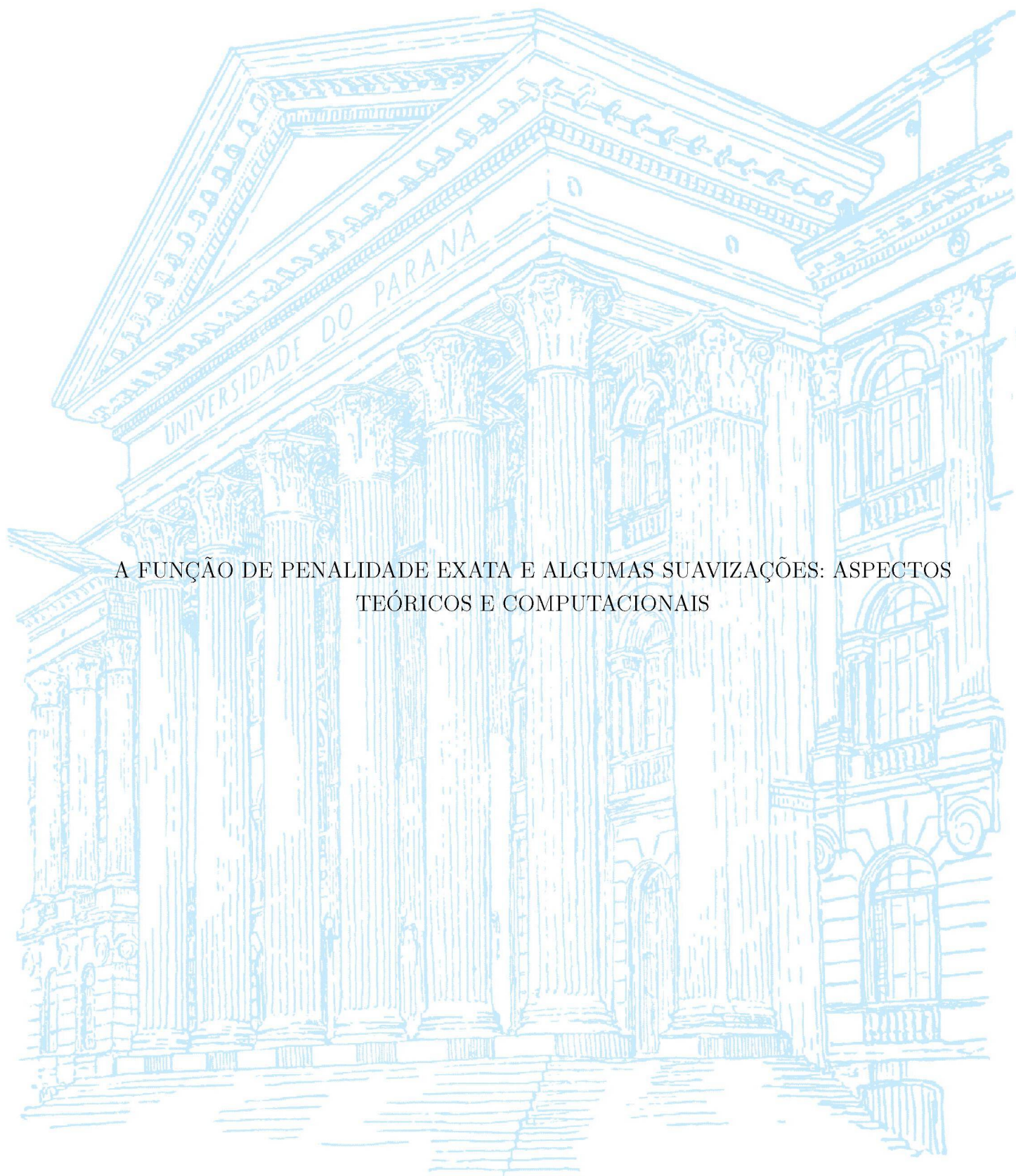


UNIVERSIDADE FEDERAL DO PARANÁ  
MARIANA DA ROSA



A FUNÇÃO DE PENALIDADE EXATA E ALGUMAS SUAVIZAÇÕES: ASPECTOS  
TEÓRICOS E COMPUTACIONAIS

CURITIBA  
2022

UNIVERSIDADE FEDERAL DO PARANÁ  
MARIANA DA ROSA

A FUNÇÃO DE PENALIDADE EXATA E ALGUMAS SUAVIZAÇÕES: ASPECTOS  
TEÓRICOS E COMPUTACIONAIS

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Matemática, Setor de Ciências Exatas, Universidade Federal do Paraná, como requisito parcial à obtenção do Título de Mestre em Matemática Aplicada.

Orientador: Prof. Dr. Ademir Alves Ribeiro.

CURITIBA  
2022

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)  
UNIVERSIDADE FEDERAL DO PARANÁ  
SISTEMA DE BIBLIOTECAS – BIBLIOTECA DE CIÊNCIA E TECNOLOGIA

Rosa, Mariana da

A função de penalidade exata e algumas suavizações: aspectos teóricos e computacionais / Mariana da Rosa. – Curitiba, 2022.

1 recurso on-line : PDF.

Dissertação (Mestrado) - Universidade Federal do Paraná, Setor de Ciências Exatas, Programa de Pós-Graduação em Matemática.

Orientador: Ademir Alves Ribeiro

1. Otimização matemática. 2. Algoritmos. 3. Funções (Matemática). 4. Penalidade exata. I. Universidade Federal do Paraná. II. Programa de Pós-Graduação em Matemática. III. Ribeiro, Ademir Alves. IV. Título.

Bibliotecário: Elias Barbosa da Silva CRB-9/1894

## ATA DE SESSÃO PÚBLICA DE DEFESA DE MESTRADO PARA A OBTENÇÃO DO GRAU DE MESTRA EM MATEMÁTICA

No dia sete de março de dois mil e vinte e dois às 20:30 horas, na sala [https://meet.jit.si/DefesaMestrado\\_UFPRR\\_Mariana\\_da\\_Rosa](https://meet.jit.si/DefesaMestrado_UFPRR_Mariana_da_Rosa), remoto, foram instaladas as atividades pertinentes ao rito de defesa de dissertação da mestranda **MARIANA DA ROSA**, intitulada: **A função de penalidade exata e algumas suavizações: aspectos teóricos e computacionais.**, sob orientação do Prof. Dr. ADEMIR ALVES RIBEIRO. A Banca Examinadora, designada pelo Colegiado do Programa de Pós-Graduação MATEMÁTICA da Universidade Federal do Paraná, foi constituída pelos seguintes Membros: ADEMIR ALVES RIBEIRO (UNIVERSIDADE FEDERAL DO PARANÁ), CLAUDIA ALEJANDRA SAGASTIZÁBAL (UNIVERSIDADE ESTADUAL DE CAMPINAS), REGINA SANDRA BURACHIK (UNIVERSITY OF SOUTH AUSTRALIA). A presidência iniciou os ritos definidos pelo Colegiado do Programa e, após exarados os pareceres dos membros do comitê examinador e da respectiva contra argumentação, ocorreu a leitura do parecer final da banca examinadora, que decidiu pela APROVAÇÃO. Este resultado deverá ser homologado pelo Colegiado do programa, mediante o atendimento de todas as indicações e correções solicitadas pela banca dentro dos prazos regimentais definidos pelo programa. A outorga de título de mestra está condicionada ao atendimento de todos os requisitos e prazos determinados no regimento do Programa de Pós-Graduação. Nada mais havendo a tratar a presidência deu por encerrada a sessão, da qual eu, ADEMIR ALVES RIBEIRO, lavrei a presente ata, que vai assinada por mim e pelos demais membros da Comissão Examinadora.

CURITIBA, 07 de Março de 2022.

Assinatura Eletrônica

08/03/2022 06:58:40.0

ADEMIR ALVES RIBEIRO

Presidente da Banca Examinadora

Assinatura Eletrônica

08/03/2022 08:39:11.0

CLAUDIA ALEJANDRA SAGASTIZÁBAL

Avaliador Externo (UNIVERSIDADE ESTADUAL DE CAMPINAS)

Assinatura Eletrônica

08/03/2022 23:19:47.0

REGINA SANDRA BURACHIK

Avaliador Externo (UNIVERSITY OF SOUTH AUSTRALIA)



MINISTÉRIO DA EDUCAÇÃO  
SETOR DE CIÊNCIAS EXATAS  
UNIVERSIDADE FEDERAL DO PARANÁ  
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO MATEMÁTICA -  
40001016041P1

## TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação MATEMÁTICA da Universidade Federal do Paraná foram convocados para realizar a arguição da dissertação de Mestrado de **MARIANA DA ROSA** intitulada: **A função de penalidade exata e algumas suavizações: aspectos teóricos e computacionais.**, sob orientação do Prof. Dr. ADEMIR ALVES RIBEIRO, que após terem inquirido a aluna e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de mestra está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 07 de Março de 2022.

Assinatura Eletrônica

08/03/2022 06:58:40.0

ADEMIR ALVES RIBEIRO

Presidente da Banca Examinadora

Assinatura Eletrônica

08/03/2022 08:39:11.0

CLAUDIA ALEJANDRA SAGASTIZÁBAL

Avaliador Externo (UNIVERSIDADE ESTADUAL DE CAMPINAS)

Assinatura Eletrônica

08/03/2022 23:19:47.0

REGINA SANDRA BURACHIK

Avaliador Externo (UNIVERSITY OF SOUTH AUSTRALIA)

Coordenação PPGMA, Centro Politécnico, UFPR - CURITIBA - Paraná - Brasil

CEP 81531990 - Tel: (41) 3361-3026 - E-mail: [pgmat@ufpr.br](mailto:pgmat@ufpr.br)

Documento assinado eletronicamente de acordo com o disposto na legislação federal Decreto 8539 de 08 de outubro de 2015.

Gerado e autenticado pelo SIGA-UFPR, com a seguinte identificação única: 159892

Para autenticar este documento/assinatura, acesse <https://www.prppg.ufpr.br/siga/visitante/autenticacaoassinaturas.jsp> e insira o código 159892

## AGRADECIMENTOS

Agradeço aos meus pais Vilson e Fatima e aos meus irmãos Jean e Vilcineia pelo afeto, amor e encorajamento para que eu sempre seguisse os meus objetivos.

Ao meu parceiro Alexandre, por cada instante que estive ao meu lado, também pelas contribuições no desenvolvimento deste trabalho.

