bar{u}^k) - \alpha_k c(x^k)^t \tilde{\mu}^k + \sum_{i=1}^m \mu_i \alpha_k |c_i(x^k)| = \\
& hpred_k(\bar{u}^k) + \alpha_k \sum_{i=1}^m \{\mu_i |c_i(x^k)| - \tilde{\mu}_i^k c_i(x^k)\} \geq \\
& hpred_k(\bar{u}^k) + \alpha_k \sum_{i=1}^m \{(|\tilde{\mu}_i^k| + \pi) |c_i(x^k)| - \tilde{\mu}_i^k c_i(x^k)\} \geq \\
& hpred_k(\bar{u}^k) + \pi \alpha_k \sum_{i=1}^m |c_i(x^k)| = hpred_k(\bar{u}^k) + \pi \alpha_k \|c(x^k)\|_1. \quad \bullet
\end{aligned}$$

O seguinte teorema demonstra que o algoritmo está bem definido, no sentido que cada iteração interna terminará com um passo aceitável depois de um número finito de iterações.

TEOREMA IV.1. Suponhamos que as hipóteses padrões se cumprem. Então, a menos que algum ponto x^k satisfaça as condições necessárias de primeira ordem

para uma solução de (IV.1), cada iteração interna do algoritmo terminará depois de um número finito de iterações.

Demonstração.

Consideremos um ponto qualquer x^k (fixo) gerado pelo algoritmo, o qual não satisfaz as condições necessárias de primeira ordem para uma solução de (IV.1). Então $c(x^k) \neq 0$ ou $Z_k^t g^k \neq 0$. Consideremos primeiro o teste

$$\frac{ared_k(d^k)}{pred_k(d^k)} \geq \eta,$$

o qual é feito em cada iteração interna, com valores decrescentes do raio de região de confiança Δ_k (para o mesmo x^k temos $\Delta_k \downarrow 0$). Consideremos algum

$$0 < \Delta_k \leq \min\{x_i^k - l_i, u_i - x_i^k\}, \quad \forall i = 1, \dots, n.$$

Então

$$D = \Delta_k I \text{ e } D^{-1} = \frac{1}{\Delta_k} I. \quad (\text{IV.13})$$

Suponhamos primeiro que $\|c(x^k)\|_1 > 0$. Pelo lema IV.7, tem-se

$$pred_k(d^k) \geq hpred_k(\bar{u}^k) + \pi\alpha_k \|c(x^k)\|_1 \geq \pi\alpha_k \|c(x^k)\|_1,$$

já que $hpred_k(\bar{u}^k) \geq 0$. Pelo estágio (6) do algoritmo e por (IV.13) temos que

$$\alpha_k \geq \min\left\{1, \theta \frac{\rho}{\|\bar{v}^k\|}\right\} = \min\left\{1, \theta \frac{\rho}{\|D^{-1}v^k\|}\right\} = \min\left\{1, \theta \frac{\rho\Delta_k}{\|v^k\|}\right\}.$$

Logo, para Δ_k suficientemente pequeno e por (IV.7) temos

$$\begin{aligned} pred_k(d^k) &\geq \pi \min\left\{1, \theta \frac{\rho\Delta_k}{\|v^k\|}\right\} \|c(x^k)\|_1 = \pi \min\left\{\|c(x^k)\|_1, \theta\rho \frac{\|c(x^k)\|_1}{\|v^k\|} \Delta_k\right\} \geq \\ &\pi \min\left\{\|c(x^k)\|_1, \frac{\theta\rho}{\gamma_1} \Delta_k\right\} = \frac{\pi\theta\rho}{\gamma_1} \Delta_k. \end{aligned}$$

Por outra parte, pelo lema IV.4 e por (IV.3.4) temos que

$$|ared_k(d^k) - pred_k(d^k)| \leq \kappa \|d\|^2 \leq \kappa\rho^2 \Delta_k^2,$$

para alguma constante $\kappa > 0$. Logo,

$$\left| \frac{ared_k(d^k)}{pred_k(d^k)} - 1 \right| \leq \frac{\kappa\rho^2 \Delta_k^2}{pred_k(d^k)} \leq \frac{\kappa\rho\gamma_1}{\pi\theta} \Delta_k.$$

Então, para Δ_k suficientemente pequeno temos que

$$\left| \frac{ared_k(d^k)}{pred_k(d^k)} - 1 \right| \leq (1 - \eta), \text{ o qual implica que } \frac{ared_k(d^k)}{pred_k(d^k)} \geq \eta.$$

Portanto, depois de um número finito de iterações internas, o passo será aceito.

Suponhamos agora que $\|c(x^k)\|_1 = 0$. Então, por hipóteses, $\|Z_k^t g^k\| > 0$. Por (IV.7) temos que $v(x^k) = 0$ e portanto $\tilde{g}^k = \bar{g}^k$ e $\tilde{\Delta}_k = \rho$. Pelo lema IV.5, e tomando Δ_k suficientemente pequeno tem-se que

$$\begin{aligned} hpred_k(\bar{u}^k) &\geq \frac{1}{2} \|\bar{Z}_k^t \bar{g}^k\| \min \left\{ \rho, \frac{\|\bar{Z}_k^t \bar{g}^k\|}{\|\bar{Z}_k^t B_k \bar{Z}_k\|} \right\} = \\ &\frac{1}{2} \|\bar{Z}_k^t D g^k\| \min \left\{ \rho, \frac{\|\bar{Z}_k^t D g^k\|}{\|\bar{Z}_k^t D B_k D \bar{Z}_k\|} \right\} = \frac{1}{2} \Delta_k \|\bar{Z}_k^t g^k\| \min \left\{ \rho, \frac{\|\bar{Z}_k^t g^k\|}{\Delta_k \|\bar{Z}_k^t B_k \bar{Z}_k\|} \right\} \geq \\ &\frac{1}{2} \|\bar{Z}_k^t g^k\| \min \left\{ \rho \Delta_k, \frac{\|\bar{Z}_k^t g^k\|}{\beta} \right\} = \frac{1}{2} \|\bar{Z}_k^t g^k\| \rho \Delta_k, \end{aligned}$$

já que pelas hipóteses padrões $\|B_k\| \leq \beta$ para todo k , as colunas da matriz \bar{Z}_k são ortonormais e para Δ_k suficientemente pequeno, se $Z_k^t g^k \neq 0$ então $\bar{Z}_k^t g^k \neq 0$. Pelo lema IV.7, temos:

$$pred_k(d^k) \geq hpred_k(\bar{u}^k) + \pi \alpha_k \|c(x^k)\|_1 = hpred_k(\bar{u}^k) \geq \frac{1}{2} \|\bar{Z}_k^t g^k\| \rho \Delta_k.$$

Novamente, pelo lema IV.4 e por (IV.3.4) temos que

$$|ared_k(d^k) - pred_k(d^k)| \leq \kappa \|d\|^2 \leq \kappa \rho^2 \Delta_k^2,$$

para alguma constante $\kappa > 0$. Logo,

$$\left| \frac{ared_k(d^k)}{pred_k(d^k)} - 1 \right| \leq \frac{\kappa \rho^2 \Delta_k^2}{pred_k(d^k)} \leq \frac{2\kappa\rho}{\|\bar{Z}_k^t g^k\|} \Delta_k \leq (1 - \eta),$$

para Δ_k suficientemente pequeno.

Em conseqüência $\frac{ared_k(d^k)}{pred_k(d^k)} \geq \eta$ e assim o passo será aceito. •

COROLÁRIO IV.3. Suponhamos que as hipóteses padrões se cumprem. Então

i. $\{\phi(x^k)\}$ é estritamente decrescente.

ii. $\phi(\hat{x}^1) = \phi(\hat{x}^2)$ para qualquer par de pontos de acumulação \hat{x}^1, \hat{x}^2 de $\{x^k\}$.

Demonstração.

i. Pelo teorema IV.1, no final da k -ésima iteração do algoritmo, depois de um número finito de iterações internas, o teste

$$\frac{ared_k(d^k)}{pred_k(d^k)} \geq \eta,$$

é satisfeito. Temos portanto que

$$\phi(x^k) - \phi(x^{k+1}) = \phi(x^k) - \phi(x^k + d^k) = ared_k(d^k) \geq \eta pred_k(d^k) > 0.$$

Portanto, $\phi(x^k) > \phi(x^{k+1}) \forall k$.

ii. Pelo corolário IV.1 e por i. deste corolário, temos que todos os pontos de acumulação da seqüência $\{x^k\}$ estão contidos em \bar{H} . Por ser ϕ contínua e o conjunto \bar{H} compacto, então ϕ é limitada em \bar{H} . Temos portanto que $\{\phi(x^k)\}$ é limitada inferiormente em \bar{H} . Sejam \hat{x}^1 e \hat{x}^2 dois pontos de acumulação de $\{x^k\}$. Então existem subseqüências $\{x^{k_i}\}$ e $\{x^{k_j}\}$ tais que

$$\{x^{k_i}\} \longrightarrow \hat{x}^1 \quad \text{e} \quad \{x^{k_j}\} \longrightarrow \hat{x}^2.$$

Pela continuidade de ϕ temos que

$$\{\phi(x^{k_i})\} \longrightarrow \phi(\hat{x}^1) \quad \text{e} \quad \{\phi(x^{k_j})\} \longrightarrow \phi(\hat{x}^2).$$

Temos então que $\phi(\hat{x}^1)$ e $\phi(\hat{x}^2)$ são pontos de acumulação de $\{\phi(x^k)\}$. Por outra parte, a seqüência $\{\phi(x^k)\}$ é limitada inferiormente e estritamente decrescente e portanto é convergente. Em conseqüência, $\phi(\hat{x}^1) = \phi(\hat{x}^2)$. •

Note-se que se o ponto inicial x^0 da seqüência $\{x^k\}$ gerada pelo algoritmo, está em H_1^0 então, pelo lema IV.1 e pelo corolário IV.3, temos que a seqüência inteira $\{x^k\}$ está contida em \bar{H} . Como por hipóteses, o conjunto \bar{H} é compacto, temos que o conjunto de pontos de acumulação de $\{x^k\}$ é não vazio e está contido em \bar{H} .

LEMA IV.8. Suponhamos que as hipóteses padrões se cumprem. Então

$$pred_k(d^k) \longrightarrow 0.$$

Demonstração.

Pelo teorema IV.1, no final da k -ésima iteração do algoritmo, depois de um número finito de iterações internas, o teste

$$\frac{ared_k(d^k)}{pred_k(d^k)} \geq \eta,$$

é satisfeito. Temos portanto que para todo k se cumpre:

$$ared_k(d^k) \geq \eta pred_k(d^k) > 0,$$

o qual implica que

$$\phi(x^k) - \phi(x^{k+1}) \geq \eta pred_k(d^k) > 0.$$

Como a seqüência $\{\phi(x^k)\}$ é convergente, então $\phi(x^k) - \phi(x^{k+1}) \longrightarrow 0$ e em conseqüência tem-se que $pred_k(d^k) \longrightarrow 0$. •

COROLÁRIO IV.4. Suponhamos que as hipóteses padrões se cumprem. Então

$$hpred_k(\bar{u}^k) \longrightarrow 0 \text{ e } \alpha_k c(x^k) \longrightarrow 0.$$

Demonstração.

Pelo lema IV.7 temos que

$$pred_k(d^k) \geq hpred_k(\bar{u}^k) + \pi \alpha_k \|c(x^k)\|_1.$$

Como, $\forall k \alpha_k \geq 0$, pelo lema IV.5, $hpred_k(\bar{u}^k) \geq 0$ e pelo lema anterior $pred_k(d^k) \longrightarrow 0$, então temos que $hpred_k(\bar{u}^k) \longrightarrow 0$ e $\alpha_k c(x^k) \longrightarrow 0$. •

No seguinte teorema vamos demonstrar que qualquer ponto de acumulação da seqüência gerada pelo algoritmo, satisfaz a condição (IV.2.1) de Karush-Kuhn-Tucker.

TEOREMA IV.2. Suponhamos que as hipóteses padrões se cumprem e seja \hat{x} um ponto de acumulação de $\{x^k\}$. Se $\hat{\Delta} > 0$ então

$$c(\hat{x}) = 0.$$

Demonstração.

Seja \hat{x} um ponto de acumulação de $\{x^k\}$ e $K \subset \mathcal{N}$ tal que $x^k \xrightarrow{K} \hat{x}$ e $\Delta^k \xrightarrow{K} \hat{\Delta}$, sendo $\hat{\Delta} > 0$. Pelas hipóteses padrões temos que as matrizes $A(x)$ e $(A(x)^t A(x))^{-1}$ são limitadas em S e por construção a matriz D também é limitada em S . Pelo corolário IV.2 temos que a matriz $\hat{A}^t \hat{D}^2 \hat{A}$ é não singular. Portanto, existe uma constante σ tal que $0 < \|DA_k[A_k^t D^2 A_k]^{-1}\| < \sigma, \forall k \in K$. Pelo passo 6 do algoritmo temos que

$$\alpha_k = \min\left\{1, \theta \frac{\rho}{\|D^{-1}v^k\|}\right\} = \min\left\{1, \theta \frac{\rho}{\|DA_k[A_k^t D^2 A_k]^{-1}c(x^k)\|}\right\}.$$

Logo

$$0 \leq \min\{\|c(x^k)\|, \theta \frac{\rho}{\sigma}\} \leq \min\{\|c(x^k)\|, \theta \frac{\rho}{\|DA_k[A_k^t D^2 A_k]^{-1}\|}\} \leq \alpha_k c(x^k).$$

Mas, pelo corolário IV.4, $\alpha_k c(x^k) \rightarrow 0$ e conseqüentemente temos que $c(\hat{x}) = 0$. •

O teorema IV.2 junto com o corolário IV.1 garantem que qualquer ponto de acumulação da seqüência gerada pelo algoritmo é viável para o problema IV.1. Este resultado junto com as hipóteses padrões garantem que todos os pontos de acumulação da seqüência gerada pelo algoritmo são regulares.

COROLÁRIO IV.5. Suponhamos que as hipóteses padrões se cumprem. Se

$$\lim_{\substack{k \rightarrow +\infty \\ k \in K}} \Delta_k \neq 0 \text{ então } \lim_{\substack{k \rightarrow +\infty \\ k \in K}} \bar{v}^k = 0 \text{ e } \lim_{\substack{k \rightarrow +\infty \\ k \in K}} \alpha_k = 1.$$

Demonstração.

Por hipótese temos que $\lim_{\substack{k \rightarrow +\infty \\ k \in K}} \Delta_k \neq 0$. Então, pelo teorema IV.2,

$$\lim_{\substack{k \rightarrow +\infty \\ k \in K}} c(x^k) = 0. \text{ Portanto, temos que}$$

$$\lim_{\substack{k \rightarrow +\infty \\ k \in K}} \bar{v}^k = \lim_{\substack{k \rightarrow +\infty \\ k \in K}} -\bar{A}_k[\bar{A}_k^t \bar{A}_k]^{-1} c(x^k) = 0 \text{ e}$$

$$\lim_{\substack{k \rightarrow +\infty \\ k \in K}} \alpha_k = \lim_{\substack{k \rightarrow +\infty \\ k \in K}} \min\left\{1, \theta \frac{\rho}{\|\bar{v}^k\|}\right\} = 1,$$

já que por hipóteses as matrizes \bar{A}_k , $[\bar{A}_k^t \bar{A}_k]^{-1}$ são limitadas. •

A seguir demonstraremos que qualquer ponto de acumulação da seqüência gerada pelo algoritmo, satisfaz a condição (IV.2.3) de Karush-Kuhn-Tucker.

TEOREMA IV.3. Suponhamos que as hipóteses padrões se cumprem e seja \hat{x} um ponto de acumulação de $\{x^k\}$. Então existem $\hat{\lambda} \in \mathcal{R}^m$ e $\hat{\gamma} \in \mathcal{R}^n$, tais que as subsequências $\{\lambda^k\}_{k \in K}$ e $\{\gamma^k\}_{k \in K}$ convergem a $\hat{\lambda}$ e $\hat{\gamma}$ respectivamente e

$$\nabla f(\hat{x}) = \sum_{i=1}^m \hat{\lambda}_i \nabla c_i(\hat{x}) + \hat{\gamma}.$$

Demonstração.

De acordo com o visto na seção IV.4, $\forall k > 0$ existem $\lambda^k \in \mathcal{R}^m$ e $\gamma^k \in \mathcal{R}^n$ tais que $\nabla f(x^k) = A_k \lambda^k + \gamma^k$, onde

$$\lambda^k = (A_k^t D^2 A_k)^{-1} A_k^t D^2 \nabla f(x^k) \text{ e}$$

$$\gamma^k = (I - A_k (A_k^t D^2 A_k)^{-1} A_k^t D^2) \nabla f(x^k) = \nabla f(x^k) - A_k \lambda^k.$$

Seja \hat{x} um ponto de acumulação de $\{x^k\}$ e $K \subset \mathcal{N}$ tal que $x^k \xrightarrow{K} \hat{x}$ e $\Delta^k \xrightarrow{K} \hat{\Delta}$. Pela continuidade de ∇f e de A temos que $\nabla f(x^k) \xrightarrow{K} \nabla f(\hat{x})$ e $A_k \xrightarrow{K} \hat{A}$.

Se $\hat{\Delta} > 0$ então, por [i.] do corolário IV.2, temos que a matriz $\hat{A}^t \hat{\Delta}^2 \hat{A}$ é não singular, onde $\hat{D} = D(\hat{\Delta}, \hat{x})$. Por outro lado temos que

$$D_{k_{ii}} = \min\{\Delta_k, x_i^k - l_i, u_i - x_i^k\} \xrightarrow{K} \min\{\hat{\Delta}, \hat{x}_i^k - l_i, u_i - \hat{x}_i^k\} = \hat{D}_{ii}.$$

Logo, $D_k \xrightarrow{K} \hat{D}$. Temos portanto que

$$\lambda^k \xrightarrow{K} (\hat{A}^t \hat{\Delta}^2 \hat{A})^{-1} \hat{A}^t \hat{\Delta}^2 \nabla f(\hat{x}),$$

$$\gamma^k \xrightarrow{K} (I - \hat{A}(\hat{A}^t \hat{\Delta}^2 \hat{A})^{-1} \hat{A}^t \hat{\Delta}^2) \nabla f(\hat{x}).$$

Seja $\hat{\lambda} = (\hat{A}^t \hat{\Delta}^2 \hat{A})^{-1} \hat{A}^t \hat{\Delta}^2 \nabla f(\hat{x})$ e $\hat{\gamma} = \nabla f(\hat{x}) - \hat{A} \hat{\lambda}$. Então $\hat{\lambda}$ e $\hat{\gamma}$ assim definidos satisfazem (IV.2.3).

Suponhamos agora que $\hat{\Delta} = 0$.

$\forall k \in K$ temos que $(A_k^t D^2 A_k)^{-1} A_k^t D^2 \nabla f(x^k) = (A_k^t \tilde{D}^2 A_k)^{-1} A_k^t \tilde{D}^2 \nabla f(x^k)$, onde $\tilde{D} = \frac{1}{\Delta_k} D$.

Por [ii.] do corolário IV.2, temos que a matriz $\hat{A}^t \tilde{D}^2 \hat{A}$ é não singular, onde \tilde{D} uma matriz diagonal cujos elementos são $\tilde{D}_{ii} = \min \left\{ 1, \lim_{\substack{k \rightarrow +\infty \\ k \in K}} \frac{x_i^k - l_i}{\Delta_k}, \lim_{\substack{k \rightarrow +\infty \\ k \in K}} \frac{u_i - x_i^k}{\Delta_k} \right\}$.

