

Penalização Exata com Subproblemas Restritos

Silvia Martini de Holanda Janesch

**Tese apresentada ao Curso de Doutorado
do Instituto de Matemática, Estatística e
Computação Científica da Universidade
Estadual de Campinas, como requisito
parcial para a obtenção do Título de
Doutora em Matemática Aplicada.**

Área: Otimização

Orientador: Prof. Dr. José Mário Martínez

Co-Orientador: Prof. Dr. Lúcio T. Santos

**Campinas
IMECC - UNICAMP
1998**

9903303

UNIDADE	BC
N.º CHAMADA:	
V. Ex.	
TOMADA	BC/36268
PRON.	229/99
C. D.	<input checked="" type="checkbox"/>
PREÇO	R\$ 11,00
DATA	23/01/99
N.º CPU	

CM-00120435-1

**FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO IMECC DA UNICAMP**

Janesch, Silvia Martini de Holanda

J254p

Penalização exata com subproblemas restritos / Silvia Martini de Holanda Janesch -- Campinas, [S.P. :s.n.], 1998.

Orientador : José Mário Martínez, Lúcio T. Santos

Tese (doutorado) - Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

I. Otimização matemática. 2. Programação não-linear. I. Martínez, José Mário. II. Santos, Lúcio Túnes. III. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. VI. Título.

Penalização Exata com Subproblemas Restritos

Este exemplar corresponde à redação final da tese devidamente corrigida e defendida por Silvia Martini de Holanda Janesch e aprovada pela comissão julgadora.

Campinas, 28 de outubro de 1998



**Prof. Dr. José Mário Martínez
Orientador**

Tese apresentada ao Instituto de Matemática Estatística e Computação Científica da Universidade Estadual de Campinas, como requisito parcial para a obtenção do Título de Doutora em Matemática Aplicada.

Tese de Doutorado defendida e aprovada em 28 outubro de 1998

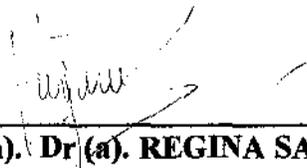
Pela Banca Examinadora composta pelos Profs. Drs.



Prof (a). Dr (a). JOSÉ MÁRIO MARTÍNEZ PÉREZ



Prof (a). Dr (a). NELSON MACULAN FILHO



Prof (a). Dr (a). REGINA SANDRA BURACHIK



Prof (a). Dr (a). MÁRCIA APARECIDA GOMES RUGGIERO



Prof (a). Dr (a). JOÃO FREDERICO DA COSTA AZEVEDO MEYER

Resumo

Apresentamos resultados gerais de penalização externa e exata. Estendemos o teorema clássico de penalização exata para o caso onde os subproblemas penalizados permanecem restritos. Introduzimos um algoritmo para resolver problemas de programação não linear baseado na função de penalização exata L_1 , onde penalizamos somente as restrições não lineares. Para resolver os subproblemas penalizados não suaves desenvolvemos um algoritmo de região de confiança. Ilustramos o método de penalização com região de confiança através de exemplos simples. Testes numéricos comparando o método de penalização com região de confiança com o algoritmo BOX-QUACAN foram efetuados em 3 conjuntos de problemas. Abordamos o problema global de Lennard-Jones e propomos gerar bons pontos iniciais para este problema usando a solução de um subproblema restrito.

Abstract

We present the classical results for the exact and the exterior penalty problems. We extend the classic exact penalty function theorem for the case where the penalty subproblems remain constrained. We introduce an algorithm for solving nonlinear programming problems based on the L_1 exact penalty function for which only the nonlinear constraints are penalized. For solving the nonsmooth penalty subproblems we develop a trust region algorithm. We illustrate the penalty method with trust region with simple examples. Numerical experiments comparing the penalty method with BOX-QUACAN algorithm were realized in three sets of problems. We attack Lennard Jones's global problem and we propose to generate good starting points for this problem using the solution of a constrained subproblem.

AGRADECIMENTOS

À FAPESP, pelo apoio financeiro durante o curso.

Ao Professor Martínez, pela sua orientação e dedicação durante a elaboração deste trabalho.

Ao Professor Lúcio, pela sua orientação e motivação na parte teórica do trabalho.

Ao meu esposo Oscar, e ao meu filho Ricardo, pela compreensão e incentivo.

Índice

Introdução	i
1 Penalização externa e penalização exata	1
1.1 Introdução	1
1.2 Resultados gerais de penalização	5
1.3 Exatidão de $P(x, \rho)$	15
2 Método de penalização com subproblemas restritos	23
2.1 Introdução	23
2.2 Algoritmo principal	24
3 Visualização através de problemas simples	38

4 Aplicações e resultados numéricos	58
4.1 Formulação dos problemas	58
4.2 Experimentos numéricos	62
4.3 Comentários	70
5 Alguns problemas de minimização global	71
5.1 Introdução	71
5.2 Problema de Lennard-Jones	73
5.3 Problema tipo Lennard-Jones	76
5.4 Comentários	81
Conclusões	82
Bibliografia	84

Introdução

Problemas de programação não linear (PNL) aparecem em muitos problemas da vida real. São os casos, por exemplo, de problemas de fluxo de potência ótimo em redes de transmissão de energia elétrica, problemas de controle e problemas de alocação de recursos.

O problema em que estamos interessados é

$$\begin{array}{ll} \text{minimizar} & f(x) \\ \text{sujeita a} & h(x) = 0 \\ & g(x) \leq 0 \\ & x \in \Omega, \end{array} \quad (0.1)$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $h : \mathbb{R}^n \rightarrow \mathbb{R}^{m_1}$, $g : \mathbb{R}^n \rightarrow \mathbb{R}^{m_2}$, $f, h, g \in C^2$ e Ω representa um conjunto de restrições simples, como por exemplo uma caixa, uma bola ou um polítopo.

Existem várias formas de resolver o problema de programação não linear. Uma maneira é aplicar, por exemplo, Métodos de Direções Factíveis. A filosofia destes métodos é mover de um ponto factível para um ponto factível melhor (ponto que reduz o valor da função objetivo). Outra forma de resolver o problema (0.1) é aplicar os Métodos de Lagrange. Estes são baseados na resolução de um sistema de equações e (possivelmente) inequações que constituem as condições necessárias de otimalidade para o problema de otimização. Uma terceira aproximação para resolver (PNL) são os métodos de penalização, baseados na eliminação de restrições do problema original usando funções de penalização. Os métodos clássicos de penalização consistem em resolver (PNL) através de uma seqüência irrestrita de subproblemas penalizados nos

quais, para garantir convergência global dos métodos, é exigido que a seqüência de parâmetros penalizadores cresça ilimitadamente. Assim, uma seqüência infinita de subproblemas deve ser resolvida.

Uma forma de evitar o processo seqüencial infinito dos métodos de penalização clássicos é usar as funções de penalização exata. Neste caso o (PNL) é resolvido por meio de um único subproblema de minimização irrestrita ou uma seqüência finita de subproblemas irrestritos. Infelizmente a maioria das funções de penalização exata são não diferenciáveis e técnicas convencionais de otimização suave irrestrita não podem ser aplicadas para resolver os subproblemas. Em geral, a teoria de penalização exata está desenvolvida para o caso onde todo o conjunto de restrições do problema original é penalizado, e assim os subproblemas envolvidos são irrestritos. Neste trabalho resolvemos o (PNL) usando uma função de penalização exata não suave, onde penalizamos somente as restrições não lineares de forma que os subproblemas permanecem restritos, e tratamos o problema da não diferenciabilidade através da técnica de região de confiança.

Os objetivos do trabalho são exibir a teoria de penalização exata para o caso onde os subproblemas permanecem restritos, apresentar um algoritmo para resolver o (PNL) baseado na função de penalização exata não suave L_1 , onde penalizamos somente as restrições não lineares, e verificar a consistência do algoritmo proposto através da comparação com o algoritmo BOX (Lagrangeano aumentado com restrições tipo caixa) desenvolvido por Friedlander-Martínez e Santos. Este algoritmo tem sofrido várias comparações com o pacote LANCELOT de Coon-Gould-Toint e se mostrou competitivo.

No Capítulo 1 apresentamos os principais resultados da teoria de penalização e fornecemos novos resultados da teoria. Entre eles, mostramos que o teorema clássico da função de penalização exata para o caso irrestrito permanece válido para o caso onde mantemos os subproblemas penalizados restritos ao conjunto Ω usando condições de 2ª ordem. A importância de tal resultado é que podemos desenvolver algoritmos combinando as vantagens de explorar a estrutura simples do conjunto Ω e usar as funções de penalização exata.

No Capítulo 2 apresentamos um método para resolver o problema (0.1). A idéia do método é aproveitar a estrutura do conjunto Ω e aplicar a função de penalização

exata L_1 às restrições não lineares. Para resolver os subproblemas penalizados não suaves desenvolvemos um algoritmo de região de confiança onde são apresentadas várias funções modelo para aproximar a função penalizada não suave. As funções modelo são também não suaves com a propriedade de conter um número menor de pontos não diferenciáveis. A quantidade de pontos não diferenciáveis das funções modelo está associada às constantes de tolerância usadas nas definições dos modelos.

No Capítulo 3 ilustramos o método proposto (método de penalização com região de confiança) através de problemas simples em que são exibidos os subproblemas em cada iteração do método. Em particular, para dois problemas consideramos diferentes valores para as constantes de tolerância e analisamos o comportamento dos subproblemas envolvidos no algoritmo de região de confiança.

No Capítulo 4 apresentamos os resultados numéricos da comparação entre o método de penalização com região de confiança e o algoritmo BOX. Foram testados três conjuntos de problemas. Um conjunto envolveu restrições não lineares de desigualdade e igualdade, e restrições de canalização. Os outros dois conjuntos envolveram somente restrições não lineares de desigualdade.

No Capítulo 5, motivados pelos resultados do algoritmo BOX na versão com desigualdade, abordamos o problema de configuração de pontos (descobrir localização de pontos no plano ou espaço) baseado na função potencial de Lennard-Jones. A heurística proposta para resolver o problema consiste em usar a solução de um problema de programação não linear com restrições de desigualdade como ponto inicial. Testes numéricos usando o algoritmo BOX e o algoritmo de penalização com região de confiança são descritos.

Finalmente apresentamos as conclusões do trabalho e sugestões para pesquisas futuras.

Capítulo 1

Penalização Externa e Penalização Exata

1.1 Introdução

Vários problemas da vida real podem ser formulados como problemas de programação não linear. Por exemplo, problemas de controle ótimo, projeto estrutural, projeto mecânico, redes elétricas, administração de recursos hidráulicos, alocação estocástica de recursos. O trabalho de Bongartz-Conn-Gould-Toint [4] (CUTE) relaciona uma coleção de problemas práticos e acadêmicos, oriundos de diversas áreas e aplicações, cuja formulação conduz a um problema de programação não linear.

O problema de programação não linear (PNL) consiste na minimização de uma função objetivo sujeita a um conjunto de restrições. Usualmente este conjunto de restrições envolve igualdades e desigualdades. Muitas vezes é possível tirar vantagens da estrutura das restrições e desenvolver algoritmos eficientes. É o caso, por exemplo, de restrições do tipo caixa, bolas ou restrições lineares.

O Problema

O problema em que estamos interessados é o seguinte:

$$\begin{aligned} & \text{minimizar} && f(x) \\ & \text{sujeita a} && h(x) = 0 \\ & && g(x) \leq 0 \\ & && x \in \Omega, \end{aligned} \tag{1.1}$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $h : \mathbb{R}^n \rightarrow \mathbb{R}^{m_1}$, $g : \mathbb{R}^n \rightarrow \mathbb{R}^{m_2}$, $f, h, g \in C^2$ e Ω representa um conjunto de restrições simples, como por exemplo uma caixa, uma bola ou um politopo.

Existem vários métodos disponíveis para resolver (PNL). O trabalho recente de Conn-Gould-Toint [10] fornece uma visão de aproximações atuais para resolver o (PNL), incluindo métodos interiores e exteriores, algoritmos baseados em região de confiança e busca linear.

Os métodos de penalização externa, usualmente referidos somente como métodos de penalização, consistem em resolver problemas de programação não linear através da resolução de uma seqüência de problemas irrestritos ou com restrições simples. As restrições são colocadas na função objetivo via um parâmetro de penalização, ρ , de forma a punir qualquer violação das restrições. Em geral, os métodos de penalização externa geram uma seqüência de pontos infactíveis cujo limite (se existir) é uma solução do problema original. Na maioria dos métodos de penalização, a factibilidade é somente atingida na solução. Porém existem funções de penalização que geram uma seqüência de pontos que podem estar no interior ou no exterior da região factível (ver [16, 34]).

As funções de penalização que aparecem com freqüência na literatura de otimização são: a Função de Penalização de Perda Quadrática, $P_2(x, \rho) = f(x) + \rho (\|h(x)\|_2^2 + \|g^+(x)\|_2^2)$, e a Função de Penalização L_1 , $P_1(x, \rho) = f(x) + \rho (\|h(x)\|_1 + \|g^+(x)\|_1)$, onde $\rho > 0$ é o parâmetro de penalização e $g_j^+(x) = \max\{0, g_j(x)\}$, $j = 1, \dots, m_2$.

Na teoria desenvolvida para os métodos de penalização clássicos, a seqüência

de parâmetros penalizadores deve ser ilimitada para podermos garantir convergência global. Dificuldades computacionais de mal condicionamento aparecem quando temos que resolver problemas penalizados para valores grandes de ρ . Uma alternativa para contornar esta situação é a aproximação discutida por Martínez e Santos [32] para $P_2(x, \rho)$. A idéia do trabalho é determinar uma direção de Newton, usando um procedimento que introduz novas variáveis no sistema não linear dado pelas condições de otimalidade de primeira ordem para um minimizador local de $P_2(x, \rho)$, isto é, $\nabla P_2(x, \rho) = 0$. Outra dificuldade dos métodos de penalização é a suposição de que minimizadores globais dos problemas penalizados devam ser calculados. Na prática, não obtemos minimizadores globais. Os algoritmos usualmente geram uma seqüência de minimizadores locais, ou simplesmente pontos estacionários.

Outra classe de métodos que resolve o (PNL) são os métodos de penalização exata. A filosofia destes métodos é resolver o (PNL) por meio de um único problema de minimização irrestrita. Ligeiramente falando, uma função de penalização exata para o problema (1.1) é uma função $P(x, \rho)$, onde $\rho > 0$ é o parâmetro de penalização, com a propriedade de que existe um limite inferior $\bar{\rho} > 0$ tal que para $\rho > \bar{\rho}$ qualquer minimizador local do (PNL) é também um minimizador local do problema penalizado. Definição mais rigorosa de função de penalização exata pode ser encontrada em [11]. As funções de penalização exata podem ser divididas em duas classes: funções de penalização exata continuamente diferenciáveis e funções de penalização exata não diferenciáveis. As funções de penalização continuamente diferenciáveis foram introduzidas por Fletcher [17] para problemas com restrições de igualdade e por Glad e Polak [20] para problemas com restrições de desigualdade; contribuições adicionais podem ser encontradas em Di Pillo e Grippo [11, 12]. As funções de penalização não diferenciáveis foram introduzidas por Zangwill [38] e Pietrzykowski [36], e várias pesquisas tem sido desenvolvidas neste sentido (ver [7, 8, 9, 13, 18]).

Uma classe de funções de penalização exata não diferenciável associada à (1.1) para $\Omega = \mathbb{R}^n$ foi analisada por Charalambous [7] em 1978,

$$P_p(x, \alpha, \beta) = f(x) + \left(\sum_{i=1}^{m_1} [\alpha_i |h_i(x)|]^p + \sum_{j=1}^{m_2} [\beta_j g_j^+(x)]^p \right)^{\frac{1}{p}} \quad (1.2)$$

onde $p \geq 1$, $\alpha_i, \beta_j > 0$, $i = 1, \dots, m_1$ e $j = 1, \dots, m_2$. Para $p = 1$ e considerando todos os parâmetros de penalização iguais a ρ , temos a função de penalização L_1 ,

introduzida por Pietrzykowski [36] em 1969,

$$P_1(x, \rho) = f(x) + \rho \left(\sum_{i=1}^{m_1} |h_i(x)| + \sum_{j=1}^{m_2} g_j^+(x) \right). \quad (1.3)$$

Para $p = \infty$, temos

$$P_\infty(x, \alpha, \beta) = f(x) + \max_{\substack{1 \leq i \leq m_1 \\ 1 \leq j \leq m_2}} \{\alpha_i |h_i(x)|, \beta_j g_j^+(x)\}, \quad (1.4)$$

a função de penalização exata Minimax de Bandler e Charalambous [1].

Pietrzykowski mostrou que a função (1.3) é exata no sentido de existir um $\rho > 0$ finito tal que qualquer minimizador regular de (PNL) é também um minimizador local do problema penalizado irrestrito. Em 1970, Luenberger [28] mostrou que, usando hipótese de convexidade, o limite inferior para ρ é igual ao maior multiplicador de Lagrange em valor absoluto, associado ao problema não linear. Em 1978, Charalambous [7, 8] generalizou o resultado de Luenberger para a função de penalização L_1 , assumindo as condições suficientes de segunda ordem para o (PNL).

Outros trabalhos importantes envolvendo a função de penalização L_1 são os trabalhos de Charalambous [8], Coleman e Conn [9] e Ben-tal e Zowe [5]. Estes trabalhos apresentam as condições de otimalidade para minimização da função L_1 . Contribuições adicionais envolvendo a função L_1 podem ser encontradas em [14, 26, 38]. A função de penalização L_1 aparece com frequência em técnicas de programação quadrática seqüencial, onde é usada como função de mérito (ver [6, 30]).

Em 1979, Han e Mangasarian [22] introduziram uma classe de funções de penalização associada ao problema (1.1) com $\Omega = \mathbb{R}^n$,

$$P_Q(x, \rho) = f(x) + \rho Q(\|h(x), g^+(x)\|) \quad (1.5)$$

onde ρ é o parâmetro de penalização, $\|\cdot\|$ é qualquer norma vetorial em $\mathbb{R}^{m_1+m_2}$, e $Q: \mathbb{R}_+ \rightarrow \mathbb{R}_+$ uma função convexa tal que

$$Q(t) = 0 \Leftrightarrow t = 0 \text{ e } 0 < Q'(0^+) = \lim_{t \rightarrow 0^+} \frac{Q(t) - Q(0)}{t} < \infty.$$

Observe que para $Q(\alpha) = \alpha$ e $\|\cdot\| = \|\cdot\|_1$ temos a função de penalização L_1 , $P_1(x, \rho)$; para $Q(\alpha) = \alpha$ e $\|\cdot\| = \|\cdot\|_p$ temos a classe de funções estudada por Charalambous [7] com pesos iguais; para $Q(\alpha) = \alpha$ e $\|\cdot\| = \|\cdot\|_\infty$ temos a função de penalização Minimax.

Ao longo do capítulo, para $u \in \mathbb{R}^{m_1}$ e $v \in \mathbb{R}^{m_2}$, a notação $\|u, v\|$ significa $\|(u_1, \dots, u_{m_1}, v_1, \dots, v_{m_2})^T\|$.

Por simplicidade e sem perda de generalidade restringimos nossa atenção para funções de penalização exata (1.5) com $Q(\alpha) = \alpha$, isto é,

$$P(x, \rho) = f(x) + \rho \|h(x), g^+(x)\|. \quad (1.6)$$

A função $P(x, \rho)$ penaliza somente as restrições g e h do problema original, mantendo os problemas penalizados restritos ao conjunto Ω , isto é, nosso subproblema é dado por

$$\begin{array}{ll} \text{minimizar} & P(x, \rho) = f(x) + \rho P(x) \\ \text{sujeita a} & x \in \Omega. \end{array} \quad (1.7)$$

Os objetivos deste capítulo são apresentar os principais resultados de métodos de penalização para problemas de programação não linear, e mostrar que o teorema clássico da função de penalização exata para o caso irrestrito permanece válido quando mantemos os subproblemas penalizados restritos ao conjunto Ω . A consequência de tal resultado é que podemos aplicar algoritmos eficientes já conhecidos, que aproveitam a estrutura dos problemas penalizados. Como exemplo, podemos citar o trabalho recente de Martínez e Moretti [31] que resolve problema de otimização não suave restrito a um polítopo.

1.2 Resultados Gerais de Penalização

Os métodos que usam funções de penalização transformam o problema original em um único problema irrestrito ou em uma seqüência de problemas irrestritos. Em

geral, uma função de penalização deve acarretar uma penalização positiva para pontos infactíveis e não penalizar pontos factíveis.

Nesta seção, apresentamos resultados gerais de penalização destacando os principais resultados associados às funções de penalização $P(x, \rho)$ e $P_1(x, \rho)$. A teoria de penalização exata desenvolvida para resolver problemas de programação não linear é, em geral, somente apresentada para o caso onde os problemas penalizados são irrestritos. Neste capítulo, desenvolvemos uma extensão da teoria de penalização para o caso onde os problemas penalizados estão restritos ao conjunto Ω .

Vamos considerar os problemas penalizados da seguinte forma:

$$\begin{aligned} \text{minimizar } P(x, \rho_k) &= f(x) + \rho_k P(x) \\ \text{sujeita a } x &\in \Omega, \end{aligned} \tag{1.8}$$

onde

$$P(x) = \|h(x), g^+(x)\|,$$

com $\rho_k > 0$ e $\|\cdot\|$ é uma norma qualquer em $\mathbb{R}^{m_1+m_2}$.

Inicialmente, apresentamos o teorema de convergência global. A prova deste resultado para o caso onde os problemas penalizados são irrestritos pode ser encontrada em [29]. Para o caso onde os problemas penalizados estão restritos a um conjunto fechado, a prova pode ser encontrada em [3]. Dado que os problemas (1.8) se enquadram no segundo caso, vamos reproduzir a demonstração de [3].

Teorema 1: *Seja $\{x_k\}$, $k = 0, 1, 2, \dots$, uma seqüência de minimizadores globais do problema (1.8), e $0 < \rho_k < \rho_{k+1}$, $\rho_k \rightarrow +\infty$. Então, todo ponto limite da seqüência $\{x_k\}$ é um minimizador global do problema (1.1).*

Prova: Seja \bar{x} um ponto limite de $\{x_k\}$. Pela definição de x_k temos que

$$P(x_k, \rho_k) = f(x_k) + \rho_k P(x_k) \leq f(x) + \rho_k P(x) \text{ para todo } x \in \Omega. \tag{1.9}$$

Sejam f^* o valor ótimo associado ao problema original e $S = \{x \in \mathbb{R}^n | h(x) = 0 \text{ e } g(x) \leq 0\}$. Temos

$$f^* = \min_{x \in S \cap \Omega} f(x) = \min_{x \in S \cap \Omega} P(x, \rho_k).$$

Portanto, tomando o ínfimo do lado direito de (1.9) sobre $x \in \Omega \cap S$ obtemos

$$P(x_k, \rho_k) = f(x_k) + \rho_k P(x_k) \leq f^*.$$

Passando o limite superior na relação acima e usando a continuidade de f , segue que

$$f(\bar{x}) + \limsup_{k \rightarrow \infty} \rho_k P(x_k) \leq f^*. \quad (1.10)$$

De (1.10), $P(x_k) \geq 0$ e $\rho_k \rightarrow +\infty$, temos que

$$P(x_k) \rightarrow 0.$$

Pela continuidade de P ,

$$P(\bar{x}) = 0$$

isto é, $\bar{x} \in S$. Mas $\{x_k\} \subset \Omega$ e Ω é um conjunto fechado, assim $\bar{x} \in \Omega$. Logo, \bar{x} é factível para o problema (1.1). De (1.10) temos

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

Portanto, \bar{x} é um minimizador global para o problema (1.1). \square

Observe que a demonstração do teorema usa fortemente o fato do conjunto Ω ser fechado.

Uma limitação do teorema acima aparece quando assumimos que os problemas penalizados têm solução global para todo ρ_k . Os exemplos abaixo mostram que não há garantia de chegarmos a um minimizador do problema original se não seguirmos minimizadores globais dos problemas penalizados.