Aos meus orientadores Professores Ademir e Elizabeth, agradeço pela paciência, pelas instigantes discussões que promoveram muito aprendizado, em especial as aulas de otimização que nunca esquecerei. Agradeço pela orientação deste trabalho, pela amizade, motivação e pelos conselhos em momentos de aflição.

Às Professoras Claudia e Regina, pela leitura cuidadosa e pelas valiosas contribuições a este trabalho e sugestões para trabalhos futuros. Agradeço também pelas palestras no Seminário Brasileiro de Otimização, que proporcionaram a inspiração e convicção em prosseguir os meus estudos em otimização.

Aos meus amigos e colegas, pelas risadas, apoio e compreensão, especialmente, a Maria Eduarda que tornou a caminhada do mestrado mais prazerosa. Certamente aprendemos muita matemática juntas e compartilhamos momentos muito especiais, mesmo remotamente; à Angela e ao João que tive o privilégio de conviver durante o período de isolamento e sempre estiveram ao meu lado.

À todos os Professores do PPGM-UFPR e do curso de Licenciatura em Matemática - Unioeste/Cascavel que fizeram parte da minha história e contribuíram com a minha formação.

À CAPES pelo financiamento deste projeto.

Àqueles que de alguma forma cooperaram, permitindo a realização deste trabalho, o meu muito obrigada.

## RESUMO

Uma das estratégias para resolver problemas de otimização com restrições através de métodos de penalidade externa faz uso de funções de penalidade não diferenciáveis, o que limita o uso de eficientes algoritmos suaves para resolver os subproblemas. Em vista desta dificuldade, técnicas de suavizar vêm sendo amplamente propostas na literatura. Neste trabalho, estudamos a teoria de métodos de penalidade externa, mostrando suas principais propriedades e discutindo a dicotomia entre funções de penalidade exata e suaves. Trazemos de forma unificada resultados e métodos baseados em funções que suavizam e aproximam uma função de penalidade exata. Por fim, apresentamos resultados numéricos de comparação do desempenho computacional dos métodos estudados na resolução de problemas da biblioteca CUTEst.

**Palavras-chave:** *Métodos de penalidade externa, Função de penalidade exata, Função de penalidade suavizada, Experimentos numéricos*

## ABSTRACT

One of the strategies to solve constrained optimization problems through external penalty methods utilizes non-differentiable penalty functions, which limits the use of efficient smooth algorithms to solve the subproblems. In light of this difficulty, a great number of smoothing techniques have been proposed in the literature. In this work, we study the theory of exterior penalty methods. We show their main properties and discuss the dichotomy between exact and smooth penalty functions. We present on a unified manner results and methods based on functions that smooth and approximate an exact penalty function. Finally, we present numerical results comparing the computational performance of the studied methods on the resolution of problems from the CUTEst library.

**Keywords:** *Exterior penalty methods, Exact penalty function, Smoothed penalty function, Numerical experiments.*

# SUMÁRIO

<b>Introdução</b>	<b>1</b>
<b>1 Métodos de penalidade</b>	<b>4</b>
1.1 O método . . . . .	4
1.1.1 O problema de penalidade . . . . .	4
1.1.2 O algoritmo de penalidade e sua convergência global . . . . .	9
1.1.3 Outras considerações . . . . .	12
1.2 Funções de penalidade . . . . .	14
1.2.1 Exemplos . . . . .	14
1.2.2 Uma estimativa local de inviabilidade . . . . .	17
1.3 Taxa de convergência . . . . .	22
<b>2 Penalização exata e suave</b>	<b>29</b>
2.1 Método de penalidade exata . . . . .	29
2.1.1 O problema de penalidade exata . . . . .	30
2.1.2 Condições necessárias e condições suficientes . . . . .	33
2.2 Método de penalidade suave . . . . .	41
2.2.1 Construindo funções suaves . . . . .	42
2.2.2 Multiplicadores de Lagrange e o mal condicionamento . . . . .	45
<b>3 Funções de aproximação suave</b>	<b>50</b>
3.1 Definições e propriedades . . . . .	50
3.2 Relações entre as soluções dos problemas . . . . .	52
3.3 Algoritmo de penalidade suavizado . . . . .	55
3.4 Exemplos de aproximação suave . . . . .	60
<b>4 Experimentos numéricos</b>	<b>69</b>
4.1 Algoritmos e critérios de parada . . . . .	69
4.1.1 Parâmetros iniciais . . . . .	71
4.2 Metodologia de comparação . . . . .	72
4.3 Resultados numéricos . . . . .	73
4.3.1 Resultados obtidos com o critério de parada (4.2) . . . . .	73
4.3.2 Resultados obtidos com o critério de parada (4.3) . . . . .	78
4.3.3 Comentários gerais . . . . .	81
<b>Conclusões</b>	<b>84</b>

<b>Referências</b>	<b>84</b>
<b>A Resultados numéricos: comparação entre os algoritmos</b>	<b>89</b>
A.1 Resultados com o critério de parada (4.2) . . . . .	89
A.2 Resultados com o critério de parada (4.3) . . . . .	95

# Introdução

Inúmeras aplicações em diferentes áreas como engenharias, setor energético, transporte, economia e inteligência artificial recaem em problemas de otimização. Para resolver um problema dessa natureza, considerado *difícil*, vários métodos têm como filosofia substituí-lo por uma sequência de problemas mais *fáceis*. Como por exemplo, o método de Newton para um sistema de equações não lineares, em que uma solução é aproximada resolvendo-se sucessivos sistemas de equações lineares.

Outros exemplos são as abordagens de Lagrangiano Aumentado, Programação Quadrática Sequencial e métodos de penalidade interna ou externa [3, 15, 22, 23, 26] que transformam um problema de otimização com restrições em uma sequência de problemas irrestritos ou com restrições apenas lineares ou quadráticas.

Neste trabalho, estamos particularmente interessados nos métodos de função de penalidade externa para resolver problemas de otimização com restrições de desigualdade

$$\begin{aligned} &\text{minimizar } f(x) \\ &\text{sujeito a } g_i(x) \leq 0, \quad i = 1, \dots, m. \end{aligned} \tag{1}$$

em que  $f, g_i : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $i = 1, \dots, m$ , são funções suaves. Esses métodos, também chamados simplesmente de métodos de penalidade, consideram uma sequência de problemas *mais fáceis* que consistem na minimização irrestrita de uma função penalizada. Tal função é definida pela soma da função objetivo do problema original e um termo de penalização da inviabilidade. Este termo consiste no produto de um parâmetro de penalidade, que é acrescido a cada novo problema, por uma função de penalidade, fixada para todos os problemas *mais fáceis*. Desta forma, distintas escolhas desta função nos fornecem diferentes métodos de penalidade.

Uma das mais clássicas é a função de penalidade exata  $\ell_1$  ou simplesmente penalidade exata, que foi introduzida por Zangwill [40]. Esta função é uma medida de inviabilidade obtida pela composição da norma da soma com a função  $\max\{0, \cdot\}$  aplicada às restrições. Zangwill mostrou um importante resultado relativo ao método de penalidade  $\ell_1$ : se o problema original é convexo e o problema irrestrito tem um minimizador local, então uma solução deste problema, com um parâmetro de penalidade suficientemente grande, é também uma solução para o problema com restrições. Todas as funções de penalidade

com essa propriedade são chamadas de exatas [12, 31].

Por outro lado, a função objetivo do problema de penalidade exata não é suave, visto que a máximo carece de diferenciabilidade na origem. Do ponto de vista computacional, isto limita o uso de eficientes algoritmos de minimização, tais como os métodos quase-Newton, na resolução dos problemas irrestritos. Uma forma de superar esta dificuldade consiste em suavizar a função máximo. Em particular, na área de otimização o quadrado dessa função é amplamente utilizado em vista da sua simplicidade. Tal suavização, no contexto de métodos de penalidade, resulta na função  $\ell_2$  também chamada penalidade quadrática.

No entanto, esse novo método de penalidade é assintótico, no sentido que não há como garantir que uma solução do problema irrestrito seja de fato uma solução do problema original. Um dos problemas relacionados a isto, é que as funções penalizadas tornam-se cada vez mais mal condicionadas, à medida que o parâmetro de penalidade fica muito grande [15, 23, 34].

Diante dessas dificuldades, diversos autores introduziram famílias de funções que, além de suavizar, aproximam a função máximo (veja por exemplo, Bertsekas [3], Zang [13], Gonzaga e Castillo [10], e Pinar e Zenios [32]). Em geral, tais famílias podem ser expressas através de uma única função definida em  $\mathbb{R} \times \mathbb{R}_+$ , na qual a segunda variável determina a precisão da aproximação. Neste trabalho, nos referimos a esta suavização como *função de aproximação suave*.

Deste modo, um novo método para resolver o problema com restrições é obtido, no qual os problemas considerados *mais fáceis* são irrestritos e suaves, cuja função objetivo é construída substituindo a função máximo pela função de aproximação suave com um parâmetro de aproximação fixado. Assim, quanto mais próximo de zero o parâmetro de suavização estiver, menor será o erro de aproximação entre as funções penalizadas exata e suavizadas e, portanto, espera-se que uma solução do problema de penalidade suavizado esteja mais próxima de ser uma solução do problema *difícil* (1).

Tendo em vista a discussão apresentada e que um número considerável de métodos que suavizam uma função de penalidade exata foram propostos nas últimas duas décadas [7, 20, 35, 38, 39], apresentamos neste trabalho um estudo, tanto do ponto de vista teórico como computacional, de métodos de penalidade, em especial dos métodos de penalização exata e algumas de suas suavizações. Para isto, organizamos este trabalho em quatro capítulos.

No *Capítulo 1*, motivamos e definimos o método de penalidade e provamos sua convergência global. Em seguida, definimos funções de penalidade de interesse neste trabalho e mostramos que algumas delas satisfazem uma estimativa local de inviabilidade conhecida como *condição de error bound*. Na Seção 1.3, trazemos uma estimativa local para a taxa de convergência do método de penalidade. Em especial, definimos função de penalidade exata e provamos que métodos de penalidade diferenciáveis não são exatos.

O *Capítulo 2* apresenta uma discussão mais detalhada dos métodos de penalidade exata [37] e suave [22]. Na Seção 2.1, trazemos condições necessárias e suficientes e estendemos resultados apresentados em [15] para penalização exata. Na Seção 2.2, mostramos formalmente que as condições apresentadas em [22] de fato podem ser utilizadas para construir funções de penalidade suaves, para as quais os multiplicadores de Lagrange são muito bem aproximados. Por fim, ilustramos o mal condicionamento dos problemas de penalidade quadrática.