Por outro lado temos que

$$\tilde{D}_{kii} = \min \left\{ 1, \frac{x_i^k - l_i}{\Delta_k}, \frac{u_i - x_i^k}{\Delta_k} \right\} \xrightarrow{K} \min \left\{ 1, \lim_{\substack{k \rightarrow +\infty \\ k \in K}} \frac{x_i^k - l_i}{\Delta_k}, \lim_{\substack{k \rightarrow +\infty \\ k \in K}} \frac{u_i - x_i^k}{\Delta_k} \right\}.$$

Logo, $D_k \xrightarrow{K} \tilde{D}$. Temos portanto que

$$\lambda^k \xrightarrow{K} (\hat{A}^t \tilde{D}^2 \hat{A})^{-1} \hat{A}^t \tilde{D}^2 \nabla f(\hat{x}),$$

$$\gamma^k \xrightarrow{K} (I - \hat{A}(\hat{A}^t \tilde{D}^2 \hat{A})^{-1} \hat{A}^t \tilde{D}^2) \nabla f(\hat{x}).$$

Seja $\hat{\lambda} = (\hat{A}^t \tilde{D}^2 \hat{A})^{-1} \hat{A}^t \tilde{D}^2 \nabla f(\hat{x})$ e $\hat{\gamma} = \nabla f(\hat{x}) - \hat{A} \hat{\lambda}$. Então $\hat{\lambda}$ e $\hat{\gamma}$ assim definidos satisfazem (IV.2.3). •

Agora só falta demonstrar que o vetor $\hat{\gamma}$ do teorema IV.3 satisfaz as condições (IV.2.4)-(IV.2.6).

No seguinte lema mostramos uma útil relação entre os vetores γ^k , \bar{g}^k e $\bar{Z}_k^t \bar{g}^k$.

LEMA IV.9. Suponhamos que as hipóteses padrões se cumprem. Então em cada iteração do algoritmo temos que

$$D_k \gamma^k = P_{\bar{A}_k^t} \bar{g}^k = \bar{Z}_k \bar{Z}_k^t \bar{g}^k.$$

Demonstração.

Pela definição de γ^k temos que

$$\begin{aligned} D_k \gamma^k &= D_k (I - A_k (A_k^t D^2 A_k)^{-1} A_k^t D^2) \nabla f(x^k) \\ &= (D_k - D_k A_k (A_k^t D^2 A_k)^{-1} A_k^t D^2) \nabla f(x^k) \\ &= (I - \bar{A}_k (\bar{A}_k^t \bar{A}_k)^{-1} \bar{A}_k^t) D_k \nabla f(x^k) \\ &= (I - \bar{A}_k (\bar{A}_k^t \bar{A}_k)^{-1} \bar{A}_k^t) \bar{g}^k = P_{\bar{A}_k^t} \bar{g}^k. \end{aligned}$$

Vamos agora demonstrar a segunda igualdade.

$\bar{g}^k = P_{\bar{A}_k^t} \bar{g}^k + A_k w$, para algum $w \in \mathcal{R}^m$. Multiplicando esta equação por \bar{Z}_k^t , temos $\bar{Z}_k^t \bar{g}^k = \bar{Z}_k^t P_{\bar{A}_k^t} \bar{g}^k$. Por outro lado, $P_{\bar{A}_k^t} \bar{g}^k$ pertence ao espaço nulo de \bar{A}_k^t e em conseqüência existe $y \in \mathcal{R}^{n-m}$ tal que $P_{\bar{A}_k^t} \bar{g}^k = \bar{Z}_k y$. Logo, $\bar{Z}_k^t \bar{g}^k = \bar{Z}_k^t P_{\bar{A}_k^t} \bar{g}^k = \bar{Z}_k^t \bar{Z}_k y = y$. Portanto, $P_{\bar{A}_k^t} \bar{g}^k = \bar{Z}_k \bar{Z}_k^t \bar{g}^k$. •

Como conseqüência do lema IV.9, temos que a matriz $\bar{Z}_k \bar{Z}_k^t$ também pode ser utilizada como matriz de projeção no espaço nulo da matriz \bar{A}_k^t .

No seguinte teorema vamos demonstrar que o vetor $\hat{\gamma}$ do teorema IV.3 satisfaz a condição (IV.2.4), isto é, a condição de complementaridade.

TEOREMA IV.4. Suponhamos que as hipóteses padrões se cumprem e seja $\hat{\gamma}$ como no teorema IV.3. Se $\hat{\Delta} > 0$ então $\hat{\gamma}$ satisfaz a condição (IV.2.4).

Demonstração.

Sob a hipótese de ser $\hat{\Delta} > 0$, provar que $\hat{\gamma}$ satisfaz a condição (IV.2.4), é equivalente a demonstrar que $D_k \gamma^k \xrightarrow{K} 0$. Pelo lema IV.9 temos que $D_k \gamma^k = \bar{Z}_k \bar{Z}_k^t \bar{g}^k$ e por ser a matriz \bar{Z}_k ortonormal, temos que $D_k \gamma^k \xrightarrow{K} 0$ se e somente se $\bar{Z}_k^t \bar{g}^k \xrightarrow{K} 0$. Pelo corolário IV.4, temos que $hpred_k(\bar{u}^k) \rightarrow 0$. Logo,

$$\|\bar{Z}_k^t \bar{g}^k\| \min \left\{ \tilde{\Delta}_k, \frac{\|\bar{Z}_k^t \bar{g}^k\|}{\|\bar{Z}_k^t \bar{B}_k \bar{Z}_k\|} \right\} \rightarrow 0. \quad (\text{IV.14})$$

Pelo corolário IV.5, temos que $\bar{v}^k \xrightarrow{K} 0$ e $\alpha_k \xrightarrow{K} 1$. Então, por ser a matriz \bar{B}_k limitada, temos que $\bar{g}^k = \bar{g}^k + \alpha_k \bar{B}_k \bar{v}^k \xrightarrow{K} \bar{g}^k$ e $\tilde{\Delta}_k = \sqrt{\rho^2 - \alpha_k^2 \|\bar{v}^k\|^2} \xrightarrow{K} \rho$, o que junto (IV.14) implica que $\bar{Z}_k^t \bar{g}^k \xrightarrow{K} 0$, completando a demonstração. •

Agora só falta demonstrar que os sinais do vetor $\hat{\gamma}$ do teorema IV.3 são corretos. Nos resultados obtidos até agora não foi preciso colocar a hipótese de que a seqüência gerada pelo algoritmo fosse convergente, mas para poder demonstrar que as condições (IV.2.5)-(IV.2.6) são satisfeitas por $\hat{\gamma}$, vamos precisar de tal hipótese. Portanto, no que segue, suporemos que a seqüência gerada pelo algoritmo converge ao ponto \hat{x} e que $\hat{\Delta} > 0$.

Vamos supor que existe $i \in \{i = 1, \dots, n\}$ tal que $\hat{\gamma}_i \neq 0$. Pelo teorema IV.4, temos que $\forall i \in \bar{N} \quad \hat{\gamma}_i = 0$. Portanto, $\hat{\gamma}_i \neq 0$ somente para $i \in N$. Consideremos os seguintes conjuntos:

$$N_l = \{i \in N / \hat{x}_i = l_i \text{ e } \hat{\gamma}_i \neq 0\} \text{ e } N_u = \{i \in N / \hat{x}_i = u_i \text{ e } \hat{\gamma}_i \neq 0\}.$$

Devemos demonstrar que

$$\hat{\gamma}_i > 0 \quad \forall i \in N_l \quad \text{e} \quad \hat{\gamma}_i < 0 \quad \forall i \in N_u.$$

Para isto, vamos demonstrar primeiramente o seguinte lema:

LEMA IV.10. Existe $\tilde{k} > 0$ tal que se $k \geq \tilde{k}$ então $\forall i \in N_l \cup N_u \quad d_i^k \gamma_i^k < 0$.

Demonstração.

Pelas condições de Karush-Kuhn-Tucker para o subproblema quadrático (IV.6), $\forall k \in \mathcal{N}$ existe $\psi_k \in \mathcal{R}$ tal que

$$\bar{Z}_k^t(\bar{g}^k + \alpha_k \bar{B}_k \bar{v}^k) + \bar{Z}_k^t \bar{B}_k \bar{Z}_k \bar{u}^k = -\psi_k \bar{u}^k \quad (\text{IV.15})$$

$$\psi_k \geq 0.$$

Por definição temos que $\bar{d}^k = \alpha_k \bar{v}^k + \bar{Z}_k \bar{u}^k$. Logo, pelo lema IV.9 e por (IV.15) temos

$$\begin{aligned} -\psi_k \bar{d}^k &= -\psi_k \alpha_k \bar{v}^k - \psi_k \bar{Z}_k \bar{u}^k \\ &= -\psi_k \alpha_k \bar{v}^k + \bar{Z}_k \bar{Z}_k^t (\bar{g}^k + \alpha_k \bar{B}_k \bar{v}^k) + \bar{Z}_k \bar{Z}_k^t \bar{B}_k \bar{Z}_k \bar{u}^k \\ &= -\psi_k \alpha_k \bar{v}^k + \bar{Z}_k \bar{Z}_k^t \bar{g}^k + \alpha_k \bar{Z}_k \bar{Z}_k^t \bar{B}_k \bar{v}^k + \bar{Z}_k \bar{Z}_k^t \bar{B}_k \bar{Z}_k \bar{u}^k \\ &= -\psi_k \alpha_k \bar{v}^k + D_k \gamma^k + \alpha_k \bar{Z}_k \bar{Z}_k^t \bar{B}_k \bar{v}^k + \bar{Z}_k \bar{Z}_k^t D_k B_k D_k \bar{Z}_k \bar{u}^k. \end{aligned}$$

Temos então que

$$-\psi_k \bar{d}^k - D_k \gamma^k = -\psi_k \alpha_k \bar{v}^k + \alpha_k \bar{Z}_k \bar{Z}_k^t \bar{B}_k \bar{v}^k + \bar{Z}_k \bar{Z}_k^t D_k B_k D_k \bar{Z}_k \bar{u}^k.$$

Por hipótese, a seqüência $\{x^k\}$ é convergente, o que implica que $d^k \rightarrow 0$ e como $d^k = \alpha_k D_k \bar{v}^k + D_k \bar{Z}_k \bar{u}^k$ e $\alpha_k D_k \bar{v}^k \rightarrow 0$ então $D_k \bar{Z}_k \bar{u}^k \rightarrow 0$. Temos então que $-\psi_k \bar{d}^k - D_k \gamma^k \rightarrow 0$. Logo, $-\psi_k D_k^{-1} \bar{d}^k \rightarrow \hat{\gamma}$, ou equivalentemente, $\psi_k D_k^{-2} \bar{d}^k \rightarrow -\hat{\gamma} \neq 0$. Portanto, existe $\tilde{k} > 0$ tal que $\forall k \geq \tilde{k}$ e

$$\forall i \in N_l \cup N_u \quad d_i^k \gamma_i^k < 0. \quad \bullet\bullet$$

O último teorema completa a prova que os sinais de $\hat{\gamma}_N$ são corretos e portanto \hat{x} satisfaz as condições necessárias de otimalidade de Karush-Kuhn-Tucker.

TEOREMA IV.5. Suponhamos que as hipóteses padrões se cumprem e seja $\hat{\gamma}$ como no teorema IV.3. Se $\hat{\Delta} > 0$ então $\hat{\gamma}_{N_l} > 0$ e $\hat{\gamma}_{N_u} < 0$.

Demonstração.

Ab absurdo. Suponhamos que existe um $i \in N_l$ tal que $\hat{\gamma}_i < 0$. Pelo lema IV.1, temos que $\forall k \in \mathcal{N} \quad x_i^k > l_i$. Como $\gamma_i^k \rightarrow \hat{\gamma}_i$ então existe $\bar{k} > 0$ tal que $\forall k \geq \bar{k} \quad \gamma_i^k < 0$ e pelo lema IV.10, temos que existe $\tilde{k} > 0$ tal que $\forall k \geq \tilde{k} \quad d_i^k \gamma_i^k < 0$. Logo, para $\forall k \geq \max\{\tilde{k}, \bar{k}\}$ tem-se que $d_i^k > 0$ e conseqüentemente, $x_i^{k+1} = x_i^k + d_i^k > x_i^k$, o que contradiz o fato de que $x_i^k \rightarrow \hat{x}_i = l_i$. Portanto, $\forall i \in N_l$ tem-se que $\hat{\gamma}_i > 0$.

Analogamente demonstra-se que $\forall i \in N_u \quad \hat{\gamma}_i < 0$. \bullet

Capítulo V

Testes Computacionais

V.1 Introdução

Neste capítulo, apresentamos os testes computacionais realizados para avaliar o desempenho do algoritmo proposto no capítulo IV deste trabalho. A linguagem computacional escolhida foi a linguagem C, pelo seu alto grau de portabilidade, permitir dispor de um modo mais eficiente dos recursos de memória e possuir poderosas estruturas de dados e mecanismos de controle de fluxo eficientes. O equipamento utilizado foi uma estação de trabalho IBM AIX Version 3 for RISC System/6000.

V.2 Testes Realizados

Nesta seção apresentamos os resultados dos testes realizados. Os problemas testados foram tomados de [30] e [38]. Os pontos iniciais, os quais denotaremos por x^0 , são os mesmos das referências. Denotaremos por x^* o ponto para o qual o algoritmo convergiu.

Nos primeiros 13 testes realizados, o desempenho do algoritmo IV.1 foi comparado com o obtido por 6 códigos apresentados em [30]. Estes códigos são:

Código	Autor	Método
VF02AD	Powell	Aproximação quadrática
OPRQP	Bartholomew-Biggs	Aproximação quadrática
GRGA	Abadie	Gradiente reduzido generalizado
VF01A	Fletcher	Multiplicadores
FUNMIN	Kraft	Multiplicadores
FMIN	Kraft, Lootsma	Penalidade

Uma informação mais detalhada sobre estes programas pode ser achada em [45].

Para cada teste realizado apresentamos o enunciado do problema, o ponto inicial e o ponto ao qual o algoritmo convergiu. Fornecemos também o valor da função objetivo no ótimo, o número de iterações empregadas pelo algoritmo até a convergência e o número de iterações internas efetuadas, isto é, o número total de vezes que foi executado o laço do passo 3 ao passo 12 do algoritmo até obter-se um passo aceito em cada iteração. Os resultados dos testes realizados, assim como o desempenho do algoritmo IV.1, são mostrados nas tabelas e figuras a seguir.

Nas tabelas apresentamos os resultados obtidos pelo algoritmo IV.1 e os obtidos em [30] pelos códigos anteriormente citados. Na primeira linha de cada tabela, denotada por ALG. IV.1, são mostrados os resultados do nosso algoritmo. Nas linhas seguintes expomos os resultados obtidos pelos códigos da referência.

As abreviaturas usadas nas tabelas são descritas a seguir:

T.E. : tempo de execução em segundos.

N.A.F.O. : número de avaliações da função objetivo.

N.A.REST. : número de avaliações das restrições (cada restrição é contada).

N.A.GRAD. : número de avaliações do gradiente da função objetivo.

N.A.JAC. : número de avaliações dos gradientes das restrições (cada restrição é contada).

VALOR F.O. : valor da função objetivo no ponto x^* .

VIOL. REST.: soma das violações das restrições no ponto x^* $\left(\sum_{i=1}^m |c_i(x^*)|\right)$.

Os tempos exibidos nas tabelas não podem ser comparados diretamente, pois os resultados apresentados em [30] foram obtidos utilizando-se um computador Telefunken TR440, e são mostrados somente a título de informação.

Nas figuras mostramos o desempenho do nosso algoritmo nos testes realizados. Tal desempenho é mostrado em cada uma delas através de 4 gráficos:

valores de f: mostra a variação da diferença entre os valores da função objetivo e o valor ótimo do teste respectivo, em escala logarítmica, no decurso das iterações. Este valor tende a zero.

inviabilidade: mostra a variação da soma das violações das restrições, em escala logarítmica, no decurso das iterações. Este valor tende a zero.

fc. de merito: mostra a variação da diferença entre os valores da função de mérito e o valor ótimo do teste respectivo, em escala logarítmica, no decurso das iterações. Este valor tende a zero.

raio da bola: mostra a variação do parâmetro Δ , o qual é calculado no passo 2 do algoritmo IV.1, no decurso das iterações.

A seguir, apresentamos os testes realizados:

Teste No. 1 (problema 10 de [30]).

minimizar $x_1 - x_2$

sujeito a $-3x_1^2 + 2x_1x_2 - x_2^2 + 1 \geq 0$

$$x^0 = (-10, 10)^t$$

$$x^* = (0, 1)^t$$

valor da função objetivo no ótimo: -1

número de iterações realizadas: 8

número de iterações internas: 8

CODIGO	T.E. (s)	N.A. F.O.	N.A. REST.	N.A. GRAD.	N.A. JAC.	VALOR F.O.	VIOL. REST.
ALG. IV.1	0.15	9	9	9	9	-1.0000000	.46E-06
VF02AD	.60	12	12	12	12	-1.0000001	.19E-06
OPRQP	.45	32	32	31	31	-1.0000047	.93E-05
GRGA	1.96	260	320	67	31	-1.0000000	.0
VF01A	.35	70	70	70	70	-1.0000000	.53E-07
FUNMIN	.92	225	225	52	52	-1.0000000	.13E-07
FMIN	2.29	289	290	54	54	-0.99999985	.0

Teste No. 1

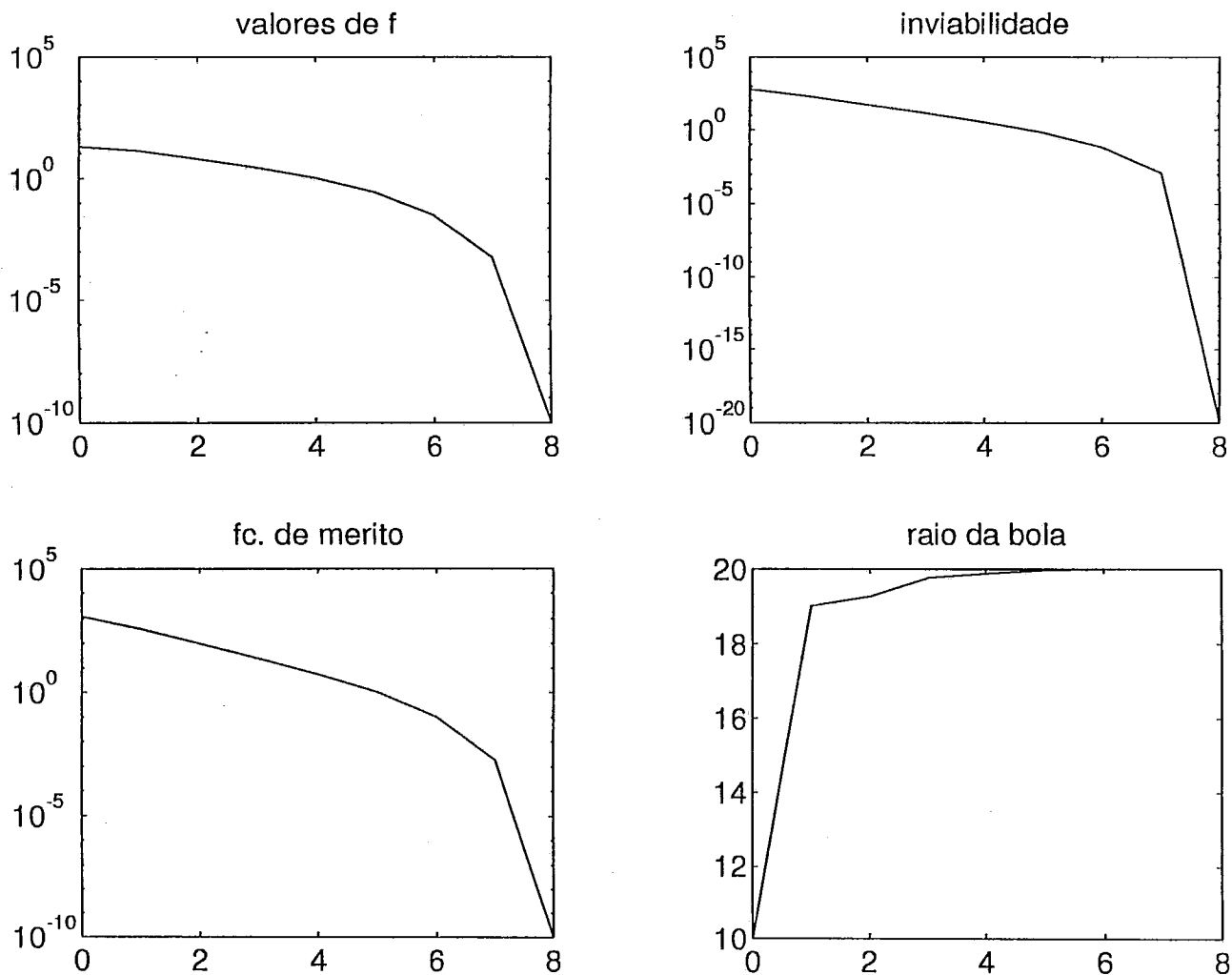


Figura V.1: Problema No. 10 de [30].