Exemplo 1: Considere o problema $\min 0$ s.a $x^3 + 3x^2 + 3 = 0$. Para a função de penalização L_1 , $x_k = 0$ é um minimizador local do problema penalizado para todo $k > 0$. Mas o único ponto limite ($x^* = 0$) é infactível para o problema original.

Exemplo 2: Considere o problema $\min -|x|^{1.5}$ s.a $0 \leq x \leq 1$. Usando a função de penalização de perda quadrática, $x_k = -9/(16\rho_k^2)$ é um minimizador local do problema penalizado. O ponto limite da seqüência ($x^* = 0$) é factível, mas x^* é maximizador do problema original.

Outra questão que devemos considerar é o que acontece com o ponto limite (se existir) de uma seqüência de pontos que são apenas estacionários dos problemas penalizados. Para a função de penalização quadrática, quando assumimos regularidade no ponto limite, é provado que este ponto limite é um ponto estacionário do problema original no caso $\Omega = \mathbb{R}^n$ (ver [3]). Para a função de penalização L_1 mostramos que este resultado permanece válido, um de nossos objetivos, assumimos factibilidade e regularidade no ponto limite. Para provarmos o resultado, apresentamos as definições das derivadas direcionais e a condição de otimalidade necessária para minimização de uma função φ .

Definição 1: *As derivadas direcionais de primeira e segunda ordem de uma função contínua $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$ em $x \in \mathbb{R}^n$ ao longo da direção $d \in \mathbb{R}^n$ são dadas por*

$$D\varphi(x; d) = \lim_{t \rightarrow 0^+} \frac{\varphi(x + td) - \varphi(x)}{t}$$

$$D^2\varphi(x; d) = \lim_{t \rightarrow 0^+} \frac{D\varphi(x + td; d) - D\varphi(x; d)}{t}$$

se os limites existirem.

Teorema 2: (Condição Necessária de 1ª ordem) *Se x^* é minimizador local da função $\varphi(x)$ então $D\varphi(x^*; d) \geq 0$ para todo $d \in \mathbb{R}^n$.*

Prova: Suponha por absurdo que existe $d \in \mathbb{R}^n$, tal que

$$D\varphi(x^*; d) = -\beta < 0. \quad (1.11)$$

Temos que

$$\varphi(x^* + \alpha d) = \varphi(x^*) + \alpha D\varphi(x^*; d) + o(d; \alpha) \quad (1.12)$$

onde

$$\lim_{\alpha \rightarrow 0^+} \frac{o(d; \alpha)}{\alpha} = 0. \quad (1.13)$$

De (1.13) temos que, para todo $\epsilon > 0$, existe $\alpha_1 > 0$ tal que $\alpha \leq \alpha_1$ implica que $|\frac{o(d; \alpha)}{\alpha}| \leq \epsilon$. Em particular, para $\epsilon = \frac{\beta}{2}$ e $\alpha = \alpha_1$, temos

$$|o(d; \alpha_1)| \leq \alpha_1 \frac{\beta}{2}. \quad (1.14)$$

De (1.11), (1.12) e (1.14) segue que

$$\varphi(x^* + \alpha d) \leq \varphi(x^*) + \alpha(-\beta) + \alpha \frac{\beta}{2} \leq \varphi(x^*),$$

o que contradiz o fato de x^* ser minimizador local da função $\varphi(x)$. \square

Para a função de penalização L_1 , a condição necessária de primeira ordem pode ser escrita de forma alternativa como apresentada no Teorema 3. O teorema está demonstrado em [8]. Os trabalhos de Fletcher [18] e Coleman-Coon [9] mostram o mesmo resultado exigindo independência linear do conjunto formado pelos gradientes das restrições ativas no ponto minimizador. Para provarmos o Teorema 3 precisamos do Lema 1 e seu Corolário.

Lema 1: Seja $\psi(x) = |\phi(x)|$, onde $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ e $\phi \in C^1$. Então a derivada direcional de $\psi(x)$ em $x^* \in \mathbb{R}^n$ ao longo da direção d , é dada por

$$D\psi(x^*; d) = \begin{cases} |\nabla\phi(x^*)^T d|, & \text{se } \phi(x^*) = 0 \\ -\nabla\phi(x^*)^T d, & \text{se } \phi(x^*) < 0 \\ \nabla\phi(x^*)^T d, & \text{se } \phi(x^*) > 0. \end{cases}$$

Prova: Seja $d \in \mathbb{R}^n$. Queremos mostrar que

$$\lim_{t \rightarrow 0^+} \frac{\psi(x^* + td) - \psi(x^*)}{t} = D\psi(x^*; d).$$

Temos que ϕ é diferenciável em x^* então

$$\phi(x^* + td) = \phi(x^*) + td^T \nabla\phi(x^*) + o(d; t),$$

onde

$$\lim_{t \rightarrow 0^+} \frac{o(d; t)}{t} = 0.$$

Assim,

$$\psi(x^* + td) = |\phi(x^* + td)| = |\phi(x^*) + td^T \nabla\phi(x^*) + o(d; t)|.$$

(a) Se $\phi(x^*) > 0$ então para t suficientemente pequeno a função ϕ permanece positiva. Assim,

$$\psi(x^* + td) = |\phi(x^* + td)| = \phi(x^*) + td^T \nabla\phi(x^*) + o(d; t) \text{ e}$$

$$\psi(x^*) = |\phi(x^*)| = \phi(x^*).$$

(b) Se $\phi(x^*) < 0$ então para t suficientemente pequeno a função ϕ permanece negativa. Assim,

$$\begin{aligned}\psi(x^* + td) &= |\phi(x^* + td)| = -\phi(x^*) - td^T \nabla \phi(x^*) + o(d; t) \text{ e} \\ \psi(x^*) &= |\phi(x^*)| = -\phi(x^*).\end{aligned}$$

(c) Se $\phi(x^*) = 0$ então

$$\begin{aligned}\psi(x^* + td) &= |td^T \nabla \phi(x^*) + o(d; t)| \text{ e} \\ \psi(x^*) &= 0.\end{aligned}$$

Portanto, de (a), (b) e (c) temos

$$\lim_{t \rightarrow 0^+} \frac{\psi(x^* + td) - \psi(x^*)}{t} = D\psi(x^*; d). \square$$

Corolário 1: Seja $P_h(x, \rho) = f(x) + \rho \|h(x)\|_1$ com $\rho > 0$. Então a derivada direcional de $P_h(x, \rho)$ em $x^* \in \mathbb{R}^n$ ao longo da direção d , é dada por

$$DP_h(x^*, \rho; d) = d^T \nabla f(x^*) + \rho \left(\sum_{i \in A} |d^T \nabla h_i(x^*)| + \sum_{i \notin A} \sigma_i d^T \nabla h_i(x^*) \right),$$

onde $A = A(x^*) = \{i \mid h_i(x^*) = 0, i = 1, \dots, m_1\}$ e $\sigma_i = \text{ sinal}(h_i(x^*))$, se $i \notin A$.

Teorema 3: As seguintes afirmações são equivalentes:

(i) Para todo $d \in \mathbb{R}^n$, $DP_1(x^*, \rho; d) \geq 0$.

(ii) Existem escalares $v_i \in [-1, 1]$, $i \in A \equiv A(x^*) = \{i \mid h_i(x^*) = 0, i = 1, \dots, m_1\}$ e $w_j \in [0, 1]$, $j \in B \equiv B(x^*) = \{j \mid g_j(x^*) = 0, j = 1, \dots, m_2\}$ tais que

$$\nabla f(x^*) + \rho \left(\sum_{i \in A} v_i \nabla h_i(x^*) + \sum_{i \notin A} \sigma_i \nabla h_i(x^*) + \sum_{j \in B} w_j \nabla g_j(x^*) + \sum_{j \notin B} s_j \nabla g_j(x^*) \right) = 0,$$

onde $\sigma_i = \text{sinal}(h_i(x))$, $i \notin A$
e $s_j = \begin{cases} 0, & \text{se } g_j(x) < 0 \\ 1, & \text{se } g_j(x) > 0 \end{cases} = \frac{1 + \text{sinal}(g_j(x))}{2}$ para todo $j \notin B$.

Prova: Para simplificar a demonstração assumiremos somente restrições de igualdade.

(a) Vamos assumir (ii) e mostraremos que (i) se verifica.

Suponhamos por absurdo que existe $u \in \mathbb{R}^n$ tal que $DP_1(x^*, \rho; u) < 0$, isto é,

$$u^T \nabla f(x^*) + \rho \left(\sum_{i \in A} |u^T \nabla h_i(x^*)| + \sum_{i \notin A} \sigma_i u^T \nabla h_i(x^*) \right) \leq -\beta < 0 \quad (1.15)$$

para algum β . Seja

$$y = \nabla f(x^*) + \rho \left(\sum_{i \in A} v_i \nabla h_i(x^*) + \sum_{i \notin A} \sigma_i \nabla h_i(x^*) \right),$$

onde $-1 \leq v_i \leq 1$, $i \in A$. Então

$$y^T u = u^T \nabla f(x^*) + \rho \left(\sum_{i \in A} v_i \nabla h_i(x^*)^T u + \sum_{i \notin A} \sigma_i \nabla h_i(x^*)^T u \right). \quad (1.16)$$

Somando e subtraindo $\rho \sum_{i \in A} |u^T \nabla h_i(x^*)|$ em (1.16) e usando (1.15) temos

$$y^T u \leq -\beta - \rho \left(\sum_{i \in A} \left\{ |u^T \nabla h_i(x^*)| - v_i \nabla h_i(x^*)^T u \right\} \right).$$

Temos que $v_i \in [-1, 1]$, $i \in A$, então

$$v_i \nabla h_i(x^*)^T u \leq |u^T \nabla h_i(x^*)|, \quad i \in A.$$

Assim,

$$-\sum_{i \in A} \left\{ |u^T \nabla h_i(x^*)| - v_i u^T \nabla h_i(x^*) \right\} \leq 0.$$

Segue que $y^T u \leq -\beta < 0$. Portanto, y não pode ser o vetor zero, o que contradiz (ii).

(b) Vamos assumir (i) e provaremos (ii).

Suponhamos que (ii) não é satisfeito, isto é, para todo $v_i \in [-1, 1]$, $i \in A$, o vetor

$$z = \nabla f(x^*) + \rho \left(\sum_{i \in A} v_i \nabla h_i(x^*) + \sum_{i \notin A} \sigma_i \nabla h_i(x^*) \right) \neq 0. \quad (1.17)$$

Para um x fixo, consideremos o conjunto

$$S(x) = \{w \mid w = \nabla f(x) + \rho \sum_{i \in A} v_i \nabla h_i(x), -1 \leq v_i \leq 1, i \in A\}.$$

Note que $S(x)$ é convexo e fechado. Seja

$$s = -\rho \sum_{i \notin A} \sigma_i \nabla h_i(x^*).$$

Temos que $s \notin S(x^*)$ pois, caso contrário para cada $i \in A$ existiria $v_i \in [-1, 1]$ tal que

$$s = \nabla f(x^*) + \rho \sum_{i \in A} v_i \nabla h_i(x^*),$$

isto é,

$$\nabla f(x^*) + \rho \sum_{i \in A} v_i \nabla h_i(x^*) + \rho \sum_{i \notin A} \sigma_i \nabla h_i(x^*) = 0,$$

o que contradiz (1.17).

Pelo teorema da separação, existe um vetor $q \neq 0 \in \mathbb{R}^n$ e um escalar η tal que $q^T s > \eta$ e $q^T y \leq \eta$ para todo $y \in S(x^*)$, isto é,

$$q^T \left\{ -\rho \sum_{i \notin A} \sigma_i \nabla h_i(x^*) - y \right\} > 0.$$

Seja

$$y = \left(\nabla f(x^*) + \rho \sum_{i \in A} \gamma_i \nabla h_i(x^*) \right) \in S(x^*)$$

onde $\gamma_i = \text{sinal}(\nabla h_i(x^*)^T q)$, $i \in A$. Então

$$-q^T \nabla f(x^*) - \rho \left(\sum_{i \in A} |q^T \nabla h_i(x^*)| + \sum_{i \notin A} \sigma_i q^T \nabla h_i(x^*) \right) > 0.$$

Portanto, existe uma direção $q \in \mathbb{R}^n$ tal que $DP_1(x^*, \rho; q) < 0$, o que contradiz (i).

□

Agora, podemos mostrar que o ponto limite (quando existe) de uma seqüência de pontos estacionários dos problemas penalizados irrestrito usando a função L_1 é um ponto estacionário do problema original com $\Omega = \mathbb{R}^n$, quando assumimos regularidade e factibilidade no ponto limite.

Teorema 4: *Sejam $\{x_k\}$ seqüência de pontos estacionários do problema irrestrito minimizar $P_1(x, \rho) = f(x) + \rho \|h(x), g^+(x)\|_1$ e $M(x) = \begin{bmatrix} J_h(x) \\ J_{gB}(x) \end{bmatrix}$ onde $J_h(x)$ é a matriz formada pelos vetores linhas $\nabla h_i(x)^T, i \in A(x)$, $J_{gB}(x)$ é a matriz formada pelos vetores linhas $\nabla g_j(x)^T, j \in B(x)$, e os conjuntos $A(x)$ e $B(x)$ estão definidos no Teorema 3. Assuma que $0 < \rho_k < \rho_{k+1}$ para todo k e $\rho_k \rightarrow +\infty$. Assuma também que existe uma subseqüência $\{x_k\}_K, K \subseteq \mathbb{N}$, convergindo para $x^* \in \mathbb{R}^n$ tal que $h(x^*) = 0$ e $g(x^*) \leq 0$ e que a matriz $M(x^*)$ tem posto $(m_1 + m)$, onde m é a cardinalidade de $B(x^*)$. Então x^* é um ponto estacionário do problema original com $\Omega = \mathbb{R}^n$.*

Prova: Para simplificar a demonstração assumiremos somente restrições de igualdade. Como x_k é um ponto estacionário do problema irrestrito $P_1(x, \rho)$, pelo Teorema 3, existem $\lambda_i^k \in [-1, 1]$ tais que

$$\nabla f(x_k) + \rho_k \left(\sum_{i=1}^{m_1} \lambda_i^k \nabla h_i(x_k) \right) = 0.$$

Como $J_h(x^*)$ tem posto completo, então para k suficientemente grande $J_h(x_k)$ também possui posto completo. Assim podemos escrever

$$\rho_k \lambda_k = - \left(J_h(x_k) J_h(x_k)^T \right)^{-1} (J_h(x_k) \nabla f(x_k)).$$

Da convergência $\{x_k\}_K$ para x^* temos

$$\{\rho_k \lambda_k\}_K \rightarrow \lambda^* = - \left(J_h(x^*) J_h(x^*)^T \right)^{-1} (J_h(x^*) \nabla f(x^*))$$

e podemos escrever

$$\nabla f(x^*) + \left(\sum_{i=1}^{m_1} \lambda_i^* \nabla h_i(x^*) \right) = 0.$$

Da igualdade acima e $h(x^*) = 0$ temos que, x^* é um ponto estacionário para o problema $\min f(x)$ sujeita a $h(x) = 0$, o que completa a prova. \square

O Teorema 2 apresenta a condição de otimalidade necessária de 1ª ordem para a função de penalização $P(x, \rho)$. Condições de otimalidade suficientes de 2ª ordem podem ser encontradas em Fletcher [18], onde o autor emprega o conceito de subgradiente. Para a função de penalização L_1 , a condição de otimalidade necessária de 1ª ordem pode ser escrita de forma equivalente usando os gradientes da função objetivo e das restrições do (PNL) associado. Condições suficientes de 2ª ordem para a função L_1 envolvendo as Hessianas das funções objetivo e restrições do (PNL) associado podem ser encontradas em [5, 9, 18].

Tentando buscar uma forma alternativa para expressar as condições suficientes de 2ª ordem para função $P(x, \rho)$, visando um paralelo com as condições suficientes de problemas irrestritos diferenciáveis e motivados pela demonstração do teorema da penalização exata para a função $P_p(x, \alpha, \beta)$ de Charalambous [7], escrevemos as condições suficientes da seguinte forma:

Seja x^* tal que para todo $0 \neq d \in \mathbb{R}^n$, $DP(x^*, \rho; d) \geq 0$. Se para todo d satisfazendo $DP(x^*, \rho; d) = 0$ vale $D^2P(x^*, \rho; d) > 0$ então x^* é um minimizador local da função $P(x, \rho)$.

O exemplo abaixo mostra que as condições suficientes apresentadas desta forma não são verdadeiras. Portanto, não podemos provar que um ponto x^* é minimizador local da função $P(x, \rho)$ simplesmente calculando as derivadas direcionais de primeira e segunda ordem.

Exemplo 3: Considere o problema irrestrito

$$\text{minimizar } P_1(x, \rho) = \rho| -\alpha x_1^2 + 4x_2 | + \rho|x_1^2 + (x_2 - 1)^2|.$$

Temos que $x^* = (0, 0)^T$ não é minimizador local para $\alpha > 2$ (ver [5]). Mas temos $DP_1(x^*, \rho; d) \geq 0$ para toda $d \in \mathbb{R}^2$ e para as direções $d \in \mathbb{R}^2$ tais que $DP_1(x^*, \rho; d) = 0$ temos que $D^2P_1(x^*, \rho; d) > 0$, independente do valor de α .

Queremos ressaltar que a demonstração da exatidão da função $P_p(x, \alpha, \beta)$ des-

crita por Charalambous em [7] está incorreta, porém o resultado é verdadeiro. Uma prova completa do resultado é exibida na próxima seção. Antes de escrever a prova apresentamos as condições suficientes de segunda ordem para o problema (1.1) com $\Omega = \mathbb{R}^n$ (a prova pode ser encontrada em [29]). A função Lagrangeana é definida por

$$\mathcal{L}(x, \lambda, \mu) = f(x) + h(x)^T \lambda + g(x)^T \mu,$$

e assim

$$\nabla_x \mathcal{L}(x, \lambda, \mu) = \nabla f(x) + J_h(x)^T \lambda + J_g(x)^T \mu,$$

e

$$\nabla_{xx}^2 \mathcal{L}(x, \lambda, \mu) = \nabla^2 f(x) + \sum_{i=1}^{m_1} \lambda_i \nabla^2 h_i(x) + \sum_{j=1}^{m_2} \mu_j \nabla^2 g_j(x),$$

onde

$$J_h(x) = [\nabla h_1(x), \dots, \nabla h_{m_1}(x)]^T \text{ e } J_g(x) = [\nabla g_1(x), \dots, \nabla g_{m_2}(x)]^T.$$

Teorema 5: Se $(x^*, \lambda^*, \mu^*) \in \mathbb{R}^n \times \mathbb{R}^{m_1} \times \mathbb{R}_+^{m_2}$ é tal que

$$\nabla_x \mathcal{L}(x^*, \lambda^*, \mu^*) = 0, \tag{1.18}$$

$$h(x^*) = 0, \quad g(x^*) \leq 0, \quad g(x^*)^T \mu^* = 0, \tag{1.19}$$

$$d^T \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*, \mu^*) d > 0, \tag{1.20}$$

para todo $d \in \{y \in \mathbb{R}^n \mid J_h(x^*)y = 0, \nabla g_j(x^*)^T y = 0, i \in J\}$, onde $J = \{j \mid g_j(x^*) = 0, \mu_j^* > 0\}$ então x^* é um minimizador local estrito do problema (1.1) com $\Omega = \mathbb{R}^n$.

1.3 Exatidão de $P(x, \rho)$

Para estabelecer um limite inferior para o parâmetro de penalização precisamos do conceito de norma dual. Dada qualquer norma $\|\cdot\|$ em \mathbb{R}^n existe uma norma correspondente $\|\cdot\|'$ chamada de norma dual, que é definida por

$$\|x\|' = \max_{\|y\|=1} x^T y.$$

Por exemplo, a norma dual de $\|\cdot\|_1$ é $\|\cdot\|_\infty$ e a $\|\cdot\|_2$ é a sua própria dual. Da definição acima, segue a chamada desigualdade de Cauchy generalizada,

$$|x^T y| \leq \|x\|' \cdot \|y\| \text{ para todo } x, y \in \mathbb{R}^n.$$

Várias provas de exatidão de $P(x, \rho)$ podem ser encontradas dependendo da escolha da norma. Para $\|\cdot\|_1$ ver [5, 8, 29] e para $\|\cdot\|_p$ ver [11, 22]. Há provas que não dependem da norma e trabalham com a definição de subgradiente [18] ou usam a exatidão da função de penalização L_1 [22]. No que segue apresentamos uma versão mais simples da prova [22] para a exatidão da função $P(x, \rho)$.

Teorema 6: *Seja $(x^*, \lambda^*, \mu^*) \in \mathbb{R}^n \times \mathbb{R}^{m_1} \times \mathbb{R}_+^{m_2}$ satisfazendo as condições suficientes de segunda ordem para o problema (1.1) com $\Omega = \mathbb{R}^n$ e $\|\cdot\|$ uma norma qualquer em $\mathbb{R}^{m_1+m_2}$ com $\|\cdot\|'$ sua norma dual. Então para $\rho > \bar{\rho} = \|\lambda^*, \mu^*\|'$, x^* é um minimizador local estrito da função $P(x, \rho) = f(x) + \rho\|h(x), g^+(x)\|$.*

Prova: Vamos assumir que x^* não é um minimizador local estrito de $P(x, \rho)$. Logo, existe uma seqüência $\{x^k\}$ convergindo para x^* tal que $P(x^k, \rho) \leq P(x^*, \rho)$ para todo $k \geq 0$. Assim,

$$P(x^k, \rho) - P(x^*, \rho) = f(x^k) - f(x^*) + \rho\|h(x^k), g^+(x^k)\| \leq 0. \quad (1.21)$$

Defina a seqüência $d^k = (x^k - x^*)/\|x^k - x^*\|$ e tome uma subseqüência convergente tal que $\{d^k\}$ converge para d com $\|d\| = 1$. Tomando o limite em (1.21), temos

$$\nabla f(x^*)^T d + \|J_h(x^*)d, J_{g^+}(x^*)d\| \leq 0, \quad (1.22)$$

onde

$$J_{g^+}(x^*)d = [\nabla g_1^+(x^*)^T d, \dots, \nabla g_{m_2}^+(x^*)^T d]^T,$$

com

$$\nabla g_j^+(x^*)^T d = \begin{cases} 0, & g_j(x^*) < 0 \\ \max\{\nabla g_j^+(x^*)^T d, 0\}, & g_j(x^*) = 0. \end{cases}$$

Usando a equação (1.18) em (1.22), encontramos

$$\rho\|J_h(x^*)d, J_{g^+}(x^*)d\| - (\lambda^*)^T J_h(x^*)d - (\mu^*)^T J_g(x^*)d \leq 0, \quad (1.23)$$

que implica em

$$\rho \|J_h(x^*)d, J_{g^+}(x^*)d\| - (\lambda^*)^T J_h(x^*)d - (\mu^*)^T J_{g^+}(x^*)d \leq 0,$$

dado que $(\mu^*)^T J_{g^+}(x^*)d \leq (\mu^*)^T J_{g^+}(x^*)d$. Agora, usando a desigualdade de Cauchy generalizada temos

$$(\rho - \|\lambda^*, \mu^*\|') \|J_h(x^*)d, J_{g^+}(x^*)d\| \leq 0.$$

De $\rho > \|\lambda^*, \mu^*\|'$, segue que $J_h(x^*)d = 0$ e $J_{g^+}(x^*)d = 0$. Além disso, da desigualdade (1.23), $\nabla g_j(x^*)^T d = 0$ para $j \in J = \{j | g_j(x^*) = 0, \mu_j^* > 0\}$. Da equação (1.20) do Teorema 5, segue que $d^T \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*, \mu^*)d > 0$, que implica que para todo k suficientemente grande

$$\mathcal{L}(x^k, \lambda^*, \mu^*) > \mathcal{L}(x^*, \lambda^*, \mu^*).$$

Assim,

$$\begin{aligned} P(x^k, \rho) &= f(x^k) + \rho \|h(x^k), g^+(x^k)\| \\ &> f(x^k) + \|\lambda^*, \mu^*\|' \|h(x^k), g^+(x^k)\| \\ &\geq f(x^k) + h(x^k)^T \lambda^* + (g^+(x^k))^T \mu^* \\ &\geq f(x^k) + h(x^k)^T \lambda^* + g(x^k)^T \mu^* \\ &= \mathcal{L}(x^k, \lambda^*, \mu^*) \\ &> \mathcal{L}(x^*, \lambda^*, \mu^*) \\ &= f(x^*) \\ &= P(x^*, \rho) \\ &\geq P(x^k, \rho), \end{aligned}$$

o que é uma contradição. \square

O Teorema 6 relaciona minimizadores locais do problema (1.1) (com $\Omega = \mathbb{R}^n$) com minimizadores locais do problema penalizado irrestrito. Para estabelecer uma relação entre pontos estacionários do problema não linear com pontos estacionários do problema penalizado irrestrito, precisamos calcular a derivada direcional da função $P(x, \rho)$ em pontos factíveis do problema não linear associado.