O *Capítulo 3* é dedicado aos métodos com funções de aproximação suave. Apresentamos de forma unificada os resultados e algoritmos encontrados nos trabalhos [20, 35, 38, 39]. Para este objetivo, estabelecemos condições gerais de modo que qualquer função satisfazendo tais condições herde esses resultados. Uma teoria de funções de aproximação suave para penalização exata, em espaços mais gerais do que  $\mathbb{R}^n$ , é também apresentada em [9].

Por fim, no *Capítulo 4* são discutidos resultados numéricos da resolução de problemas de otimização com restrições de desigualdade coletados da coleção CUTEst [11] pelos algoritmos implementados na linguagem de programação JULIA. Comparamos o desempenho do algoritmo de penalidade com as funções exata e quadrática, e dos métodos de suavização com as funções apresentadas em [20, 35, 38, 39]. Os resultados numéricos são apresentados no formato de tabela no Apêndice A.

# Capítulo 1

## Métodos de penalidade

O objetivo deste capítulo é estudar métodos de penalidade para o problema geral de otimização com restrições de desigualdade dado em (1). Para isto, denotamos o conjunto viável desse problema por

$$X := \{x \in \mathbb{R}^n \mid g_i(x) \leq 0, i = 1, \dots, m\},$$

o qual é fechado, visto que as funções  $g_i$ ,  $i = 1, \dots, m$ , são contínuas, e é assumido ser não vazio.

Vamos inicialmente introduzir esses métodos em uma abordagem geral, sem especificar a função de penalidade. Em seguida, na Seção 1.2, apresentamos alguns exemplos clássicos de funções de penalidade e discutimos algumas propriedades. Por fim, na Seção 1.3, abordamos aspectos da taxa de convergência dos métodos de penalidade que satisfazem uma estimativa local de inviabilidade e definimos penalização exata. As referências utilizadas na construção deste capítulo são [1, 14, 15, 22, 23, 33, 34].

### 1.1 O método

Nesta seção, motivamos e definimos o método de penalidade para o problema (1). Além de discutir propriedades dos problemas de penalidade, provaremos sob algumas hipóteses sua convergência global. Ademais, fazemos algumas considerações sobre o quão restritivas são essas hipóteses e discutimos brevemente alternativas propostas pela literatura.

#### 1.1.1 O problema de penalidade

Começamos esta seção introduzindo uma função que penaliza violações das restrições do problema (1), isto é, uma função  $P : \mathbb{R}^n \rightarrow \mathbb{R}_+$  contínua tal que para todo  $x \in \mathbb{R}^n$

$$P(x) = 0, \text{ se, e somente se, } x \in X. \tag{1.1}$$

Uma função  $P$  definida como em (1.1) é chamada de *função de penalidade para o problema (1)*.

Dada  $P$  uma função de penalidade, definimos uma família de funções  $\{\varphi_c\}_{c \in \mathbb{R}_+}$ , tal que para cada *parâmetro de penalidade*  $c \in \mathbb{R}_+$ , uma *função penalizada* é obtida como segue

$$\mathbb{R}^n \ni x \mapsto \varphi_c(x) := f(x) + cP(x)$$

Observe que fixado  $c > 0$ , a função penalizada aparece como uma perturbação da função objetivo do problema original, visto que se um ponto  $x \in \mathbb{R}^n$  é inviável, o termo  $cP(x) > 0$  penaliza a função objetivo de (1). Por outro lado, em pontos tais que  $x \in X$ , temos que  $cP(x) = 0$  e portanto nenhuma penalidade é exercida, e conseqüentemente  $\varphi_c(x) = f(x)$ . De imediato, temos os seguintes resultados.

**Proposição 1.1.** *Considere uma função de penalidade  $P$  e um ponto  $\bar{x} \in X$ . Então, para qualquer  $c > 0$  fixado,*

$$\inf\{\varphi_c(x) \mid x \in \mathbb{R}^n\} \leq f(\bar{x}). \quad (1.2)$$

*Além disso, se para algum  $\bar{c} > 0$  a função  $\varphi_{\bar{c}}$  é limitada inferiormente, então o problema (1) é limitado inferiormente e  $\varphi_c$  é limitada inferiormente para qualquer  $c > \bar{c}$ .*

*Prova.* A primeira afirmação decorre da condição (1.1) visto que

$$\inf\{\varphi_c(x) \mid x \in \mathbb{R}^n\} \leq \varphi_c(\bar{x}) = f(\bar{x}) + cP(\bar{x}) = f(\bar{x}).$$

Suponha agora que  $\varphi_{\bar{c}}$  é limitada inferiormente para algum  $\bar{c} > 0$ . Então da última relação segue que o problema (1) é limitado inferiormente, além disso para qualquer  $c > \bar{c}$  vale que

$$\varphi_{\bar{c}}(x) = f(x) + \bar{c}P(x) \leq f(x) + cP(x) = \varphi_c(x),$$

o que prova a última afirmação, visto que  $\varphi_{\bar{c}}$  é limitada inferiormente por hipótese. ■

Em particular, o resultado acima mostra que qualquer minimizador da função penalizada fornece um limitante inferior para o problema original.

Considere então o *problema de penalidade*, um problema irrestrito que depende do parâmetro de penalidade  $c > 0$  e da variável  $x \in \mathbb{R}^n$ , dado por

$$(P_c) \quad \text{minimizar } \varphi_c(x) = f(x) + cP(x) \\ \text{sujeito a } x \in \mathbb{R}^n.$$

Vamos discutir algumas relações entre a família de problemas de penalidade  $(P_c)$  e o problema (1). Com este intuito assumimos que o problema de penalidade sempre tem solução. No final da presente seção, discutimos brevemente essa suposição.

**Hipótese 1.** *Seja  $P$  uma função de penalidade. Para cada  $c > 0$  dado, existe um ponto  $x(c) \in \mathbb{R}^n$  tal que*

$$x(c) \in \operatorname{argmin}\{\varphi_c(x) \mid x \in \mathbb{R}^n\}. \quad (1.3)$$

Sob a hipótese acima, vemos que se  $x(c)$  é viável para o problema com restrições então  $P(x(c)) = 0$ , o que junto com a Proposição 1.1 mostra que

$$f(x(c)) = \varphi_c(x(c)) = \min_{\bar{x} \in \mathbb{R}^n} \varphi_c(\bar{x}) \leq f(\bar{x}), \quad (1.4)$$

para todo  $\bar{x} \in X$ , ou seja,  $x(c)$  é também uma solução de (1).

Infelizmente, as soluções de  $(P_c)$  não pertencem a  $X$  em geral. Na verdade, como sugere a nomenclatura “método de penalidade externa”, as soluções do problema de penalidade normalmente são externas ao conjunto viável do problema com restrições.

Sendo assim, a estratégia para melhorar a viabilidade de  $x(c)$  é baseada no seguinte raciocínio descrito em [22]: à medida que  $c$  aumenta, o custo da violação das restrições do problema (1) torna-se mais caro. Assim, espera-se que soluções do problema de penalidade se aproximem cada vez mais do conjunto viável e um ponto de acumulação de uma sequência de soluções dos problemas de penalidade resolva o problema original.

Antes de formalizarmos o efeito descrito acima e o método, vamos analisar em um simples problema de otimização o comportamento das funções  $f, P$  e  $\varphi$  em soluções do problema de penalidade, quando  $c$  cresce ilimitadamente.

**Exemplo 1.1.** *Considere o problema de otimização em  $\mathbb{R}$  com uma restrição de desigualdade*

$$\begin{aligned} & \text{minimizar } f(x) = x^2 \\ & \text{sujeito a } g(x) = x + 1 \leq 0. \end{aligned} \quad (1.5)$$

*Este problema tem uma única solução  $x^* = -1$ , ( $X = (-\infty, -1]$ ), com valor ótimo  $f(x^*) = 1$ .*

*Considere  $P : \mathbb{R} \rightarrow \mathbb{R}_+$  a função suave definida como*

$$P(x) = \frac{1}{2} \max\{x + 1, 0\}^2.$$

*Afirmamos que essa função satisfaz a condição (1.1) que a caracteriza como uma função de penalidade para o problema (1.5). De fato, note que*

$$x \in X = \{x \in \mathbb{R} \mid g(x) \leq 0\} \text{ se, e somente se, } x + 1 \leq 0,$$

*donde segue que  $x \in X$  se, e somente se,  $P(x) = 0$ .*

Considere agora a família de problemas de penalidade,

$$\begin{aligned} & \text{minimizar } \varphi_c(x) := x^2 + \frac{c}{2} \max\{x + 1, 0\}^2. \\ & \text{sujeito a } x \in \mathbb{R}^n. \end{aligned}$$

Fixado  $c > 0$ , a função penalizada é convexa e diferenciável. Portanto, um minimizador  $x(c)$  de  $\varphi_c$  satisfaz

$$0 = \nabla \varphi_c(x(c)) = 2x(c) + c \max\{x(c) + 1, 0\}.$$

Logo  $x(c) = -c/(2 + c) \notin X$ . Além disso, para quaisquer  $c > \bar{c} > 0$ , vale que

$$P(x(c)) = \frac{2}{(2 + c)^2} < \frac{2}{(2 + \bar{c})^2} = P(x(\bar{c})).$$

Ou seja, conforme  $c$  cresce o valor da penalidade decresce, e a distância da solução  $x(c)$  até a solução do problema original  $x^*$  fica cada vez menor, pois

$$|x(c) - x^*| = \frac{2}{2 + c} < \frac{2}{2 + \bar{c}} = |x(\bar{c}) - x^*|.$$

Em particular, dada uma sequência  $(c_k) \subset \mathbb{R}_+$  tal que  $c_k \rightarrow +\infty$ , definindo  $x^k = x(c_k)$ , vemos que

$$\lim_{k \rightarrow \infty} P(x^k) = \lim_{k \rightarrow \infty} \frac{2}{(2 + c_k)^2} \rightarrow 0 \quad \text{e} \quad \lim_{k \rightarrow \infty} x^k = x^*.$$

Por outro lado, à medida que o parâmetro de penalidade cresce, as funções  $f$  e  $\varphi_c$  também crescem

$$\varphi_c(x(c)) = \frac{c}{2 + c} > \frac{\bar{c}}{2 + \bar{c}} = \varphi_{\bar{c}}(x(\bar{c})) \quad \text{e} \quad f(x(c)) = \frac{c^2}{(2 + c)^2} > \frac{\bar{c}^2}{(2 + \bar{c})^2} = f(x(\bar{c})).$$

A Figura 1.1 mostra os minimizadores  $x^k$  das funções  $\varphi_{c_k}$ , para  $k = 1, 2, 3$ , nos quais os parâmetros de penalidade satisfazem  $c_1 < c_2 < c_3$ . Nessa figura podemos observar que o ponto  $x^3$  está mais próximo da solução  $x^*$  do que  $x^2$  e  $x^1$ . O mesmo ocorre para  $\varphi_{c_3}(x^3)$  e  $f(x^3)$  que tem valores objetivos piores em relação a  $\varphi_{c_k}(x^k)$  e  $f(x^k)$ , para  $k = 1, 2$ , porém estão mais próximos do mínimo ótimo  $f(x^*) = 1$ . Em particular, vemos que as imagens das funções penalizadas e da função objetivo do problema restrito coincidem à esquerda da linha tracejada, onde qualquer ordenada é viável para o problema (1.5), enquanto que  $x^1, x^2$  e  $x^3$  são inviáveis para o problema original.