Teste No. 2 (problema 11 de [30]).

$$\text{minimizar } (x_1 - 5)^2 + x_2^2 - 25$$

$$\text{sujeito a } -x_1^2 + x_2 \geq 0$$

$$x^0 = (4.9, 0.1)^t$$

$$x^* = (1.234773, 1.524664)^t$$

valor da função objetivo no ótimo: -8.498464223

número de iterações realizadas: 6

número de iterações internas: 6

CODIGO	T.E. (s)	N.A. F.O.	N.A. REST.	N.A. GRAD.	N.A. JAC.	VALOR F.O.	VIOL. REST.
ALG. IV.1	0.09	7	7	7	7	-8.498464	.74E-12
VF02AD	.41	9	9	9	9	-8.4984642	.64E-08
OPRQP	.37	24	24	24	24	-8.4984844	.66E-05
GRGA	1.70	186	201	112	48	-8.4984642	.0
VF01A	.18	40	40	40	40	-8.4984642	.49E-08
FUNMIN	.50	129	129	28	28	-8.4984671	.94E-06
FMIN	2.56	279	280	60	60	-8.4984702	.20E-05

Teste No. 2

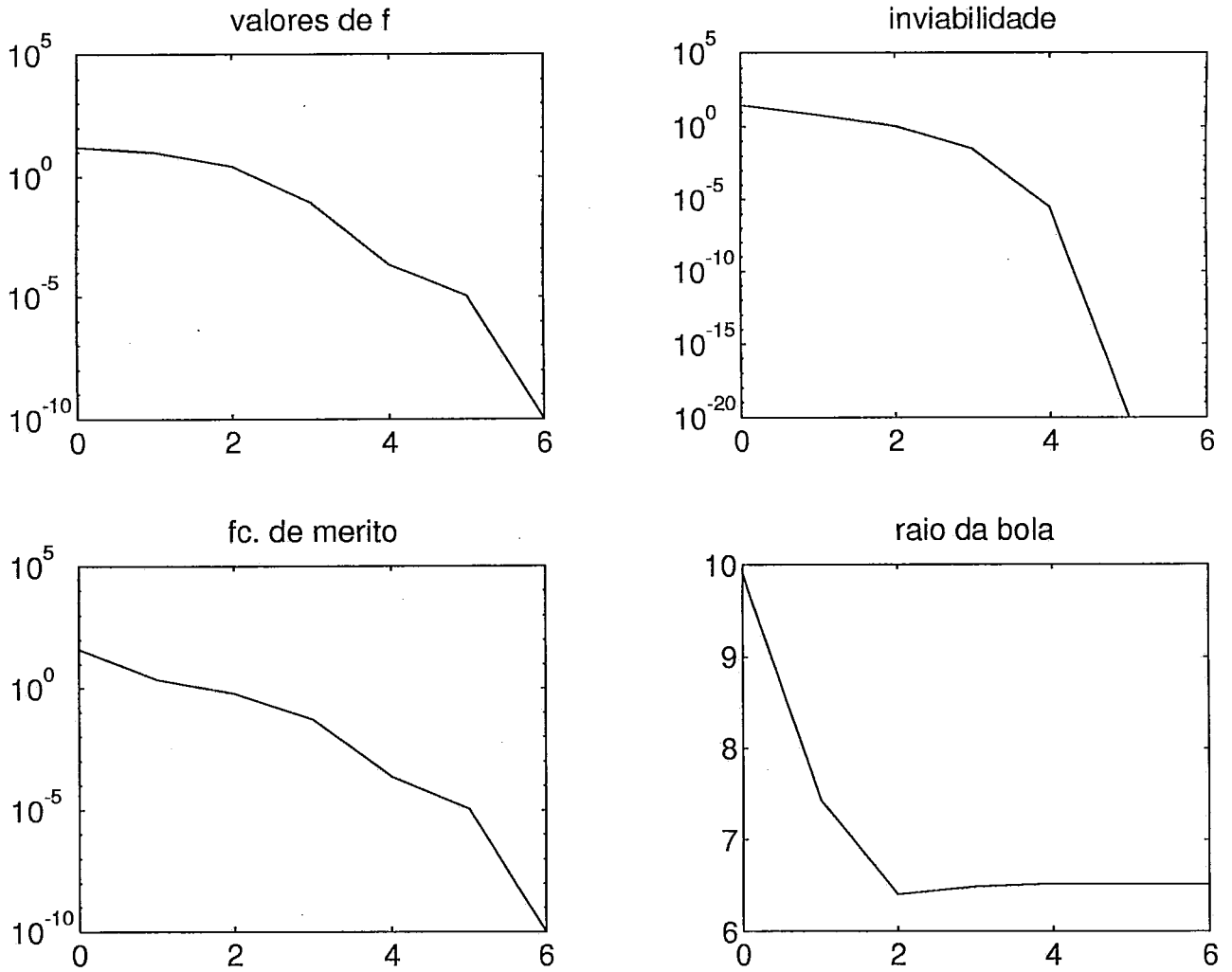


Figura V.2: Problema No. 11 de [30].

Teste No. 3 (problema 12 de [30]).

$$\text{minimizar } 0.5x_1^2 + x_2^2 - x_1x_2 - 7x_1 - 7x_2$$

$$\text{sujeito a } 25 - 4x_1^2 - x_2^2 \geq 0$$

$$x^0 = (0, 0)^t$$

$$x^* = (2, 3)^t$$

valor da função objetivo no ótimo: -30

número de iterações realizadas: 7

número de iterações internas: 7

CODIGO	T.E. (s)	N.A. F.O.	N.A. REST.	N.A. GRAD.	N.A. JAC.	VALOR F.O.	VIOL. REST.
ALG. IV.1	0.13	8	8	8	8	-30.000000	.32E-11
VF02AD	.51	12	12	12	12	-30.000000	.58E-09
OPRQP	.40	40	40	26	26	-30.000004	.76E-05
GRGA	.92	145	89	24	19	-30.000000	.0
VF01A	.30	79	79	79	63	-30.000000	.68E-07
FUNMIN	.87	203	203	43	43	-30.000001	.14E-05
FMIN	1.07	117	133	28	28	-30.000003	.53E-05

Teste No. 3

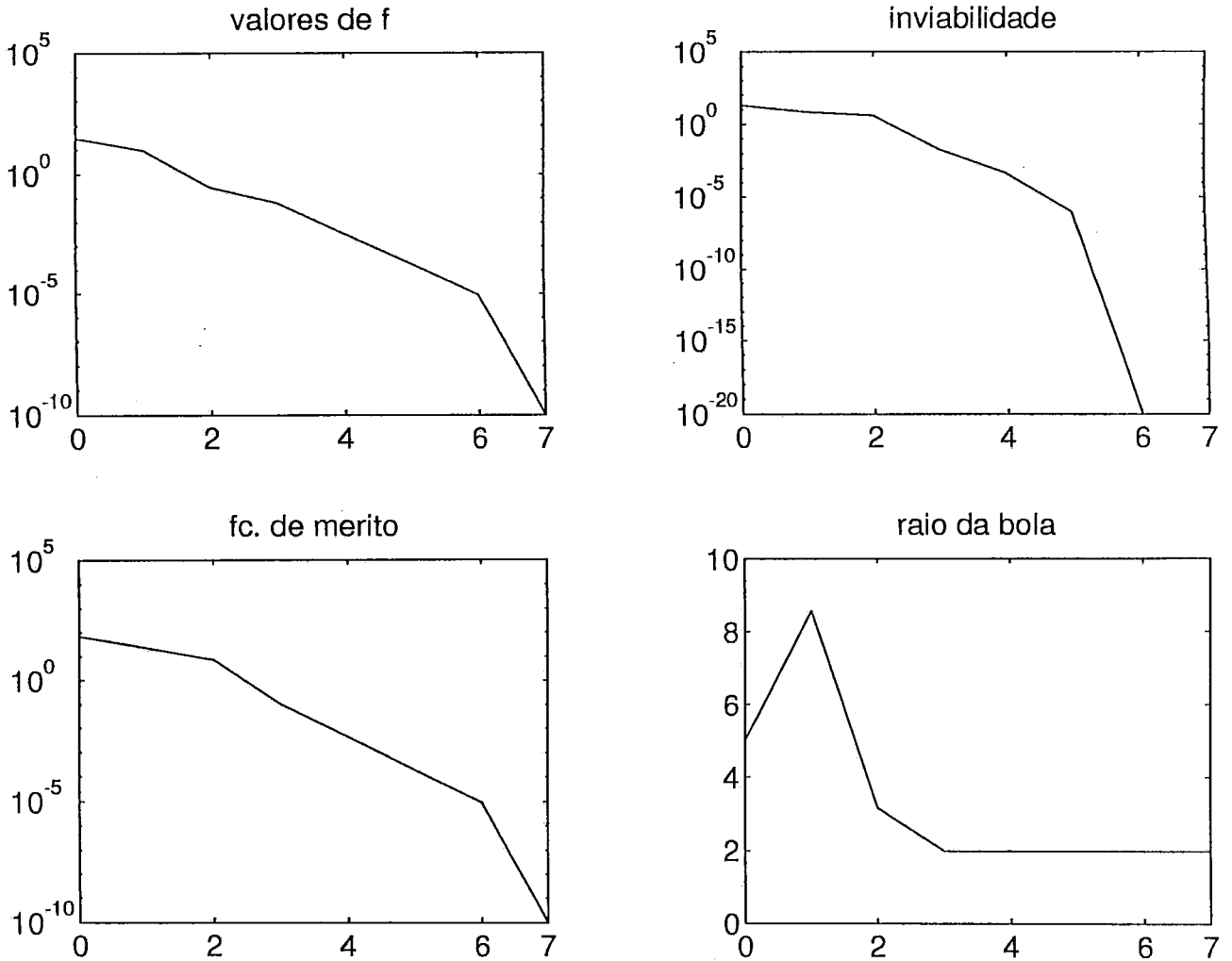


Figura V.3: Problema No. 12 de [30].

Teste No. 4 (problema 13 de [30]).

$$\text{minimizar } (x_1 - 2)^2 + x_2^2$$

$$\text{sujeito a } (1 - x_1)^3 - x_2 \geq 0$$

$$0 \leq x_1$$

$$0 \leq x_2$$

$$x^0 = (2, 2)^t$$

$$x^* = (1, 0)^t$$

valor da função objetivo no ótimo: 1

número de iterações realizadas: 36

número de iterações internas: 47

CODIGO	T.E. (s)	N.A. F.O.	N.A. REST.	N.A. GRAD.	N.A. JAC.	VALOR F.O.	VIOL. REST.
ALG. IV.1	0.59	48	48	37	37	0.9999991	.83E-19
VF02AD	2.80	45	45	45	45	1.0000001	.0
OPRQP	1.18	75	75	75	75	0.99010346	.12E-06
GRGA	.71	72	95	14	11	1.8000585	.0
VF01A	.76	143	143	143	136	0.99568269	.10E-07
FUNMIN	2.08	379	379	85	85	0.97718160	.15E-05
FMIN	13.73	1522	2224	216	216	0.83727822	.15E-03

Teste No. 4

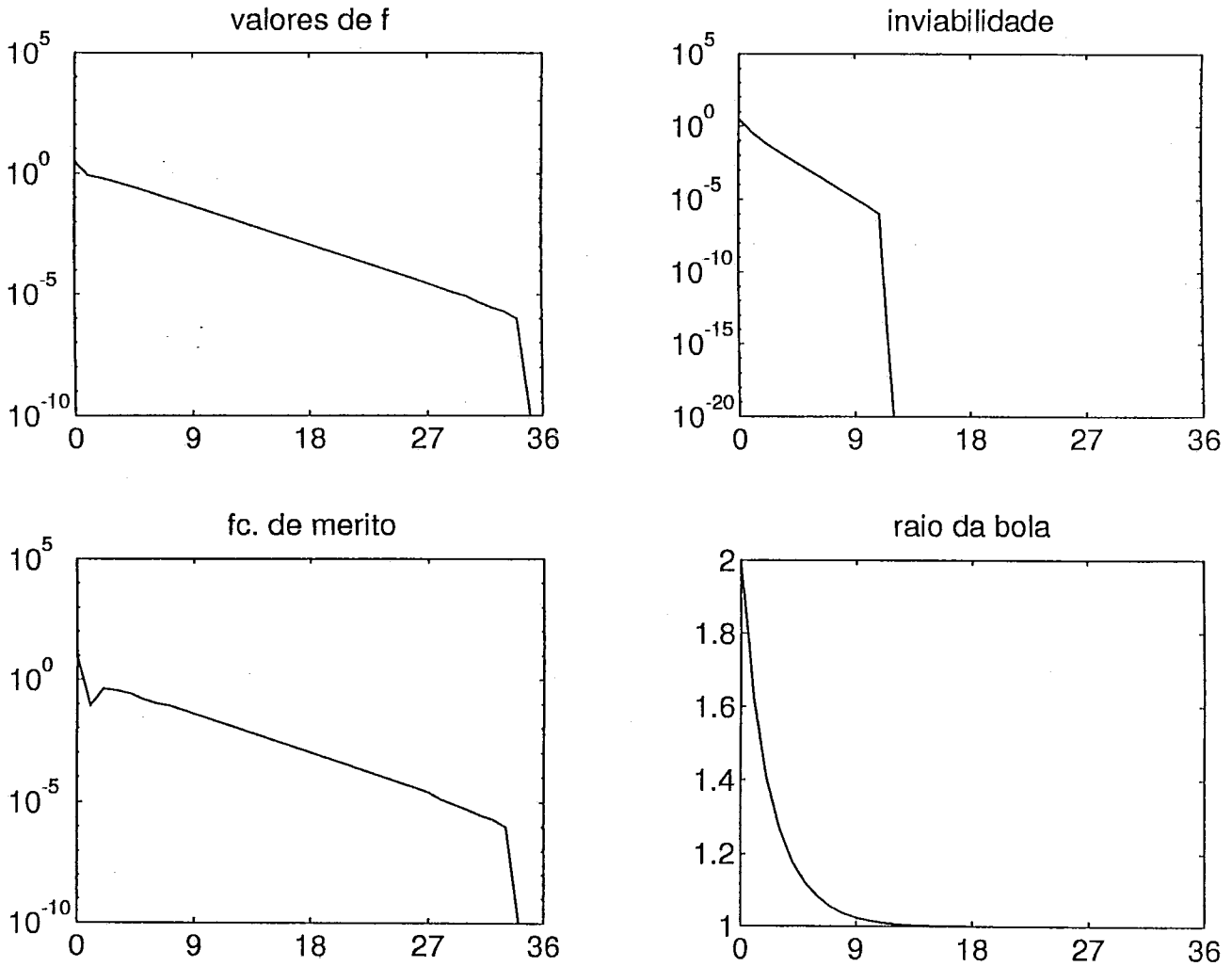


Figura V.4: Problema No. 13 de [30].

Teste No. 5 (problema 14 de [30]).

$$\text{minimizar } (x_1 - 2)^2 + (x_2 - 1)^2$$

$$\text{sujeito a } \begin{aligned} -0.25x_1^2 - x_2^2 + 1 &\geq 0 \\ x_1 - 2x_2 + 1 &= 0 \end{aligned}$$

$$x^0 = (2, 2)^t$$

$$x^* = (0.822876, 0.911438)^t$$

valor da função objetivo no ótimo: $9 - 2.875\sqrt{7}$

número de iterações realizadas: 5

número de iterações internas: 5

CODIGO	T.E. (s)	N.A. F.O.	N.A. REST.	N.A. GRAD.	N.A. JAC.	VALOR F.O.	VIOL. REST.
ALG. IV.1	0.04	6	12	6	12	1.393465	.77E-12
VF02AD	.33	6	12	6	12	1.3934650	.87E-10
OPRQP	.37	21	42	21	42	1.3934485	.95E-05
GRGA	.44	22	62	8	16	1.3934650	.0
VF01A	.16	32	64	32	64	1.3934650	.39E-08
FUNMIN	.99	196	392	46	92	1.3934650	.36E-09
FMIN	2.49	232	465	47	94	1.3934646	.30E-06

Teste No. 5

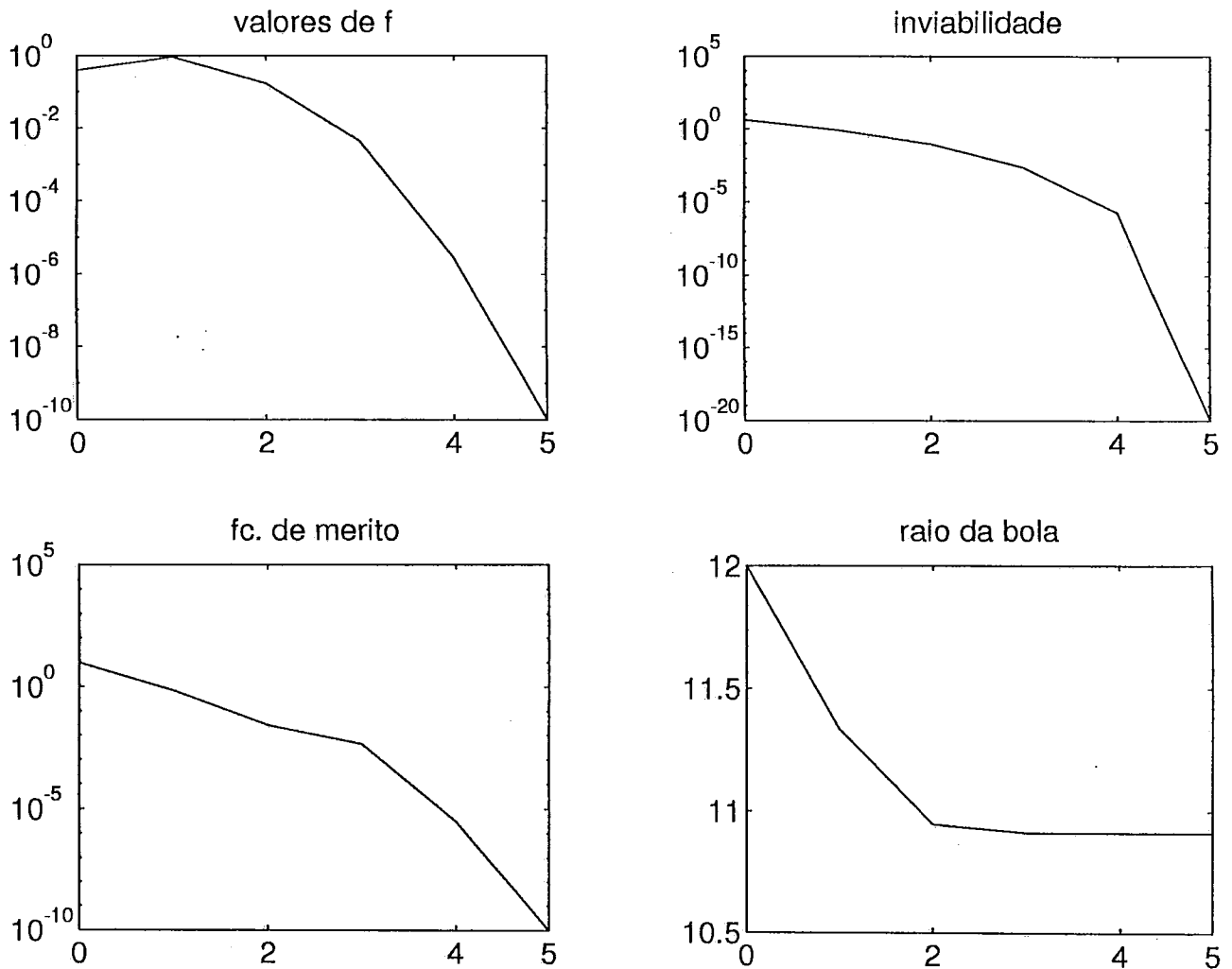


Figura V.5: Problema No. 14 de [30].