Teorema 7: *Sejam $\|\cdot\|$ uma norma qualquer em $\mathbb{R}^{m_1+m_2}$ e $P(x, \rho) = f(x) + \rho \|h(x), g^+(x)\|$, com $\rho > 0$. Então, para $x^* \in \mathbb{R}^n$ tal que $h(x^*) = 0$ e $g(x^*) \leq 0$, a*

derivada direcional de primeira ordem de $P(x, \rho)$ em x^* na direção $d \in \mathbb{R}^n$ é dada por

$$DP(x^*, \rho; d) = \nabla f(x^*)^T d + \rho \|J_h(x^*)d, J_{g_+}(x^*)d\|, \quad (1.24)$$

onde

$$J_h(x^*)d = [\nabla h_1(x^*)^T d, \dots, \nabla h_{m_1}(x^*)^T d]^T$$

$$J_{g_+}(x^*)d = [\nabla g_1^+(x^*)^T d, \dots, \nabla g_{m_2}^+(x^*)^T d]^T$$

e

$$\nabla g_j^+(x^*)^T d = \begin{cases} 0, & \text{se } g_j(x^*) < 0 \\ \max\{\nabla g_j(x^*)^T d, 0\}, & \text{se } g_j(x^*) = 0 \end{cases}$$

Prova : Seja $d \in \mathbb{R}^n$. Queremos mostrar que

$$DP(x^*, \rho; d) = \lim_{\alpha \rightarrow 0^+} \frac{P(x^* + \alpha d, \rho) - P(x^*, \rho)}{\alpha}.$$

Temos que f é diferenciável em x^* então

$$f(x^* + \alpha d) = f(x^*) + \alpha \nabla f(x^*)^T d + o(d; \alpha) \quad (1.25)$$

onde

$$\lim_{\alpha \rightarrow 0^+} \frac{o(d; \alpha)}{\alpha} = 0.$$

De h_i e g_j diferenciáveis em x^* para todo $i = 1, \dots, m_1$ e $j = 1, \dots, m_2$ temos

$$h_i(x^* + \alpha d) = h_i(x^*) + \alpha \nabla h_i(x^*)^T d + o_i(d; \alpha)$$

e

$$g_j^+(x^* + \alpha d) = \max\{0, g_j(x^*) + \alpha \nabla g_j(x^*)^T d + o_{j+m_1}(d; \alpha)\}$$

onde

$$\lim_{\alpha \rightarrow 0^+} \frac{o_i(d; \alpha)}{\alpha} = 0 \text{ para todo } i = 1, \dots, m_1 + m_2.$$

Como $h_i(x^*) = 0$ para todo $i = 1, \dots, m_1$, temos que

$$h_i(x^* + \alpha d) = \alpha \nabla h_i(x^*)^T d + o_i(d; \alpha). \quad (1.26)$$

Se $g_j(x^*) < 0$ então, para α suficientemente pequeno a função g_j permanece negativa. Assim,

$$g_j^+(x^* + \alpha d) = 0 \text{ e } g_j^+(x^*) = 0. \quad (1.27)$$

Se $g_j(x^*) = 0$ então

$$g_j^+(x^* + \alpha d) = \max\{0, \alpha \nabla g_j(x^*)^T d + o_{j+m_1}(d; \alpha)\} \quad \text{e} \quad g_j^+(x^*) = 0. \quad (1.28)$$

Combinando 1.25, ..., 1.28 temos

$$DP(x^*, \rho; d) = \lim_{\alpha \rightarrow 0^+} \frac{P(x^* + \alpha d, \rho) - P(x^*, \rho)}{\alpha},$$

o que completa a prova. \square

Teorema 8: *Sejam x^* ponto estacionário do problema (1.1) com $\Omega = \mathbb{R}^n$ e $\lambda^* \in \mathbb{R}^{m_1}$ e $\mu^* \in \mathbb{R}_+^{m_2}$ os multiplicadores de Lagrange correspondentes. Então para $\rho > \|\lambda^*, \mu^*\|'$, x^* é um ponto estacionário da função $P(x, \rho)$.*

Prova: Da expressão da derivada direcional da função $P(x, \rho)$, de x^* ponto estacionário do (PNL) com $\mu^* \geq 0$ e usando a desigualdade de Cauchy generalizada obtemos

$$\begin{aligned} DP(x^*, \rho; d) &= d^T \nabla f(x^*) + \rho \|J_h(x^*)d, J_{g_+}(x^*)d\| \\ &= -(\lambda^*)^T J_h(x^*)d - (\mu^*)^T J_{g_+}(x^*)d + \rho \|J_h(x^*)d, J_{g_+}(x^*)d\| \\ &\geq -(\lambda^*)^T J_h(x^*)d - (\mu^*)^T J_{g_+}(x^*)d + \rho \|J_h(x^*)d, J_{g_+}(x^*)d\| \\ &\geq -\|\lambda^*, \mu^*\|' \|J_h(x^*)d, J_{g_+}(x^*)d\| + \rho \|J_h(x^*)d, J_{g_+}(x^*)d\| \\ &\geq (\rho - \|\lambda^*, \mu^*\|') \|J_h(x^*)d, J_{g_+}(x^*)d\|. \end{aligned}$$

Assim, para $\rho > \|\lambda^*, \mu^*\|'$ temos que $DP(x^*, \rho; d) \geq 0$. Portanto, x^* é ponto estacionário da função $P(x, \rho)$. \square

O Teorema 6 garante que, assumindo as condições suficientes de segunda ordem, se x^* é minimizador local do (1.1) então x^* é também um minimizador local do problema penalizado. O interesse prático é com a recíproca deste teorema. Para o caso convexo, a caracterização de soluções do problema de programação não linear através de soluções do problema penalizado é inteiramente satisfatória, dado que o conjunto de soluções ótimas dos dois problemas são idênticos (ver [7]). No caso não convexo, há exemplos onde não temos equivalência entre os problemas não linear e o problema de penalização exata associado. Exemplos onde x^* resolve o problema (1.1) e não resolve o problema penalizado podem ser encontrados em [18]. Mas, como descrito por Fletcher [18], os exemplos são casos patológicos ou extremos e não

precisam preocupar o usuário. Outra possibilidade que ilustra a não equivalência entre os problemas é o Exemplo 1 apresentado na Seção 2, onde o minimizador local do problema penalizado é infactível para o problema original. De qualquer forma, estudo de propriedades que assegurem que solução local (global) do problema penalizado é solução local (global) do problema não linear é de grande interesse.

Queremos destacar que o Teorema 6 permanece verdadeiro se usarmos diferentes parâmetros de penalização para cada restrição. Não há dificuldades em reorganizar a teoria para acomodar esta troca. O teorema abaixo ilustra o uso de dois diferentes parâmetros de penalização.

Teorema 9: *Se x^* satisfaz as condições suficientes de segunda ordem para o problema (1.1) com $\Omega = \mathbb{R}^n$. Sejam $\lambda^* \in \mathbb{R}^{m_1}$, $\mu^* \in \mathbb{R}_+^{m_2}$ e η^* os vetores dos multiplicadores de Lagrange associados a h, g e ao conjunto de restrições Ω respectivamente. Seja $\|\cdot\|$ uma norma qualquer com $\|\cdot\|'$ sua norma dual. Então para $\rho > \bar{\rho} = \|\lambda^*, \mu^*\|'$, e $\rho_1 > \bar{\rho}_1 = \|\eta^*\|'$, x^* é um minimizador local estrito da função $P(x, \rho) = f(x) + \rho\|h(x), g^+(x)\| + \rho_1 R(x)$, onde $R(x)$ é a função de penalização associada ao conjunto Ω .*

O Teorema 6 mostra que a função de penalização $P(x, \rho)$ associada ao problema (1.1) com $\Omega = \mathbb{R}^n$ é exata. Um resultado análogo é dado no Teorema 9, onde o conjunto Ω foi incluído na função de penalização. No teorema seguinte, mostramos que existe um limite inferior finito para o parâmetro de penalização no caso em que o problema (1.1) é penalizado somente com respeito a h e g , e assim os problemas penalizados permanecem restritos ao conjunto Ω . Observe que, na teoria clássica para funções de penalização exata, os problemas penalizados são irrestritos.

Teorema 10 : *Seja $x^* \in \mathbb{R}^n$ satisfazendo as condições suficientes de segunda ordem para um minimizador local do problema (1.1). Sejam $\lambda^* \in \mathbb{R}^{m_1}$ e $\mu^* \in \mathbb{R}_+^{m_2}$ os multiplicadores de Lagrange correspondentes às restrições h e g , e $\|\cdot\|$ uma norma vetorial com $\|\cdot\|'$ sua norma dual. Então existe $0 < \bar{\rho} < \infty$, tal que para $\rho > \bar{\rho}$, x^* é um minimizador local estrito do problema penalizado (1.8).*

Prova : Considere o problema

$$\begin{aligned} &\text{minimizar } f(x) + \rho \|h(x), g^+(x)\| + \rho_1 R(x) \\ &\text{sujeito a } x \in \mathbb{R}^n. \end{aligned} \tag{1.29}$$

onde $\rho, \rho_1 > 0$ e $R(x)$ está definida no Teorema 9. Dado que x^* satisfaz as condições suficientes de segunda ordem do problema (1.1), do Teorema 9, existem $\bar{\rho} = \|\lambda^*, \mu^*\|'$ e $\bar{\rho}_1 = \|\eta^*\|'$, tais que para $\rho > \bar{\rho}$ e $\rho_1 > \bar{\rho}_1$, x^* é um minimizador local do problema (1.29). Portanto, $\rho > \bar{\rho}$ e $\rho_1 > \bar{\rho}_1$, existe $\epsilon \equiv \epsilon(\rho, \rho_1) > 0$, tal que para todo $x \in B(x^*, \epsilon) \equiv \{x \in \mathbb{R}^n \mid \|x - x^*\| \leq \epsilon\}$, temos

$$f(x^*) < f(x) + \rho \|h(x), g^+(x)\| + \rho_1 R(x) \text{ para todo } x \in B(x^*, \epsilon) \cap \Omega,$$

dado que $R(x) = 0$ para todo $x \in \Omega$. Assim, para $\rho > \bar{\rho} = \|\lambda^*, \mu^*\|'$, x^* é um minimizador local do problema

$$\begin{aligned} &\text{minimizar } f(x) + \rho \|h(x), g^+(x)\| \\ &\text{sujeito a } x \in \Omega, \end{aligned}$$

e a prova está completa. \square

A importância do Teorema 10 é a garantia da existência de um parâmetro de penalização que depende somente dos multiplicadores de Lagrange associado às restrições h e g , e não do conjunto Ω . Tal resultado permite desenvolvermos algoritmos para resolver (1.1) usando $P(x, \rho)$ de forma que os problemas penalizados permaneçam restritos ao conjunto Ω . Note que, dado que o limite inferior do parâmetro de penalização é finito, a seqüência de problemas penalizados que devemos resolver é também finita.

Na teoria, os métodos de penalização exata permitem resolver o problema de programação não linear através de um único problema de penalização para um parâmetro de penalização acima do valor limite. Na prática, é mais eficiente permitir pontos fora da região factível usando uma seqüência crescente do parâmetro de penalização começando abaixo do valor limite. Assim, geramos uma seqüência de pontos que são soluções dos problemas penalizados. O próximo teorema garante que, sob condições razoáveis, a seqüência gerada é constante a partir de um certo índice.

Teorema 11: *Seja $\{x_k\}$ uma seqüência de minimizadores globais do problema penalizado. Se x^* é um ponto limite de $\{x_k\}$, satisfaz as condições suficientes de*

segunda ordem do problema original, e x^* é minimizador global estrito do problema original, então para $\rho_k > \bar{\rho}$, $x_k = x^*$ para todo k a partir de certo índice k_1 .

Prova: De x^* satisfazendo as condições suficientes de segunda ordem para o problema (1.1), temos para $\rho_k > \bar{\rho}$ que

$$x^* \text{ é um minimizador local do problema (1.8),}$$

isto é, existe $k_0 \in N$ tal que para todo $k > k_0$, x^* é um minimizador local do problema penalizado. Assim, existe $\delta_{k_0} > 0$ tal que

$$f(x^*) \leq f(x) + \rho_{k_0}P(x) \text{ para todo } x \in B(x^*, \delta_{k_0}) \cap \Omega.$$

Por hipótese, x_k é minimizador global do problema penalizado (1.8) então

$$f(x_k) + \rho_k P(x_k) \leq f(x^*). \quad (1.30)$$

De x^* ponto limite de $\{x_k\}$, temos que, para $\delta_{k_0} > 0$, existe $k_1 > k_0$ tal que

$$x_k \in B(x^*, \delta_{k_0}) \cap \Omega \text{ para todo } k > k_1.$$

Assim, para $k > k_1$

$$f(x^*) \leq f(x_k) + \rho_{k_0}P(x_k). \quad (1.31)$$

De (1.31), $\rho_{k_1} > \rho_{k_0}$, $\rho_k > \rho_{k_1}$ por todo $k > k_1$ e (1.30) temos

$$f(x^*) \leq f(x_k) + \rho_{k_0}P(x_k) \leq f(x_k) + \rho_{k_1}P(x_k) \leq f(x_k) + \rho_k P(x_k) \leq f(x^*) \quad (1.32)$$

para $k > k_1$. Para que (1.32) seja satisfeita

$$\rho_{k_0}P(x_k) = \rho_{k_1}P(x_k) \text{ para todo } k > k_1.$$

Mas, $\rho_{k_1} > \rho_{k_0}$. Assim

$$P(x_k) = 0 \text{ para todo } k > k_1.$$

De (1.30) temos que

$$f(x_k) \leq f(x^*) \text{ para todo } k > k_1. \quad (1.33)$$

De x^* minimizador global do problema original, $P(x_k) = 0$ para todo $k > k_1$ e $x_k \in \Omega$ temos

$$f(x^*) \leq f(x_k) \text{ para todo } k > k_1. \quad (1.34)$$

De (1.33), (1.34) e x^* minimizador estrito temos para $\rho_k > \bar{\rho}$ que

$$x_k = x^* \text{ para todo } k > k_1,$$

o que completa a prova. \square

Capítulo 2

Método de Penalização com Subproblemas Restritos

2.1 Introdução

Os métodos de penalização externa consistem em resolver problemas de programação não linear através da resolução de uma seqüência de problemas irrestritos. As restrições são colocadas na função objetivo via um parâmetro de penalização de forma a punir sua violação. À medida que o parâmetro de penalização cresce, a aproximação entre as soluções dos problemas penalizado e original também melhora. Em geral, os métodos de penalização externa trabalham fora do conjunto factível, obtendo a factibilidade apenas na solução.

Neste capítulo o objetivo é apresentar um algoritmo para resolver problemas da forma

$$\begin{aligned} &\text{mininizar} && f(x) \\ &\text{sujeita a} && h(x) = 0 \\ & && g(x) \leq 0 \\ & && x \in \Omega, \end{aligned} \tag{2.1}$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $h : \mathbb{R}^n \rightarrow \mathbb{R}^{m_1}$, $g : \mathbb{R}^n \rightarrow \mathbb{R}^{m_2}$, $f, h, g \in C^2$ e Ω é um conjunto simples. Por exemplo Ω pode ser uma caixa, uma bola ou, talvez, um polítopo.

Nossa proposta para resolver (2.1) é aproveitar a estrutura do conjunto Ω e aplicar a função de penalização L_1 somente às restrições não lineares. Desta forma os problemas penalizados permanecem restritos ao conjunto Ω e, assim, a seqüência de problemas que devemos resolver é da forma

$$\begin{aligned} \text{minimizar} \quad & P_1(x, \rho_k) = f(x) + \rho_k P(x) \\ \text{sujeita a} \quad & x \in \Omega, \end{aligned} \quad (2.2)$$

onde

$$P(x) = \|h(x)\|_1 + \|g^+(x)\|_1 = \sum_{i=1}^{m_1} |h_i(x)| + \sum_{j=1}^{m_2} \max\{0, g_j(x)\} \quad (2.3)$$

e $\{\rho_k\}$ é uma seqüência de parâmetros penalizadores, satisfazendo $0 < \rho_k < \rho_{k+1}$ para todo k . A sistemática do algoritmo proposto é apresentada a seguir.

2.2 Algoritmo Principal

O problema de programação não linear que estamos interessados em resolver tem a propriedade que as funções f, h e $g \in C^2$. O método que descrevemos abaixo não impõe qualquer restrição sobre as funções objetivo e restrições do problema original, somente continuidade. Assim o método proposto pode ser usado em casos onde um procedimento eficiente esteja disponível para resolver o problema especificado no passo 1 do Algoritmo 1.

Algoritmo 1

Inicialização:

Dados $\rho_1 > 0$, $x_0 \in \Omega$. Seja $k=0$.

Passo 1: Resolução do problema penalizado

Utilize x_k como aproximação inicial e resolva (2.2), obtendo x_{k+1} como solução ótima

deste subproblema.

Passo 2: *Teste da factibilidade e atualização do parâmetro de penalização*

Se $P(x_{k+1}) \approx 0$ então PARE, x_{k+1} é uma solução aproximada de (2.1). Caso contrário, faça $\rho_{k+1} = 10\rho_k$, $k = k + 1$ e volte ao Passo 1.

No capítulo anterior foi discutida a teoria dos métodos de penalização clássicos. A seqüência dos parâmetros penalizadores deve crescer ilimitadamente para podermos garantir convergência global. Problemas computacionais de mal-condicionamento aparecem quando temos que resolver problemas penalizados com valores grandes de ρ . Outra dificuldade associada aos métodos de penalização é a exigência de que os minimizadores globais devam ser calculados. Na prática, os algoritmos que resolvem os problemas penalizados geram, aproximadamente, uma seqüência de minimizadores locais ou uma seqüência de pontos estacionários.

A escolha da função de penalização L_1 ao longo do trabalho, deve-se ao fato da função ser exata, isto é, existe um limite inferior finito para o parâmetro de penalização, $\bar{\rho} > 0$, tal que para $\rho > \bar{\rho}$ qualquer minimizador local do problema original é também um minimizador local do problema penalizado. Este resultado é conhecido como Teorema Clássico da Penalização Exata L_1 , que na literatura está estabelecido somente para o caso onde os problemas penalizados são irrestritos. No capítulo anterior, estendemos o resultado para o caso onde os problemas penalizados estão restritos ao conjunto Ω . A importância do novo resultado é a possibilidade de desenvolver algoritmos para resolver (2.1), usando a função de penalização L_1 , onde os problemas penalizados permanecem restritos ao conjunto Ω . Desta forma, temos garantia de resolver uma seqüência finita de problemas penalizados, evitando assim o processo seqüencial infinito dos métodos de penalização clássicos.

A função de penalização L_1 é não-suave, e assim os problemas penalizados envolvidos são não-suaves e técnicas eficientes de otimização suave não podem ser aplicadas. Podemos resolver os problemas penalizados introduzindo variáveis adicionais e transformando-os em problemas suaves. Porém, os problemas transformados apresentarão um número superior de variáveis e restrições não lineares quando comparados com o problema original. Assim, devemos buscar uma forma alternativa

para resolver os problemas penalizados.

Os métodos de região de confiança geralmente apresentam bons resultados na resolução de problemas de otimização suave com restrições simples. A filosofia destes métodos é, a cada iteração, construir uma aproximação $\psi_\ell(x)$ para a função objetivo em torno do ponto atual x_ℓ . Como o modelo $\psi_\ell(x)$ deixa de ser representativo à medida que x se afasta de x_ℓ , então confiamos em aproximar a função objetivo por ψ_ℓ numa vizinhança de x_ℓ , ou seja, no conjunto $\{x \in \Omega \mid \|x - x_\ell\| \leq \Delta\}$ onde $\Delta > 0$ e $\|\cdot\|$ é uma norma qualquer em \mathbb{R}^n . Assim, um minimizador de $\psi_\ell(x)$ na região acima é uma boa aproximação para o minimizador da função objetivo original nesta região.

Devido ao sucesso e à facilidades dos métodos de região de confiança para o caso suave, a proposta para resolver (2.2) é definir uma função modelo,

$$\psi_\ell(x) = f(x) + \rho_k \left(\sum_{i=1}^{m_1} \phi_i(x, x_\ell) + \sum_{j=1}^{m_2} \gamma_j(x, x_\ell) \right) + \frac{1}{2}(x - x_\ell)^T B_\ell (x - x_\ell),$$

que aproxime a função $P_1(x, \rho_k)$ em torno de um ponto x_ℓ e aplicar a estrutura dos algoritmos de região de confiança. A função ψ_ℓ desempenha o papel do “modelo quadrático” da otimização suave. A questão é como definir $\phi_i(x, x_\ell)$, $\gamma_j(x, x_\ell)$ e B_ℓ . A função $P_1(x, \rho_k)$ pode ter derivadas descontínuas em pontos tais que $h_i(x) = 0$ ou $g_j(x) = 0$. Assim, para a definição de $\phi_i(x, x_\ell)$ e $\gamma_j(x, x_\ell)$ devemos considerar o comportamento das funções h_i e g_j na proximidade destes pontos. Para isto, usaremos constantes $p_1, p_2, p_3, p_4 > 0$ que fornecerão tolerâncias para avaliar tal comportamento. Existem algumas possibilidades naturais para definir $\phi_i(x, x_\ell)$ que envolvem a própria função $h_i(x)$ e sua aproximação linear no ponto x_ℓ . A saber:

$$\phi_i(x, x_\ell) = \begin{cases} -h_i(x), & \text{se } h_i(x_\ell) \leq -p_1 \\ h_i(x), & \text{se } h_i(x_\ell) \geq p_2 \\ |h_i(x)|, & \text{se } -p_1 \leq h_i(x_\ell) \leq p_2, \end{cases} \quad (2.4)$$

$$\phi_i(x, x_\ell) = \begin{cases} -\nabla h_i(x_\ell)^T (x - x_\ell) - h_i(x_\ell), & \text{se } h_i(x_\ell) \leq -p_1 \\ \nabla h_i(x_\ell)^T (x - x_\ell) + h_i(x_\ell), & \text{se } h_i(x_\ell) \geq p_2 \\ |h_i(x)|, & \text{se } -p_1 \leq h_i(x_\ell) \leq p_2, \end{cases} \quad (2.5)$$

$$\phi_i(x, x_\ell) = \begin{cases} -\nabla h_i(x_\ell)^T(x - x_\ell) - h_i(x_\ell), & \text{se } h_i(x_\ell) \leq -p_1 \\ \nabla h_i(x_\ell)^T(x - x_\ell) + h_i(x_\ell), & \text{se } h_i(x_\ell) \geq p_2 \\ |\nabla h_i(x_\ell)^T(x - x_\ell) + h_i(x_\ell)|, & \text{se } p_1 \leq h_i(x_\ell) \leq p_2, \end{cases} \quad (2.6)$$

$$\phi_i(x, x_\ell) = \begin{cases} -h_i(x), & \text{se } h_i(x_\ell) \leq -p_1 \\ h_i(x), & \text{se } h_i(x_\ell) \geq p_2 \\ |\nabla h_i(x_\ell)^T(x - x_\ell) + h_i(x_\ell)|, & \text{se } -p_1 \leq h_i(x_\ell) \leq p_2. \end{cases} \quad (2.7)$$

De forma análoga à função $\phi_i(x, x_\ell)$, podemos definir alguns modelos $\gamma_j(x, x_\ell)$, que envolvem a própria função $g_j(x)$ e a sua aproximação linear no ponto x_ℓ .

$$\gamma_j(x, x_\ell) = \begin{cases} 0, & \text{se } g_j(x_\ell) \leq -p_3 \\ g_j(x), & \text{se } g_j(x_\ell) \geq p_4 \\ \max\{0, g_j(x)\}, & \text{se } -p_3 \leq g_j(x_\ell) \leq p_4, \end{cases} \quad (2.8)$$

$$\gamma_j(x, x_\ell) = \begin{cases} 0, & \text{se } g_j(x_\ell) \leq -p_3 \\ \nabla g_j(x_\ell)^T(x - x_\ell) + g_j(x_\ell), & \text{se } g_j(x_\ell) \geq p_4 \\ \max\{0, g_j(x)\}, & \text{se } -p_3 \leq g_j(x_\ell) \leq p_4, \end{cases} \quad (2.9)$$

$$\gamma_j(x, x_\ell) = \begin{cases} 0, & \text{se } g_j(x_\ell) \leq -p_3 \\ \nabla g_j(x_\ell)^T(x - x_\ell) + g_j(x_\ell), & \text{se } g_j(x_\ell) \geq p_4 \\ \max\{0, \nabla g_j(x_\ell)^T(x - x_\ell) + g_j(x_\ell)\}, & \text{se } -p_3 \leq g_j(x_\ell) \leq p_4, \end{cases} \quad (2.10)$$

$$\gamma_i(x, x_\ell) = \begin{cases} 0, & \text{se } g_j(x_\ell) \leq -p_3 \\ g_j(x), & \text{se } g_j(x_\ell) \geq p_4 \\ \max\{0, \nabla g_j(x_\ell)^T(x - x_\ell) + g_j(x_\ell)\}, & \text{se } -p_3 \leq g_j(x_\ell) \leq p_4. \end{cases} \quad (2.11)$$

Note que, dependendo da proximidade do ponto x_ℓ com os pontos de não diferenciabilidade de $P_1(x, \rho_k)$, os modelos continuam não suaves. Segue abaixo uma ilustração gráfica dos modelos (2.8) – (2.11) considerando apenas uma restrição de desigualdade não linear. Alguns pontos do gráfico da função g estão representada por “o”, o gráfico da função γ está representado por traço contínuo e “*” representa o ponto x_ℓ .