**Observação 1.2.** Em especial, o exemplo acima mostra que em geral as soluções dos problemas penalizados são pontos inviáveis para o problema original. Uma importante

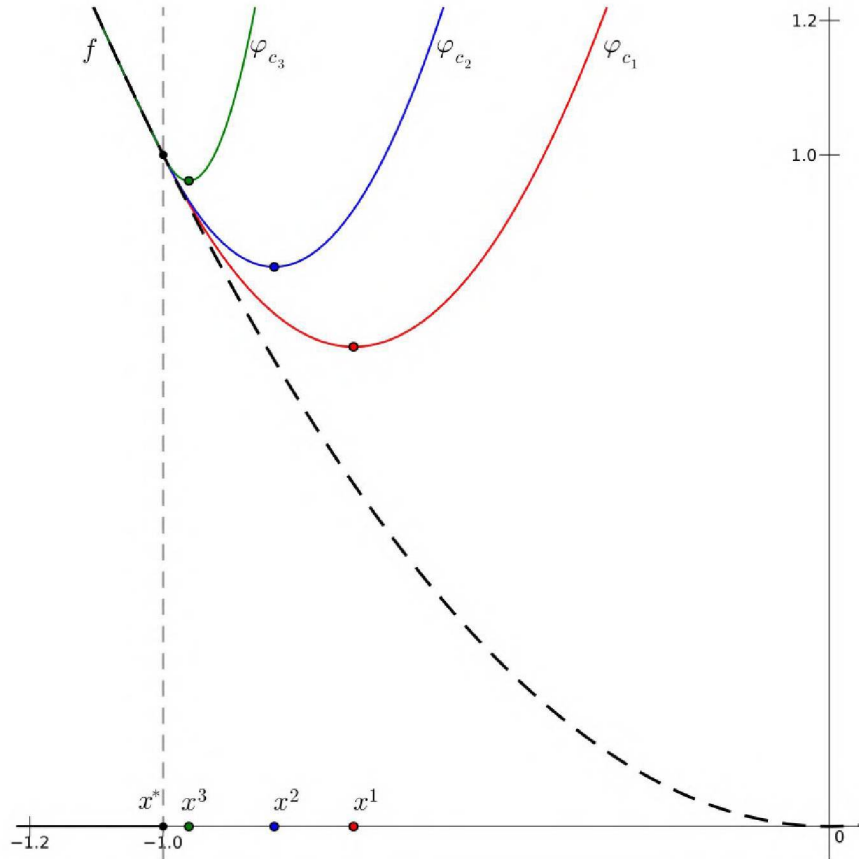


Figura 1.1: Ilustração do comportamento das soluções dos problemas de penalidade.

exceção é dada pela penalização exata, na qual uma solução do problema (1) pode ser recuperada com um valor de parâmetro  $c$  suficientemente grande, porém finito. Discutiremos com mais detalhes a penalização exata na Seção 1.3 e Seção 2.1. Em antecipação, no Exemplo 1.4 mostramos que uma solução do problema (1.5) pode ser obtida resolvendo um único problema de penalidade ao utilizarmos uma função de penalidade exata.

No Exemplo 1.1 vimos que se o parâmetro de penalidade aumenta a inviabilidade de  $x(c)$  e as imagens de  $f$  e  $\varphi_c$  nesse ponto tornam-se mais próximas do mínimo do problema com restrições. A seguir mostramos que essas propriedades são mantidas para qualquer função de penalidade utilizada. A demonstração que apresentamos é baseada em [15, Prop. 4.3.4].

**Lema 1.3.** *Seja  $P : \mathbb{R}^n \rightarrow \mathbb{R}_+$  uma função de penalidade. Então as seguintes afirmações são verdadeiras para todo  $c > \bar{c} > 0$ :*

$$P(x(c)) \leq P(x(\bar{c})); \quad (1.6)$$

$$f(x(c)) \geq f(x(\bar{c})); \quad (1.7)$$

$$\varphi_c(x(c)) \geq \varphi_{\bar{c}}(x(\bar{c})). \quad (1.8)$$

*Prova.* Sejam  $c, \bar{c} \in \mathbb{R}_+$  com  $\bar{c} < c$ . Como os pontos  $x(c)$  e  $x(\bar{c})$  são minimizadores globais das funções penalizadas  $\varphi_c$  e  $\varphi_{\bar{c}}$  respectivamente, segue que

$$f(x(c)) + \bar{c}P(x(c)) \geq f(x(\bar{c})) + \bar{c}P(x(\bar{c})), \quad (1.9)$$

$$f(x(\bar{c})) + cP(x(\bar{c})) \geq f(x(c)) + cP(x(c)). \quad (1.10)$$

Subtraindo (1.10) de (1.9) e depois simplificando, obtemos

$$(c - \bar{c})(P(x(\bar{c})) - P(x(c))) \geq 0.$$

Logo,  $P(x(\bar{c})) \geq P(x(c))$  pois  $c > \bar{c}$  o que prova (1.6). A relação (1.7) é imediata das relações (1.6) e (1.9).

Finalmente, note que como  $x(\bar{c})$  é minimizador da função  $\varphi_{\bar{c}}$  e  $0 < \bar{c} < c$ , temos

$$\varphi_{\bar{c}}(x(\bar{c})) \leq \varphi_{\bar{c}}(x(c)) = f(x(c)) + \bar{c}P(x(c)) \leq f(x(c)) + cP(x(c)) = \varphi_c(x(c)),$$

donde segue a relação (1.8). ■

### 1.1.2 O algoritmo de penalidade e sua convergência global

A relação (1.4) garante que o valor ótimo do problema original é limitado inferiormente pelo valor ótimo da função penalizada para qualquer  $c > 0$ . Em adição, o Lema 1.3 mostra que um melhor limitante inferior pode ser obtido à medida que o peso de penalidade é aumentado. Essa discussão é traduzida em termos matemáticos pela relação

$$\sup_{c \in \mathbb{R}_+} \varphi_c(x(c)) = \lim_{c_k \rightarrow +\infty} \varphi_{c_k}(x^k) \leq f^* \quad (1.11)$$

em que  $f^*$  denota o valor ótimo de (1) e  $x(c_k) = x^k$ . A pergunta natural que surge é se o valor ótimo do problema (1) é atingido no limite acima. Veremos no Teorema 1.5 que sob a hipótese da sequência  $(x^k)$  possuir um ponto de acumulação a resposta é afirmativa. Ainda mais, qualquer ponto de acumulação de  $(x^k)$  é uma solução de (1). Antes, vamos formalizar em termos algorítmicos a discussão feita até então, isto é, vamos definir um algoritmo de penalidade básico para obter um minimizador do problema (1), conforme encontrado em [15, 22, 23].

**Algoritmo 1.1.** *Método de função de penalidade básico*

DADO:  $c_0 > 0$

PARA  $k = 0, 1, 2, \dots$

CALCULE  $x^k \in \operatorname{argmin}\{\varphi_{c_k}(x) \mid x \in \mathbb{R}^n\}$

ESCOLHA  $c_{k+1} > c_k$

$k = k + 1$

**Observação 1.4.** *Martínez e Santos [23] observam que é aconselhável que  $x^{k-1}$  seja utilizado como ponto inicial do método aplicado para minimizar a função  $\varphi_{c_k}(\cdot)$ . E que, em geral, a atualização do parâmetro  $c_k$  é feita escolhendo  $c_0 = 1$  e  $c_k = 10c_{k-1}$ . Ou ainda, considerando a estratégia de tomar  $c_0$  muito grande (por exemplo,  $10^{24}$ ) e resolver um único problema de penalidade, o que nem sempre funciona segundo os autores.*

O teorema a seguir se baseia em resultados estabelecidos na literatura [15, 22, 23, 34] sobre o algoritmo de penalidade.

**Teorema 1.5.** *Seja  $(x^k)$  uma sequência gerada pelo Algoritmo 1.1 e suponha que  $c_k \rightarrow \infty$ . Então,*

- (i) *A função de penalidade converge para zero ao longo da sequência  $(x^k)$ , isto é,  $P(x^k) \rightarrow 0$ . Em particular, qualquer ponto de acumulação de  $(x^k)$  é viável para o problema (1).*
- (ii) *O problema (1) é limitado inferiormente e se  $x^* \in \mathbb{R}^n$  é um ponto de acumulação de  $(x^k)$ , então  $x^*$  é uma solução de (1) e*

$$\lim_{k \rightarrow \infty} \varphi_{c_k}(x^k) = f(x^*). \quad (1.12)$$

*Prova.* Pelo Lema 1.3, a sequência  $(\varphi_{c_k}(x^k))$  é não decrescente. Além disso, para quaisquer  $k \in \mathbb{N}$  e  $x \in X$ , temos

$$\varphi_{c_0}(x^0) \leq \varphi_{c_k}(x^k) \leq \varphi_{c_k}(x) = f(x). \quad (1.13)$$

Portanto,  $(\varphi_{c_k}(x^k))$  também é limitada e consequentemente convergente, digamos,

$$\varphi_{c_k}(x^k) \rightarrow \varphi_* = \sup\{\varphi_{c_k}(x^k) \mid k \in \mathbb{N}\}. \quad (1.14)$$

Por outro lado, novamente usando o Lema 1.3, temos que a sequência  $(f(x^k))$  também é monótona. Mais ainda,

$$f(x^0) \leq f(x^k) \leq f(x^k) + c_k P(x^k) = \varphi_{c_k}(x^k) \leq \varphi_*, \quad (1.15)$$

o que nos fornece

$$0 \leq c_k P(x^k) \leq \varphi_* - f(x^0).$$

Dividindo por  $c_k$  e passando o limite, obtemos (i).