Teste No. 6 (problema 22 de [30]).

$$\text{minimizar } (x_1 - 2)^2 + (x_2 - 1)^2$$

$$\text{sujeito a } \begin{aligned} -x_1 - x_2 + 2 &\geq 0 \\ -x_1^2 + x_2 &\geq 0 \end{aligned}$$

$$x^0 = (2, 2)^t$$

$$x^* = (1, 1)^t$$

valor da função objetivo no ótimo: 1
 número de iterações realizadas: 4
 número de iterações internas: 4

CODIGO	T.E. (s)	N.A. F.O.	N.A. REST.	N.A. GRAD.	N.A. JAC.	VALOR F.O.	VIOL. REST.
ALG. IV.1	0.04	5	10	5	10	1.0000000	.20E-08
VF02AD	.40	9	18	9	18	1.0000000	.0
OPRQP	.37	20	40	20	40	0.99998988	.15E-04
GRGA	.91	112	140	37	32	1.0000000	.0
VF01A	.14	23	46	23	46	1.0000000	.38E-08
FUNMIN	.56	134	268	32	64	1.0000000	.0
FMIN	1.76	174	351	36	69	0.99999976	.36E-06

Teste No. 6

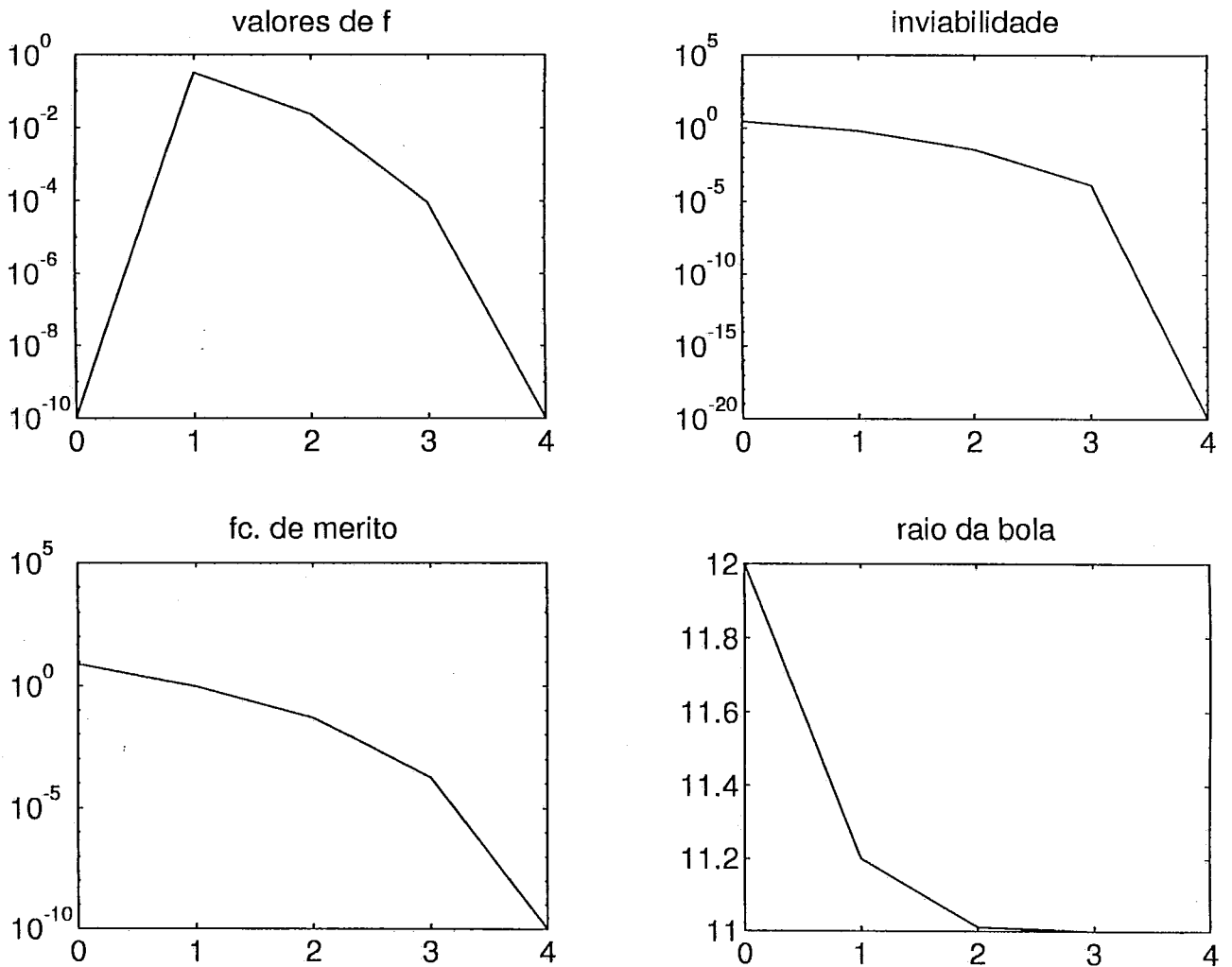


Figura V.6: Problema No. 22 de [30].

Teste No. 7 (problema 29 de [30]).

minimizar $-x_1x_2x_3$

sujeito a $-x_1^2 - 2x_2^2 - 4x_3^2 + 48 \geq 0$

$$x^0 = (1, 1, 1)^t$$

$$x^* = (4, 2.828423, 2)^t$$

valor da função objetivo no ótimo: $-16\sqrt{2}$

número de iterações realizadas: 18

número de iterações internas: 45

CODIGO	T.E. (s)	N.A. F.O.	N.A. REST.	N.A. GRAD.	N.A. JAC.	VALOR F.O.	VIOL. REST.
ALG. IV.1	1.69	46	46	19	19	-22.627416	.24E-12
VF02AD	.95	13	13	13	13	-22.627417	.0
OPRQP	.70	64	64	39	39	-22.627421	.56E-05
GRGA	2.76	310	196	84	56	-22.627417	.0
VF01A	.49	110	110	110	91	-22.627417	.0
FUNMIN	.90	196	196	45	45	-22.627417	.18E-07
FMIN	1.68	159	177	39	39	-22.627416	.0

Teste No. 7

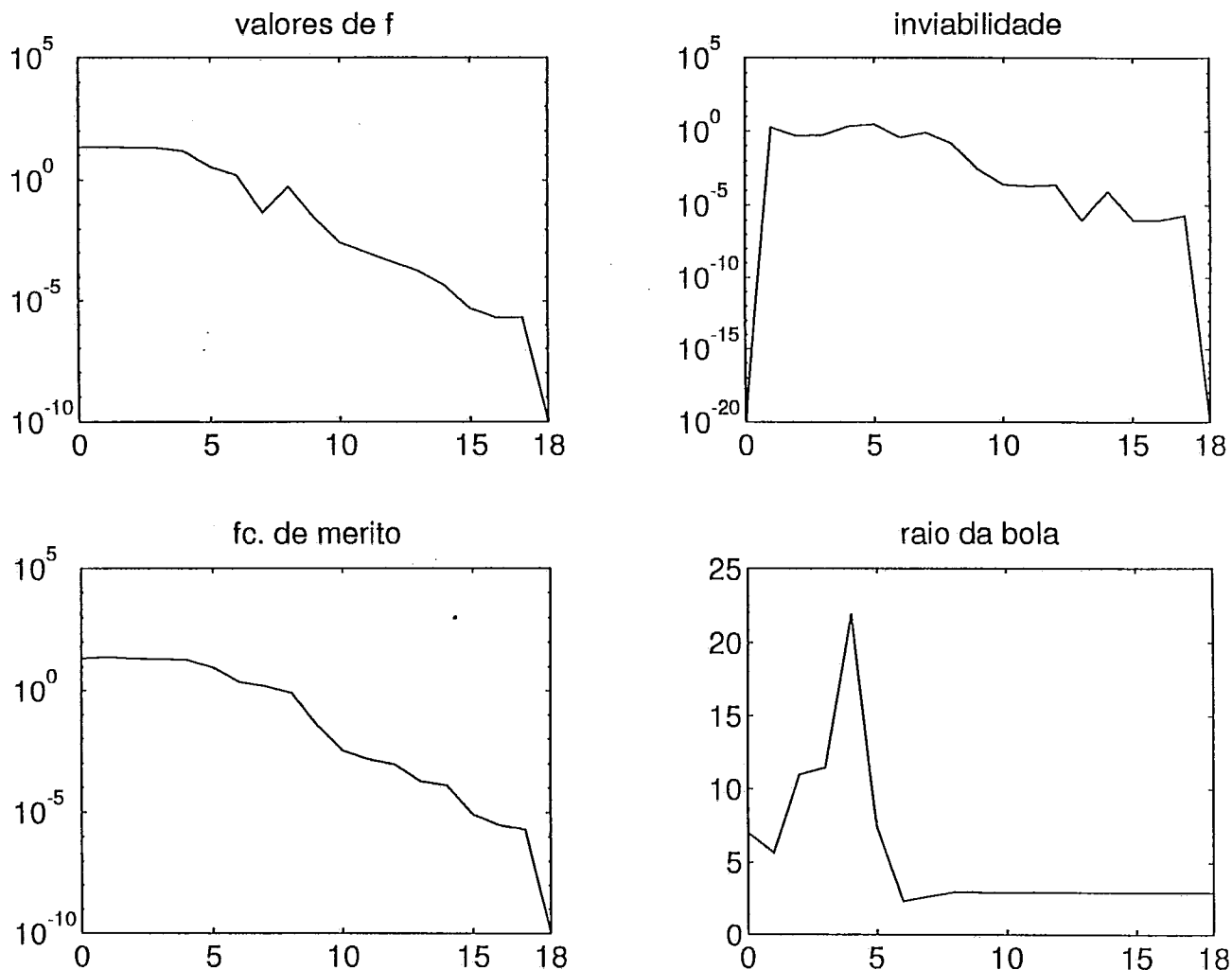


Figura V.7: Problema No. 29 de [30].

Teste No. 8 (problema 43 de [30]).

$$\text{minimizar } x_1^2 + x_2^2 + 2x_3^2 + x_4^2 - 5x_1 - 5x_2 - 21x_3 + 7x_4$$

$$\text{sujeito a } 8 - x_1^2 - x_2^2 - x_3^2 - x_4^2 - x_1 + x_2 - x_3 + x_4 \geq 0$$

$$10 - x_1^2 - 2x_2^2 - x_3^2 - 2x_4^2 + x_1 + x_4 \geq 0$$

$$5 - 2x_1^2 - x_2^2 - x_3^2 - 2x_1 + x_2 + x_4 \geq 0$$

$$x^0 = (0, 0, 0, 0)^t$$

$$x^* = (0, 1, 2, -1)^t$$

valor da função objetivo no ótimo: -44

número de iterações realizadas: 17

número de iterações internas: 20

CODIGO	T.E. (s)	N.A. F.O.	N.A. REST.	N.A. GRAD.	N.A. JAC.	VALOR F.O.	VIOL. REST.
ALG. IV.1	0.29	21	63	18	54	-44.000000	.31E-12
VF02AD	1.47	12	36	12	36	-44.000000	.35E-09
OPRQP	.54	31	93	24	72	-44.000013	.79E-05
GRGA	2.95	426	900	101	153	-44.000000	.0
VF01A	.42	66	198	66	118	-44.000000	.67E-08
FUNMIN	1.31	211	633	47	141	-44.000000	.41E-07
FMIN	5.84	366	1215	92	276	-44.000000	.0

Teste No. 8

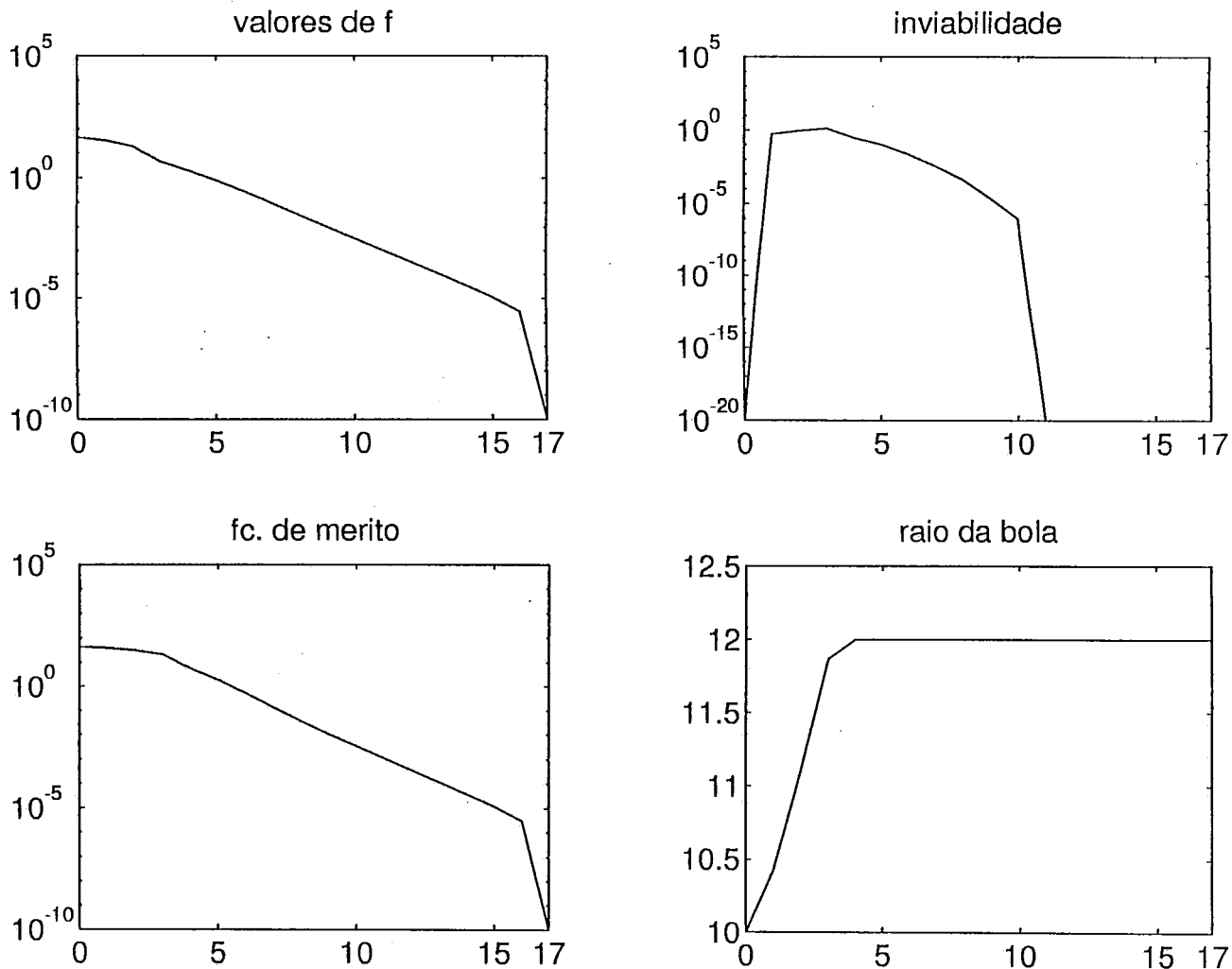


Figura V.8: Problema No. 43 de [30].

Teste No. 9 (problema 44 de [30]).

minimizar $x_1 - x_2 - x_3 - x_1x_3 + x_1x_4 + x_2x_3 - x_2x_4$

sujeito a $8 - x_1 - 2x_2 \geq 0$
 $12 - 4x_1 - x_2 \geq 0$
 $12 - 3x_1 - 4x_2 \geq 0$
 $8 - 2x_3 - x_4 \geq 0$
 $8 - x_3 - 2x_4 \geq 0$
 $5 - x_3 - x_4 \geq 0$
 $0 \leq x_i, i = 1, \dots, 4$

$$x^0 = (0, 0, 0, 0)^t$$

$$x^* = (0, 3, 0, 4)^t$$

valor da função objetivo no ótimo: -15

número de iterações realizadas: 18

número de iterações internas: 18

CODIGO	T.E. (s)	N.A. F.O.	N.A. REST.	N.A. GRAD.	N.A. JAC.	VALOR F.O.	VIOL. REST.
ALG. IV.1	1.12	19	114	19	114	-15.000000	.28E-14
VF02AD	1.15	6	36	6	36	-15.000000	.12E-09
OPRQP	.39	25	150	10	60	-15.000016	.26 E-05
GRGA	.55	28	102	13	60	-15.000000	.0
VF01A	.62	105	630	105	630	-15.000004	.94E-06
FUNMIN	2.48	284	1704	58	348	-13.066273	.29E-02
FMIN	21.26	1099	7042	58	348	-.30872255E+23	.0

Teste No. 9

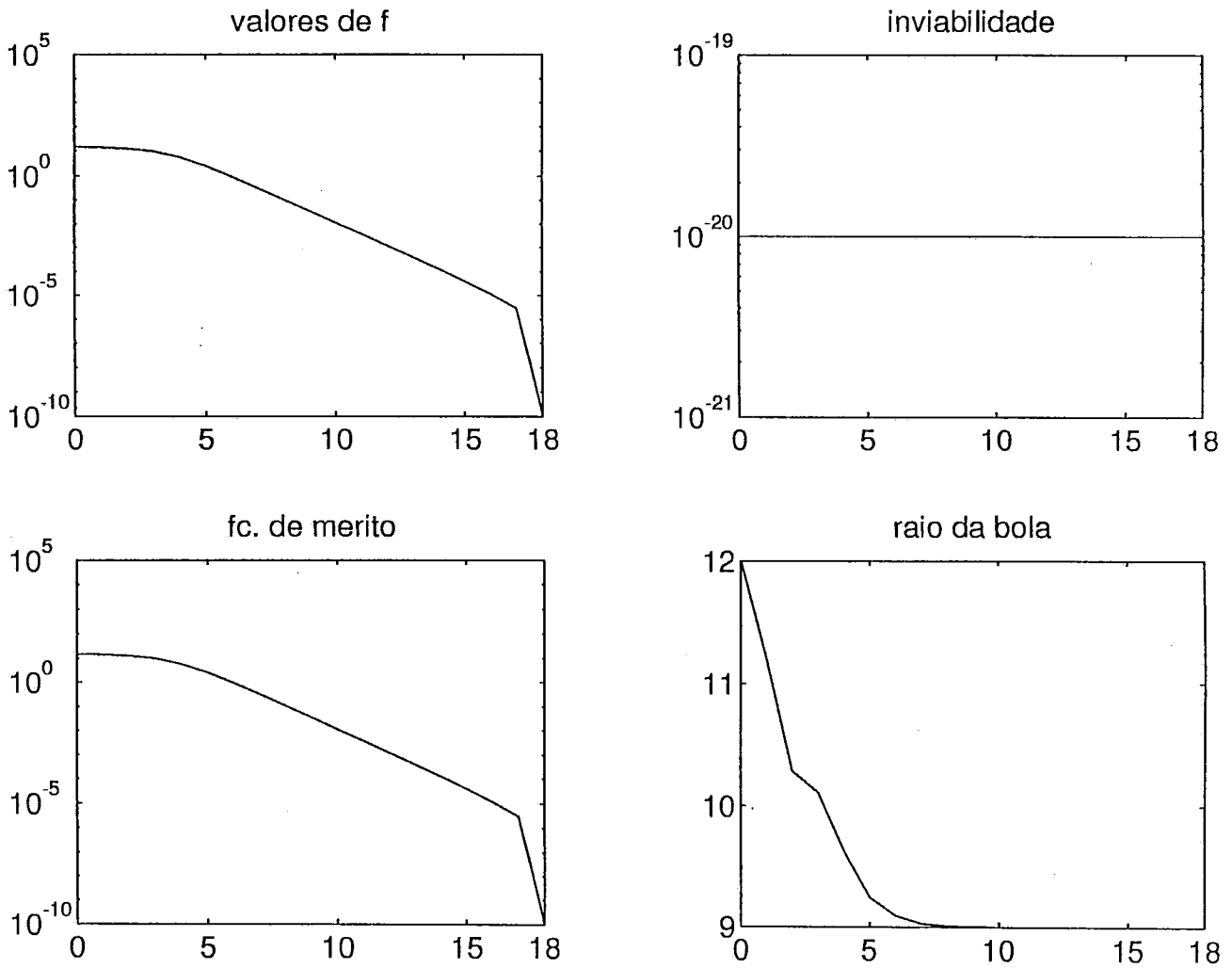


Figura V.9: Problema No. 44 de [30].