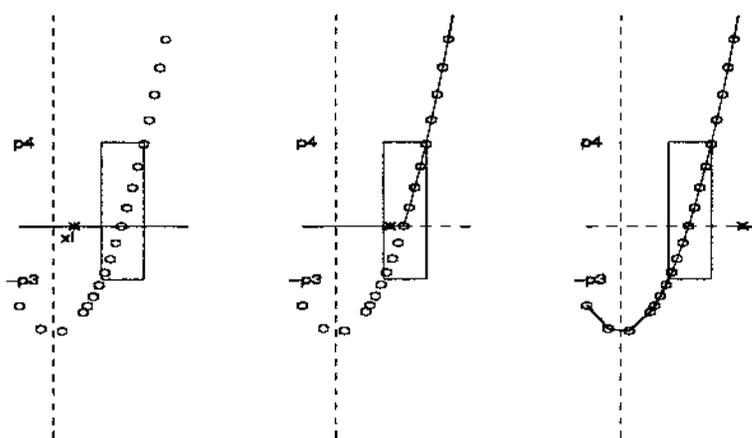


Figura 2.1: Modelo (2.8)

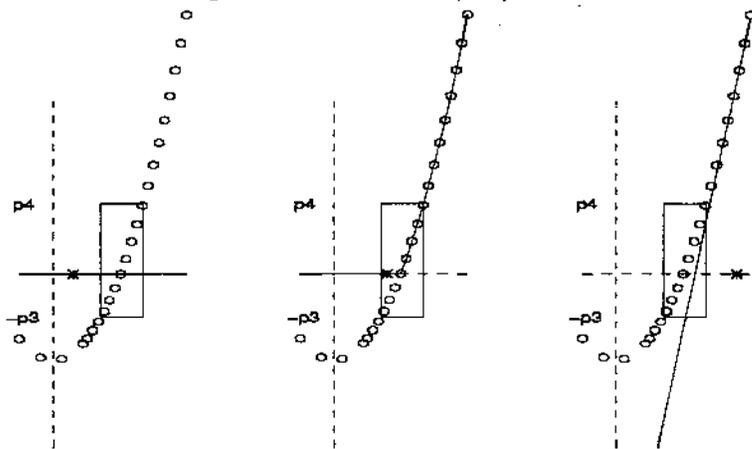


Figura 2.2: Modelo (2.9)

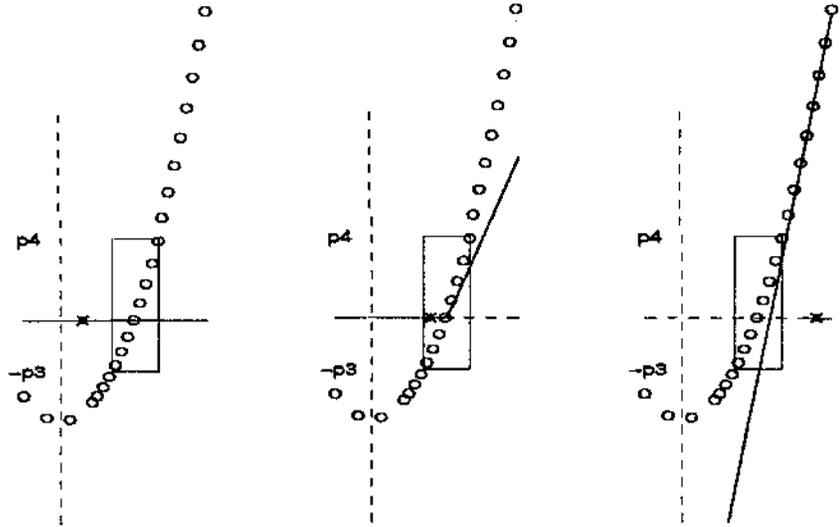


Figura 2.3: Modelo (2.10)

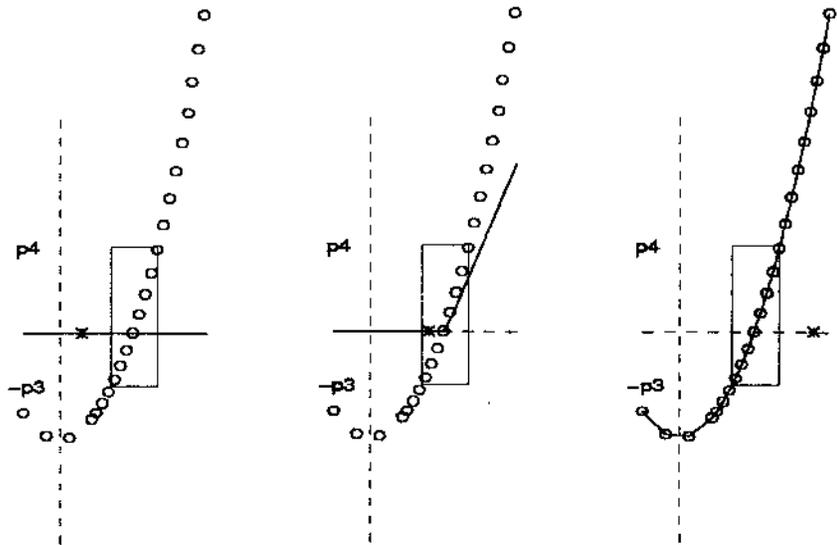


Figura 2.4: Modelo (2.11)

Por simplicidade de apresentação vamos considerar o problema original somente com restrições de desigualdade. Analisaremos os subproblemas que são envolvidos no algoritmo de região de confiança quando estabelecemos um dos modelos (2.8) – (2.11) para definir $\psi_\ell(x)$. Os objetivos da análise são verificar que tipos de subproblemas devemos resolver em cada um dos modelos e como são passadas as informações do problema original para cada subproblema.

Quando trabalhamos com o modelo (2.8) e o ponto atual está longe das regiões problemáticas (regiões próximas aos pontos de não diferenciabilidade da função $P_1(x, \rho_k)$), o subproblema que devemos resolver é um problema de otimização suave. No caso em que as restrições do problema original são fortemente satisfeitas ($g_j(x_\ell) \leq -p_3$), estas não aparecerão no subproblema. Se as restrições do problema original $g_j(x)$ são fracamente satisfeitas ($g_j(x_\ell) \geq p_4$) então as funções $g_j(x)$ são colocadas na função objetivo do subproblema. Assim, a não linearidade das restrições do problema original ficarão na função objetivo do subproblema. Portanto, os subproblemas envolvidos nestes casos são problemas com função objetivo não linear e as restrições definidas pelo conjunto Ω e pela região de confiança. Ainda considerando o modelo (2.8), se o ponto atual estiver em alguma região que satisfaça $-p_3 \leq g_j(x_\ell) \leq p_4$ então a restrição $g_j(x)$ fará parte da função objetivo do subproblema. Neste caso, o subproblema torna-se um problema de otimização não-suave. Se a condição $-p_3 \leq g_j(x_\ell) \leq p_4$ for satisfeita para todo $j = 1, \dots, m_2$ então o subproblema envolvido tem a mesma complexidade do problema penalizado que desejamos resolver, o que tornaria o modelo pouco econômico. Porém, na prática este comportamento é pouco provável. Assim, o subproblema não-suave apresenta uma quantidade menor de pontos não diferenciáveis quando comparado com o problema penalizado inicial. Desta forma, quando resolvemos o subproblema não-suave acrescentando variáveis adicionais, carregamos para o subproblema somente as restrições não lineares do problema original que satisfazem a condição $-p_3 \leq g_j(x_\ell) \leq p_4$. Portanto, o subproblema que devemos resolver é um problema de programação não linear da forma do problema original com um número inferior de restrições de desigualdade não lineares.

O modelo (2.9) tem praticamente o mesmo comportamento do modelo (2.8), porém quando as restrições do problema original são fracamente satisfeitas $g_j(x_\ell) \geq p_4$, ao invés de trabalharmos com a própria função $g_j(x)$ usamos uma aproximação linear da função $g_j(x)$ em torno do ponto atual x_ℓ . Como a função objetivo do problema original é não linear, a linearização das funções $g_j(x)$ não diminui a com-

plexidade dos subproblemas que devem ser resolvidos. A linearização das funções $g_j(x)$ somente provoca perda de informações destas funções. Assim, de forma geral, os subproblemas envolvidos pelo modelo (2.9) são problemas de programação não linear como no modelo (2.8).

O modelo (2.10) trabalha com aproximação linear das restrições $g_j(x)$ do problema original, indiferente ao ponto atual estar ou não próximo a pontos de não diferenciabilidade da função $P_1(x, \rho_k)$. No caso em que o ponto atual x_ℓ está longe das regiões problemáticas, a linearização das funções $g_j(x)$ ocasiona perda de informações das funções $g_j(x)$ e não diminui a complexidade do subproblema a ser resolvido. Neste caso, os subproblemas que devemos resolver são problemas de otimização não linear com restrições lineares e canalizações. No caso em que o ponto atual satisfaz $-p_3 \leq g_j(x_\ell) \leq p_4$ então as funções $g_j(x)$ são linearizadas e aparecem na função objetivo do subproblema, que será não-suave. Ao resolvermos o subproblema não-suave usando variáveis adicionais, o subproblema é transformado em um problema com função objetivo não linear, restrições lineares, restrições definidas pelo conjunto Ω e pela região de confiança. Assim, a perda de informações ocorrida com a linearização das funções $g_j(x)$ que satisfazem a condição $-p_3 \leq g_j(x_\ell) \leq p_4$, é compensada pela vantagem do subproblema permanecer restrito a um conjunto de restrições simples.

Quando adotamos o modelo (2.11) para definir $\psi_\ell(x)$, estamos somente linearizando as restrições $g_j(x)$ do problema original que farão parte das restrições dos subproblemas. Assim, evitamos a linearização de funções $g_j(x)$ que ficam na função objetivo dos subproblemas e conseqüentemente evitamos a perda de informações destas funções. Os subproblemas envolvidos neste modelo após acrescentar as variáveis adicionais (quando necessárias) são problemas de programação não linear com restrições lineares e canalizações.

Queremos enfatizar que, independentemente do modelo escolhido para definir $\psi_\ell(x)$, a filosofia dos modelos é não trabalhar com todas as restrições $g_j(x)$ do problema original nos subproblemas do algoritmo de região de confiança. Dependendo da posição do ponto atual, as funções $g_j(x)$ ou suas aproximações lineares são colocadas nas funções objetivo dos subproblemas, ou nas restrições destes, ou não aparecem nos subproblemas. Enfatizamos também que a definição do melhor modelo depende do problema original e dos algoritmos disponíveis para resolver os subproblemas.

A seguir, apresentamos o algoritmo de região de confiança para resolver os problemas penalizados. Note que ρ_k e x_k são fornecidos pelo problema penalizado.

Algoritmo 2 - Região de confiança

Inicialização:

Fixar $\underline{\Delta} > 0$, $\alpha \in (0, 1)$, $\ell = 1$, $x_\ell = x_k$.

Passo 1:

Escolha $\Delta \geq \underline{\Delta}$ e $B_\ell = B_\ell^T$. Defina

$$\psi_\ell(x) = f(x) + \rho_k \sum_{i=1}^{m_1} \phi_i(x, x_\ell) + \rho_k \sum_{j=1}^{m_2} \gamma_j(x, x_\ell) + \frac{1}{2}(x - x_\ell)^T B_\ell (x - x_\ell),$$

Passo 2:

Resolva aproximadamente

$$\begin{aligned} \min \quad & \psi_\ell(x) \\ \text{s.a} \quad & x \in \Omega, \\ & \|x - x_\ell\| \leq \Delta \end{aligned} \tag{2.12}$$

e obtenha \bar{x} .

Passo 3: Aceita ou rejeita o ponto

Se

$$P_1(\bar{x}, \rho_k) \leq P_1(x_\ell, \rho_k) + \alpha(\psi_\ell(\bar{x}) - \psi_\ell(x_\ell))$$

então $x_{\ell+1} = \bar{x}$, e vá ao Passo 4. Caso contrário, escolha $\Delta_{\text{novoo}} \in [0.1\|\bar{x} - x_\ell\|, 0.9\Delta]$, $\Delta = \Delta_{\text{novoo}}$ e vá ao Passo 2.

Passo 4: Critério de parada

Se

$$|P_1(x_\ell, \rho_k) - P_1(x_{\ell+1}, \rho_k)| \leq 0.01|P_1(x_{\ell+1}, \rho_k)|$$

então pare, $x_{\ell+1}$ é solução aproximada do problema penalizado. Caso contrário, vá ao Passo 1.

As matrizes B_ℓ são escolhidas simétricas. Em muitos casos é conveniente usar $B_\ell = 0$ em todas as iterações, pois isso simplifica a resolução dos subproblemas do algoritmo de região de confiança.

Queremos ressaltar que a escolha dos modelos $\phi_i(x, x_\ell)$ e $\gamma_j(x, x_\ell)$ depende dos algoritmos disponíveis para resolver os subproblemas do algoritmo de região de confiança. Se temos disponível um bom algoritmo que resolva problemas de minimização de função objetivo não linear com restrições lineares, os modelos (2.7) ou (2.11) são mais adequados. Para aplicarmos os modelos (2.4) ou (2.8) precisamos de um algoritmo que resolva problemas gerais de programação não linear. Note que se aplicarmos o algoritmo não linear diretamente ao problema original, temos a desvantagem de trabalhar com todas as restrições não lineares. Ao passo que se aplicarmos o algoritmo não linear aos subproblemas do algoritmo de região de confiança, vamos trabalhar com um número inferior de restrições não lineares do problema original, o que pode significar uma vantagem quando o número de restrições não lineares é grande.

O Algoritmo 2 explica como o Passo 1 do Algoritmo 1 é executado. O procedimento abaixo explica como executar o Passo 2 do Algoritmo 2.

Vamos assumir que os modelos (2.4) e (2.8) foram selecionados para definir a função $\psi_\ell(x)$. O primeiro passo para resolver (2.12) é transformá-lo em um problema suave, introduzido variáveis adicionais. Assim, o problema transformado que devemos resolver aproximadamente é:

$$\begin{aligned}
\text{minimizar } \bar{\psi}(x, y, z) &= f(x) + \rho_k \left(\sum_{i=1}^{n_1} y_i + \sum_{i=1+n_1}^{n_2} h_i(x) - \sum_{i=n_2+1}^{n_3} h_i(x) \right) + \\
&\rho_k \left(\sum_{j=1}^{n_4} z_j + \sum_{j=1+n_4}^{n_5} g_j(x) \right) + \frac{1}{2} (x - x_\ell)^T B_\ell (x - x_\ell) \\
\text{sujeita a } & y_i \geq h_i(x), \quad i = 1, \dots, n_1 \\
& y_i \geq -h_i(x), \quad i = 1, \dots, n_1 \\
& z_j \geq 0, \quad j = 1, \dots, n_4 \\
& z_j \geq g_j(x), \quad j = 1, \dots, n_4 \\
& x \in \Omega \\
& \|x - x_\ell\| \leq \Delta,
\end{aligned} \tag{2.13}$$

onde, sem perda de generalidade, $-p_1 \leq h_i(x_\ell) \leq p_2$ para $i = 1, \dots, n_1$, $h_i(x_\ell) \geq p_2$ para $i = n_1 + 1, \dots, n_2$, $h_i(x_\ell) \leq -p_1$ para $i = 1 + n_2, \dots, n_3$, $-p_3 \leq g_j(x_\ell) \leq p_4$ para $j = 1, \dots, n_4$, e $g_j(x_\ell) \geq p_4$ para $j = 1 + n_4, \dots, n_5$. É pouco provável que todas as restrições não lineares do problema original façam parte do conjunto de restrições do problema acima, dado que dificilmente teremos as condições $-p_1 \leq h_i(x_\ell) \leq p_2$ e $-p_3 \leq g_j(x_\ell) \leq p_4$ satisfeitas para todo $i = 1, \dots, m_1$ e para todo $j = 1, \dots, m_2$. Assim, o problema (2.13) apresentará um número menor de restrições não lineares quando comparado com o problema original.

Para resolvermos (2.13) basta usarmos o método do Lagrangeano aumentado com o algoritmo BOX-QUACAN. No que segue, descrevemos as filosofias dos algoritmos do Lagrangeano aumentado e BOX-QUACAN.

Um dos métodos clássicos para resolver o problema de programação não linear é o método do Lagrangeano aumentado proposto independentemente por Hestenes (1969) e Powell (1969). O método do Lagrangeano aumentado pode ser visto como uma combinação de funções de penalização e métodos de dualidade local. Para exemplificar o método do Lagrangeano aumentado, consideremos o seguinte problema de otimização:

$$\begin{aligned}
\text{minimizar } & \varphi(x) \\
\text{sujeita a } & \omega(x) = 0
\end{aligned} \tag{2.14}$$

onde $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}, \omega : \mathbb{R}^n \rightarrow \mathbb{R}^p, \varphi$ e $\omega \in C^1$.

A função Lagrangeana aumentada associada ao problema (2.14) é a função

$$L_\alpha(x, \lambda) = \varphi(x) + \lambda^T \omega(x) + \frac{1}{2} \alpha \|\omega(x)\|_2^2$$

para alguma constante α . O algoritmo típico de método do Lagrangeano aumentado é descrito como segue.

Algoritmo 3 - Lagrangeano Aumentado

Dados $x_0 \in \mathbb{R}^n$, $\alpha_1 > 0$, $\lambda_1 \in \mathbb{R}^p$, $k = 1$.

Passo 1: Resolva o problema irrestrito

$$\min L_{\alpha_k}(x, \lambda_k),$$

começando com x_{k-1} como ponto inicial e obtendo x_k como solução deste problema.

Passo 2: Se $\|\omega(x_k)\| > 0.1\|\omega(x_{k-1})\|$ então $\alpha_k = 10\alpha_k$.

Passo 3: Atualize λ_{k+1} , $\alpha_{k+1} = \alpha_k$, $k = k + 1$ e volte ao Passo 1.

O algoritmo BOX-QUACAN foi desenvolvido por Friedlander, Martínez e Santos (1992) para resolver problemas de otimização de grande porte com restrições de canalização, isto é, problemas da forma

$$\begin{array}{ll} \text{minimizar} & L(x) \\ \text{sujeita a} & l \leq x \leq u. \end{array} \quad (2.15)$$

onde $L : \mathbb{R}^n \rightarrow \mathbb{R}$, $L \in C^1$, $l, u \in \mathbb{R}^n$ com $l < u$.

O algoritmo BOX-QUACAN consiste em resolver (2.15) através de técnicas de região de confiança. A seguir descrevemos os passos do algoritmo.

Algoritmo 4 - BOX-QUACAN

Sejam $\Delta_{min} > 0$ e $\alpha \in (0, 1)$.

Dado x_k tal que $l \leq x_k \leq u$, obter x_{k+1} da seguinte maneira:

Passo 1: Escolher $\Delta \geq \Delta_{min}$ e $\bar{B}_k = \bar{B}_k^T$,
obter M_k tal que $\|\bar{B}_k\| \leq M_k$.

Passo 2: Resolver o problema fácil

Calcule uma solução global $s_k^Q(\Delta)$ para

$$\begin{aligned} & \text{minimizar} && Q_k(s) \equiv \nabla L(x_k)^T s + \frac{1}{2} M_k \|s\|^2 \\ & \text{sujeita a} && l \leq x_k + s \leq u \\ & && \|s\| \leq \Delta. \end{aligned} \tag{2.16}$$

Se $Q_k(s_k^Q(\Delta)) = 0$, PARE.

Passo 3: Calcule $\bar{s}_k(\Delta)$ tal que

$$\begin{aligned} \phi_k(\bar{s}_k(\Delta)) &\leq 0.1 Q_k(s_k^Q(\Delta)), \\ l &\leq x_k + \bar{s}_k(\Delta) \leq u, \\ \|\bar{s}_k(\Delta)\| &\leq \Delta \end{aligned}$$

onde

$$\phi_k(s) = \frac{1}{2} s^T \bar{B}_k s + \nabla L(x_k)^T s$$

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

Passo 4: Se

$$L(x_k + \bar{s}_k(\Delta)) \leq L(x_k) + \alpha \phi_k(\bar{s}_k(\Delta))$$

então defina $w_{k+1} = x_k + \bar{s}_k(\Delta)$, $\Delta_k = \Delta$,

escolha $\Delta_{k+1} \geq \Delta_{min}$, $\bar{B}_{k+1} \in \mathbb{R}^{n \times n}$, $\bar{B}_{k+1}^T = \bar{B}_{k+1}$ e retorne ao Passo 1 senão,
 $\Delta = \Delta_{novo}$ onde

$$\Delta_{novo} \in [0.1 \|\bar{s}_k\|, 0.9 \|\Delta\|]$$

e volte para o Passo 2.

Anteriormente, sugerimos resolver o problema (2.13) usando o método do Lagrangeano aumentado com BOX-QUACAN. A idéia é aplicar o método do Lagrangeano aumentado para as restrições não lineares e lineares, caso o conjunto Ω contenha restrições deste tipo, manter os subproblemas com restrições de canalização e aplicar o algoritmo BOX-QUACAN para resolver os subproblemas restritos.

O processo descrito para resolver o problema (2.13) pode ser aplicado diretamente ao problema original, porém todas as restrições não lineares do problema original são carregadas nos subproblemas. A idéia de aplicar o método do Lagrangeano aumentado com BOX-QUACAN somente nos subproblemas do algoritmo 2 (algoritmo de região de confiança) é não trabalhar com todas as restrições não lineares do problema original.