Para (ii), note primeiro que a limitação de (1) segue diretamente de (1.13). Suponha agora que  $x^* \in \mathbb{R}^n$  é um ponto de acumulação de  $(x^k)$ , digamos,  $x^k \xrightarrow{\mathcal{K}} x^*$  com  $\mathcal{K} \subset \mathbb{N}$  um conjunto infinito. Pelo que provamos em (i), temos  $x^* \in X$ . Portanto, usando (1.13)

e (1.15), concluímos que

$$f(x^k) \leq \varphi_{c_k}(x^k) \leq f(x^*),$$

o que resulta em  $\varphi^* = f(x^*)$ , provando (1.12). Finalmente, passando o limite em (1.13), obtemos

$$f(x^*) = \varphi^* \leq f(x)$$

para todo  $x \in X$ , o que mostra que  $x^*$  é uma solução de (1). ■

O Lema 1.3 garante que a sequência  $(\varphi_{c_k}(x^k))$  é monótona, o que junto com a limitação por  $f^*$  dada em (1.11) implica na existência do limite. Uma questão que está em aberto é se vale a igualdade na relação em (1.11) sem a hipótese de existência de ponto de acumulação da sequência  $(x^k)$ .

Observe ainda que a limitação do problema restrito decorre da limitação do problema irrestrito para algum  $c > 0$ . Mas a recíproca não é verdadeira, como mostra o próximo exemplo.

**Exemplo 1.2.** *Considere o problema de otimização dado por*

$$\begin{aligned} \text{minimizar } f(x) &= -x^2 \\ \text{sujeito a } g_1(x) &= x - 1 \leq 0, \\ g_2(x) &= -x \leq 0. \end{aligned}$$

Observe que o conjunto viável do problema acima é compacto  $X = [0, 1]$ , portanto esse problema é limitado. Por outro lado, considerando a função de penalidade  $P : \mathbb{R} \rightarrow \mathbb{R}$  definida como

$$P(x) = \sum_{i=1}^2 \max\{0, g_i(x)\},$$

vemos que para qualquer  $c > 0$  a função penalizada, dada por

$$\varphi_c(x) = f(x) + cP(x) = \begin{cases} -x^2 & 0 \leq x \leq 1, \\ -x^2 + c(x - 1) & x > 1, \\ -x^2 - cx & x \leq 0, \end{cases}$$

é ilimitada inferiormente. Portanto, os problemas de penalidade são sempre ilimitados.

Em particular, o exemplo acima mostra que o método de penalidade pode ser ineficiente se a Hipótese 1 não é satisfeita, visto que minimizadores globais do problema irrestrito podem não existir.

### 1.1.3 Outras considerações

Destacamos que a convergência global do método de penalidade é assegurada sob a Hipótese 1. Entretanto, Izmailov e Solodov [15] salientam que exceto nos casos em que as funções penalizadas são convexas ou coercivas, não há como calcular soluções globais (pelo menos em geral). Por outro lado, Luenberger e Ye [22] relatam que não é necessário obter um minimizador exato da função penalizada a cada iteração do Algoritmo 1: se um minimizador aproximado do problema de penalidade é obtido, isto é,

$$\varphi_{c_k}(x^k) \leq \min_{x \in \mathbb{R}^n} \varphi_{c_k}(x) + \varepsilon,$$

então é possível mostrar que qualquer ponto de acumulação  $\bar{x}$  da sequência  $(x^k)$  é viável para o problema original e

$$f(\bar{x}) + \varepsilon \leq f(x^*),$$

com  $x^*$  um minimizador ótimo de (1). Generalizando esta discussão, apresentamos a seguir um algoritmo de penalidade relaxado para resolver o problema (1). A condição imposta na Hipótese 1 é relaxada para

**Hipótese 2.** *Existe  $\bar{c} > 0$  tal que  $\inf\{\varphi_{\bar{c}}(x) \mid x \in \mathbb{R}^n\} > -\infty$ .*

Note que devido à Proposição 1.1,  $\varphi_c$  é limitada inferiormente para qualquer  $c_0 > \bar{c}$ , o que garante a boa definição do algoritmo abaixo.

**Algoritmo 1.2.** *Método de penalidade relaxado*

DADOS:  $c_0 > \bar{c}$ ,  $\varepsilon_0 > 0$

PARA  $k = 0, 1, 2, \dots$

CALCULE  $x^k \in \mathbb{R}^n \mid \varphi_{c_k}(x^k) \leq \inf\{\varphi_{c_k}(x) \mid x \in \mathbb{R}^n\} + \varepsilon_k$

ESCOLHA  $c_{k+1} > c_k$  e  $0 \leq \varepsilon_{k+1} \leq \varepsilon_k$

$k = k + 1$

Estabelecemos no próximo teorema a convergência do Algoritmo 1.2.

**Teorema 1.6.** *Suponha que a Hipótese 2 é satisfeita, considere uma sequência  $(x^k)$  gerada pelo Algoritmo 1.2 com  $c_k \rightarrow \infty$  e defina  $\hat{\varepsilon} := \lim_{k \rightarrow \infty} \varepsilon_k$ . Então,*

(i) *O problema (1) é limitado inferiormente.*

(ii) *Se  $\bar{x} \in \mathbb{R}^n$  é um ponto de acumulação de  $(x^k)$  e  $f_* = \inf\{f(x) \mid x \in X\}$ , então  $\bar{x}$  é viável para o problema (1) e*

$$f_* \leq f(\bar{x}) \leq f_* + \hat{\varepsilon}.$$

(iii) No contexto do item anterior, se  $\hat{\varepsilon} = 0$ , então  $\bar{x}$  é uma solução de (1) e  $f_*$  é um ponto de acumulação da sequência  $(\varphi_{c_k}(x^k))$ .

*Prova.* Considere  $z \in X$  arbitrário. Então,

$$\varphi_{c_0}(x^0) - \varepsilon_0 \leq \inf\{\varphi_{c_0}(x) \mid x \in \mathbb{R}^n\} \leq \varphi_{c_0}(z) = f(z),$$

provando (i).

Para (ii), considere  $k \in \mathbb{N}$  fixado. Dado  $\hat{x} \in X$  temos

$$\begin{aligned} f(x^k) &\leq f(x^k) + c_k P(x^k) \\ &= \varphi_{c_k}(x^k) \\ &\leq \inf\{\varphi_{c_k}(x) \mid x \in \mathbb{R}^n\} + \varepsilon_k \\ &\leq \varphi_{c_k}(\hat{x}) + \varepsilon_k \\ &= f(\hat{x}) + \varepsilon_k. \end{aligned}$$

Consequentemente,

$$f(x^k) \leq \varphi_{c_k}(x^k) \leq f_* + \varepsilon_k \leq f_* + \varepsilon_0, \quad (1.16)$$

em que  $f_* = \inf\{f(x) \mid x \in X\}$ . Portanto, para todo  $k \in \mathbb{N}$  temos que

$$0 \leq c_k P(x^k) \leq f_* + \varepsilon_0 - f(x^k). \quad (1.17)$$

Por outro lado, como  $\bar{x}$  é um ponto de acumulação de  $(x^k)$ , existe um conjunto infinito  $\mathcal{K} \subset \mathbb{N}$  tal que  $x^k \xrightarrow{\mathcal{K}} \bar{x}$ . Assim, dividindo (1.17) por  $c_k$  e passando o limite sob a subsequência, obtemos  $P(\bar{x}) = 0$ , provando a primeira afirmação de (ii). Agora, note que (1.16) implica  $f(x^k) \leq f_* + \varepsilon_k$ , o que imediatamente resulta

$$f_* \leq f(\bar{x}) \leq f_* + \hat{\varepsilon},$$

concluindo a prova de (ii).

Finalmente, note que a primeira afirmação de (iii) segue diretamente de (ii). Além disso, de (1.16) decorre

$$f(x^k) \leq \varphi_{c_k}(x^k) \leq f_* + \varepsilon_k.$$

Passando o limite na expressão acima sob a subsequência convergente, obtemos que  $\varphi_{c_k}(x^k) \xrightarrow{\mathcal{K}} f_*$ , concluindo a prova.  $\blacksquare$

Uma outra alternativa para enfraquecer a exigência da Hipótese 1 é modificar a função de penalidade. Ao invés de considerarmos o problema de penalidade um problema irrestrito, poderíamos incluir na função de penalidade apenas restrições que são difíceis de lidar, enquanto que restrições simples, tais como restrições de caixa, poderiam ser manti-

das intactas. Assim, o problema de penalidade tomaria a forma

$$\text{minimizar } f(x) + c\bar{P}(x), \quad x \in X_0, \quad (1.18)$$

em que  $\bar{P} : \mathbb{R}^n \rightarrow \mathbb{R}_+$  é uma função contínua satisfazendo  $\bar{P}(x) = 0$ , se e somente se,  $x \in X_0$  com  $X \subset X_0$  em que  $X_0$ .

Uma das principais vantagens da configuração acima é que a Hipótese 1 pode ser enfraquecida. Por exemplo que  $X_0$  seja um conjunto compacto. Como discutido em [1], essa é uma hipótese pouco restritiva na maioria dos casos práticos, uma vez que as variáveis geralmente estão limitadas inferiormente e superiormente. Os resultados que foram apresentados nesta seção permanecem válidos sob a configuração (1.18). Entretanto, as demonstrações requerem mais detalhes que omitimos neste trabalho, mas podem ser encontradas em [1, 34].

## 1.2 Funções de penalidade

Apresentamos agora exemplos de funções de penalidade que serão importantes para a discussão das próximas seções. Em seguida trazemos algumas definições e mostramos que algumas funções de penalidade satisfazem uma estimativa local de inviabilidade conhecida como *condição de error bound*.