Teste No. 10 (problema 100 de [30]).

$$\text{minimizar } (x_1 - 10)^2 + 5(x_2 - 12)^2 + x_3^4 + 3(x_4 - 11)^2 + 10x_5^6 + 7x_6^2 + x_7^4 - 4x_6x_7 - 10x_6 - 8x_7$$

$$\begin{aligned} \text{sujeito a } \quad & 127 - 2x_1^2 - 3x_2^4 - x_3 - 4x_4^2 - 5x_5 \geq 0 \\ & 282 - 7x_1 - 3x_2 - 10x_3^2 - x_4 + x_5 \geq 0 \\ & 196 - 23x_1 - x_2^2 - 6x_6^2 + 8x_7 \geq 0 \\ & -4x_1^2 - x_2^2 + 3x_1x_2 - 2x_3^2 - 5x_6 + 11x_7 \geq 0 \end{aligned}$$

$$x^0 = (1, 2, 0, 4, 0, 1, 1)^t$$

$$x^* = (2.330499, 1.951372, -0.477541, 4.365726, -0.624487, 1.038131, 1.594227)^t$$

valor da função objetivo no ótimo: 680.6300573

número de iterações realizadas: 17

número de iterações internas: 37

CODIGO	T.E. (s)	N.A. F.O.	N.A. REST.	N.A. GRAD.	N.A. JAC.	VALOR F.O.	VIOL. REST.
ALG. IV.1	2.98	38	152	18	72	680.630059	.21E-11
VF02AD	5.29	20	80	20	80	680.63006	.76E-07
OPRQP	1.09	49	196	31	124	680.63005	.76 E-05
GRGA	1.94	213	504	33	104	680.63006	.0
VF01A	1.25	128	512	128	176	680.63004	.24E-04
FUNMIN	9.10	1037	4148	263	1052	705.30863	.0
FMIN	10.17	544	2273	109	436	680.63006	.14E-05

Teste No. 10

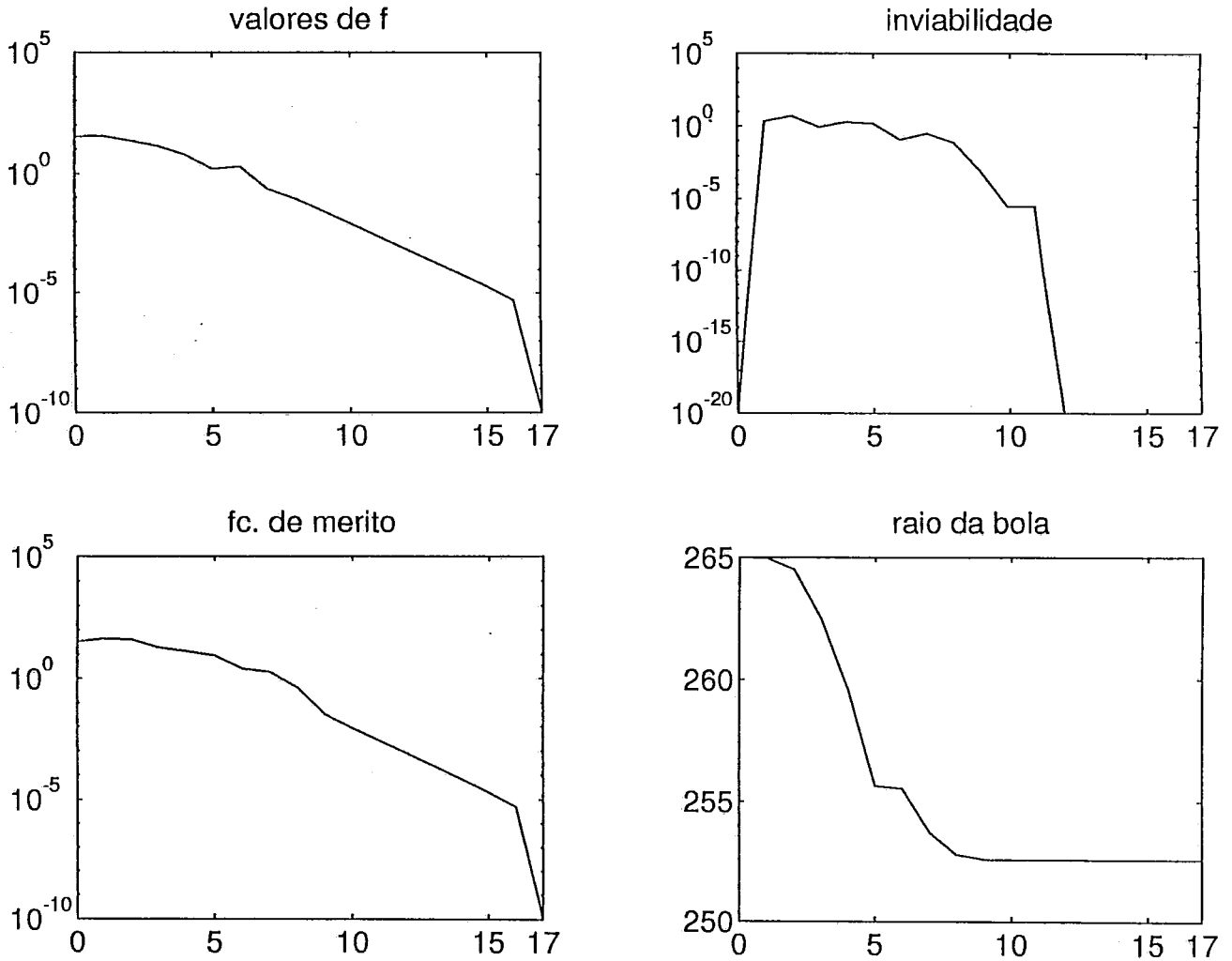


Figura V.10: Problema No. 100 de [30].

Teste No. 11 (problema 108 de [30]).

minimizar $-0.5(x_1x_4 - x_2x_3 + x_3x_9 - x_5x_9 + x_5x_8 - x_6x_7)$

sujeito a

$$1 - x_3^2 - x_4^2 \geq 0$$

$$1 - x_9^2 \geq 0$$

$$1 - x_5^2 - x_6^2 \geq 0$$

$$1 - x_1^2 - (x_2 - x_9)^2 \geq 0$$

$$1 - (x_1 - x_5)^2 - (x_2 - x_6)^2 \geq 0$$

$$1 - (x_1 - x_7)^2 - (x_2 - x_8)^2 \geq 0$$

$$1 - (x_3 - x_5)^2 - (x_4 - x_6)^2 \geq 0$$

$$1 - (x_3 - x_7)^2 - (x_4 - x_8)^2 \geq 0$$

$$1 - x_7^2 - (x_8 - x_9)^2 \geq 0$$

$$x_1x_4 - x_2x_3 \geq 0$$

$$x_3x_9 \geq 0$$

$$-x_5x_9 \geq 0$$

$$x_5x_8 - x_6x_7 \geq 0$$

$$0 \leq x_9$$

$$x^0 = (1, 1, 1, 1, 1, 1, 1, 1, 1)^t$$

$$x^* = (0.01579084, -0.99987548, 0.87381296, -0.48626252, 0.01579079, -0.99987548, 0.87381298, -0.48626247, 0.00000000)^t$$

valor da função objetivo no ótimo: -0.8660254038

número de iterações realizadas: 68

número de iterações internas: 69

CODIGO	T.E. (s)	N.A. F.O.	N.A. REST.	N.A. GRAD.	N.A. JAC.	VALOR F.O.	VIOL. REST.
ALG. IV.1	5.11	70	910	69	897	-0.8660256	.18E-05
VF02AD	14.17	9	117	9	117	-0.69701242	.34E-02
OPRQP	2.86	51	663	39	507	-0.86602552	.82E-06
GRGA	21.85	678	14976	176	1027	-0.67498144	.0
VF01A	3.58	201	2613	201	893	-0.86666678	.47E-02
FUNMIN	7.99	546	7098	123	1599	-0.67498144	.0
FMIN	46.47	984	13566	210	1964	-0.86588938	.19E-03

Teste No. 11

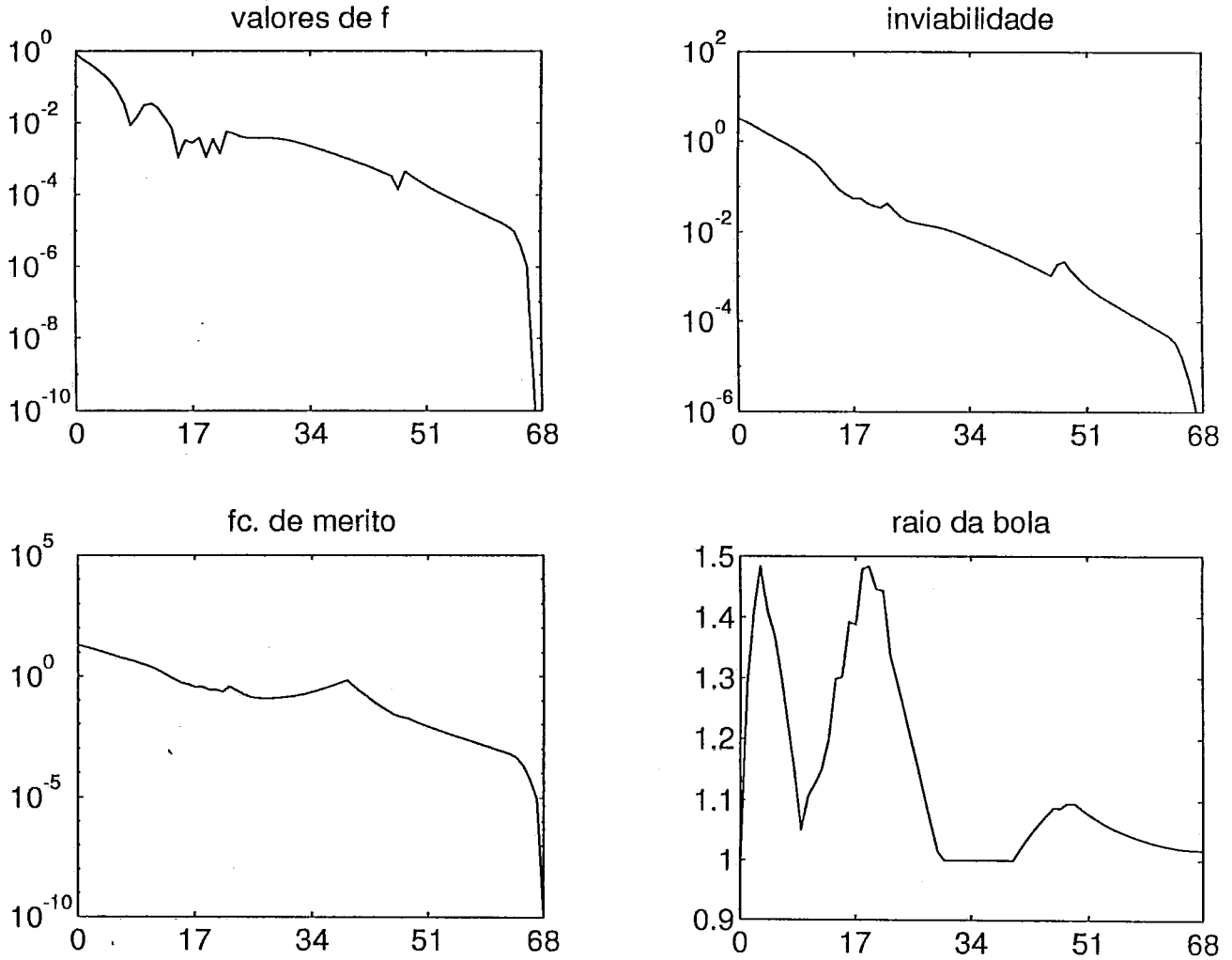


Figura V.11: Problema No. 108 de [30].

Teste No. 12 (problema 113 de [30]).

$$\text{minimizar } x_1^2 + x_2^2 + x_1x_2 - 14x_1 + 16x_2 + (x_3 - 10)^2 + 4(x_4 - 5)^2 + (x_5 - 3)^2 + 2(x_6 - 1)^2 + 5x_7^2 + 7(x_8 - 11)^2 + 2(x_9 - 10)^2 + (x_{10} - 7)^2 + 45$$

$$\begin{aligned} \text{sujeito a } & 105 - 4x_1 - 5x_2 + 3x_7 - 9x_8 \geq 0 \\ & -10x_1 + 8x_2 + 17x_7 - 2x_8 \geq 0 \\ & 8x_1 - 2x_2 - 5x_9 + 2x_{10} + 12 \geq 0 \\ & -3(x_1 - 2)^2 - 4(x_2 - 3)^2 - 2x_3^2 + 7x_4 + 120 \geq 0 \\ & -5x_1^2 - 8x_2 - (x_3 - 6)^2 + 2x_4 + 40 \geq 0 \\ & -0.5(x_1 - 8)^2 - 2(x_2 - 4)^2 - 3x_5^2 + x_6 + 30 \geq 0 \\ & -x_1^2 - 2(x_2 - 2)^2 + 2x_1x_2 - 14x_5 + 6x_6 \geq 0 \\ & 3x_1 - 6x_2 - 12(x_9 - 8)^2 + 7x_{10} \geq 0 \end{aligned}$$

$$x^0 = (2, 3, 5, 5, 1, 2, 7, 3, 6, 10)^t$$

$$x^* = (2.17199637, 2.36368298, 8.77392572, 5.09598449, 0.99065476, 1.43057398, 1.32164421, 9.82872581, 8.28009167, 8.37592666)^t$$

valor da função objetivo no ótimo: 24.3062091

número de iterações realizadas: 44

número de iterações internas: 52

CODIGO	T.E. (s)	N.A. F.O.	N.A. REST.	N.A. GRAD.	N.A. JAC.	VALOR F.O.	VIOL. REST.
ALG. IV.1	2.81	53	424	45	360	24.306209	.19E-13
VF02AD	12.40	15	120	15	120	24.306209	.16E-07
OPRQP	1.94	30	240	28	224	24.306193	.13E-04
GRGA	3.59	336	1040	46	232	24.306209	.0
VF01A	2.91	201	1608	201	860	24.285322	.19E-01
FUNMIN	14.57	1183	9464	269	2152	24.897161	.0
FMIN	32.22	906	8089	189	1512	24.306207	.56E-04

Teste No. 12

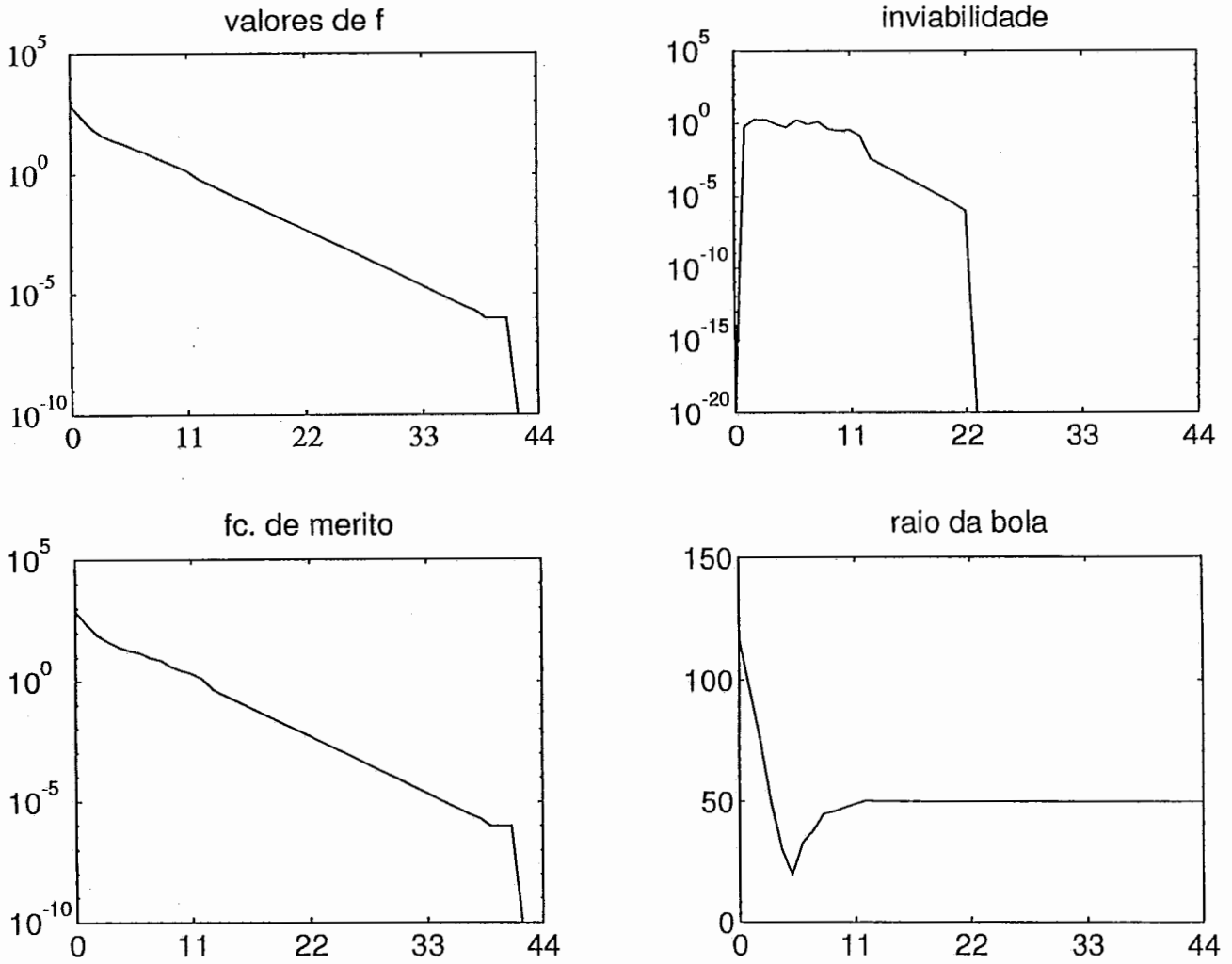


Figura V.12: Problema No. 113 de [30].