Capítulo 3

Visualização através de problemas simples

No capítulo anterior apresentamos o algoritmo de penalização L_1 , o Algoritmo 1, que envolve uma seqüência de subproblemas restritos não suaves e o algoritmo de região de confiança, o Algoritmo 2, que resolve os subproblemas penalizados. Neste capítulo o objetivo é ilustrar os algoritmos através de problemas simples, exibindo os subproblemas envolvidos em cada iteração dos algoritmos. Os problemas considerados foram extraídos do conjunto de problemas testes de Hock e Schittkowski [25] e de Bazaraa [2]. Os pontos iniciais utilizados foram os sugeridos pelos autores. Em alguns casos, os pontos iniciais sugeridos são ineficazes com respeito ao conjunto de restrições simples dos problemas considerados.

Para ilustrar os Algoritmos 1 e 2 precisamos definir alguns parâmetros. No Algoritmo 1, usaremos o parâmetro de penalização inicial igual a 1 e o atualizaremos multiplicando-o por 10. Consideraremos um ponto como factível quando a norma 1 das restrições não lineares do problema original for menor ou igual a 10^{-5} . No Algoritmo 2 escolheremos o modelo (2.8) para aproximar a função penalizada. As constantes de tolerâncias usadas na definição do modelo serão tomadas iguais a 1. Ainda, no algoritmo de região de confiança usaremos $B_t \equiv 0$ e $\underline{\Delta} = 10$. A norma escolhida para definir a região de confiança é a $\|\cdot\|_\infty$. Esta escolha deve-se ao fato de

que a interseção da caixa original com a caixa de confiança também resulta numa caixa.

Problema 1:

Consideremos o problema

$$\begin{aligned} &\text{minimizar} && f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2 \\ &\text{sujeita a} && g_1(x) = -x_1 - x_2^2 \leq 0 \\ &&& g_2(x) = -x_1^2 - x_2 \leq 0 \\ &&& -0.5 \leq x_1 \leq 0.5 \\ &&& x_2 \leq 1. \end{aligned} \tag{3.1}$$

A solução ótima para este problema é $(0.5, 0.25)$.

Resolução gráfica

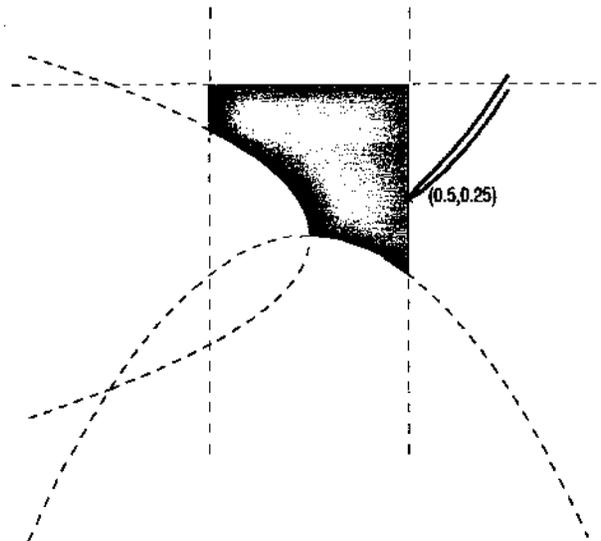


Figura 3.1:

Ao aplicar o Algoritmo 1 para resolver o problema acima, o primeiro subpro-

blema penalizado que devemos resolver é

$$\begin{aligned}
 &\text{minimizar } P_1(x, \rho_1) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2 + 1 \max\{0, -x_1 - x_2^2\} \\
 &\quad \quad \quad + 1 \max\{0, -x_1^2 - x_2\} \\
 &\text{sujeita a } \quad -0.5 \leq x_1 \leq 0.5 \\
 &\quad \quad \quad x_2 \leq 1.
 \end{aligned} \tag{3.2}$$

Resolução gráfica

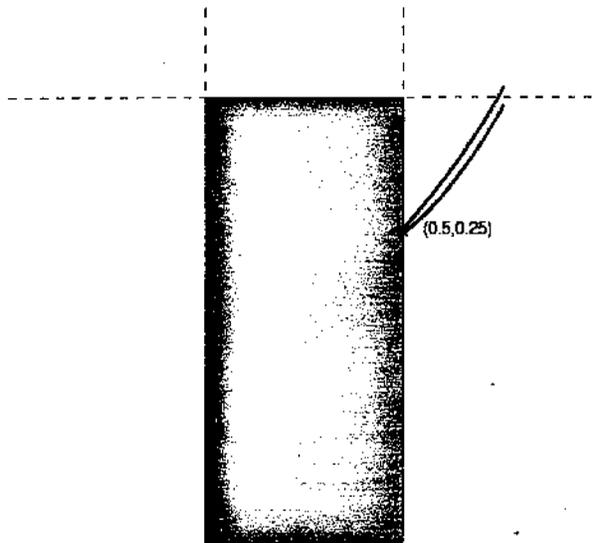


Figura 3.2:

Observe que $(0.5, 0.25)$ é solução deste problema. Para resolver o subproblema penalizado aplicando o algoritmo proposto, Algoritmo 2, e usando $(-2, 1)$ como ponto inicial, devemos considerar um modelo para aproximar a função objetivo $P_1(x, \rho_1)$ que seja válido próximo ao ponto $x^0 = (-2, 1)$. Agora,

$$g_1(-2, 1) = 1 \quad \text{e} \quad g_2(-2, 1) = -5.$$

Logo, o modelo para representar a penalização associada à $g_1(x)$ é $\max\{0, g_1(x)\}$ e o modelo para representar a penalização associada à $g_2(x)$ é 0. Portanto, o primeiro subproblema do algoritmo de região de confiança que devemos resolver é

$$\begin{aligned}
 &\text{minimizar } 100(x_2 - x_1^2)^2 + (1 - x_1)^2 + 1 \max\{0, -x_1 - x_2^2\} \\
 &\text{sujeita a } \quad -0.5 \leq x_1 \leq 0.5 \\
 &\quad \quad \quad x_2 \leq 1 \\
 &\quad \quad \quad \|x - x^0\| \leq 10.
 \end{aligned} \tag{3.3}$$

Resolução gráfica

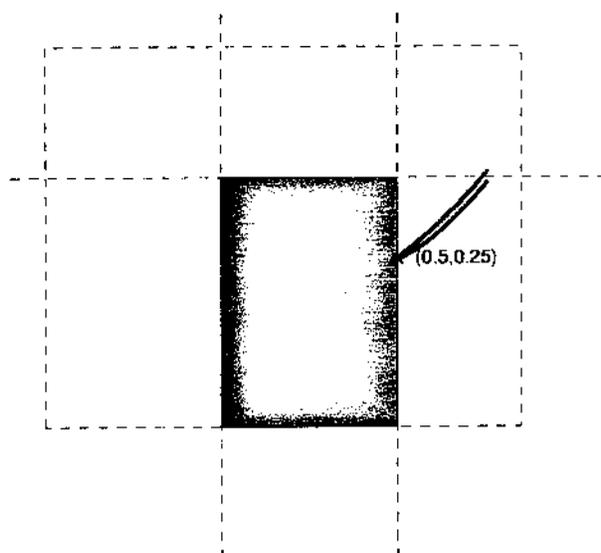


Figura 3.3:

Podemos transformar o problema (3.3) em um problema diferenciável, a saber:

$$\begin{aligned} &\text{minimizar} && 100(x_2 - x_1^2)^2 + (1 - x_1)^2 + 1z_1 \\ &\text{sujeita a} && z_1 \geq -x_1 - x_2^2 \\ &&& z_1 \geq 0 \\ &&& -0.5 \leq x_1 \leq 0.5 \\ &&& x_2 \leq 1 \\ &&& \|x - x^0\| \leq 10. \end{aligned} \tag{3.4}$$

A solução de (3.3) é $(0.5, 0.25)$. Claramente esta é a solução do problema penalizado e também do problema original. Entretanto, não podemos detectar isto ainda. Podemos garantir somente que $(0.5, 0.25)$ representa uma melhor aproximação para solução do subproblema penalizado. Assim devemos continuar resolvendo (3.2), para isto montamos um modelo para representar a função objetivo $P_1(x, \rho_1)$ próximo ao novo ponto $x^1 = (0.5, 0.25)$. Agora,

$$g_1(0.5, 0.25) = -0.5625 \quad \text{e} \quad g_2(0.5, 0.25) = -0.5.$$

Logo, os modelos para representarem as penalizações associadas às $g_1(x)$ e $g_2(x)$ são $\max\{0, g_1(x)\}$ e $\max\{0, g_2(x)\}$. Portanto o modelo para aproximar a função objetivo

de (3.2) é ela mesma, e o próximo subproblema que devemos resolver é

$$\begin{aligned}
 &\text{minimizar} && 100(x_2 - x_1^2)^2 + (1 - x_1)^2 + 1 \max\{0, -x_1 - x_2^2\} + \\
 & && 1 \max\{0, -x_1^2 - x_2\} \\
 &\text{sujeita a} && -0.5 \leq x_1 \leq 0.5 \\
 & && x_2 \leq 1 \\
 & && \|x - x^1\| \leq 10.
 \end{aligned} \tag{3.5}$$

Resolução gráfica

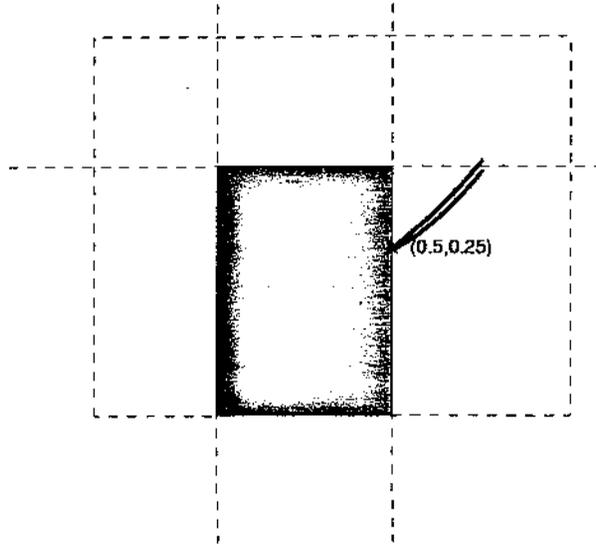


Figura 3.4:

Transformando (3.5) em um problema diferenciável temos:

$$\begin{aligned}
 &\text{minimizar} && 100(x_2 - x_1^2)^2 + (1 - x_1)^2 + 1z_1 + 1z_2 \\
 &\text{sujeita a} && z_1 \geq -x_1 - x_2^2 \\
 & && z_2 \geq -x_1^2 - x_2 \\
 & && -0.5 \leq x_1 \leq 0.5 \\
 & && x_2 \leq 1 \\
 & && z_1 \geq 0 \\
 & && z_2 \geq 0 \\
 & && \|x - x^1\| \leq 10.
 \end{aligned} \tag{3.6}$$

Note que apesar da função objetivo do subproblema (3.5) ser exatamente a mesma função objetivo do problema que desejamos resolver, o conjunto de restrições de (3.5) inclui a caixa de confiança.

A solução de (3.5) coincide com a solução do primeiro subproblema do algoritmo de região de confiança. Logo, uma solução do problema penalizado foi encontrada. Uma vez obtida uma solução do subproblema penalizado, devemos continuar aplicando o algoritmo de penalização, testando se a solução do subproblema penalizado é factível com respeito o problema original. Temos que $(0.5, 0.25)$ é factível. Portanto, o algoritmo de penalização pára com uma solução do problema original.

Queremos salientar que as escolhas das constantes p_3 e p_4 necessárias para definir o modelo para a função objetivo não suave, determinam os subproblemas envolvidos no algoritmo de região de confiança. Vamos ilustrar este fato analisando os subproblemas envolvidos no algoritmo de região de confiança, resolvendo o subproblema penalizado (3.2) usando o mesmo ponto inicial $x^0 = (-2, 1)$ e tomando p_3 e p_4 iguais à 0.1. Inicialmente devemos definir um modelo para função objetivo de (3.2) próximo ao ponto $(-2, 1)$. Agora,

$$g_1(-2, 1) = 1 \quad \text{e} \quad g_2(-2, 1) = -5.$$

Logo, os modelos para representarem as penalizações associadas à $g_1(x)$ e $g_2(x)$ com os novos valores de p_3 e p_4 são respectivamente $g_1(x)$ e 0. Portanto, o primeiro subproblema neste caso é

$$\begin{aligned} &\text{minimizar} && 100(x_2 - x_1^2)^2 + (1 - x_1)^2 + 1(-x_1 - x_2^2) \\ &\text{sujeita a} && -0.5 \leq x_1 \leq 0.5 \\ &&& x_2 \leq 1 \\ &&& \|x - x^0\| \leq 10. \end{aligned} \tag{3.7}$$

A solução deste subproblema é $(0.5, 0.25)$. Novamente, apesar da solução do subproblema ser a solução do problema penalizado, ainda não podemos detectar isto. Podemos garantir que $(0.5, 0.25)$ representa uma melhor aproximação para solução do subproblema penalizado. Assim, devemos continuar resolvendo (3.2) definindo um

modelo válido para função objetivo de (3.2) próximo ao ponto $x^1 = (0.5, 0.25)$. Agora,

$$g_1(0.5, 0.25) = -0.5625 \quad \text{e} \quad g_2(0.5, 0.25) = -0.5.$$

Com p_3 e p_4 iguais a 0.1, os modelos para representarem as penalizações associadas a $g_1(x)$ é 0 e a $g_2(x)$ é também 0. Portanto, o próximo subproblema do algoritmo de região de confiança é:

$$\begin{aligned} &\text{minimizar} && 100(x_2 - x_1^2)^2 + (1 - x_1)^2 \\ &\text{sujeita a} && -0.5 \leq x_1 \leq 0.5 \\ &&& x_2 \leq 1 \\ &&& \|x - x^1\| \leq 10. \end{aligned} \tag{3.8}$$

A solução de (3.8) é a mesma do problema (3.7). Portanto, uma solução do problema penalizado (3.2) foi encontrada.

Com os novos valores de p_3 e p_4 , os subproblemas envolvidos para resolver o problema penalizado (3.2) foram subproblemas mais simples, subproblemas diferenciáveis com restrições apenas de canalizações. Ao passo que os subproblemas diferenciáveis para resolver o problema penalizado (3.2) usando p_3 e p_4 iguais a 1, os problemas (3.4) e (3.7), envolveram restrições não lineares.

Problema 2:

Consideremos o problema

$$\begin{aligned} &\text{minimizar} && f(x) = x_1 + x_2 \\ &\text{sujeita a} && g_1(x) = x_1^2 + x_2^2 - 1 \leq 0 \\ &&& x_1, x_2 \geq 0. \end{aligned} \tag{3.9}$$

Resolução gráfica

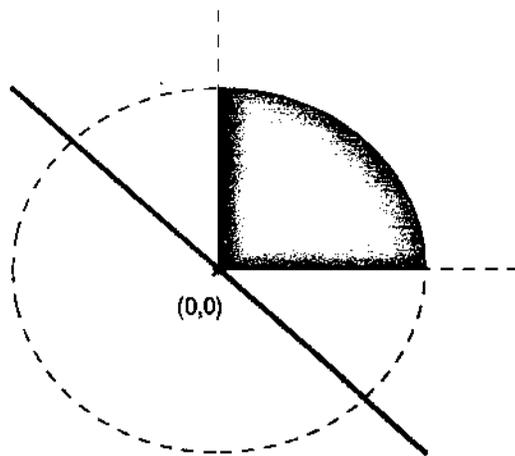


Figura 3.5:

Claramente, a solução deste problema é $(0,0)$. Aplicando o Algoritmo 1 para resolver (3.9), o primeiro subproblema penalizado associado é

$$\begin{aligned} &\text{minimizar} && P_1(x, \rho_1) = x_1 + x_2 + 1 \max\{0, x_1^2 + x_2^2 - 1\} \\ &\text{sujeita a} && x_1, x_2 \geq 0. \end{aligned} \quad (3.10)$$

Resolução gráfica

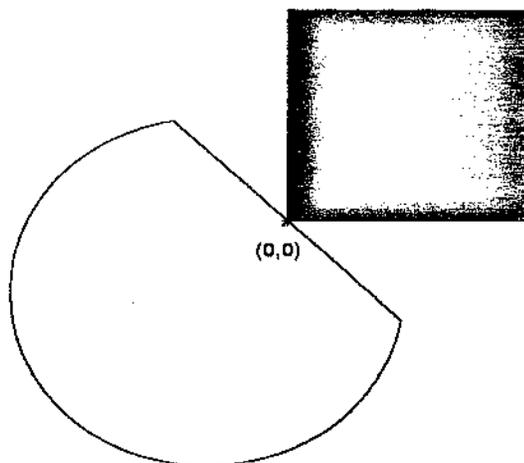


Figura 3.6:

A solução deste problema é $(0, 0)$. Para resolver (3.10) aplicando o Algoritmo 2 e usando $(1, 1)$ como ponto inicial, precisamos considerar um modelo para a função objetivo de (3.10) que seja válido próximo ao ponto $x^0 = (1, 1)$. Assim,

$$g_1(1, 1) = 1.$$

Neste caso, o modelo para representar a penalização associada à $g_1(x)$ é $\max\{0, g_1(x)\}$. Portanto, o modelo para a função objetivo de (3.10) é ela mesma e o subproblema associado é

$$\begin{aligned} &\text{minimizar} && x_1 + x_2 + 1 \max\{0, x_1^2 + x_2^2 - 1\} \\ &\text{sujeita a} && \|x - x^0\| \leq 10 \\ &&& x_1, x_2 \geq 0. \end{aligned} \tag{3.11}$$

A solução deste subproblema é $(0, 0)$. O valor da função objetivo do problema penalizado neste ponto é suficientemente menor do que o valor da função no ponto inicial $x^0 = (1, 1)$. Isto significa que $(0, 0)$ representa uma melhor aproximação para solução do problema penalizado. Mas ainda não podemos confirmar $(0, 0)$ como solução do problema penalizado, pois o critério de convergência estabelecido pelo Algoritmo 2 não é satisfeito. Assim, devemos continuar resolvendo o problema penalizado montando um modelo válido perto do novo ponto $x^1 = (0, 0)$. Agora,

$$g_1(0, 0) = -1.$$

Novamente, o modelo para representar a penalização associada à $g_1(x)$ é $\max\{0, g_1(x)\}$ e portanto o subproblema associado é

$$\begin{aligned} &\text{minimizar} && x_1 + x_2 + 1 \max\{0, x_1^2 + x_2^2 - 1\} \\ &\text{sujeita a} && \|x - x^1\| \leq 10 \\ &&& x_1, x_2 \geq 0. \end{aligned} \tag{3.12}$$

A solução de (3.12) é a mesma solução do subproblema anterior. Logo, uma solução do subproblema penalizado (3.10) foi encontrada pelo Algoritmo 2. Como $(0, 0)$ é factível com respeito o problema original, então uma solução ótima do problema original foi encontrada.

Como no Problema 1, vamos resolver o problema penalizado (3.10) aplicando o Algoritmo 2 e usando as constantes p_3 e p_4 iguais a 0.1. O objetivo é analisar os subproblemas envolvidos.

Aplicando o Algoritmo 2 para resolver (3.10), vamos montar um modelo para a função objetivo de (3.10) perto do ponto $x^0 = (1, 1)$. Temos

$$g_1(1, 1) = 1.$$

Com os novos valores de p_3 e p_4 , o modelo para representar a penalização associada à $g_1(x)$ é a própria $g_1(x)$. Logo, o subproblema associado é

$$\begin{aligned} &\text{minimizar} && x_1 + x_2 + 1(x_1^2 + x_2^2 - 1) \\ &\text{sujeita a} && \|x - x^0\| \leq 10 \\ &&& x_1, x_2 \geq 0. \end{aligned} \tag{3.13}$$

Resolução gráfica

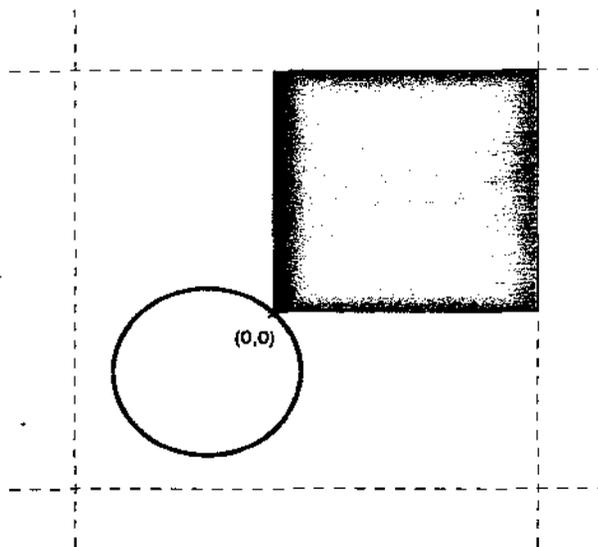


Figura 3.7:

A solução deste subproblema é $(0,0)$. Esta solução faz decrescer suficientemente o valor da função objetivo do problema penalizado quando comparado com o valor da função objetivo no ponto inicial $(1, 1)$. Assim, devemos aceitar o ponto $(0,0)$ como uma melhor aproximação para solução do problema penalizado. Claramente, $(0,0)$ é a solução do problema penalizado, mas não podemos garantir isto ainda. Devemos continuar resolvendo o problema penalizado definindo um modelo para aproximar $P_1(x, \rho_1)$ perto do novo ponto $x^1 = (0,0)$. Temos

$$g_1(0, 0) = -1.$$

Logo, o modelo para representar a penalização associada à $g_1(x)$ é 0 e o subproblema associado é

$$\begin{aligned} & \text{minimizar} && x_1 + x_2 \\ & \text{sujeita a} && \|x - x^1\| \leq 10 \\ & && x_1, x_2 \geq 0. \end{aligned} \tag{3.14}$$

Resolução gráfica

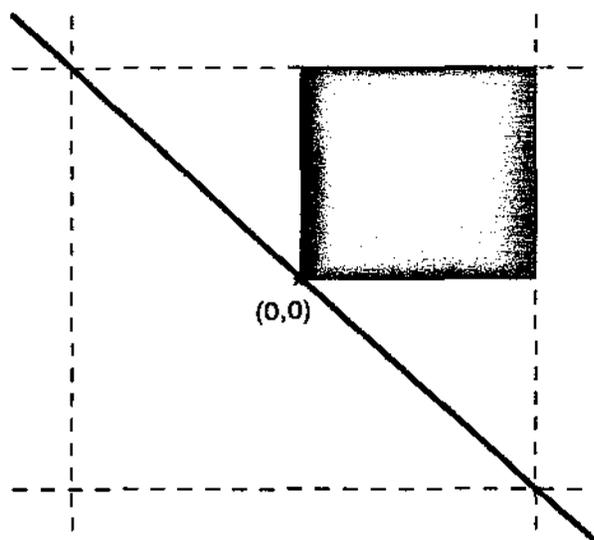


Figura 3.8:

A solução de (3.14) é a mesma do subproblema (3.13), isto é $(0,0)$. Portanto, uma solução do problema penalizado foi encontrada.

Comparando as figuras 3.6 e 3.7 temos que a solução ótima da função penalizada (figura 3.6) coincide com a solução ótima da função modelo (figura 3.7). Notamos que perto do ponto ótimo, o comportamento da função penalizada e da função modelo difere bastante. Portanto a função modelo usada para aproximar a função penalizada deve ser melhorada. Agora, comparando as figuras 3.6 e 3.8 temos que as soluções ótimas coincidem e o comportamento da função modelo perto da solução ótima é o mesmo da função penalizada. Desta forma temos que a função modelo da figura 3.8 representa uma boa aproximação para a função penalizada.

Quando resolvemos o subproblema penalizado (3.10) usando as constantes p_3 e p_4 iguais a 1, os subproblemas envolvidos foram muito parecidos com o problema que desejávamos resolver. Assim, não houve vantagem significativa na forma proposta para resolver o subproblema penalizado. Agora, quando usamos p_3 e p_4 iguais a 0.1, os subproblemas envolvidos foram subproblemas diferenciáveis com restrições simples. Portanto, as escolhas das constantes p_3 e p_4 influenciam diretamente na determinação dos subproblemas e conseqüentemente na sua simplificação. Logo, escolhas adequadas para as constantes p_3 e p_4 devem ser consideradas.