### 1.2.1 Exemplos

Considere a função auxiliar

$$\begin{aligned} [\cdot]_+ : \mathbb{R} &\rightarrow \mathbb{R}_+ \\ t &\mapsto [t]_+ = \max\{0, t\}. \end{aligned}$$

Utilizando a função  $[\cdot]_+$  podemos dizer se um ponto é, ou não, viável para o problema (1), visto que

$$[g_i(x)]_+ = 0, \quad \forall i = 1, \dots, m, \quad \text{se, e somente se, } x \in X.$$

Portanto, temos uma caracterização do conjunto viável  $X$  que nos permite definir diversas funções de penalidade para o problema (1) como vemos a seguir.

#### A função de penalidade exata $\ell_1$

Uma das mais clássicas funções de penalidade da literatura, conhecida como função de *penalidade exata*  $\ell_1$  ou simplesmente penalidade exata foi introduzida por Zangwill [40],

a qual é definida como

$$\ell_1(x) := \sum_{i=1}^m [g_i(x)]_+ = \|[g(x)]_+\|_1, \quad (1.19)$$

em que  $[g(x)]_+$  denota o vetor em  $\mathbb{R}^m$  cuja  $i$ -ésima entrada é dada por  $[g_i(x)]_+$ .

Observamos que a função de penalidade  $\ell_1$  já foi introduzida no Exemplo 1.2 e o termo “exata” é devido à propriedade de poder obter um minimizador do problema original via o método de penalidade em que  $P = \ell_1$ . Essa discussão será feita com mais detalhes nas Seções 1.3 e 2.1.

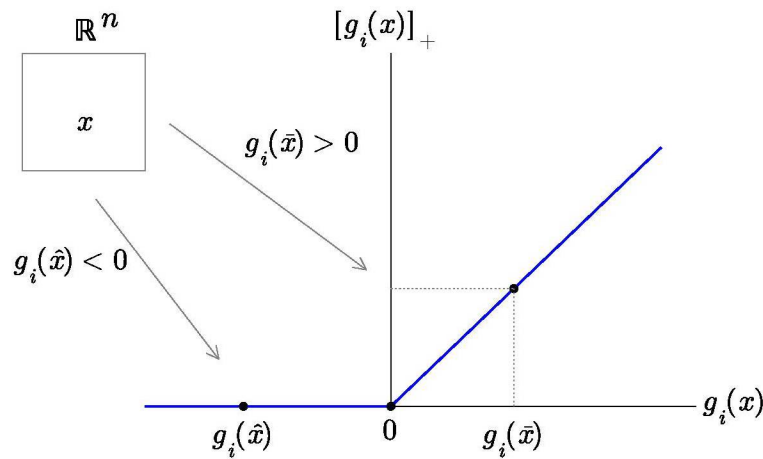


Figura 1.2: Representação da função  $[g_i(\cdot)]_+$ .

Normalmente pontos ótimos para um problema de otimização com restrições são tais que alguma das restrições é ativa, isto é,  $g_i(x) = 0$  para algum  $i \in \{1, \dots, m\}$ , e como ilustrado na Figura 1.2 a função de penalidade exata não é diferenciável nesses pontos.

### A função de penalidade $\ell_2$

Uma forma de superar a falta de diferenciabilidade da função  $\ell_1$  é suavizar a função auxiliar  $[\cdot]_+$  utilizando a função  $t \mapsto \frac{1}{2}([t]_+)^2$ .

Considere então a *função de penalidade suave*  $\ell_2$ , também chamada de *penalidade quadrática*,

$$\ell_2(x) := \sum_{i=1}^m \frac{1}{2} [g_i(x)]_+^2 = \frac{1}{2} \|[g(x)]_+\|_2^2. \quad (1.20)$$

Essa função já foi apresentada no Exemplo 1.1. Alguns autores como [1, 15, 23] definem a função de penalidade quadrática sem a constante  $1/2$ . Neste trabalho, seguimos a

definição dada em [22, 34] que permite eliminar a constante ao calcularmos a derivada

$$t \mapsto \frac{d}{dt} \left( \frac{1}{2}([t]_+)^2 \right) = [t]_+.$$

Na Figura 1.3 é comparada a penalização imposta pela função  $\max\{0, \cdot\}$  (representada em azul) e a suavização  $\max\{0, \cdot\}^2$  (representada em vermelho) em pontos  $x \in \mathbb{R}^n$  tais que  $|g_i(x)| < 1/2$ . Note que diferentemente de  $[g_i(\cdot)]_+$ , a função  $([g_i(\cdot)]_+)^2$  é suave em pontos  $x \in \mathbb{R}^n$  que possuem alguma restrição ativa. Observe ainda, que em um ponto  $\bar{x} \notin X$  para o qual  $g_i(\bar{x}) > 0$ , o valor de penalidade imposta pela função suavizada é menor do que a dada pela função não suave, isto é,  $[g_i(\bar{x})]_+ > ([g_i(\bar{x})]_+)^2$ .

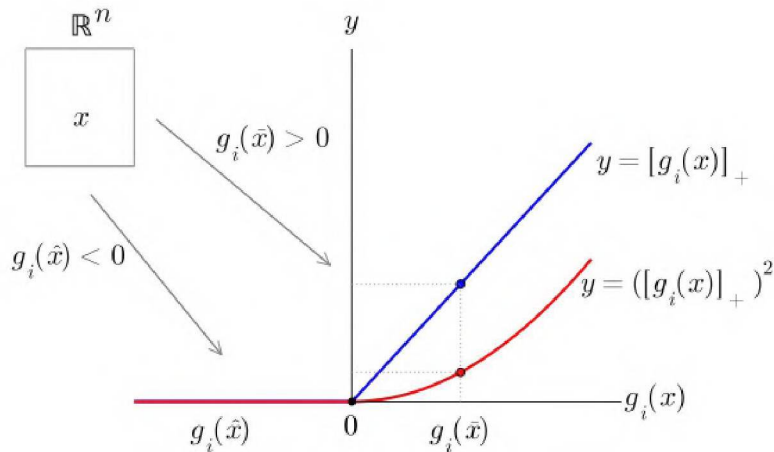


Figura 1.3: Comparação entre o valor de penalidade das funções  $[\cdot]_+$  e  $([\cdot]_+)^2$ .

### Uma família de funções de penalidade

Mais geralmente, considerando  $r > 0$  e  $s > 0$  ambos positivos podemos definir a seguinte função de penalidade

$$P(x) := \frac{1}{r} \left( \sum_{i=1}^m ([g_i(x)]_+)^r \right)^s.$$

Em particular, se  $r = s = 1$  a função de penalidade  $\ell_1$  é obtida, enquanto que considerando  $r = s = 2$  recuperamos a função de penalidade  $\ell_2$ . Destacamos ainda, as funções de penalidade  $\ell_p$  obtidas considerando  $r = p$  e  $s = 1$

$$\ell_p(x) = \frac{1}{p} \sum_{i=1}^m ([g_i(x)]_+)^p.$$

Em particular, se  $p \geq 1$  a função  $\ell_p(x) = \frac{1}{p} \| [g(x)]_+ \|_p^p$ , em que  $\| \cdot \|_p$  é a norma  $p$  em  $\mathbb{R}^m$ .

Outras funções de penalidade podem ser definidas. Por exemplo, [22, Exe. 2] analisa uma classe de funções de penalidade dada por  $x \mapsto x^T \Gamma x$ , com  $\Gamma \in \mathbb{R}^{n \times n}$  uma matriz definida positiva. Uma classe mais geral de funções de penalidade é também considerada em [22, 23] dada por  $P(x) = \psi([g(x)]_+)$  em que  $\psi$  é tal que  $P$  é contínua e satisfaz a condição (1.1) que a caracteriza como uma função de penalidade.

## 1.2.2 Uma estimativa local de inviabilidade

O principal resultado que apresentamos nesta subseção diz respeito a uma estimativa local de inviabilidade conhecida como *condição error bound* [24]. Esta propriedade desempenha um importante papel em derivar relações de dualidade, na análise de sensibilidade/estabilidade, nas estimativas de limite de erro e taxa de convergência de métodos computacionais [36]. Tal resultado será utilizado na próxima seção para estabelecermos uma taxa de convergência dos métodos de penalidade. Antes, algumas definições são importantes. As primeiras delas dizem respeito a condições de qualificação das restrições.

**Definição 1.7.** Dizemos que  $x^* \in X$  satisfaz a condição de qualificação de Mangasarian-Fromovitz (MFCQ), quando existe  $d \in \mathbb{R}^n$  tal que

$$\nabla g_i(x^*)^T d < 0 \quad \forall i \in I(x^*) := \{i = 1, \dots, m \mid g_i(x^*) = 0\}. \quad (1.21)$$

Uma definição equivalente para MFCQ, obtida usando o Teorema das Alternativas, é a seguinte: A única solução  $(x, \mu) \in \mathbb{R}^n \times \mathbb{R}^r$ , com  $r = |I(x^*)|$ , do sistema

$$\sum_{i \in I(x^*)}^m \mu_i \nabla g_i(x^*) = 0 \quad \mu_i \geq 0 \quad \forall i \in I(x^*) \quad (1.22)$$

é a solução nula (veja [14, Proposição 4.1.3]).

É conhecido que se um minimizador local  $x^* \in X$  do problema (1) satisfaz alguma condição de qualificação (CQ), por exemplo MFCQ, então existe um multiplicador  $\mu^* \in \mathbb{R}_+^m$  tal que

$$\nabla f(x^*) + \sum_{i=1}^m \mu_i^* \nabla g_i(x^*) = 0, \quad (1.23)$$

$$g_i(x^*) \mu_i^* = 0, \quad i = 1, \dots, m. \quad (1.24)$$

As condições (1.23) e (1.24) juntas são conhecidas como *condições KKT* (Karush-Kuhn-Tucker), o ponto  $x^*$  é dito um *ponto KKT*,  $(x^*, \mu^*)$  é chamado de *par KKT*, e a condição (1.24) é chamada de *condição de complementaridade*. Para mais detalhes recomendamos [14, 33].

Uma propriedade interessante de MFCQ é que esta CQ é preservada localmente, conforme vemos no próximo resultado.