Teste No. 13 (problema 118 de [30]).

$$\text{minimizar } \sum_{k=0}^4 (2.3x_{3k+1} + 0.0001x_{3k+1}^2 + 1.7x_{3k+2} + 0.0001x_{3k+2}^2 + 2.2x_{3k+3} + 0.00015x_{3k+3}^2)$$

$$\begin{aligned} \text{sujeito a } & 0 \leq x_{3k+1} - x_{3k-2} + 7 \leq 13 \\ & 0 \leq x_{3k+2} - x_{3k-1} + 7 \leq 14 \\ & 0 \leq x_{3k+3} - x_{3k} + 7 \leq 13 \\ & k = 1, \dots, 4 \\ & x_1 + x_2 + x_3 - 60 \geq 0 \\ & x_4 + x_5 + x_6 - 50 \geq 0 \\ & x_7 + x_8 + x_9 - 70 \geq 0 \\ & x_{10} + x_{11} + x_{12} - 85 \geq 0 \\ & x_{13} + x_{14} + x_{15} - 100 \geq 0 \\ & 8 \leq x_1 \leq 21 \\ & 43 \leq x_2 \leq 57 \\ & 3 \leq x_3 \leq 16 \\ & 0 \leq x_{3k+1} \leq 90 \\ & 0 \leq x_{3k+2} \leq 120 \\ & 0 \leq x_{3k+3} \leq 60 \\ & k = 1, \dots, 4 \end{aligned}$$

$$x^0 = (20, 55, 15, 20, 60, 20, 20, 60, 20, 20, 60, 20, 20, 60, 20)^t$$

$$x^* = (8, 49, 3, 1, 56, 0, 1, 63, 6, 3, 70, 12, 5, 77, 18)^t$$

valor da função objetivo no ótimo: 664.8204500

número de iterações realizadas: 71

número de iterações internas: 71

CODIGO	T.E. (s)	N.A. F.O.	N.A. REST.	N.A. GRAD.	N.A. JAC.	VALOR F.O.	VIOL. REST.
ALG. IV.1	9.43	72	1224	72	1224	664.82045	.62E-13
VF02AD	****	**	****	**	***	*****	*****
OPRQP	5.71	35	1015	25	725	664.82044	.26E-05
GRGA	3.64	35	3422	23	377	665.37276	.0
VF01A	****	**	****	**	***	*****	*****
FUNMIN	****	**	****	**	***	*****	*****
FMIN	****	**	****	**	***	*****	*****

O grau de dificuldade deste teste pode ser evidenciado pelo fato de que os códigos VF02AD, VF01A, FUNMIN e FMIN não conseguiram convergir. Por outro lado,

mesmo tendo convergido, o código GRGA encontrou uma solução com valor maior da função objetivo. A solução obtida pelo código OPRQP tem uma precisão menor àquela obtida pelo nosso algoritmo. Este fato pode ser constatado ao comparar-se os valores da violação das restrições e da função objetivo das respectivas soluções.

Teste No. 13

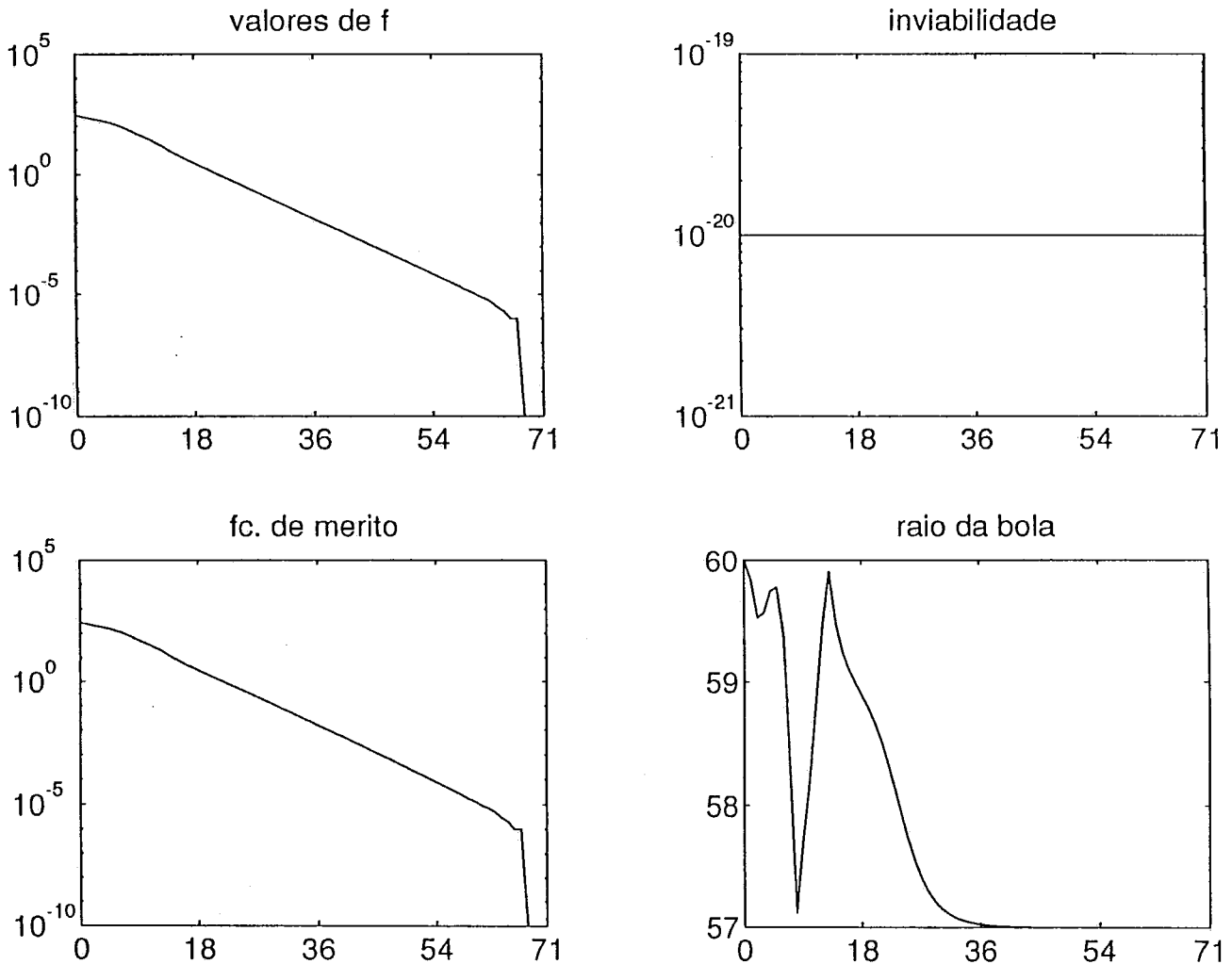


Figura V.13: Problema No. 118 de [30].

Os seguintes testes foram tomados de [38]. Eles formam parte de um conjunto de problemas que foram utilizados para avaliar o desempenho do programa MINOS/AUGMENTED, uma extensão do código de otimização MINOS, cujo objetivo é resolver problemas tanto com restrições lineares como não lineares. Os resultados aqui apresentados, foram obtidos utilizando o MINOS 5.1, uma versão mais recente do MINOS/AUGMENTED. Nos testes realizados com o MINOS 5.1 foram empregadas duas estratégias: Newton e lagrangiano aumentado com parâmetro de penalidade $\rho \geq 0$. Na segunda estratégia os valores do parâmetro de penalidade utilizados foram: $\rho = 0$, $\rho = 10$ e $\rho = 100$. Mas detalhes sobre o MINOS 5.1 podem ser achados em [39]. Os testes com o MINOS 5.1 foram realizados em um microcomputador 486/50 MHz. Por esta razão, os tempos de execução do algoritmo IV.1 e do MINOS 5.1 exibidos nas tabelas a seguir não são diretamente comparáveis, e são mostrados somente a título de informação.

O teste No. 9 de [38] é um problema altamente não linear e não convexo, para o qual 4 mínimos locais tem sido determinados, todos os quais foram achados pelo nosso algoritmo, tal como é mostrado nos seguintes 5 resultados apresentados.

Teste No. 14 (problema 9 de [38]).

$$\text{minimizar } (x_1 - 1)^2 + (x_1 - x_2)^2 + (x_2 - x_3)^3 + (x_3 - x_4)^4 + (x_4 - x_5)^4$$

$$\begin{aligned} \text{sujeito a } \quad x_1 + x_2^2 + x_3^3 &= 3\sqrt{2} + 2 \\ x_2 - x_3^2 + x_4 &= 2\sqrt{2} - 2 \\ x_1 x_5 &= 2 \end{aligned}$$

$$x^0 = (1, 1, 1, 1, 1)^t$$

$$x^* = (1.11663469, 1.22044054, 1.53778551, 1.97277085, 1.79109606)^t$$

valor da função objetivo no ótimo: 0.02931083

número de iterações realizadas: 6

número de iterações internas: 7

CODIGO	T.E. (s)	N.A. F.O.	N.A. REST.	N.A. GRAD.	N.A. JAC.	VALOR F.O.	VIOL. REST.
ALG. IV.1	0.08	8	24	7	21	0.0293108	.80E-07
MINOS c/ Newton	0.93	39	27	39	27	0.0293108	.13E-08
MINOS c/ $\rho = 0$	1.21	63	189	63	189	0.0293108	.15E-13
MINOS c/ $\rho = 10$	0.99	40	120	40	120	0.0293108	.27E-12
MINOS c/ $\rho = 100$	1.16	55	165	55	165	0.0293108	.14E-14

Teste No. 14

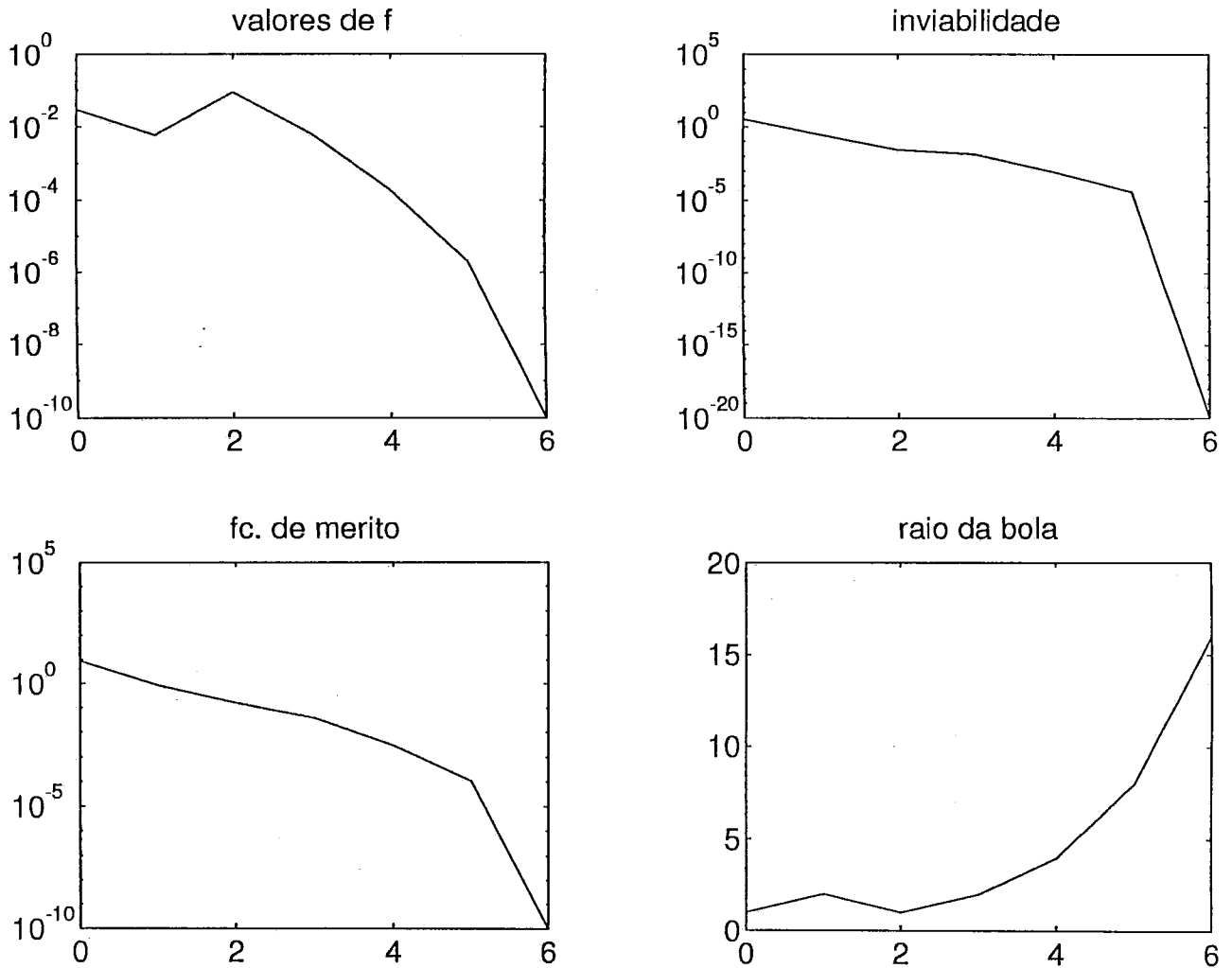


Figura V.14: Problema No. 9 de [38].

Teste No. 15 (problema 9 de [38]).

$$\text{minimizar } (x_1 - 1)^2 + (x_1 - x_2)^2 + (x_2 - x_3)^3 + (x_3 - x_4)^4 + (x_4 - x_5)^4$$

$$\begin{aligned} \text{sujeito a } \quad x_1 + x_2^2 + x_3^3 &= 3\sqrt{2} + 2 \\ x_2 - x_3^2 + x_4 &= 2\sqrt{2} - 2 \\ x_1 x_5 &= 2 \end{aligned}$$

$$x^0 = (2, 2, 2, 2, 2)^t$$

$$x^* = (1.11663472, 1.22044070, 1.53778544, 1.97277049, 1.79109602)^t$$

valor da função objetivo no ótimo: 0.02931083

número de iterações realizadas: 5

número de iterações internas: 5

CODIGO	T.E. (s)	N.A. F.O.	N.A. REST.	N.A. GRAD.	N.A. JAC.	VALOR F.O.	VIOL. REST.
ALG. IV.1	0.09	6	18	6	18	0.0293108	.41E-07
MINOS c/ Newton	0.83	24	24	24	24	0.0293108	.18E-10
MINOS c/ $\rho = 0$	0.77	25	75	25	75	0.0293108	.94E-14
MINOS c/ $\rho = 10$	1.04	34	102	34	102	0.0293108	.10E-10
MINOS c/ $\rho = 100$	1.21	52	156	52	156	0.0293108	.98E-15

Teste No. 15

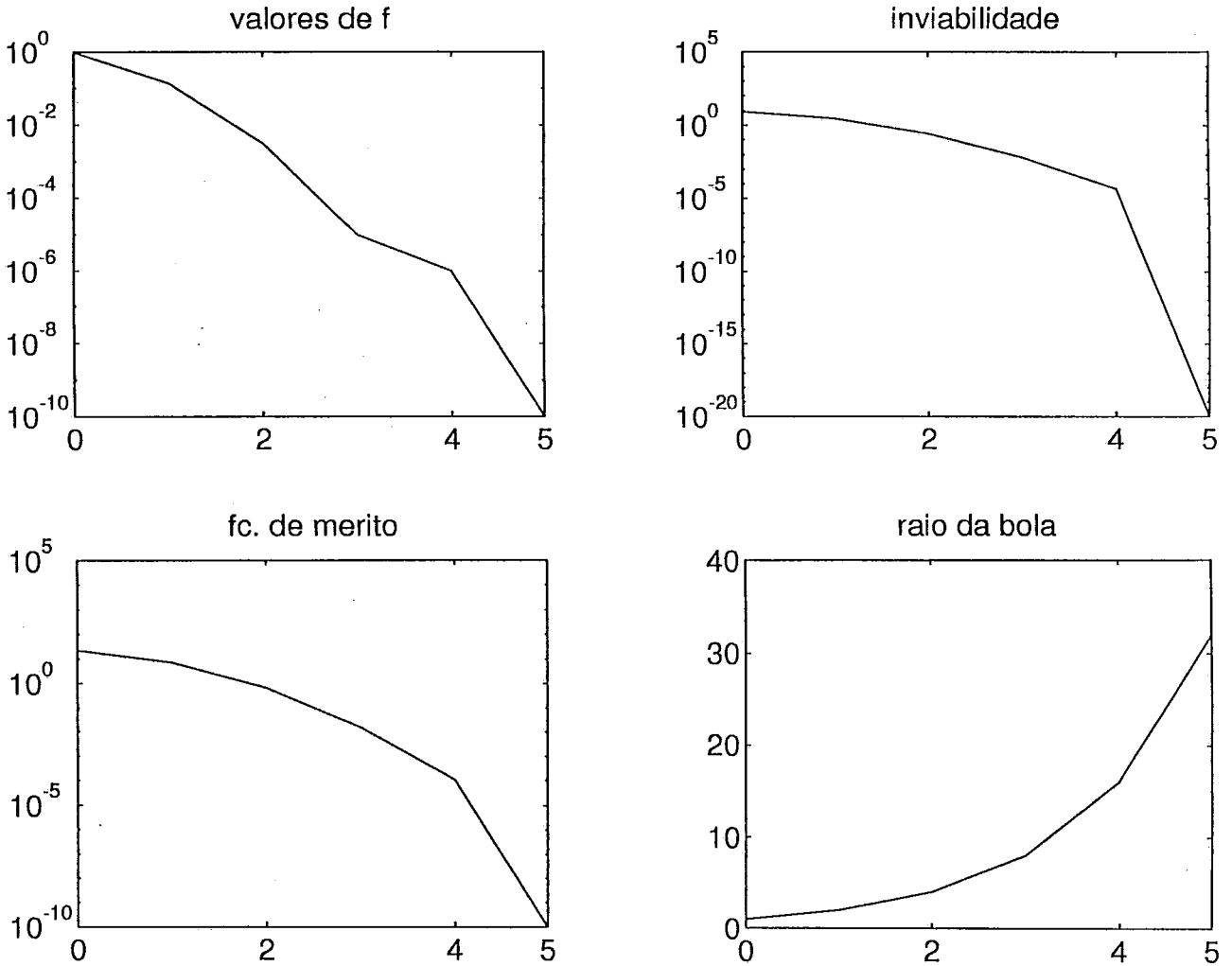


Figura V.15: Problema No. 9 de [38].