Problema 3:

Consideremos o problema

$$\begin{aligned}
 &\text{minimizar} && f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2 \\
 &\text{sujeita a} && g_1(x) = x_1 - x_2^2 \leq 0 \\
 &&& g_2(x) = -x_1^2 + x_2 \leq 0 \\
 &&& -0.5 \leq x_1 \leq 0.5 \\
 &&& x_2 \leq 1.
 \end{aligned} \tag{3.15}$$

Resolução gráfica

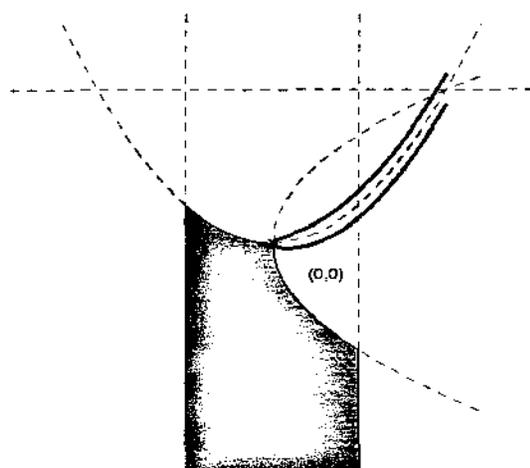


Figura 3.9:

A solução deste problema é $(0,0)$. Aplicando o Algoritmo 1 para resolver (3.15)

devemos penalizar $g_1(x)$ e $g_2(x)$. Assim, o primeiro subproblema penalizado é:

$$\begin{aligned} \text{minimizar} \quad & P_1(x, \rho_1) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2 + 1 \max\{0, x_1 - x_2\} \\ & + 1 \max\{0, -x_1^2 + x_2\} \\ \text{sujeita a} \quad & -0.5 \leq x_1 \leq 0.5 \\ & x_2 \leq 1. \end{aligned} \tag{3.16}$$

A solução deste subproblema é $(0.5, 0.25)$. Para resolvermos o problema (3.16) aplicando o Algoritmo 2 e usando $x^0 = (-2, 1)$ como ponto inicial, devemos considerar um modelo para função objetivo de (3.16) que seja válido perto do ponto x^0 . Temos

$$g_1(-2, 1) = -3 \quad \text{e} \quad g_2(-2, 1) = -3.$$

Logo, os modelos para representar as penalizações associadas às $g_1(x)$ e $g_2(x)$ são 0. Portanto, o primeiro subproblema do algoritmo de região de confiança é:

$$\begin{aligned} \text{minimizar} \quad & 100(x_2 - x_1^2)^2 + (1 - x_1)^2 \\ \text{sujeita a} \quad & -0.5 \leq x_1 \leq 0.5 \\ & x_2 \leq 1 \\ & \|x - x^0\| \leq 10. \end{aligned} \tag{3.17}$$

Resolução gráfica

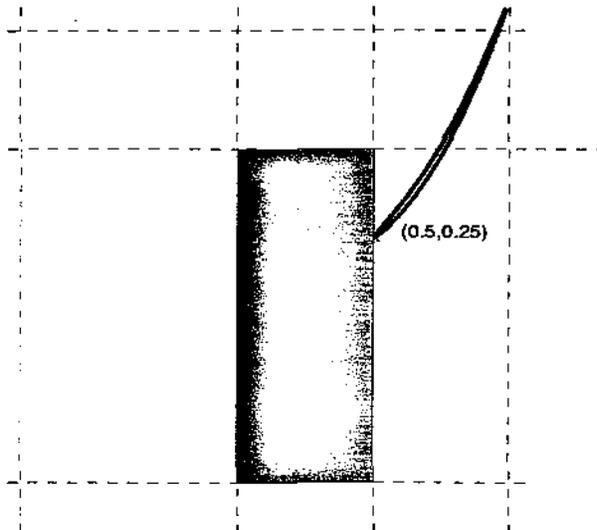


Figura 3.10:

A solução de (3.17) é $(0.5, 0.25)$. Esta solução é solução do problema penalizado, porém não podemos detectar isto. Podemos somente garantir que $(0.5, 0.25)$ representa uma melhor aproximação para a solução do problema (3.16). Assim, devemos

continuar resolvendo (3.16) e para isto montamos um modelo para aproximar a função $P_1(x, \rho_1)$ perto do ponto $(0.5, 0.25)$. Agora,

$$g_1(0.5, 0.25) = 0.4375 \quad \text{e} \quad g_2(0.5, 0.25) = 0.$$

Logo, os modelos para representarem as penalizações associadas à $g_1(x)$ é o $\max\{0, g_1(x)\}$ e à $g_2(x)$ é $\max\{0, g_2(x)\}$. Assim, o modelo para a função objetivo do problema penalizado é ela mesma. Portanto, o subproblema associado é

$$\begin{aligned} \text{minimizar} \quad & 100(x_2 - x_1^2)^2 + (1 - x_1)^2 + 1 \max\{0, x_1 - x_2^2\} \\ & + 1 \max\{0, -x_1^2 + x_2\} \\ \text{sujeita a} \quad & -0.5 \leq x_1 \leq 0.5 \\ & x_2 \leq 1 \\ & \|x - x^1\| \leq 10. \end{aligned} \tag{3.18}$$

A solução deste problema é a mesma do problema anterior. Consequentemente, uma solução do problema penalizado (3.16) foi encontrada. Agora, devemos continuar aplicando o algoritmo de penalização testando a factibilidade no ponto obtido. Temos:

$$\|g^+(0.5, 0.25)\|_1 = 0.4375 \geq 10^{-5}.$$

Assim, devemos aumentar o parâmetro de penalização e resolver o seguinte subproblema penalizado

$$\begin{aligned} \text{minimizar} \quad & P_1(x, \rho_2) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2 + 10 \max\{0, x_1 - x_2^2\} \\ & + 10 \max\{0, -x_1^2 + x_2\} \\ \text{sujeita a} \quad & -0.5 \leq x_1 \leq 0.5 \\ & x_2 \leq 1. \end{aligned} \tag{3.19}$$

Aplicando o Algoritmo 2 para resolver o subproblema acima e usando $x^0 = (0.5, 0.25)$ como ponto inicial, devemos montar um modelo para aproximar $P_1(x, \rho_2)$ perto do ponto x^0 . Temos:

$$g_1(0.5, 0.25) = 0.4375 \quad \text{e} \quad g_2(0.5, 0.25) = 0.$$

Logo, o modelo para representar a penalização associada à $g_1(x)$ é $\max\{0, g_1(x)\}$ e o modelo para representar a penalização associada à $g_2(x)$ é $\max\{0, g_2(x)\}$. Portanto,

o subproblema associado é

$$\begin{aligned}
 &\text{minimizar} && 100(x_2 - x_1^2)^2 + (1 - x_1)^2 + 10 \max\{0, x_1 - x_2^2\} + \\
 & && 10 \max\{0, -x_1^2 + x_2\} \\
 &\text{sujeita a} && -0.5 \leq x_1 \leq 0.5 \\
 & && x_2 \leq 1 \\
 & && \|x - x^0\| \leq 10.
 \end{aligned} \tag{3.20}$$

A solução deste problema é $(0, 0)$. O valor da função $P_1(x, \rho_2)$ no ponto $(0, 0)$ é menor quando comparado com o valor de $P_1(x, \rho_2)$ no ponto $(0.5, 0.25)$. Neste caso, aceitamos $(0, 0)$ como uma aproximação melhor para a solução do problema (3.19). No entanto não podemos garantir $(0, 0)$ como solução do problema penalizado, pois o critério de convergência não é satisfeito. Assim devemos continuar resolvendo o subproblema (3.19) usando $x^1 = (0, 0)$ como ponto inicial, definindo um modelo válido para $P_1(x, \rho_2)$ perto do ponto x^1 . Agora,

$$g_1(x^1) = 0 \quad \text{e} \quad g_2(x^1) = 0.$$

Logo, os modelos para representarem as penalizações associadas às $g_1(x)$ e $g_2(x)$ são $\max\{0, g_1(x)\}$ e $\max\{0, g_2(x)\}$ respectivamente. Portanto, o subproblema é

$$\begin{aligned}
 &\text{minimizar} && 100(x_2 - x_1^2)^2 + (1 - x_1)^2 + 10 \max\{0, x_1 - x_2^2\} \\
 & && + 10 \max\{0, x_1^2 - x_2\} \\
 &\text{sujeita a} && -0.5 \leq x_1 \leq 0.5 \\
 & && x_2 \leq 1 \\
 & && \|x - x^1\| \leq 10.
 \end{aligned} \tag{3.21}$$

A solução deste problema é $(0, 0)$. Com esta solução confirmamos $(0, 0)$ como solução do problema penalizado. Temos que

$$\|g^+(x^1)\|_1 = 0.$$

Logo x^1 é factível com respeito o problema original e portanto, solução do problema original.

Problema 4:

Consideremos o problema

$$\begin{aligned} &\text{minimizar} && f(x) = (x_1 - 2)^2 + x_2^2 \\ &\text{sujeita a} && g_1(x) = x_1 - (1 - x_2)^2 \leq 0 \\ &&& x_1, x_2 \geq 0. \end{aligned} \tag{3.22}$$

A solução deste problema é $(1, 0)$.

Resolução gráfica

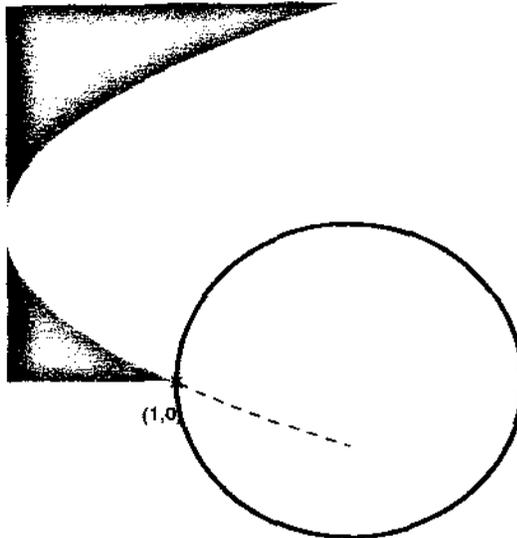


Figura 3.11:

Para resolver (3.22) aplicando o Algoritmo 1, o primeiro subproblema penalizado que devemos resolver é

$$\begin{aligned} &\text{minimizar} && P_1(x, \rho_1) = (x_1 - 2)^2 + x_2^2 + 1 \max\{0, x_1 - (1 - x_2)^2\} \\ &\text{sujeita a} && x_1, x_2 \geq 0. \end{aligned} \tag{3.23}$$

A solução deste subproblema é $(1.5, 0)$. Para resolvermos o subproblema penalizado usando o Algoritmo 2, devemos construir um modelo para aproximar a função objetivo de (3.23) que seja válido perto do ponto inicial $x^0 = (-2, 2)$. Agora,

$$g_1(-2, 2) = -3.$$

Logo, o modelo para representar a penalização associada à $g_1(x)$ é 0. Portanto, o primeiro subproblema do algoritmo de região de confiança é

$$\begin{aligned} &\text{minimizar} && (x_1 - 2)^2 + x_2^2 \\ &\text{sujeita a} && x_1, x_2 \geq 0 \\ &&& \|x - x^0\| \leq 10. \end{aligned} \tag{3.24}$$

Resolução gráfica

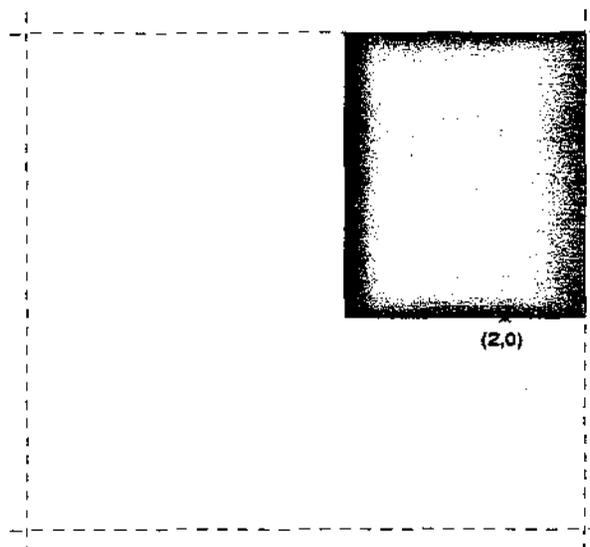


Figura 3.12:

A solução do subproblema acima é $(2,0)$. O ponto $(2,0)$ representa uma melhor aproximação para solução do subproblema penalizado do que o ponto $(-2,2)$, porém $(2,0)$ não é a solução ótima. Assim, devemos continuar resolvendo o problema penalizado usando $x^1 = (2,0)$ para definir o modelo para aproximar a função objetivo do problema penalizado. Temos que

$$g_1(2,0) = 1.$$

Logo, o modelo para representar a penalização associada à $g_1(x)$ é $\max\{0, g_1(x)\}$ e o subproblema que devemos resolver é:

$$\begin{aligned} &\text{minimizar} && (x_1 - 2)^2 + x_2^2 + 1 \max\{0, x_1 - (1 - x_2)^2\} \\ &\text{sujeita a} && x_1, x_2 \geq 0 \\ &&& \|x - x^1\| \leq 10. \end{aligned} \tag{3.25}$$

Resolução gráfica

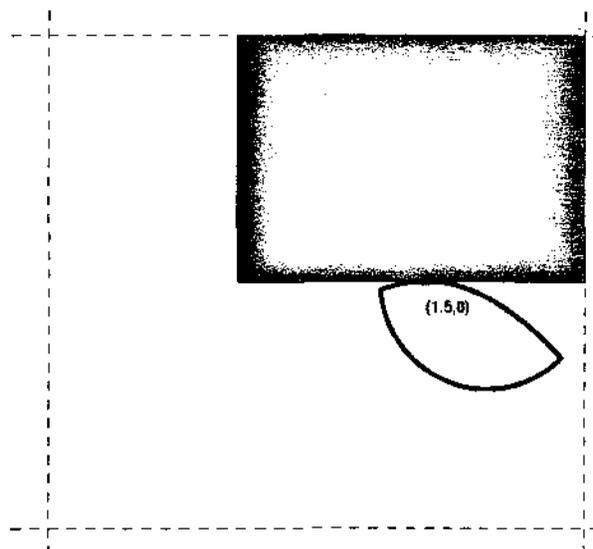


Figura 3.13:

A solução deste subproblema é $(1.5, 0)$ que é também solução do problema penalizado, porém o critério de convergência ainda não é satisfeito. Assim, devemos continuar resolvendo o subproblema penalizado usando $x^2 = (1.5, 0)$ como aproximação inicial, definindo um modelo para a função objetivo perto do ponto x^2 . Agora,

$$g_1(1.5, 0) = 0.5.$$

Logo, o modelo para representar a penalização associado à $g_1(x)$ é o $\max\{0, g_1(x)\}$ e o subproblema envolvido é

$$\begin{aligned} &\text{minimizar} && (x_1 - 2)^2 + x_2^2 + 1 \max\{0, x_1 - (1 - x_2)^2\} \\ &\text{sujeita a} && x_1, x_2 \geq 0 \\ &&& \|x - x^2\| \leq 10. \end{aligned} \quad (3.26)$$

A solução deste subproblema é exatamente a mesma solução do subproblema anterior. Desta forma, uma solução do subproblema penalizado foi encontrada, $(1.5, 0)$. Temos que

$$\|g_1^+(1.5, 0)\|_1 = 0.5 \geq 10^{-5},$$

isto significa que a factibilidade não foi satisfeita. Assim, devemos continuar resolvendo o problema original, para isto atualizamos o parâmetro de penalização e resolvemos o seguinte problema:

$$\begin{aligned} &\text{minimizar} && P_1(x, \rho_2) = (x_1 - 2)^2 + x_2^2 + 10 \max\{0, x_1 - (1 - x_2)^2\} \\ &\text{sujeita a} && x_1, x_2 \geq 0. \end{aligned} \quad (3.27)$$

A solução deste problema é $(1,0)$. Para aplicar o Algoritmo 2 para resolver o problema penalizado, devemos considerar um modelo válido para função objetivo perto do ponto inicial, neste momento $x^0 = (1.5, 0)$. Agora,

$$g_1(1.5, 0) = 0.5.$$

Assim, o modelo para representar $g_1(x)$ é o $\max\{0, g_1(x)\}$ e o subproblema associado é

$$\begin{aligned} &\text{minimizar} && (x_1 - 2)^2 + x_2^2 + 10 \max\{0, x_1 - (1 - x_2)^2\} \\ &\text{sujeita a} && x_1, x_2 \geq 0 \\ &&& \|x - x^0\| \leq 10. \end{aligned} \tag{3.28}$$

A solução deste subproblema é $(1, 0)$. Claramente, esta é a solução do problema original e também do problema (3.26). Entretanto, não podemos detectar isto ainda. Podemos garantir somente que o ponto $(1, 0) = x^1$ representa uma melhor aproximação para solução do problema penalizado (3.28). Assim devemos executar mais uma iteração do Algoritmo 2 usando x^1 como ponto inicial, e considerando um modelo para função objetivo do problema penalizado perto do ponto x^1 . Agora,

$$g_1(1, 0) = 0.$$

Logo, o modelo para representar $g_1(x)$ é o $\max\{0, g_1(x)\}$ e o subproblema associado é

$$\begin{aligned} &\text{minimizar} && (x_1 - 2)^2 + x_2^2 + 10 \max\{0, x_1 - (1 - x_2)^2\} \\ &\text{sujeita a} && x_1, x_2 \geq 0 \\ &&& \|x - x^1\| \leq 10. \end{aligned} \tag{3.29}$$

A solução deste problema é $(1, 0)$, que é também a solução do problema anterior. Logo, $(1, 0)$ é a solução do problema penalizado (3.26). Agora, devemos continuar aplicando algoritmo de penalização testando a factibilidade do ponto $(1, 0)$. Temos:

$$\|g_1^+(1, 0)\|_1 = 0.$$

Portanto, $(1, 0)$ é solução do problema original.

Os Problemas 1 e 2 foram apresentados com os objetivos de ilustrarem uma situação onde, a solução do problema original é obtida em apenas uma iteração do algoritmo de penalização, e mostrarem que as escolhas das constantes p_3 e p_4 definem os subproblemas que devem ser resolvidos para gerar uma solução do subproblema penalizado.

Os Problemas 3 e 4 ilustram o caso onde devemos atualizar o parâmetro de penalização no Algoritmo 1, isto é, quando a solução ótima do problema original não é encontrada na primeira iteração do Algoritmo 1.

Capítulo 4

Aplicações e Resultados Numéricos

Neste capítulo apresentamos um problema prático ligado à engenharia, o Fluxo de Potência Ótimo (FPO), o problema simples de encontrar o menor círculo que contém um dado conjunto de pontos em \mathbb{R}^2 , e o problema de encontrar p pontos em \mathbb{R}^3 cuja distância entre cada par de pontos é maior ou igual a 1 e determinar a menor bola centrada na origem que os contém. Estes problemas foram resolvidos usando o método de penalização com região de confiança (Algoritmos 1 e 2) e o algoritmo Lagrangeano aumentado com BOX-QUACAN, que chamaremos apenas de algoritmo BOX. Inicialmente apresentamos as formulações dos problemas e posteriormente os resultados dos experimentos.

4.1 Formulação dos Problemas

O Fluxo de Potência Ótimo

O problema de fluxo de potência ótimo consiste em otimizar o estado de operação estática de um sistema geração-transmissão de energia elétrica. O estado ótimo de operação deve satisfazer três exigências básicas:

- 1) Minimizar um critério de operação;
- 2) Atender a demanda de potência do sistema;
- 3) Manter os controles, variáveis e funções do sistema dentro de limites tolerados.

Estas exigências podem ser formuladas, respectivamente, como:

- minimização de uma função objetivo, como por exemplo: custo de geração ou perdas de potência nos componentes de transmissão;
- restrições de igualdade. Estas restrições são as equações básicas do Fluxo de Carga que resultam das Leis de Kirchhoff aplicadas ao circuito de potências; e
- restrições de desigualdade: limites nos valores das grandezas do sistema, que representam restrições físicas dos componentes e escassez de recursos. Por exemplo, limites inferiores e superiores de tensão, geração de potência (ativa e reativa), ângulo de transformadores defasadores, fluxo máximo em linhas de transmissão.

O Fluxo de Potência Ótimo é aplicado em sistemas (redes) com N barras (nós). Na formulação mais simples do problema (formulação básica), a cada barra da rede são associadas quatro variáveis, sendo que duas delas entram no problema como dados e duas como incógnitas:

V_k - magnitude da tensão (barra k)

θ_k - ângulo da tensão

P_k - geração líquida de potência ativa

Q_k - injeção de potência reativa.

Dependendo de quais variáveis entram como dados e quais são consideradas como incógnitas, definem-se três tipos de barras:

PQ - são dados P_k e Q_k , e calculados V_k e θ_k ;

PV - são dados P_k e V_k , e calculados Q_k e θ_k ;

Referência - são dados V_k e θ_k e calculados P_k e Q_k .

A barra de referência tem dupla função: fornece a referência angular do sistema e é usada para fechar o balanço da potência do sistema. O ângulo na barra de referência é zero.

As variáveis potência ativa e reativa são determinadas aplicando as equações de Fluxo de Carga. Após algumas simplificações nas equações, o problema básico con-

siste em encontrar as magnitudes das tensões e os ângulos nas barras. A formulação do problema (FPO) usada no trabalho é:

Vetor de Variáveis

O vetor $x = (V, \theta)$ de variáveis do problema é composto por:

- magnitudes das tensões (V) em todas as barras;
- ângulo das tensões (θ) nas barras PV e PQ.

Função Objetivo

A função objetivo utilizada neste trabalho minimiza as perdas de potências ativas nos componentes de transmissão e é dada por:

$$f(\theta, V) = \sum_{\text{ramos}} g_{km}(V_k^2 + V_m^2 - 2V_k V_m \cos \theta_{km})$$

onde:

g_{km} é condutância do ramo $k - m$ que liga a barra k à barra m ;

V_k, V_m são as magnitudes das tensões nas barras k e m ;

$\theta_{km} = \theta_k - \theta_m$ é diferença angular entre as tensões das barras k e m .

Restrições de Igualdade

As restrições de igualdade são as equações do fluxo de carga, que vêm das Leis de Kirchhoff aplicadas à malha do sistema elétrico. As expressões após algumas simplificações são:

$$P_{gk} - P_{ck} - V_k \sum_{m \in K} V_m (G_{km} \cos \theta_{km} + B_{km} \text{sen} \theta_{km}) = 0 \quad (4.1)$$

e

$$Q_{gk} - Q_{ck} - V_k \sum_{m \in K} V_m (G_{km} \text{sen} \theta_{km} - B_{km} \cos \theta_{km}) = 0, \quad (4.2)$$

onde:

P_{gk} e P_{ck} são as potências ativas geradas e consumidas na barra k respectivamente;

Q_{gk} e Q_{ck} são as potências reativas geradas e consumidas na barra k respectivamente;

G_{km} e B_{km} são elementos da matriz de admitância nodal e

K é o conjunto de barras vizinhas à barra k mais a própria barra k .

As equações (4.1) e (4.2) são aplicadas às barras PQ e a equação (4.1) às barras do tipo PV.