**Proposição 1.8.** *Suponha que  $x^* \in X$  satisfaz MFCQ. Então existe  $\delta > 0$  tal que MFCQ é satisfeita em qualquer ponto  $x \in B(x^*, \delta) \cap X$ .*

*Prova.* Suponha que  $|I(x^*)| < m$ . Sejam  $I = \{1, \dots, m\} \setminus I(x^*)$  e  $\gamma = \max\{g_i(x^*) \mid i \in I\}$ . Assim,

$$g_i(x^*) \leq \gamma < 0$$

para todo  $i \in I$ . A continuidade de  $g$  e o fato de  $I$  ser finito garantem que existe  $\delta > 0$  tal que  $g_i(x) < 0$  para todos  $x \in B(x^*, \delta) \cap X$  e  $i \in I$ . (Isto pode ser visto notando que para cada  $i \in I$  existe um  $\delta_i > 0$  tal que  $g_i(x) < 0$  para todo  $x \in B(x^*, \delta_i) \cap X$ . Basta então tomar  $\delta = \min\{\delta_i \mid i \in I\}$ ). Portanto,  $I(x) \subset \{1, \dots, m\} \setminus I = I(x^*)$  para todo  $x \in B(x^*, \delta) \cap X$ . Por outro lado, se  $|I(x^*)| = m$  é claro que  $I(x) \subset I(x^*)$ .

Agora, como  $x^*$  satisfaz MFCQ, existe  $d \in \mathbb{R}^n$  tal que  $\nabla g_i(x^*)^T d < 0$  para todo  $i \in I(x^*)$ . Definindo  $\beta = \max\{\nabla g_i(x^*)^T d \mid i \in I(x^*)\}$  e  $\zeta_i : \mathbb{R}^n \rightarrow \mathbb{R}$  por  $\zeta_i(x) = \nabla g_i(x)^T d$ , temos que

$$\zeta_i(x^*) \leq \beta < 0$$

para todo  $i \in I(x^*)$ . Usando o mesmo argumento anterior, reduzindo  $\delta$  se necessário, podemos concluir que

$$\nabla g_i(x)^T d = \zeta_i(x) < 0$$

para todos  $x \in B(x^*, \delta) \cap X$  e  $i \in I(x) \subset I(x^*)$ . Isto significa que qualquer ponto em  $B(x^*, \delta) \cap X$  satisfaz MFCQ. ■

Outra condição de qualificação que será importante no desenvolvimento deste trabalho, é mais restritiva do que a condição de MFCQ, no sentido de que menos problemas satisfazem essa CQ, mas garante a unicidade do multiplicador  $\mu$  tal que  $(x, \mu)$  é um par KKT.

**Definição 1.9.** *Um ponto  $x^* \in X$  satisfaz a condição de independência linear (LICQ) se o conjunto*

$$\{\nabla g_i(x^*) \mid i \in I(x^*)\}$$

*é linearmente independente.*

Para estabelecermos a condição de *error bound* vamos precisar de dois lemas técnicos nos quais consideramos a norma Euclidiana. O primeiro lema é baseado em [21, Cor. 4] e mostra que uma função continuamente diferenciável é *uniformemente diferenciável*.

**Lema 1.10.** *Sejam  $\xi : \mathbb{R}^n \rightarrow \mathbb{R}$  uma função continuamente diferenciável e  $\bar{x} \in \mathbb{R}^n$ . Dado  $\varepsilon > 0$  existe  $\delta > 0$  tal que*

$$|\xi(x) - \xi(y) - \nabla \xi(y)^T (x - y)| \leq \varepsilon \|x - y\|$$

para quaisquer  $x, y \in B(\bar{x}, \delta)$ . Em particular,

$$\nabla \xi(y)^T(x - y) \leq \varepsilon \|x - y\| + \xi(x) - \xi(y). \quad (1.25)$$

*Prova.* Dado  $\varepsilon > 0$  considere  $\delta > 0$  tal que

$$\|\nabla \xi(x) - \nabla \xi(\bar{x})\| \leq \frac{\varepsilon}{2}$$

para todo  $x \in B(\bar{x}, \delta)$ . Assim,

$$\|\nabla \xi(x) - \nabla \xi(y)\| \leq \|\nabla \xi(x) - \nabla \xi(\bar{x})\| + \|\nabla \xi(\bar{x}) - \nabla \xi(y)\| \leq \varepsilon$$

para quaisquer  $x, y \in B(\bar{x}, \delta)$ .

Agora, fixado  $y \in B(\bar{x}, \delta)$ , defina  $h : \mathbb{R}^n \rightarrow \mathbb{R}$  por  $h(x) = \xi(x) - \nabla \xi(y)^T x$ . Assim,

$$\|\nabla h(x)\| = \|\nabla \xi(x) - \nabla \xi(y)\| \leq \varepsilon$$

para todo  $x \in B(\bar{x}, \delta)$ . Portanto, denotando  $d = x - y$ , usando o fato de que

$$y + \theta d = (1 - \theta)y + \theta x \in B(\bar{x}, \delta)$$

e o teorema do valor médio, obtemos

$$\begin{aligned} |\xi(x) - \xi(y) - \nabla \xi(y)^T(x - y)| &= |h(x) - h(y)| \\ &= |\nabla h(y + \theta d)^T d| \\ &\leq \varepsilon \|x - y\|, \end{aligned}$$

sendo que a desigualdade segue da desigualdade de Cauchy-Schwarz. Finalmente, (1.25) é obtido utilizando a desigualdade triangular.  $\blacksquare$

Para o próximo lema, fixado  $\bar{x} \in \mathbb{R}^n$ , considere o problema

$$\begin{aligned} &\text{minimizar} && \|x - \bar{x}\| \\ &\text{sujeito a} && g(x) \leq 0, \end{aligned} \quad (1.26)$$

o qual sempre tem solução, já que o seu conjunto viável  $X$  é fechado e  $\|\cdot\|$  é uma função coerciva. Note que se  $\bar{x} \in X$ , então  $\bar{x}$  é a solução do problema com valor ótimo igual a zero. Por outro lado, se  $\bar{x} \notin X$ , então a função objetivo deste problema é diferenciável em qualquer solução  $\bar{x}_*$  e, se uma condição de qualificação for satisfeita, vale a relação de KKT

$$\frac{\bar{x}_* - \bar{x}}{\|\bar{x}_* - \bar{x}\|} + \sum_{i \in I(\bar{x}_*)} \mu_i \nabla g_i(\bar{x}_*) = 0 \quad (1.27)$$

para algum  $\mu \in \mathbb{R}_+^m$ .

**Lema 1.11.** *Suponha que  $x^* \in X$  satisfaz MFCQ. Então existem  $\delta, M > 0$  tais que dado  $\bar{x} \in B(x^*, \delta) \setminus X$ , qualquer solução  $\bar{x}_*$  do problema correspondente (1.26) cumpre MFCQ e qualquer multiplicador associado  $\mu \in \mathbb{R}_+^m$  satisfaz  $\|\mu\| \leq M$ .*

*Prova.* Por contradição, suponha que o resultado não seja válido. Então existem sequências  $(\bar{x}^k) \subset \mathbb{R}^n \setminus X$  e  $(\bar{x}_*^k) \subset X$  tais que  $\|\bar{x}^k - x^*\| < 1/k$ ,  $\bar{x}_*^k$  é uma solução de (1.26) com  $\bar{x} = \bar{x}^k$  e  $\bar{x}_* = \bar{x}_*^k$  ou deixa de cumprir MFCQ ou cumpre a qualificação mas algum multiplicador associado  $\mu^k \in \mathbb{R}_+^m$  é tal que  $\|\mu^k\| > k$ . Como

$$\|\bar{x}_*^k - x^*\| \leq \|\bar{x}_*^k - \bar{x}^k\| + \|\bar{x}^k - x^*\| \leq \|x^* - \bar{x}^k\| + \|\bar{x}^k - x^*\| < \frac{2}{k},$$

assim a Proposição 1.8 garante que  $\bar{x}_*^k$  satisfaz MFCQ para todo  $k \in \mathbb{N}$  suficientemente grande, logo devemos ter pela hipótese de contradição que  $\|\mu^k\| \rightarrow \infty$ . Sem perda de generalidade, assumimos que

$$\frac{\mu^k}{\|\mu^k\|} \rightarrow \mu,$$

em que  $\mu \in \mathbb{R}_+^m$  é um vetor unitário. Além disto, como  $I(\bar{x}_*^k) \subset I(x^*)$  temos que  $\mu_i^k = 0$  para todo  $i \in I(x^*) - I(\bar{x}_*^k)$ , disto e de (1.27) segue que

$$0 = \frac{\bar{x}_*^k - \bar{x}^k}{\|\bar{x}_*^k - \bar{x}^k\|} + \sum_{i \in I(\bar{x}_*^k)} \mu_i^k \nabla g_i(\bar{x}_*^k) = \frac{\bar{x}_*^k - \bar{x}^k}{\|\bar{x}_*^k - \bar{x}^k\|} + \sum_{i \in I(x^*)} \mu_i^k \nabla g_i(\bar{x}_*^k).$$

Dividindo ambos os lados por  $\|\mu^k\|$  e tomando o limite quando  $k \rightarrow \infty$ , obtemos

$$\sum_{i \in I(x^*)} \mu_i \nabla g_i(x^*) = 0,$$

o que implica em  $\mu = 0$  pela condição de MFCQ (1.22), gerando uma contradição com  $\|\mu\| = 1$ . ■

Estamos agora em condições de provar o resultado mais importante desta seção: sob a condição de Mangasarian-Fromovitz uma estimativa local de inviabilidade pode ser obtida. Esta estimativa é conhecida como *condição de error bound* [24].

**Teorema 1.12.** *Suponha que  $x^* \in X$  satisfaz a condição de MFCQ. Então existem constantes  $M > 0$  e  $\delta > 0$  tais que*

$$\text{dist}(\bar{x}, X) \leq M \max_{i \in I(x^*)} \{[g_i(\bar{x})]_+\} \quad (1.28)$$

para todo  $\bar{x} \in B(x^*, \delta)$ .

*Prova.* Sejam  $\delta_1, M_1 > 0$  cumprindo as condições do Lema 1.11. Aplicando o Lema 1.10

para  $0 < \varepsilon < 1/(M_1 m)$ , segue que existe  $\delta_2 > 0$  tal que

$$\nabla g_i(y)^T(x - y) \leq \varepsilon \|x - y\| + g_i(x) - g_i(y) \quad (1.29)$$

para quaisquer  $x, y \in B(x^*, \delta_2)$  e  $i \in I(x^*)$ .