Teste No. 16 (problema 9 de [38]).

$$\text{minimizar } (x_1 - 1)^2 + (x_1 - x_2)^2 + (x_2 - x_3)^3 + (x_3 - x_4)^4 + (x_4 - x_5)^4$$

$$\text{sujeito a } x_1 + x_2^2 + x_3^3 = 3\sqrt{2} + 2$$

$$x_2 - x_3^2 + x_4 = 2\sqrt{2} - 2$$

$$x_1 x_5 = 2$$

$$x^0 = (-1, 3, -0.5, -2, -3)^t$$

$$x^* = (-0.70339277, 2.63570260, -0.09636228, -1.79798998, -2.84336160)^t$$

valor da função objetivo no ótimo: 44.02207172

número de iterações realizadas: 4

número de iterações internas: 4

CODIGO	T.E. (s)	N.A. F.O.	N.A. REST.	N.A. GRAD.	N.A. JAC.	VALOR F.O.	VIOL. REST.
ALG. IV.1	0.06	5	15	5	15	44.022074	.25E-06
MINOS c/ Newton	1.53	53	48	53	16	44.022071	.41E-09
MINOS c/ $\rho = 0$	0.93	39	117	39	117	44.022071	.91E-15
MINOS c/ $\rho = 10$	0.83	33	99	33	99	44.022071	.37E-11
MINOS c/ $\rho = 100$	0.94	40	120	40	120	44.022071	.55E-13

Teste No. 16

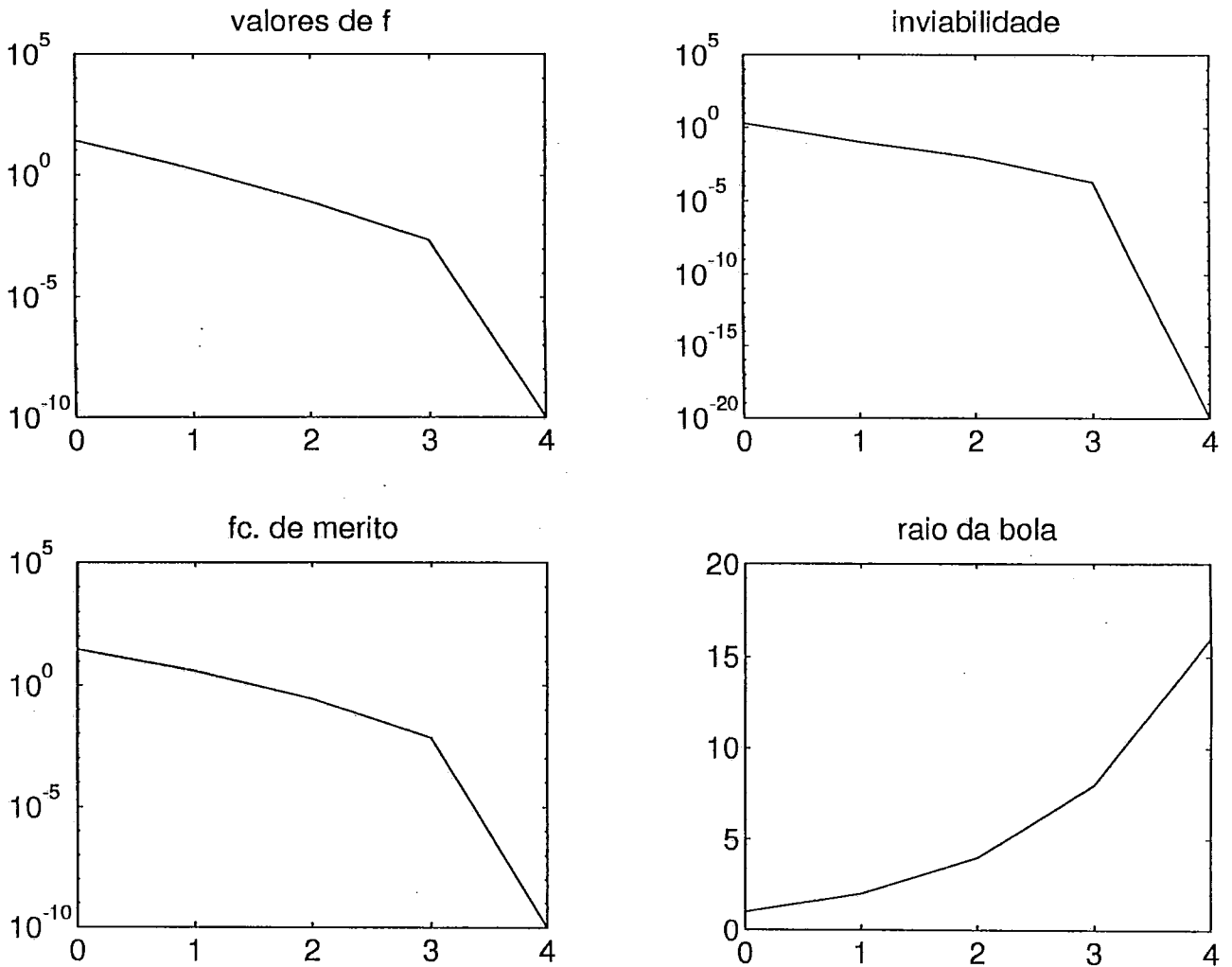


Figura V.16: Problema No. 9 de [38].

Teste No. 17 (problema 9 de [38]).

$$\text{minimizar } (x_1 - 1)^2 + (x_1 - x_2)^2 + (x_2 - x_3)^3 + (x_3 - x_4)^4 + (x_4 - x_5)^4$$

$$\begin{aligned} \text{sujeito a } \quad x_1 + x_2^2 + x_3^3 &= 3\sqrt{2} + 2 \\ x_2 - x_3^2 + x_4 &= 2\sqrt{2} - 2 \\ x_1 x_5 &= 2 \end{aligned}$$

$$x^0 = (-1, 2, 1, -2, -2)^t$$

$$x^* = (-1.27305304, 2.41035430, 1.19485910, -0.15423903, -1.57102645)^t$$

valor da função objetivo no ótimo: 27.87190522

número de iterações realizadas: 5

número de iterações internas: 5

CODIGO	T.E. (s)	N.A. F.O.	N.A. REST.	N.A. GRAD.	N.A. JAC.	VALOR F.O.	VIOL. REST.
ALG. IV.1	0.06	6	18	6	18	27.871906	.59E-06
MINOS c/ Newton	1.38	72	42	72	42	0.0293108	.92E-10
MINOS c/ $\rho = 0$	0.93	23	69	23	69	27.871905	.69E-14
MINOS c/ $\rho = 10$	0.87	27	81	27	81	27.871905	.48E-15
MINOS c/ $\rho = 100$	0.99	33	99	33	99	27.871905	.61E-15

Teste No. 17

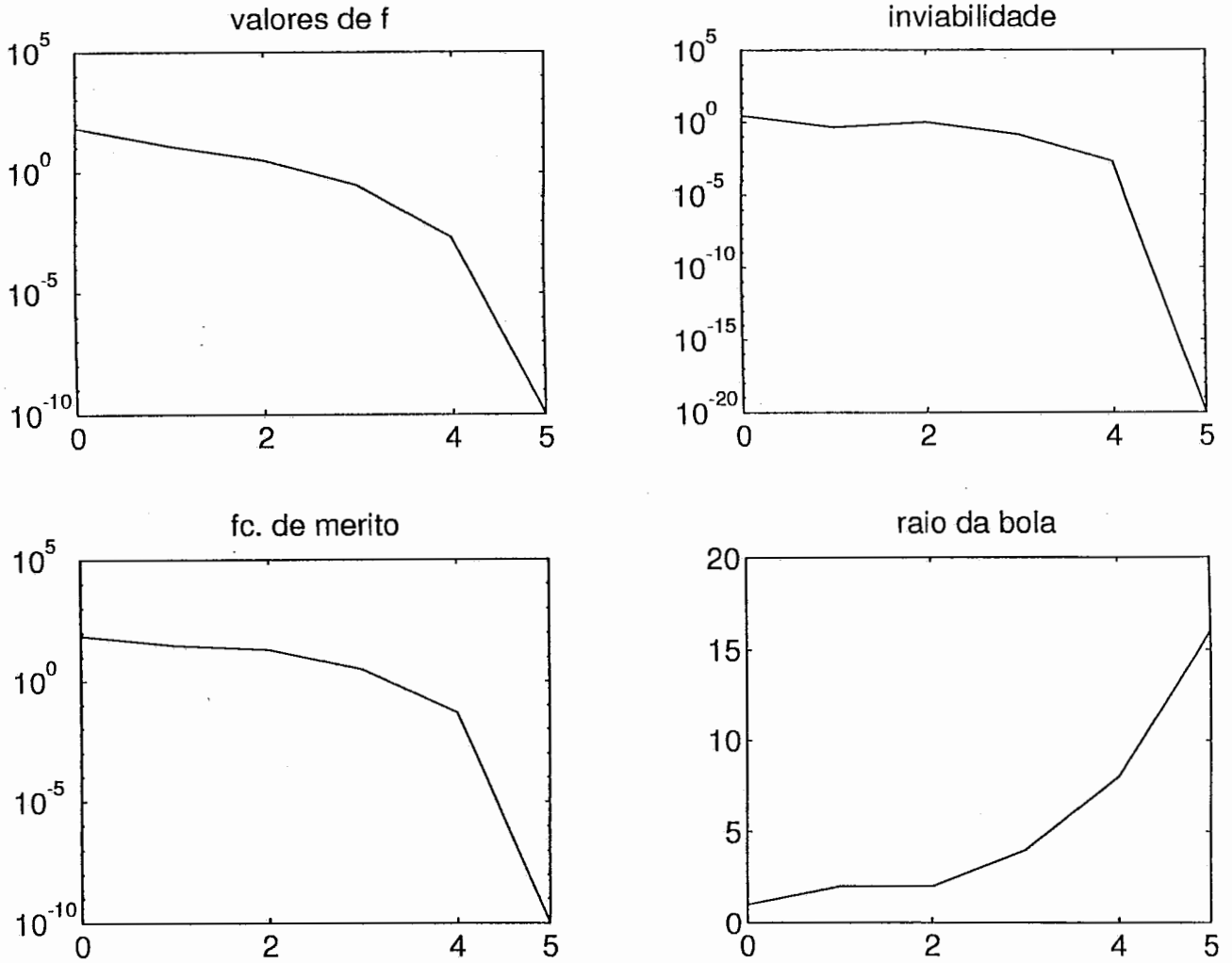


Figura V.17: Problema No. 9 de [38].

Teste No. 18 (problema 9 de [38]).

$$\text{minimizar } (x_1 - 1)^2 + (x_1 - x_2)^2 + (x_2 - x_3)^3 + (x_3 - x_4)^4 + (x_4 - x_5)^4$$

$$\text{sujeito a } x_1 + x_2^2 + x_3^3 = 3\sqrt{2} + 2$$

$$x_2 - x_3^2 + x_4 = 2\sqrt{2} - 2$$

$$x_1 x_5 = 2$$

$$x^0 = (-2, -2, -2, -2, -2)^t$$

$$x^* = (-2.79087124, -3.00413870, 0.20537583, 3.87474506, -0.71662210)^t$$

valor da função objetivo no ótimo: 607.03551529

número de iterações realizadas: 14

número de iterações internas: 17

CODIGO	T.E. (s)	N.A. F.O.	N.A. REST.	N.A. GRAD.	N.A. JAC.	VALOR F.O.	VIOL. REST.
ALG. IV.1	0.19	18	54	15	45	607.035515	.60E-12
MINOS c/ Newton	****	**	**	**	**	*****	*****
MINOS c/ $\rho = 0$	1.48	77	240	77	240	52.9025796	.12E-12
MINOS c/ $\rho = 10$	1.48	88	267	88	267	0.02931083	.27E-11
MINOS c/ $\rho = 100$	0.99	46	138	46	138	607.035515	.25E-11

Neste teste o MINOS 5.1, com a estratégia de Newton, não convergiu.

Teste No. 18

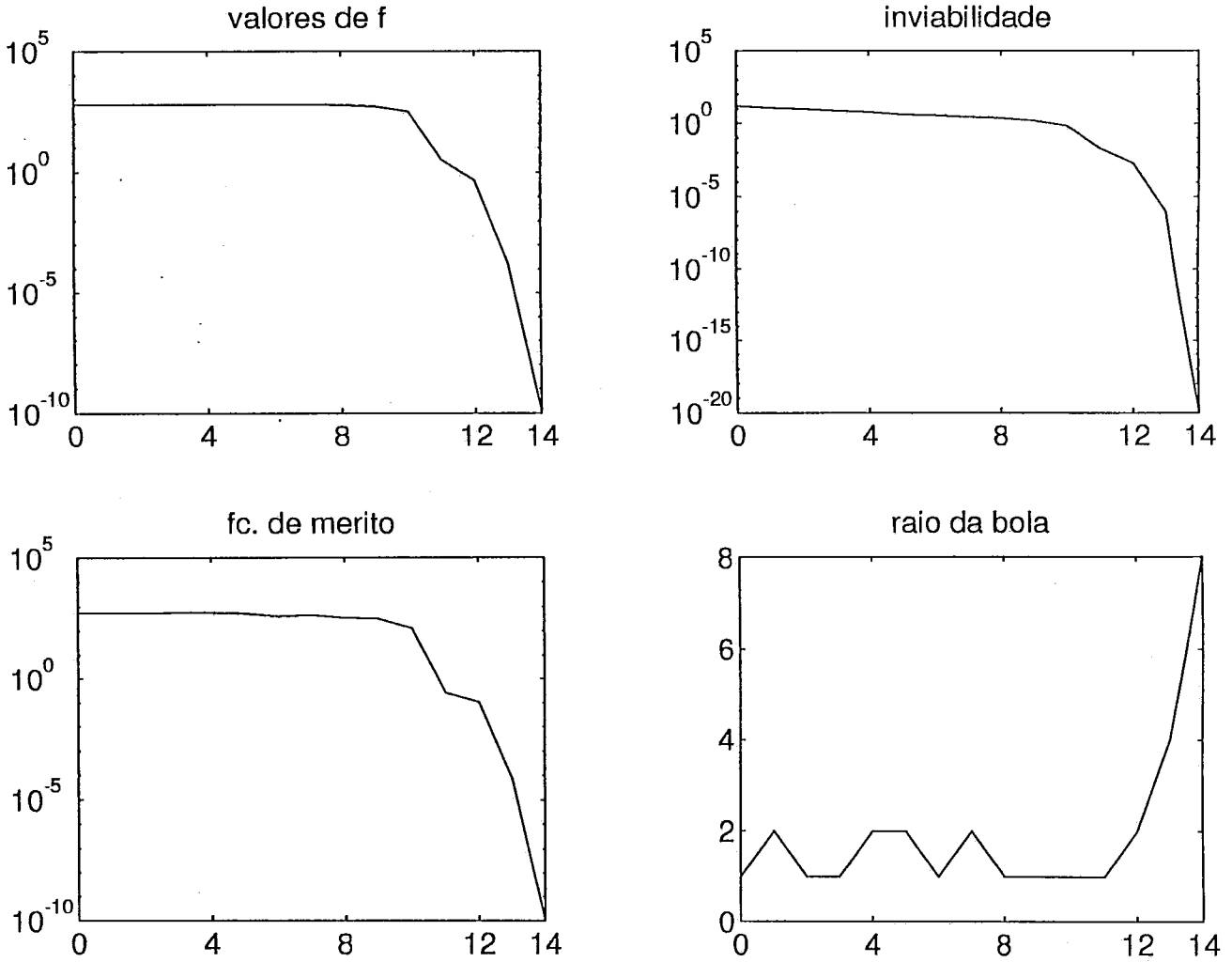


Figura V.18: Problema No. 9 de [38].

O último teste apresentado é outro exemplo altamente não linear. Na resolução deste problema, algoritmos de barreira e outros métodos que mantêm a viabilidade durante a sua execução apresentam grande dificuldade de convergência.

Teste No. 19 (problema 10 de [38]).

$$\text{minimizar } 10x_1x_4 - 6x_3x_2^2 + x_2x_1^3 + 9\text{sen}(x_5 - x_3) + x_5^4x_4^2x_2^3$$

$$\text{sujeito a } x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 \leq 20$$

$$x_1^2x_3 + x_4x_5 \geq -2$$

$$x_2^2x_4 + 10x_1x_5 \geq 5$$

$$x^0 = (1, 1, 1, 1, 1)^t$$

$$x^* = (-0.08145159, 3.69237731, 2.48741142, 0.37713393, 0.17398304)^t$$

valor da função objetivo no ótimo: -210.40801186

número de iterações realizadas: 30

número de iterações internas: 47

CODIGO	T.E. (s)	N.A. F.O.	N.A. REST.	N.A. GRAD.	N.A. JAC.	VALOR F.O.	VIOL. REST.
ALG. IV.1	3.63	48	144	31	93	-210.407821	.75E-06
MINOS c/ Newton	****	***	***	***	***	*****	*****
MINOS c/ $\rho = 0$	****	***	***	***	***	*****	*****
MINOS c/ $\rho = 10$	1.92	152	465	152	465	-210.407817	.73E-14
MINOS c/ $\rho = 100$	1.21	87	261	87	261	-210.407817	.53E-12

Neste teste o MINOS 5.1, com as estratégias de Newton e lagrangiano aumentado com $\rho = 0$, não convergiu.

Teste No. 19

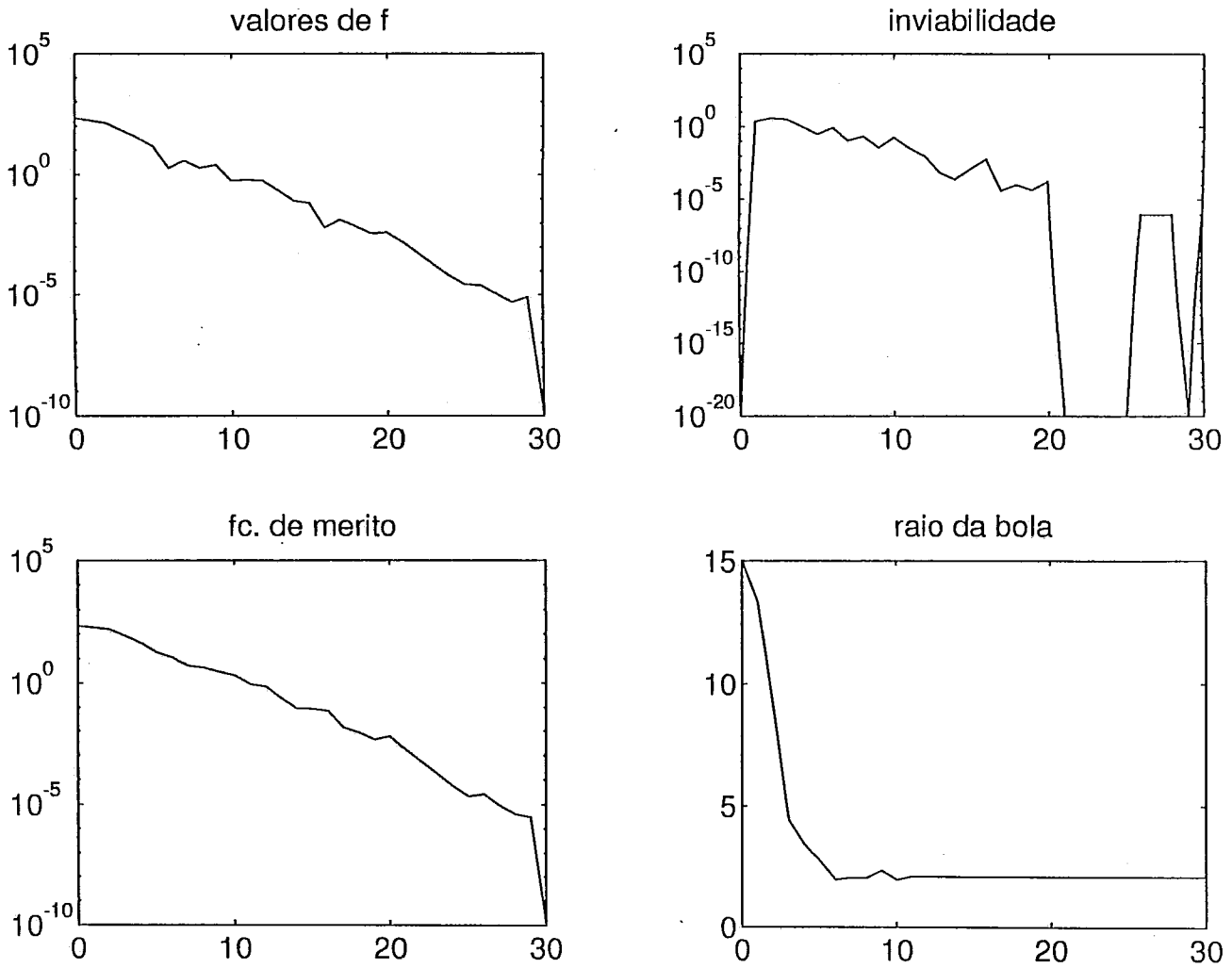


Figura V.19: Problema No. 10 de [38].