Restrições de Desigualdade

De modo geral as limitações nas variáveis são devidas a limitações físicas dos aparelhos, limites de segurança e limites de estabilidade. Neste trabalho iremos considerar o problema (FPO) onde a geração de potência ativa é pré-fixada, assim as restrições de desigualdade funcionais consideradas dizem respeito aos limites de geração de potência reativa nas barras PV e referência. Estas restrições de desigualdade são formuladas como:

$$V_k \sum_{m \in K} V_m (G_{km} \text{sen} \theta_{km} - B_{km} \text{cos} \theta_{km}) + (Q_c - Q_{gmax}) \leq 0,$$

$$-V_k \sum_{m \in K} V_m (G_{km} \text{sen} \theta_{km} - B_{km} \text{cos} \theta_{km}) + (Q_{gmin} - Q_c) \leq 0,$$

onde:

Q_c é potência reativa consumida na barra considerada,

Q_{gmin} é a quantidade mínima de potência reativa gerada,

Q_{gmax} é a quantidade máxima de potência reativa gerada e

K é o conjunto de barras vizinhas à barra k mais a própria barra k .

Consideramos também as canalizações somente sobre as magnitudes das tensões:

$$V_i^{min} \leq V_i \leq V_i^{max}.$$

Os valores de V_i^{min} e V_i^{max} usados nos experimentos foram 0.95 e 1.05, ou 0.90 e 1.10, ou 0.95 e 1.12.

Se uma rede contém p barras, sendo p_1 do tipo PQ e p_2 do tipo PV então o problema de programação não linear associado, segundo a formulação acima, apresentará $(2p - 1)$ variáveis, $(2p_1 + p_2)$ restrições de igualdade, $(2p_2 + 2)$ restrições de desigualdade e p restrições de canalizações.

O Problema do Menor Círculo

Vamos supor que foram dados m_2 pontos em \mathbb{R}^2 , $(y_{11}, y_{12}), (y_{21}, y_{22}), \dots, (y_{m_2,1}, y_{m_2,2})$. O problema do menor círculo consiste em encontrar o círculo de raio mínimo que contém os m_2 pontos. Equacionando o problema temos o seguinte problema de programação não linear:

$$\begin{aligned}
 &\text{minimizar} && x_3 \\
 &\text{sujeita a} && (x_1 - y_{11})^2 + (x_2 - y_{12})^2 - x_3 \leq 0 \\
 & && (x_1 - y_{21})^2 + (x_2 - y_{22})^2 - x_3 \leq 0 \\
 & && \vdots \\
 & && (x_1 - y_{m_2,1})^2 + (x_2 - y_{m_2,2})^2 - x_3 \leq 0.
 \end{aligned} \tag{4.3}$$

O Problema da Menor Bola

O problema da menor bola consiste em encontrar p pontos em \mathbb{R}^3 , y_1, \dots, y_p , tais que a distância entre cada par de pontos é maior ou igual a 1 e determinar a bola de mínimo raio centrada na origem que os contém. Equacionando o problema temos:

$$\begin{aligned}
 &\text{minimizar} && r \\
 &\text{sujeita a} && \|y_i\|_2^2 - r \leq 0 \text{ para todo } i = 1, \dots, p \\
 & && 1 - \|y_i - y_j\|_2^2 \leq 0 \text{ para } i \neq j.
 \end{aligned} \tag{4.4}$$

4.2 Experimentos Numéricos

Nesta seção apresentamos os resultados experimentais obtidos da implementação dos Algoritmos 1 e 2 para os problemas de (FPO) que envolve restrições não lineares de igualdade e desigualdade e canalizações e, os problemas do menor círculo e menor bola que envolvem somente restrições de desigualdade não lineares. Comparamos o tempo de CPU entre o método de penalização com região de confiança (Algoritmos 1 e 2) e o Algoritmo BOX.

Os Algoritmos 1 e 2 foram implementados em FORTRAN 77 usando cálculos com dupla precisão. Em nossos experimentos usamos o parâmetro de penalização inicial igual a 1 e o atualizamos multiplicando-o por 10. A execução do algoritmo de penalização foi implementada para ser interrompida quando uma solução for encontrada, isto é, a norma 1 das restrições originais penalizadas é menor ou igual a 10^{-5} ou quando o número de iterações permitido for extrapolado, superior a 10 iterações. Os modelos escolhidos para definir a função $\psi_\ell(x)$ no algoritmo de região de confiança foram os modelos não lineares não suaves, (2.4) e (2.8). As constantes p_1, p_2, p_3 e p_4 usadas na definição dos modelos foram tomadas iguais a 1. Ainda no algoritmo de região de confiança usamos $B_\ell \equiv 0$ e $\underline{\Delta} = 10$. O valor inicial do raio de confiança em cada iteração foi $\underline{\Delta}$. A redução no raio de confiança foi $\frac{1}{8}\Delta$ se $\Delta = 10$ e $\frac{1}{2}\Delta$ caso contrário. O algoritmo de região de confiança foi implementado para ser interrompido com fracasso quando o número de iterações for superior a 100 ou quando Δ for muito pequeno, $\Delta \leq 10^{-5}$. Queremos destacar que no algoritmo de região de confiança somente as iterações de sucesso são enumeradas, isto é, $x^\ell \neq x^{\ell+1}$ para todo ℓ .

O algoritmo BOX é uma implementação de métodos de Lagrangeano aumentado desenvolvida no grupo de otimização do Departamento de Matemática Aplicada da UNICAMP sob direção de J.M. Martínez. Sua principal subrotina está baseada no método introduzido por Friedlander, Martínez e Santos em [19]. Esta subrotina resolve problemas gerais de otimização e permite tratar as restrições de desigualdade não lineares de duas formas. Uma delas trata explicitamente cada desigualdade não linear e a outra alternativa consiste em adicionar a cada restrição de desigualdade não linear uma variável de folga e reduzi-la a forma de igualdade mais o limite. Quando todas as restrições não lineares são tratadas como restrições de igualdade, a função Lagrangeana associada ao problema considerado torna-se duas vezes diferenciável. Ao passo que se tratarmos as restrições de desigualdade explicitamente perdemos esta propriedade e temos a função Lagrangeana aumentada somente uma vez diferenciável.

Todos os testes foram executados em uma SUN Workstation SPARC4, usando o compilador FORTRAN 77, no Laboratório do Departamento de Matemática Aplicada da UNICAMP.

Foram realizados testes em redes com as seguintes características:

Rede	Total de Barras	N ^o barras PQ	N ^o barras PV	N ^o ramos
R1	2	1	0	1
R2	5	3	1	6
R3	14	9	4	20
R4	30	24	5	41

Usando a formulação descrita acima, os problemas de programação não linear envolvidos possuem as seguintes características:

Rede	n	m_1	m_2	N ^o de canalizações
R1	3	2	2	2
R2	9	7	4	5
R3	27	22	10	14
R4	59	53	12	30

onde n é o número de variáveis, m_1 é o número de restrições de igualdade não lineares, m_2 é o número de restrições de desigualdade não lineares.

Para todos os problemas (FPO) usamos o ponto inicial normalmente usado para resolver este tipo de problema na Engenharia, isto é, 1 para as variáveis associadas às magnitudes das tensões e 0 para as variáveis associadas aos ângulos. Os dados iniciais e os valores ótimos das redes foram obtidos do conjunto de problemas de [37]. Em nossos experimentos, para todas as redes, os valores ótimos encontrados tanto pelo método de penalização com região de confiança como pelo algoritmo BOX, coincidiram com os valores ótimos de [37].

A tabela abaixo mostra os resultados obtidos para as redes.

Tabela 3		
Rede	TMP	TBox
R1	0.132	0.02
R2	17.672	0.05
R3	14.317	1.646
R4	141.236	10.448

onde:

TMP é o tempo consumido pelo método de penalização com região confiança, em segundos e

TBox é o tempo consumido pelo algoritmo BOX, em segundos, para resolver o de fluxo de potência ótimo.

Claramente, o desempenho do algoritmo BOX para resolver o problema de fluxo de potência ótimo foi superior ao do método de penalização com região de confiança. As soluções de todos os problemas testados, usando o método de penalização com região de confiança, foram encontradas em apenas uma iteração do algoritmo de penalização. O alto custo do método se concentrou na resolução dos subproblemas penalizados. Para as redes R_1 , R_2 e R_4 o número de iterações do algoritmo de região de confiança foi pequeno, porém o ponto inicial usado forçou todas as restrições de igualdade e quase todas as desigualdade não lineares a participarem das funções modelo que aproximam as funções penalizadas não suaves. Assim, as funções modelo mantiveram quase a mesma quantidade de pontos não diferenciáveis das funções penalizadas. Ao acrescentarmos variáveis adicionais para transformar as funções modelo em funções suaves, acabamos tendo que resolver subproblemas com uma quantidade maior de variáveis e restrições não lineares de desigualdade do que o problema original. Para a rede R_2 foram executadas 6 iterações do algoritmo de região de confiança. Os primeiros subproblemas envolvidos para resolver o problema penalizado associado apresentaram uma quantidade menor de restrições do que o problema original. Porém à medida que foram sendo gerados pontos mais próximos do ótimo, mais restrições do problema original foram também participando do conjunto de restrições dos subproblemas, e portanto aumentando a complexidade dos subproblemas.

Para o problema de menor círculo testamos um conjunto de problemas. O número de restrições não lineares variou de 5 a 1000. Os pontos foram gerados

aleatoriamente no intervalo $[0,1]$. A quantidade de pontos gerados em cada problema representa o número de restrições de desigualdade não lineares (m_2). Aplicamos o algoritmo de penalização com região de confiança considerando diferentes valores para as constantes p_3 e p_4 , usadas na definição dos modelos para aproximarem as funções penalizadas. Aplicamos também o algoritmo BOX tratando as restrições de desigualdade não lineares de duas formas: acrescentando variáveis de folga e tratando explicitamente as restrições de desigualdade. A tabela abaixo mostra os tempos consumidos considerando as diferentes situações.

Tabela 4

m_2	TMP1	TMP2	TBoxh	TBoxg
5	0.2	0.102	0.112	0.015
10	0.186	0.281	0.259	0.017
20	0.261	1.005	1.213	0.020
30	0.493	0.778	2.140	0.020
40	0.691	0.897	4.094	0.022
50	0.821	1.067	6.819	0.023
60	0.397	1.096	21.662	0.014
70	0.652	0.745	35.488	0.020
80	1.265	1.271	68.049	0.026
90	2.265	0.897	92.070	0.037
100	2.717	1.026	116.559	0.043
200	11.953	2.103	1936	0.054
300	12.196	2.523	4862	0.045
400	20.289	4.367	8957	0.054
500	33.289	5.710	19764	0.066
1000	1076.23	17.972	*	0.15

onde:

TMP1 representa o tempo consumido pelo método de penalização com região de confiança quando usamos p_3 e p_4 iguais a 1;

TMP2 representa o tempo consumido pelo método de penalização com região de confiança quando usamos p_3 e p_4 iguais a 0.01;

TBoxh representa o tempo consumido pelo algoritmo BOX, quando acrescentamos variáveis de folga as restrições de desigualdade;

TBoxg representa o tempo consumido pelo algoritmo BOX, quando tratamos ex-

plicitamente as restrições de desigualdade, e

* significa que o tempo de processamento tornou-se inaceitável.

Observando os tempos consumidos pelo método de penalização com região de confiança usando diferentes valores para as constantes p_3 e p_4 temos que o desempenho do método é fortemente afetado pelas escolhas das constantes. Quando usamos $p_3 = p_4 = 1$, as soluções dos problemas foram obtidas em apenas uma iteração do algoritmo de penalização. Na maioria dos problemas testados, foram executadas 18 iterações do algoritmo de região de confiança. Deste total, as 16 primeiras iterações envolveram subproblemas simples com 3 variáveis e o conjunto de restrições foi formado apenas pela caixa de confiança. Em todos os problemas testados, a última iteração do algoritmo de região de confiança envolveu todas as restrições do problema original. Agora, quando usamos $p_3 = p_4 = 0.01$ as soluções dos problemas foram encontradas em 1 ou 2 iterações do algoritmo de penalização. Para resolver os subproblemas penalizados foram necessárias, na maioria dos problemas, 30 iterações do algoritmo de região de confiança. Apesar do número de iterações do algoritmo de região de confiança neste caso ter sido superior, o tempo de processamento foi inferior. O que ocorreu foi que ao colocarmos valores menores para p_3 e p_4 restringimos mais a participação das restrições do problema original na função modelo. Assim, os subproblemas envolvidos no algoritmo de região de confiança foram subproblemas sem restrição não linear ou com poucas restrições.

Analisando a Tabela 4 temos que o desempenho do algoritmo BOX, quando tratamos explicitamente as restrições de desigualdade, foi superior as demais alternativas de resolução. O método de penalização com região de confiança torna-se competitivo quando comparamos com o algoritmo BOX na versão com igualdade.

O gráfico abaixo dá uma idéia resumida dos tempos do CPU consumidos pelo método de penalização com região de confiança e o algoritmo BOX nas versões com igualdade (Boxh) e desigualdade (Boxg).

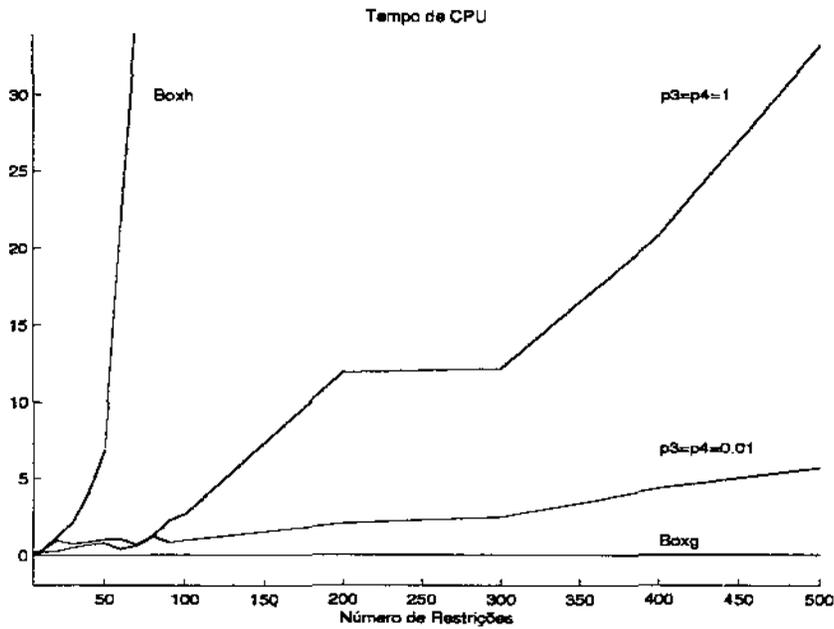


Figura 4.1:

Para o problema de menor bola, resolvemos um conjunto de problemas onde o número de pontos em \mathbb{R}^3 (np), a ser determinados, variou de 2 a 30. As características dos problemas não lineares são $(3np + 1)$ variáveis e $(np + \frac{np(np-1)}{2})$ restrições de desigualdade não lineares. Para este conjunto de problemas aplicamos o método de penalização com região de confiança usando diferentes valores para p_3 e p_4 , e aplicamos também o algoritmo BOX na versão com igualdade e desigualdade. Os tempos de processamento estão listados na tabela abaixo.

Tabela 5

np	TMP1	TMP2	Tboxh	Tboxg
2	0.049	0.5	0.027	0.022
3	0.1	1.82	0.108	0.05
4	0.27	5.39	0.281	0.08
5	6.07	13.43	1.001	0.64
6	1.33	39.14	1.464	0.27
7	2.86	64.01	4.281	0.68
8	4.75	124.67	8.470	0.61
9	14.03	237.61	12.97	3.45
10	13.73	510	28.48	1.48
15	224	4989	741	9.85
20	1456	3451	10498	22.63
30	49046	41331	96062	361

Para o conjunto de problemas de menor bola, o algoritmo BOX na versão com desigualdade apresentou melhor desempenho entre as alternativas de resolução executadas. O método de penalização com região de confiança usando $p_3 = p_4 = 1$ resolveu o conjunto de problemas em poucas iterações do algoritmo de região de confiança, porém todas as restrições dos problemas originais foram envolvidas nos subproblemas. Agora, quando as constantes p_3 e p_4 foram tomadas iguais a 0.1 foram necessárias várias iterações do algoritmo de região de confiança para chegar às soluções ótimas. Neste caso, as informações das restrições do problema original foram introduzidas gradualmente nos subproblemas do algoritmo de região de confiança, de forma que os subproblemas envolvidos apresentaram uma quantidade menor de restrições do que o problema original considerado. Para os problemas de menor bola com grande quantidade de restrições, o algoritmo BOX na versão com igualdade foi a forma menos eficiente de resolução.

4.3 Comentários

Os resultados numéricos nos mostram a eficiência do algoritmo BOX na versão com desigualdade para problemas com restrições de desigualdade. O desempenho do algoritmo BOX está fortemente associado a forma como são tratadas as restrições do problema original. Para os problemas testados o algoritmo BOX deixa de ser atraente quando as restrições do problema original são tratadas como restrições de igualdade.

Para todos os problemas testados o método de penalização com região de confiança se mostrou consistente no sentido de resolver o problema. Porém, não se mostrou eficiente quando comparado com algoritmo BOX na versão com desigualdade. Para escolhas adequadas para as constantes de tolerâncias (p_3 e p_4) a filosofia do método de penalização com região de confiança foi verificada, isto é, foram resolvidas seqüências de subproblemas com poucas restrições não lineares. As escolhas das constantes dependem da ordem de grandeza das restrições do problema original.

Do ponto de vista prático, o tempo consumido pelo método de penalização com região de confiança se concentrou na resolução dos subproblemas penalizados, que foram resolvidos usando uma função modelo para aproximar a função penalizada. Nos testes realizados usamos o modelo não linear (2.8). Assim os subproblemas do Algoritmo 2, mesmo tendo poucas restrições, foram problemas não lineares, o que exigiu tempo de processamento. Cabe investigar o desempenho do método de penalização com região de confiança para problemas com muitas restrições de desigualdade, usando um modelo linear para aproximar as penalizações associadas as restrições que estão nas regiões críticas. Desta forma, as restrições dos subproblemas do Algoritmo 2 serão lineares, o que pode economizar tempo de processamento.

Capítulo 5

Alguns Problemas de Minimização Global

5.1 Introdução

O problema de configuração molecular consiste em encontrar uma configuração de átomos ou moléculas cuja energia potencial é mínima. A energia potencial da molécula é modelada pela soma de energias potenciais entre pares de átomos. Encontrar a posição no espaço para p átomos ou moléculas x^1, \dots, x^p consiste em encontrar o minimizador global da função potencial de Lennard-Jones

$$F(X) = \sum_{i=1}^{p-1} \sum_{j=i+1}^p f(\|x^i - x^j\|) \quad (5.1)$$

onde

$$f(z) = \frac{1}{z^{12}} - \frac{2}{z^6}.$$

Encontrar o minimizador global da função acima pode ser extremamente difícil devido ao grande número de minimizadores locais. Para dar uma idéia da dificuldade

do problema, Hoare e McInnes [23] descobriram 988 minimizadores locais distintos para o problema com 13 átomos ($p = 13$). Segundo Northby [35] é possível que o número de minimizadores locais cresça mais rápido do que qualquer função linear que dependa de p . Dado que não temos bons algoritmos globais para resolver problemas de otimização, um grande número de pesquisas tem sido desenvolvidas no sentido de gerar configurações iniciais (pontos iniciais) que permitam, a partir destas configurações, conduzir a minimizadores globais usando algoritmos locais.

Na literatura são conhecidos vários trabalhos onde são relatados os menores valores ótimos (recordes), valores conhecidos até o momento, para o problema (5.1) para diferentes p . Hoare e Pal [24] encontraram recordes para o problema (5.1) para p variando de 6 até 46. Gockenback, Kearsley e Symes [21] resolveram o problema (5.1) reformulando-o como um problema de programação não linear com restrições lineares e listaram os recordes para p entre 5 e 30. Northby [35] resolveu (5.1) apresentando configurações iniciais baseadas em estruturas do tipo icosaedral. Com este tipo de heurística o autor consegue recordes para p entre 13 e 147. O forte das heurísticas usadas por Northby é que tais heurísticas existam, o que pode não acontecer em geral.

Neste capítulo, motivados pelo bom desempenho do algoritmo Box na versão com desigualdade apresentado no capítulo 4, formulamos e resolvemos um problema de configuração planar usando uma função tipo Lennard-Jones, a saber:

$$\Phi(X) = \sum_{i \neq j} \phi_{ij}(\|x^i - x^j\|)$$

onde $X = (x^1, x^2, \dots, x^p)$,

$$\phi_{ij}(\|x^i - x^j\|) = \left[\frac{d_{ij}}{\|x^i - x^j\|} \right]^{12} - 2 \left[\frac{d_{ij}}{\|x^i - x^j\|} \right]^6$$

e d_{ij} representa a distância pré-definida entre os pontos x^i e x^j .

Este problema também tem muitos minimizadores locais, porém não possui uma geometria bem estudada de forma que possamos garantir a existência de heurística como a de Northby para resolvê-lo. Uma maneira de tratar o problema tipo Lennard-Jones é rodar um programa várias vezes usando diferentes pontos iniciais aleatórios e ficar com o menor valor da função. Nossa proposta para resolver

este problema é também rodar um programa várias vezes, porém usando bons pontos iniciais. Estes pontos serão obtidos de um problema de programação não linear com restrições de desigualdade. Para verificar a consistência da idéia de gerar bons pontos iniciais, vamos inicialmente abordar o problema clássico de Lennard-Jones dado que conhecemos os valores globais para diferentes p .

5.2 Problema de Lennard-Jones

Nesta seção o objetivo é apresentar uma forma de gerar bons pontos iniciais para o problema clássico de Lennard-Jones, isto é,

$$\text{minimizar } F(X) = \sum_{i=1}^{p-1} \sum_{j=i+1}^p f(\|x^i - x^j\|)$$

onde

$$f(z) = \frac{1}{z^{12}} - \frac{2}{z^6}.$$

Note que a função $f(z)$ vai para o infinito rapidamente quando z tende a 0, tem o seu valor mínimo em $z = 1$, e vai para zero quando z tende a infinito. Isto significa que a situação ótima para os pontos x^i e x^j seria que todos eles estivessem a uma distância 1 entre si, o que é impossível. Por outro lado, ter pontos perto um do outro é ruim porque f cai muito rapidamente desde infinito até -1 . Porém, ter pontos a uma distância maior que 1 é ruim, mas nem tanto. Assim, vamos buscar pontos iniciais para o problema acima resolvendo um problema de programação não linear que garanta que os pontos estejam tão juntos quanto possível porém com a restrição que a distância entre qualquer par de pontos seja maior ou igual a 1, isto é, resolvendo o problema

$$\begin{aligned} &\text{minimizar } \sum_{i=1}^{p-1} \sum_{j=i+1}^p \|x^i - x^j\|^2 \\ &\text{sujeita a } \|x^i - x^j\| \geq 1 \text{ para } i \neq j. \end{aligned} \tag{5.2}$$

Para acomodar a idéia apresentada, implementamos um programa em FORTRAN 77 (LJ1) que resolve o problema (5.2), usando o algoritmo Box na versão com desigualdade, e em seguida resolve o problema de Lennard-Jones também usando o algoritmo Box. Implementamos outro programa (LJ2) que resolve (5.1) usando como ponto inicial, ponto aleatório definido no intervalo -3 a 3 . Implementamos um terceiro programa (LJ3) que resolve o problema restrito (5.2) usando o algoritmo de penalização com região de confiança e em seguida resolve o problema de Lennard-Jones usando o Box. O objetivo de resolver o problema restrito de duas formas é efetuar uma nova comparação entre o Box na versão com desigualdade e o método de penalização com região de confiança.

Dado que conhecemos os valores globais para o problema de Lennard-Jones para diferentes valores de p , rodamos os programas até atingirem os valores ótimos ou até extrapolarem o tempo máximo de 3 horas de processamento. Registramos o tempo, em segundos, e o número de vezes que foi necessário rodar os programas para encontrar os valores globais.

Na Tabela 1 relatamos os resultados dos experimentos numéricos para o problema de Lennard-Jones para p variando de 4 a 30. Todos os testes foram executados em uma SUN Workstation SPARC4, usando o compilador FORTRAN 77, no Laboratório do Departamento de Matemática Aplicada da UNICAMP. As colunas T1, T2 e T3 listam os tempos consumidos pelos programas LJ1, LJ2 e LJ3 respectivamente. As colunas N1 e N3 representam o número de vezes que foram resolvidos o problema (5.2) seguido do problema de Lennard-Jones pelos programas LJ1 e LJ3 e N3 representa o número de vezes que o problema de Lennard-Jones foi resolvido pelo programa LJ2. Ainda, * significa que o valor ótimo conhecido não foi atingido no prazo máximo de 3 horas de processamento.

Tabela 1

p	T1	T2	T3	N1	N2	N3
4	0.05	0.11	0.16	1	1	1
5	0.08	0.17	1.16	1	1	1
6	0.10	219	1.07	1	773	1
7	0.19	0.33	3.26	1	1	1
8	15.35	4.29	*	1	7	*
9	1.16	0.7	196.81	1	1	9
10	1.04	8.03	47.16	1	6	1
11	2.97	261.89	284.35	1	121	1
12	25.24	75.59	270.56	2	26	1
13	5.20	399.87	430	1	99	1
14	6.44	39.78	695	1	8	1
15	5.95	486.93	*	1	71	*
16	308.26	78.73	*	24	10	*
17	1015	2350	*	95	234	*
18	44.56	2757	5573	3	218	6
19	197.66	7750	3803	11	512	2
20	80.87	3744	5712	4	194	1
21	123.83	*	9274	5	*	4
22	186.71	*	*	8	*	*
23	48.49	*	*	2	*	*
24	31.41	537	*	1	14	*
25	134.85	*	*	4	*	*
26	148.20	*	*	3	*	*
27	*	*	*	*	*	*
28	10632	*	*	141	*	*
29	498	*	*	7	*	*
30	*	*	*	*	*	*

Inicialmente vamos comparar o algoritmo Box com o algoritmo de penalização com região de confiança. Analisando as colunas T1, T3, N1 e N3 confirmamos o bom desempenho do algoritmo Box na versão com desigualdade. Para vários valores de p , o programa LJ3 não localizou o valor global do problema de Lennard-Jones. Este comportamento deve-se ao fato de se investir muito tempo no processamento do problema restrito, não sendo possível rodar muitas vezes o programa LJ3.

Agora, comparando as colunas N1, N2 e N3 temos que a idéia de gerar bons pontos iniciais para o problema de Lennard-Jones através da solução do problema restrito (5.2) se verifica. O número de vezes que rodamos os programas LJ1 e LJ3 foi inferior ao programa LJ2. Uma comparação das colunas T1 e T2 mostra que é mais rápido resolver o problema de Lennard-Jones usando pontos iniciais vindo das soluções do problema restrito (usando um bom algoritmo para resolver o problema restrito, é claro) do que usar pontos iniciais aleatórios. Portanto, as soluções do problema (5.2) representam bons pontos iniciais para o problema de Lennard-Jones. Agora, vamos estender a técnica para o problema de Lennard-Jones mais geral.

5.3 Problema Tipo Lennard-Jones

Consideremos o problema de descobrir a localização de p pontos no plano, de maneira que as distâncias ótimas entre eles ao invés de 1, como no caso de Lennard-Jones, são valores dados d_{ij} . Suponhamos, mais ainda, que estimar a localização dos pontos superestimando as distâncias seja ruim, mas que subestimar é pior. Uma função que ajusta este problema é a função tipo Lennard-Jones, isto é,

$$\Phi(X) = \sum_{i \neq j} \phi_{ij}(\|x^i - x^j\|) \quad (5.3)$$

onde $X = (x^1, x^2, \dots, x^p)$,

$$\phi_{ij}(\|x^i - x^j\|) = \left[\frac{d_{ij}}{\|x^i - x^j\|} \right]^{12} - 2 \left[\frac{d_{ij}}{\|x^i - x^j\|} \right]^6$$

e d_{ij} representa a distância pré-definida entre os pontos x^i e x^j .

De fato esta função reflete as exigências sobre as distâncias entre os pontos. Note que a função cresce moderadamente quando a distância entre x^i e x^j é maior que d_{ij} e vai para o infinito quando a distância vai à zero.

Um problema prático ilustrando a aplicação da função tipo Lennard-Jones é descrito abaixo.

Aplicação: Problema de Agrupamento de Pessoas

O problema de agrupamento de pessoas vem da psicometria e consiste em definir grupos de pessoas de uma classe para executar uma determinada tarefa. Para que a tarefa seja executada da melhor forma possível é desejável que somente pessoas que tenham bom relacionamento sejam mantidas no mesmo grupo. Para saber o grau de relacionamento entre as pessoas é perguntado a cada pessoa a distância que a separa das demais, e montada uma matriz de distâncias. Note que o ideal para formar os grupos é manter as distâncias definidas pelas pessoas, o que em geral é impossível. Por outro lado, se subestimar as distâncias corremos o risco de colocar pessoas que se detestam no mesmo grupo, o que pode ser um desastre. Ao passo que, ao superestimar as distâncias podemos simplesmente correr o risco de afastar pessoas que se gostam. Assim, podemos representar o problema de agrupamento de pessoas usando o modelo acima onde cada pessoa é representada por um ponto em \mathbb{R}^2 ou \mathbb{R}^3 . A solução deste problema fornecerá uma visualização planar ou espacial da classe que auxiliará o consultor nas definições dos grupos.

A função tipo Lennard-Jones apresenta muitos minimizadores locais. Uma maneira para resolvê-lo é rodar um programa muitas vezes usando diferentes pontos iniciais aleatórios e ficar com o menor valor da função. Nossa proposta para resolver este problema de maneira eficiente é fornecer bons pontos iniciais. Para isto devemos resolver o seguinte problema restrito,

$$\begin{aligned} & \text{minimizar} \quad \sum_{i=1}^{p-1} \sum_{j=i+1}^p \|x^i - x^j\|^2 \\ & \text{sujeita a} \quad \|x^i - x^j\| \geq d_{ij}, \text{ para todo } i \neq j. \end{aligned} \tag{5.4}$$

A idéia de resolver (5.4) é que assim colocamos os pontos o mais próximos possível com a restrição de que as distâncias sejam maiores ou iguais a d_{ij} . Que é exatamente o que a função tipo Lennard-Jones deseja.

Testes Numéricos

Para ilustrar a idéia de gerar bons pontos iniciais para o problema tipo Lennard-Jones usando a solução do problema restrito (5.4), vamos aplicar o problema de agrupamento de pessoas a uma classe formada por 10 indivíduos. Os dados apresentados na tabela abaixo não são reais, uma vez que o acesso as fichas de consultas psicológicas não é permitido. Teoricamente, os dados são obtidos pelo consultor através de testes psicológicos que são aplicados a cada pessoa da classe visando conhecer o grau de proximidade entre cada par de indivíduos (p_i, p_j) , e assim é montada a matriz de distâncias.

Tabela 2

	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9	p_{10}
p_1	0	1	3	2	2	1	4	1	1	2
p_2	1	0	5	5	1	4	2	3	2	3
p_3	2	1	0	1	4	3	2	1	5	5
p_4	2	1	3	0	1	4	1	1	3	1
p_5	1	2	2	3	0	5	2	4	3	1
p_6	1	2	2	1	1	0	1	3	2	1
p_7	3	4	1	1	2	5	0	5	3	1
p_8	1	4	3	2	2	1	3	0	1	2
p_9	1	5	3	2	2	3	3	1	0	3
p_{10}	5	3	3	1	2	3	1	2	3	0

Vamos representar cada pessoa por um ponto em \mathbb{R}^2 e resolver o problema tipo Lennard-Jones associado, isto é,

$$\text{Minimizar } \sum_{i \neq j} \phi_{ij}(\|x^i - x^j\|) \text{ com } i, j = 1, \dots, 10. \quad (5.5)$$

A solução gráfica deste problema auxiliará o consultor nas definições dos grupos.

Para resolver o problema acima, implementamos um programa em FORTRAN 77, TLJ1, que resolve o problema restrito (5.4) usando o algoritmo Box e em seguida resolve o problema tipo Lennard-Jones. Criamos um segundo programa

TLJ2, que chama a rotina Box para resolver o problema tipo Lennard-Jones usando ponto inicial aleatório. Implementamos um terceiro programa TLJ3, que resolve o problema restrito (5.4) usando o algoritmo de penalização com região de confiança e em seguida resolve o problema tipo Lennard-Jones usando o algoritmo Box.

Rodamos os programas dez vezes e registramos na Tabela 3, os tempos em segundos e os menores valores encontrados para função (5.5). Todos os testes foram executados em uma SUN Workstation SPARC4, usando o compilador FORTRAN 77, no Laboratório do Departamento de Matemática Aplicada da UNICAMP.

T1	T2	T3	F1	F2	F3
10.94	47.15	1369	-39.22	-29.98	-37.25

As colunas T1, T2 e T3 fornecem os tempos consumidos pelos programas TLJ1, TLJ2 e TLJ3 respectivamente, e as colunas F1, F2 e F3 listam os menores valores de (5.5) encontrados pelos programas TLJ1, TLJ2 e TLJ3.

Comparando as colunas T1 e T3 da tabela acima temos novamente comprovado o bom desempenho do algoritmo box na versão com desigualdade sobre o algoritmo de penalização com região de confiança. Agora, comparando as colunas T1 e T2 temos que o programa TLJ1, mesmo investindo tempo na resolução do problema restrito, foi mais rápido e gerou menor valor para a função (5.5) do que o programa TLJ2. Isto significa que as soluções do problema restrito representam bons pontos iniciais para o problema tipo Lennard-Jones.

As figuras abaixo fornecem as soluções encontradas pelos programas TLJ1, TLJ2 e TLJ3.

Modelo Lennard-Jones – ponto inicial vindo do problema restrito

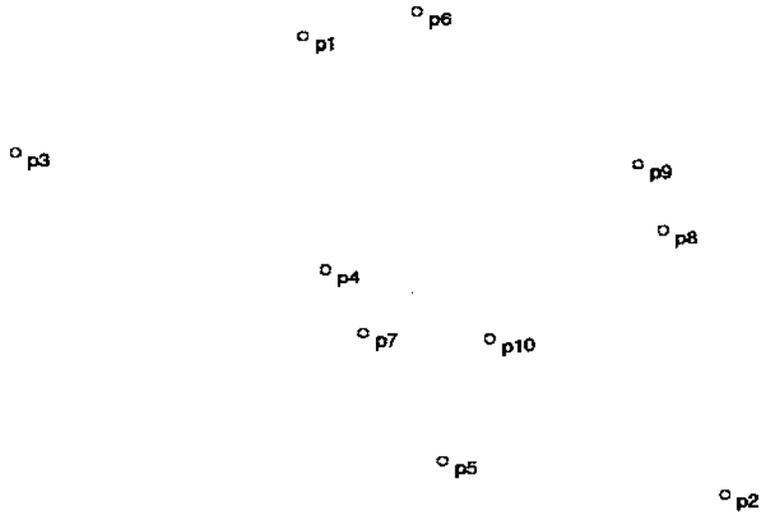


Figura 5.1:

Modelo Lennard-Jones – ponto inicial aleatório

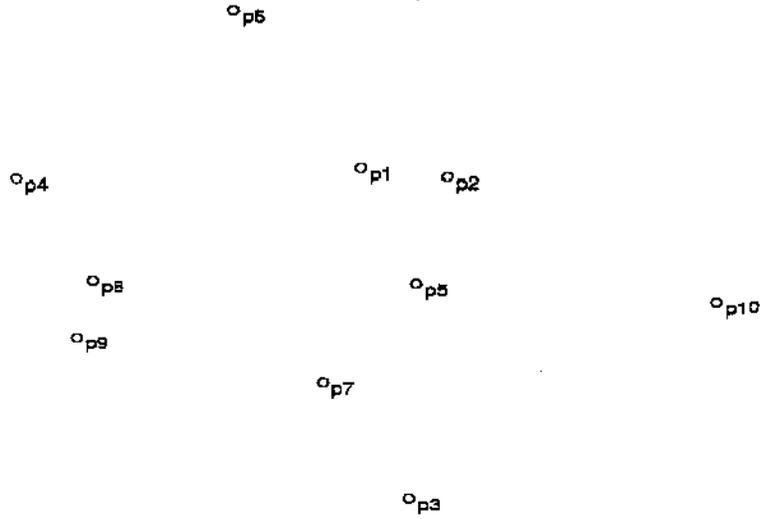


Figura 5.2:

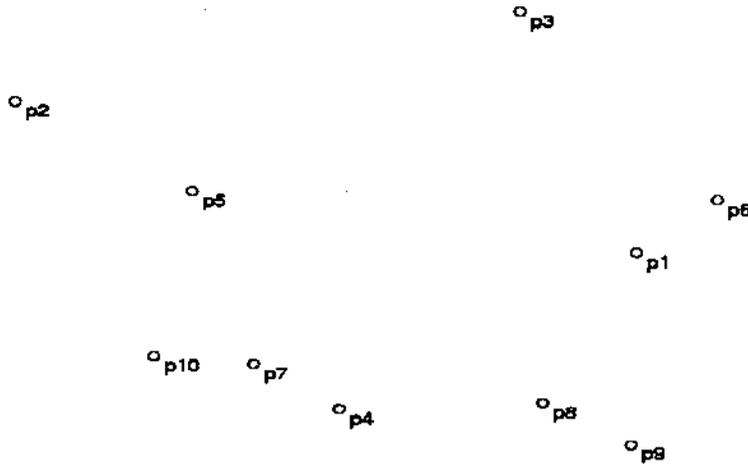


Figura 5.3:

5.4 Comentários

Nossa principal preocupação neste capítulo foi apresentar uma forma de gerar bons pontos iniciais para o problema de Lennard-Jones. A idéia foi usar a solução de um problema restrito montado de forma a refletir as exigências da função de Lennard-Jones. Os experimentos numéricos mostraram que o objetivo foi alcançado, pois é mais rápido resolver o problema de Lennard-Jones resolvendo antes o problema restrito do que resolver diretamente o problema de Lennard-Jones usando pontos iniciais aleatórios.

Para ilustrar uma aplicação da função tipo Lennard-Jones usamos um problema vindo da psicometria, o problema de agrupamento de pessoas. Outra alternativa usada para resolver o problema de agrupamento de pessoas é a minimização da função STRESS, $\sum_{i \neq j} (d_{ij} - \|x^i - x^j\|)^2$. Uma dificuldade associada a este modelo é que superestimar as distâncias têm o mesmo nível de gravidade que subestimá-las. Isto é, o modelo trata da mesma maneira os pontos que estão mais perto do que deviam e os pontos que estão mais longe do que deveriam. Já o modelo de Lennard-Jones trata de maneira diferenciada, punindo de forma mais severa os pontos que estão mais perto do que deveriam, do que os pontos que estão longe.

Conclusões

A vantagem em trabalhar com funções de penalização exata é que não há a necessidade de se resolver uma seqüência infinita de subproblemas penalizados evitando, assim, os problemas computacionais de mal condicionamento que aparecem quando temos que resolver subproblemas penalizados com valores grandes para o parâmetro de penalização.

Neste trabalho apresentamos uma classe de funções de penalização exata, $P(x, \rho)$. Mostramos (Teorema 10) que, ao penalizarmos somente as restrições não lineares do (PNL), mantendo os subproblemas penalizados restritos a um conjunto Ω , a classe de funções $P(x, \rho)$ continua sendo exata. Este resultado sugere usar uma função de penalização exata com subproblemas restritos. Embora resolver problemas restritos seja mais complexo do que resolver problemas irrestritos, aproximações tratando subproblemas restritos permitem tirar vantagens da geometria do conjunto de restrições Ω . Esta idéia possibilita usar algoritmos estáveis de programação não linear com restrições simples (ver [19],[31]) para resolver os subproblemas penalizados. Além disso, novos algoritmos podem ser desenvolvidos explorando propriedades particulares da função de penalização exata escolhida e do conjunto de restrições. Nossos resultados de penalização exata com subproblemas restritos estão também apresentados no trabalho [27].

Para resolver o (PNL) propomos um algoritmo baseado na função de penalização exata L_1 onde penalizamos somente as restrições não lineares. Desta forma os subproblemas penalizados permanecem restritos ao conjunto Ω . Para resolver os subproblemas penalizados aplicamos a técnica de região de confiança onde a função modelo usada para aproximar o subproblema penalizado foi também não

diferenciável. Os experimentos numéricos mostraram que o método de penalização com região de confiança é consistente no sentido de resolver o (PNL), porém não é eficiente quando comparado com o algoritmo Box que resolve o (PNL) usando o método do Lagrangeano aumentado. O trabalho recente de Y. Yuan [39] apresenta um algoritmo para resolver o problema de programação não linear com $\Omega = \mathbb{R}^n$ baseado na função de penalização exata L_∞ e na técnica de região de confiança. É interessante investigar a possibilidade de estender este algoritmo, tanto do ponto de vista teórico como prático, para o (PNL) mais geral procurando manter os subproblemas envolvidos restritos ao conjunto Ω , e assim tirar vantagem da estrutura deste conjunto.

Abordamos ainda o problema de minimização global da função de Lennard-Jones onde propomos gerar bons pontos iniciais para este problema usando a solução de um problema com restrições. Os resultados numéricos mostraram que é mais eficiente resolver antes o problema restrito do que resolver diretamente o problema de Lennard-Jones usando pontos iniciais aleatórios. Ilustramos o modelo de Lennard-Jones através de um problema de agrupamento de pessoas que pode também ser resolvido usando técnicas de escalamento multidimensional. Acreditamos que é interessante comparar o modelo de Lennard-Jones com técnicas de escalamento multidimensional para este tipo de problema.

Bibliografia

- [1] J.W. BANDLER & C. CHARALAMBOUS, Nonlinear programming using mini-max techniques, *Journal of Optimization Theory and Applications*, **13**, 679–686, 1974.
- [2] M.S. BAZARAA & C.M. SHETTY, *Nonlinear Programming Theory and Algorithms*, John Wiley & Sons, 1979.
- [3] D.P. BERTSEKAS, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, 1982.
- [4] I. BONGARTZ, A.R. CONN, N.I. GOULD, & L. TOINT, CUTE: Constrained and Unconstrained Testing Environment. *Research Report RC 18860*, IBM T. J. Watson Research Center, Yorktown Heights, USA, 1993.
- [5] A. BEN-TAL, J. ZOWE, Necessary and sufficient optimality conditions for a class of nonsmooth minimization problems, *Mathematical Programming*, **24**, 70–91, 1982.
- [6] R.M. CHAMBERLAIN, Some examples of cycling in variable metric methods for constrained minimization, *Mathematical Programming*, **16**, 378–383, 1979.
- [7] C. CHARALAMBOUS, A lower bound for the controlling parameters of the exact penalty functions, *Mathematical Programming*, **15**, 278–290, 1978.
- [8] C. CHARALAMBOUS, On the condition for optimality of the nonlinear L_1 problem, *Mathematical Programming*, **17**, 123–135, 1979.
- [9] T.F. COLEMAN & A.R. CONN, Second-order conditions for an exact penalty function, *Mathematical Programming*, **19**, 178–185, 1980.

- [10] A.R. CONN, N. GOULD, L. TOINT, Large-scale Nonlinear Constrained Optimization: a Current Survey, *Algorithms for Continuous Optimization*, 282–332, 1994.
- [11] G. DI PILLO & L. GRIPPO, A continuously exact penalty function for nonlinear programming problems with inequality constraints, *SIAM Journal on Control and Optimization*, **23**, 72–84, 1985.
- [12] G. DI PILLO & L. GRIPPO, An exact penalty method with global convergence properties for nonlinear programming problems, *Mathematical Programming*, **36**, 1–18, 1986.
- [13] G. DI PILLO, L. GRIPPO, Exact penalty functions in constrained optimization, *SIAM Journal on Control and Optimization*, **27**, 1333–1360, 1989.
- [14] S. DOLECKI & S. ROLEWICZ, Exact penalty for local minima, *SIAM Journal on Control and Optimization*, **17**, 111–124, 1979.
- [15] H.W. DOMMEL & W. TINNEY Optimal Power Flow Solutions *IEEE transactions on PAS*, **87**, 1866–1876, 1968.
- [16] V.R. EVANS & F.J. GOULD, Stability and exponential penalty function techniques in nonlinear programming, *Institute of Statistics Mimeo Series*, 723, University of North Carolina at Chapel Hill, 1970.
- [17] R. FLETCHER, A class of methods for nonlinear programming with termination and convergence properties, J. Abadie, ed., *Integer and Nonlinear Programming*, North-Holland, Amsterdam, 157–173, 1970.
- [18] R. FLETCHER, *Practical Methods of Optimization*, Wiley, New York, 1987.
- [19] A. FRIEDLANDER, J.M. MARTÍNEZ & S.A. SANTOS, A new trust region algorithm for bound constrained minimization, *Journal of Applied Mathematics and Optimization*, **30**, 235–266, 1994.
- [20] T. GLAD & E. POLAK, A multiplier method with automatic limitation of penalty growth, *Mathematical Programming*, **17**, 140–155, 1979.
- [21] GOCKENBACH, KEARSLEY & SYMES, An infeasible point method for minimizing the Lennard-Jones potential, *Computational Optimization and Applications*, **8**, 273–286, 1997.

- [22] S.P. HAN & O.L. MANGASARIAN, Exact penalty functions in nonlinear programming, *Mathematical Programming*, **17**, 251–269, 1979.
- [23] M.R. HOARE & J.A. MCINNES, Morphology and statistical statics of simple microclusters, *Adv. Phys.*, **32**, 791–821, 1983.
- [24] M.R. HOARE & P. PAL, Physical cluster mechanics: Statics and energy surfaces for monatomic systems, *Adv. Phys.*, **20**, 161–196, 1981.
- [25] W. HOCK & K. SCHITTKOWSKI, *Test Examples for Nonlinear Programming Codes Lect. Notes Economics and Mathematics Systems*, Springer, Berlin, 1981.
- [26] S.P. HOWE, New conditions for exactness of a simple penalty function, *SIAM Journal on Control and Optimization*, **11**, 378–381, 1973.
- [27] S.M.H. JANESCH & L.T. SANTOS, Exact penalty methods with constrained subproblems, *Investigacion Operativa*, **7**, 55–65, 1997.
- [28] D. LUENBERGER, Control problem with kinks, *IEEE Transaction on Automatic Control*, **15**, 570–574, 1970.
- [29] D. LUENBERGER, *Linear and Nonlinear Programming*, Addison-Wesley, New York, 1984.
- [30] N. MARATOS, Exact penalty functions algorithms for finite-dimensional and control optimizations problems, Ph.D. Thesis, University of London, England, 1978.
- [31] J.M. MARTÍNEZ & A.C. MORETTI, A trust region method for minimization of nonsmooth functions with linear constraints, *Mathematical Programming*, **76**, 431–449, 1997.
- [32] J.M. MARTÍNEZ & L.T. SANTOS, Some new theoretical results on recursive quadratic programming algorithms, *Journal of Optimization Theory and Applications*, **97**, 435–454, 1998.
- [33] A. MONTICELLI, *Fluxo de Carga em Redes de Energia Elétrica*, Edgard Blücher, São Paulo, 1983.
- [34] F.H. MURPHY, A class of exponential penalty functions, *SIAM Journal on Control and Optimization*, **12**, 679–686, 1974.

- [35] J.A. NORTHBY, Structure and binding of Lennard-Jones cluster: $13 \leq N \leq 147$, *J. Chem. Phys.*, **87**, 6166–6177, 1987.
- [36] T. PIETRZYKOWSKI, An exact potential method for constrained maxima, *SIAM Journal Numerical Analysis*, **6**, 299–304, 1969.
- [37] M. TREVISAN, *Restrições Funcionais de Desigualdade no FPO Newton Tratadas pelo Método da Continuação*. Campinas, 1995. Tese de Mestrado - FEE, UNICAMP.
- [38] W.I. ZANGWILL, Nonlinear programming via penalty functions, *Management Sciences*, **13**, 344–358, 1967.
- [39] Y. YUAN, On the convergence of a new trust region algorithm, *Numerische Mathematik*, **70**, 515–539, 1995.