Teste No. 20 (problema 10 de [38]).

$$\text{minimizar } 10x_1x_4 - 6x_3x_2^2 + x_2x_1^3 + 9\text{sen}(x_5 - x_3) + x_5^4x_4^2x_2^3$$

$$\text{sujeito a } x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 \leq 20$$

$$x_1^2x_3 + x_4x_5 \geq -2$$

$$x_2^2x_4 + 10x_1x_5 \geq 5$$

$$x^0 = (1.0910, -3.1740, 1.2140, -1.6140, 2.1340)^t$$

$$x^* = (1.47963308, -2.63661310, 1.05466824, -1.61151320, 2.67388423)^t$$

valor da função objetivo no ótimo: -2500.584086

número de iterações realizadas: 16

número de iterações internas: 17

CODIGO	T.E. (s)	N.A. F.O.	N.A. REST.	N.A. GRAD.	N.A. JAC.	VALOR F.O.	VIOL. REST.
ALG. IV.1	0.41	18	54	17	51	-2500.584086	.28E-11
MINOS c/ Newton	****	***	***	***	***	*****	*****
MINOS c/ $\rho = 0$	2.53	184	570	184	570	-5550.230385	.0
MINOS c/ $\rho = 10$	1.65	135	511	135	511	-2500.584488	.64E-12
MINOS c/ $\rho = 100$	0.93	51	153	51	153	-2500.584488	.91E-13

Neste teste o MINOS 5.1, com a estratégia de Newton, não convergiu.

Teste No. 20

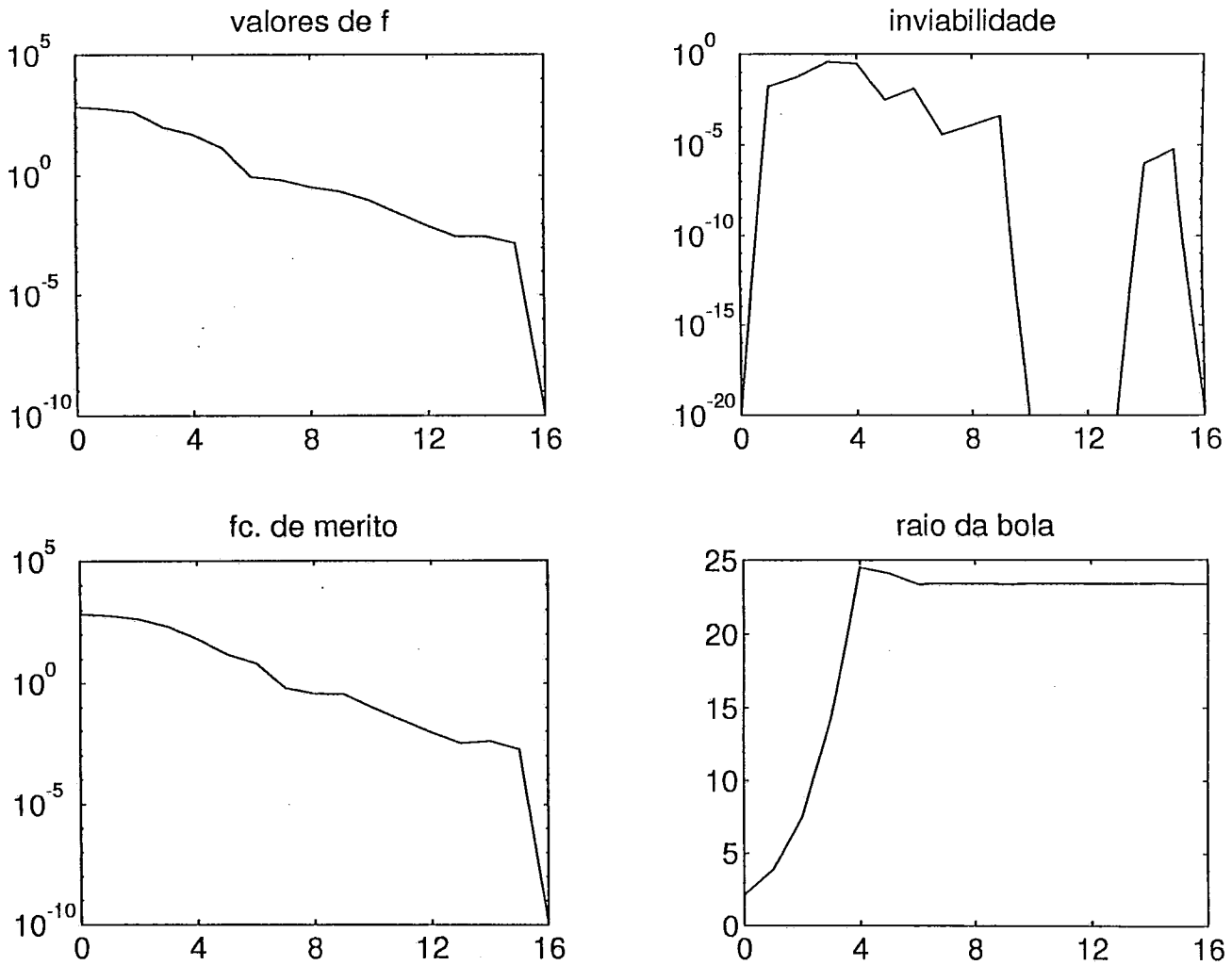


Figura V.20: Problema No. 10 de [38].

Capítulo VI

Conclusões

Neste trabalho, apresentamos primeiramente um novo algoritmo de região de confiança, para resolver problemas de minimização de funções não lineares sujeitas somente a restrições de caixa. Descrevemos a região de confiança utilizada, a qual tem um formato elíptico, distintamente dos outros métodos que utilizam uma região de confiança esférica e apresentamos os resultados de convergência global. Logo, generalizamos o algoritmo apresentado, para poder resolver o problema de programação não linear quando temos tanto restrições não lineares de igualdade como restrições de caixa. Mostramos como é feito o cálculo das estimativas dos multiplicadores de Karush-Kuhn-Tucker e a escolha da função de mérito e também, para este caso apresentamos os resultados de convergência global.

Apresentamos também os resultados numéricos obtidos com a implementação do algoritmo proposto no capítulo IV. Para tal fim, foram realizados vinte testes computacionais. Os resultados obtidos nos primeiros treze testes realizados foram comparados com os obtidos pelos seis códigos de otimização apresentados em [30]. Os resultados dos últimos sete testes foram comparados com os obtidos pelo código de otimização MINOS 5.1. Este estudo comparativo mostrou que o algoritmo teve um bom desempenho na prática, porém em alguns testes observou-se uma certa lentidão na convergência do algoritmo, quando os pontos gerados pelo algoritmo estavam perto do ótimo. Embora tivéssemos que exigir a convexidade da função objetivo para demonstrar a convergência global no caso de problemas só com restrições de caixa, entre os problemas resolvidos, estavam problemas altamente não lineares e não convexos, o que mostrou a robustez do algoritmo proposto.

Alguns tópicos que poderiam ser abordados e pesquisados em futuros trabalhos são:

- Estudar a utilização de uma aproximação da matriz hessiana da função objetivo, no algoritmo apresentado no capítulo III.
- Estudar a convergência global do algoritmo do capítulo III, sem a hipótese da função objetivo ser convexa.
- Fazer uma implementação eficiente do algoritmo do capítulo III, explorando o fato do problema a ser resolvido ter somente restrições de caixa.
- Para poder demonstrar a convergência global no capítulo IV, foi preciso impor duas hipóteses fortes: que o raio da região de confiança não tende a zero e que a seqüência gerada pelo algoritmo é convergente. Um tema de uma futura pesquisa seria achar hipóteses menos restritivas que garantam a convergência global do algoritmo proposto.
- Estudar a possível ocorrência do efeito Maratos no algoritmo do capítulo IV, devido à não diferenciabilidade da função de mérito ϕ_μ e o uso de técnicas para evitar tal efeito, como uma correção de segunda ordem no passo gerado pelo algoritmo.
- Fazer uma implementação mais eficiente do algoritmo do capítulo IV, de maneira a permitir uma convergência mais rápida quando os pontos da seqüência gerada pelo algoritmo estiverem perto do ponto ótimo.

Referências Bibliográficas

- [1] BERTSEKAS, D., Projected Newton Methods for Optimization Problems with Simple Constraints, Siam Journal on Control and Optimization, Vol. 20, No. 2, March 1982, 221-246.
- [2] BONNANS, J. and M. BOUHTOU, The Trust Region Affine interior Point Algorithm for Convex and Nonconvex Quadratic Programming, Technical Report, INRIA, Rocquencourt, France, July 1993.
- [3] BYRD, R., R. SCHNABEL and G. SHULTZ, A Trust Region Algorithm for Nonlinearly Constrained Optimization, Siam Journal on Numerical Analysis, Vol. 24, No. 5, October 1987, 1152-1170.
- [4] BYRD, R., R. SCHNABEL and G. SHULTZ, Approximate Solutions of the Trust Region Problem by Minimization over Two-Dimensional Subspaces, Mathematical Programming, Vol. 40, 1988, 247-263.
- [5] CELIS, M., J. DENNIS and R. TAPIA, A Trust Region Strategy for Nonlinear Equality Constrained Optimization, in P. BOGGS, R. BYRD and R. SCHNABEL editors Numerical Optimization 1984, SIAM, Philadelphia, Pennsylvania, 1985, 71-82.
- [6] CHAMBERLAIN, R., C. LEMARECHAL, H. PEDERSEN and M. POWELL, The Watchdog Technique for Forcing Convergence in Algorithms for Constrained Optimization, in A. G. BUCKLEY and J. L. GOFFIN editors Algorithms for Constrained Minimization of Smooth Nonlinear Functions, Mathematical Programming Study 16, North Holland, Amsterdam, 1982.

- [7] COLEMAN, T. and A. CONN, Nonlinear Programming via an Exact Penalty Function: Asymptotic Analysis, Mathematical Programming, Vol. 24, 1982, 123-136.
- [8] COLEMAN, T. and A. CONN, Nonlinear Programming via an Exact Penalty Function: Global Analysis, Mathematical Programming, Vol. 24, 1982, 137-161.
- [9] COLEMAN, T. and Y. LI, An Interior Trust Region Approach for Nonlinear Minimization Subject to Bounds, Technical Report TR93-1342, Department of Computer Science, Cornell University, Ithaca, New York, 1993.
- [10] CONN, A., N. GOULD and PH. TOINT, Global Convergence of a Class of Trust Region Algorithms for Optimization with Simple Bounds, Siam Journal on Numerical Analysis, Vol. 25, No. 2, April 1988, 433-460.
- [11] DENNIS, J., M. EL-ALEM and M. MACIEL, A Global Convergence Theory for General Trust-Region-Based Algorithms for Equality Constrained Optimization, Technical Report TR92-28, Department of Mathematical Sciences, Rice University, Houston, Texas, 1992.
- [12] DENNIS, J. and R. SCHNABEL, Numerical Methods for Unconstrained Optimization and Nonlinear Equations, Prentice-Hall, Englewood Cliffs, N.J., 1983.
- [13] DENNIS, J. and R. SCHNABEL, A View of Unconstrained Optimization, in G. L. NEMHAUSER, A. H. G. RINNOOY KAN and M. J. TODD editors Handbooks in Operations Research and Management Science, Elsevier Science Publishers B. V., New York, N.Y., Vol. 1, 1989, 1-72.
- [14] DENNIS, J. and L. VICENTE, Trust-Region Interior-Point Algorithms for Minimization Problems with Simple Bounds, Technical Report TR94-42, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, 1994.
- [15] DENNIS, J., M. HEINKENSCHLOSS and L. VICENTE, Trust-Region Interior-Point SQP Algorithms for a Class of Nonlinear Programming Problems, Technical Report TR94-45, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, 1994.

- [16] FLETCHER, R., An l_1 Penalty Method for Nonlinear Constraints, in P. BOGGS, R. BYRD and R. SCHNABEL editors Numerical Optimization 1984, SIAM, Philadelphia, Pennsylvania, 1985, 26-40.
- [17] FLETCHER, R., Practical Methods of Optimization, John Wiley and Sons, New York, N.Y., second edition, 1987.
- [18] FLETCHER, R., Second Order Corrections for Nondifferentiable Optimization in G. A. WATSON editors Numerical Analysis Proceedings, Dundee 1981, Lecture Notes in Mathematics 912, Springer-Verlag, Berlin, 1982.
- [19] FLETCHER, R., An Exact Penalty Function for Nonlinear Programming with Inequalities, Mathematical Programming, Vol. 5, 1973, 129-150.
- [20] GAY, D., Subroutines for Unconstrained Minimization Using a Model/Trust-Region approach, ACM Transactions on Mathematical Software, Vol. 9, 1983, 503-524.
- [21] GILL, P., W. MURRAY and M. WRIGHT, Practical Optimization, Academic Press Inc, New York, N.Y., 1981.
- [22] GILL, P., W. MURRAY, M. SAUNDERS and M. WRIGHT, Constrained Nonlinear Programming, in G. L. NEMHAUSER, A. H. G. RINNOOY KAN and M. J. TODD editors Handbooks in Operations Research and Management Science, Elsevier Science Publishers B. V., New York, N.Y., Vol. 1, 1989, 171-210.
- [23] GOLDFELDT S., R. QUANDT and H. TROTTER, Maximization by Quadratic Hill-Climbing, Econometrica, Vol. 34, 1966, 541-551.
- [24] GONZAGA, C., An Interior Trust Region Method for Linearly Constrained Optimization, Committee on Algorithms Newsletter, No. 19, August 1991, 55-65.
- [25] GONZAGA, C., Algoritmos de Pontos Interiores para Programação linear, 17° Colóquio Brasileiro de Matemática, IMPA, CNPq, Rio de Janeiro, 1989.
- [26] GONZAGA, C., Path-Following Methods for Linear Programming, Siam Review, Vol. 34, No. 2, June 1992, 167-224.

- [27] GONZAGA, C. and L. CARLOS, A Primal Affine-Scaling Algorithm for Linearly Constrained Convex Programs, Internal Report ES-238/90, COPPE - Universidade Federal do Rio de Janeiro, Rio de Janeiro, R.J., 1990.
- [28] HAN, S., A Globally Convergent Method for Nonlinear Programming, J. Optimization Theory and Applications, No. 22, 1977, 297-310.
- [29] HESTENES, M., Multiplier and Gradient Methods, J. Optimization Theory and Applications, No. 4, 1969, 303-320.
- [30] HOCK, W. and K. SCHITTKOWSKI, Test Examples for Nonlinear Programming Codes, Lecture Notes in Economics and Mathematical Systems 187, Springer-Verlag, Berlin, 1981.
- [31] KUHN, H. and A. TUCKER, Nonlinear Programming, Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, California, 1951, 481-492.
- [32] LEVENBERG, K., A Method for the Solution of Certain Problems in Least Squares, Quarterly Journal of Applied Mathematics, Vol. 2, 1944, 164-168.
- [33] MARQUARDT, D., An Algorithm for Least-Squares Estimation of Nonlinear Parameters, Siam Journal on Applied Mathematics, Vol. 11, 1963, 431-441.
- [34] McCORMICK, G., Second Order Conditions for Constrained Minima, Siam Journal on Applied Mathematics, Vol. 15, No. 3, May 1967, 641-652.
- [35] MORÉ, J., Recent Developments in Algorithms and Software for Trust Region Method, in A. BACHEM, M. GROTSCHHEL and B. KORTE editors Mathematical Programming. The state of the Art, Springer-Verlag, New York, N.Y., 1983, 258-287.
- [36] MORÉ, J. and D. SORENSEN, Computing a Trust Region Step, Siam Journal Sci. Stat. Comput., Vol. 4, No. 3, September 1983, 553-572.
- [37] MORÉ, J., B. GARBOW and K. HILLSTROM, Fortran Subroutines for Testing Unconstrained Optimization Software, ACM Transactions on Mathematical Software, Vol. 7, 1981, 136-140.

- [38] MURTAGH, B. and M. SAUNDERS, A Projected Lagrangian Algorithm and its Implementation for Sparse Nonlinear Constraints, Mathematical Programming Study, No. 16, 1982, 84-117.
- [39] MURTAGH, B. and M. SAUNDERS, MINOS 5.1 User's Guide, Technical Report SOL 83-20R, Stanford University, Stanford, California, Revised January 1987.
- [40] POWELL, M., Convergence Properties of a Class of Minimization Algorithms, in O. MANGASARIAN, R. MEYER and S. ROBINSON editors Nonlinear Programming 2, Academic Press Inc, New York, N.Y., 1975.
- [41] POWELL, M., A Method for Nonlinear Constraints in Minimization Problems, in R. FLETCHER editor Optimization, Academic Press Inc, New York, N.Y., 1969, 283-298.
- [42] POWELL, M., The Convergence of Variable Metric Methods for Nonlinearly Constrained Optimization Calculations, in O. MANGASARIAN, R. MEYER and S. ROBINSON editors Nonlinear Programming 3, Academic Press Inc, New York, N.Y., 1978, 27-63.
- [43] POWELL, M. and Y. YUAN, A Trust Region Algorithm for Equality Constrained Optimization, Report DAMTP 1986-NA2, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, England, 1986.
- [44] SAGARA, N. and M. FUKUSHIMA, A Hybrid Method for the Nonlinear Least Squares Problem with Simple Bounds, Journal of Computational and Applied Mathematics, Vol. 36, 1991, 149-157.
- [45] SCHITTKOWSKI, K., Nonlinear Programming Codes - Information, tests, performance, Lecture Notes in Economics and Mathematical Systems 183, Springer-Verlag, Berlin, 1980.
- [46] SCHITTKOWSKI, K., The Nonlinear Programming Method of Wilson, Han and Powell with an Augmented Lagrangian Type Line Search function, Numerische Mathematik, Vol. 38, 1981, 83-114.

- [47] SCHNABEL R., J. KOONTZ and B. WEISS, A Modular System of Algorithms of Unconstrained Minimization, ACM Transactions on Mathematical Software, Vol. 11, 1985, 419-440.
- [48] SHULTZ, G., R. SCHNABEL and R. BYRD, A Family of Trust Region-Based Algorithms for Unconstrained Minimization with Strong Global Convergence Properties, Siam Journal on Numerical Analysis, Vol. 22, No. 1, February 1985, 47-67.
- [49] VARDI, A., A Trust Region Algorithm for Equality Constrained Minimization: Convergence Properties and Implementation, Siam Journal on Numerical Analysis, Vol. 22, No. 3, June 1985, 575-591.
- [50] WRIGHT, M., Numerical Methods for Nonlinearly Constrained Optimization, Ph.D. Thesis, Stanford University, Stanford, California, 1976.