Defina  $\delta = \min\{\delta_1, \delta_2\}/2$  e tome  $\bar{x} \in B(x^*, \delta)$  arbitrário. Caso  $\bar{x} \in X$ , a relação (1.28) é trivial. Suponha então  $\bar{x} \in B(x^*, \delta) \setminus X$ . Pelo Lema 1.11, qualquer solução  $\bar{x}_*$  do problema correspondente (1.26) cumpre MFCQ e qualquer multiplicador associado  $\mu \in \mathbb{R}_+^m$  satisfaz  $\|\mu\| \leq M_1$ . Além disso,

$$\|\bar{x}_* - x^*\| \leq \|\bar{x}_* - \bar{x}\| + \|\bar{x} - x^*\| \leq \|x^* - \bar{x}\| + \|\bar{x} - x^*\| < 2\delta \leq \delta_2.$$

Por outro lado, como  $I(\bar{x}_*) \subset I(x^*)$ , temos

$$\frac{\bar{x}_* - \bar{x}}{\|\bar{x}_* - \bar{x}\|} + \sum_{i \in I(x^*)} \mu_i \nabla g_i(\bar{x}_*) = \frac{\bar{x}_* - \bar{x}}{\|\bar{x}_* - \bar{x}\|} + \sum_{i \in I(\bar{x}_*)} \mu_i \nabla g_i(\bar{x}_*) = 0.$$

Assim, de (1.29) e o fato de que  $\mu_i g_i(\bar{x}_*) = 0$  para todo  $i \in I(x^*)$ , visto que  $I(\bar{x}_*) \subset I(x^*)$ , obtemos

$$\begin{aligned} \|\bar{x} - \bar{x}_*\| &= (\bar{x} - \bar{x}_*)^T \frac{\bar{x} - \bar{x}_*}{\|\bar{x} - \bar{x}_*\|} \\ &= \sum_{i \in I(x^*)} \mu_i \nabla g_i(\bar{x}_*)^T (\bar{x} - \bar{x}_*) \\ &\leq \sum_{i \in I(x^*)} \mu_i \varepsilon \|\bar{x} - \bar{x}_*\| + \sum_{i \in I(x^*)} \mu_i (g_i(\bar{x}) - g_i(\bar{x}_*)) \\ &= \sum_{i \in I(x^*)} \mu_i \varepsilon \|\bar{x} - \bar{x}_*\| + \sum_{i \in I(x^*)} \mu_i g_i(\bar{x}) \\ &\leq \sum_{i \in I(x^*)} M_1 \varepsilon \|\bar{x} - \bar{x}_*\| + \sum_{i \in I(x^*)} M_1 [g_i(\bar{x})]_+ \\ &\leq m M_1 \varepsilon \|\bar{x} - \bar{x}_*\| + m M_1 \max_{i \in I(x^*)} \{[g_i(x)]_+\}, \end{aligned}$$

sendo que na penúltima desigualdade usamos o fato de que  $\mu_i \leq \|\mu\| \leq M_1$ . Portanto,

$$(1 - m M_1 \varepsilon) \|\bar{x} - \bar{x}_*\| \leq m M_1 \max_{i \in I(x^*)} \{[g_i(x)]_+\},$$

donde segue que

$$\text{dist}(\bar{x}, X) = \|\bar{x}_* - \bar{x}\| \leq M \max_{i \in I(x^*)} \{[g_i(x)]_+\},$$

para  $M = \frac{m M_1}{1 - m M_1 \varepsilon} > 0$ . ■

Do Teorema 1.12 podemos derivar uma estimativa local para a inviabilidade das funções  $\ell_p$  que será de grande utilidade na Seção 1.3 para obtermos uma estimativa de convergência para o método de penalidade  $\ell_p$ .

**Corolário 1.13.** *Sob as hipóteses do Teorema 1.12, existem constantes  $M_p > 0$  e  $\varepsilon_p > 0$  tais que*

$$\text{dist}(x, X) \leq \begin{cases} M_p \ell_p(x)^{1/p} & p > 1, \\ M_p \ell_p(x) & p \leq 1. \end{cases}$$

para todo  $x \in B(x^*, \varepsilon_p)$ .

*Prova.* Pelo Teorema 1.12 existem constantes  $M, \varepsilon > 0$  tais que

$$\text{dist}(x, X) \leq M \max_{i \in I(x^*)} \{[g_i(x)]_+\} \leq M \| [g(x)]_+ \|_\infty.$$

Da relação acima e a da equivalência entre normas em  $\mathbb{R}^m$ , para cada função  $\ell_p$  em que  $p \geq 1$ , existe uma constante  $M_p > 0$  tal que

$$\text{dist}(x, X) \leq M_p \| [g(x)]_+ \|_p = M_p \ell_p(x)^{1/p},$$

para todo  $x \in B(x^*, \varepsilon_p)$  com  $\varepsilon_p = \varepsilon$ .

Para  $p < 1$  observe que a função  $[g_i(\cdot)]_+$  é contínua e  $[g_i(x^*)]_+ = 0$ . Então existe  $\bar{\varepsilon} > 0$  tal que  $\bar{\varepsilon} \leq \varepsilon$  e se  $x \in B(x^*, \bar{\varepsilon})$  então  $[g_i(x)]_+ \leq 1$  para todo  $i = 1, \dots, m$ . Consequentemente,

$$\text{dist}(x, X) \leq M \sum_{i=1}^m [g_i(x)]_+ \leq \frac{M}{p} \sum_{i=1}^m ([g_i(x)]_+)^p = M \ell_p(x),$$

para todo  $x \in B(x^*, \varepsilon_p)$ , tomando  $\varepsilon_p = \varepsilon$  e  $M_p = M$  segue o resultado desejado.  $\blacksquare$

### 1.3 Taxa de convergência

Nesta seção discutimos uma estimativa local da taxa de convergência dos métodos de penalidade dada em [15]. Esta taxa de convergência é estabelecida em relação a valores da função objetivo, para os quais uma estimativa local da violação de viabilidade é conhecida, isto é:

**Hipótese 3.** *Existem uma vizinhança  $U$  de uma solução dada do problema (1) e constantes  $\nu > 0$  e  $M > 0$  tais que*

$$\text{dist}(x, X) \leq M (P(x))^\nu \tag{1.30}$$

para todo  $x \in U$ .

Observe que a relação (1.30) garante que a distância entre um ponto dado  $x \in U$  e o

conjunto viável  $X$  é limitada pela medida da violação imposta pela função de penalidade em  $x$ .

No Teorema 1.5 mostramos que qualquer ponto de acumulação de uma sequência gerada pelo Algoritmo 1.1, converge para um minimizador global do problema com restrições. Entretanto, do ponto de vista computacional, mais geralmente, é esperado que o algoritmo gere soluções locais aproximadas. Tendo isto em vista, o teorema a seguir estabelece uma estimativa de taxa de convergência do método de penalidade para o caso em que os minimizadores convergem para uma solução local de (1). Em especial, esse resultado mostra algumas situações em que a convergência é finita.

**Teorema 1.14.** [15, Teo. 4.3.12]. *Seja  $f$  Lipschitz contínua em uma vizinhança  $U$  de  $x^*$  com constante Lipschitz  $L > 0$ . Suponha que  $x^*$  é uma solução local do problema (1) e que a função de penalidade  $P$  satisfaz a Hipótese 3 para algum  $\nu > 0$  e  $M > 0$ . Se as sequências  $(c_k) \subset \mathbb{R}_+$  e  $(x^k) \subset \mathbb{R}^n$  satisfazem para todo  $k \in \mathbb{N}$ ,*

$$\varphi_{c_k}(x^k) \leq \varphi_{c_k}(x^*), \quad c_k \rightarrow +\infty, \quad e \quad x^k \rightarrow x^*, \quad (1.31)$$

então, para todo  $k$  suficientemente grande as seguintes afirmações são válidas:

(i) Se  $\nu < 1$ , então

$$0 \leq f(x^*) - f(x^k) \leq \left( \frac{(LM)^{1/\nu}}{c_k} \right)^{\nu/(1-\nu)}, \quad (1.32)$$

e

$$\text{dist}(x^k, X) \leq M \left( \frac{LM}{c_k} \right)^{\nu/(1-\nu)}. \quad (1.33)$$

(ii) Se  $\nu \geq 1$ , então

$$f(x^k) = f(x^*) \quad e \quad x^k \in X.$$

Em particular, se  $x^*$  é um minimizador local estrito para (1), então  $x^k = x^*$ .

*Prova.* Seja  $k \in \mathbb{N}$ . Da hipótese e do fato que  $x^* \in X$ , segue que

$$\varphi_{c_k}(x^k) \leq \varphi_{c_k}(x^*) = f(x^*). \quad (1.34)$$

Vamos mostrar a afirmação (i). Note que da relação (1.34) e da Hipótese 3 temos que

$$f(x^*) \geq \varphi_{c_k}(x^k) = f(x^k) + c_k P(x^k) \geq f(x^k) + c_k \left( \frac{\text{dist}(x^k, X)}{M} \right)^{1/\nu} \quad (1.35)$$

para todo  $x^k \in U$ .

Como  $X$  é um conjunto fechado, existe a projeção de  $x^k$  sobre  $X$  (do fato que  $X$  é fechado garantimos a existência da projeção, mas não podemos garantir a unicidade sem a hipótese de  $X$  ser convexo). Assim, associe para cada  $x^k$  um vetor de projeção  $z^k \in X$ .

Por hipótese a sequência  $(x^k)$  converge para  $x^*$ , conseqüentemente  $(z^k)$  também converge para  $x^*$ , então para  $k$  suficientemente grande temos que  $z^k \in U \cap X$ , donde com a hipótese de que  $f$  é Lipschitz contínua em  $U$  e o ponto  $x^*$  é um minimizador local obtemos

$$f(x^*) - f(x^k) \leq f(z^k) - f(x^k) \leq L\|z^k - x^k\|_2 = L \text{ dist}(x^k, X). \quad (1.36)$$

Logo, juntando (1.35) e (1.36) obtemos

$$f(x^*) - f(x^k) \geq c_k \left( \frac{\text{dist}(x^k, X)}{M} \right)^{1/\nu} \quad (1.37)$$

$$\geq c_k \left( \frac{f(x^*) - f(x^k)}{LM} \right)^{1/\nu}. \quad (1.38)$$

Considere agora o conjunto de índices  $J := \{k \in \mathbb{N} \mid f(x^k) \neq f(x^*)\}$ . De (1.35) e da definição de  $J$  segue que

$$f(x^k) < f(x^*),$$

para todo  $k \in J$ . Reescrevendo (1.38) temos

$$\frac{(LM)^{1/\nu}}{c_k} \geq (f(x^*) - f(x^k))^{(1-\nu)/\nu} > 0, \quad (1.39)$$

para todo  $k \in J$  suficientemente grande. Por outro lado, se  $k \notin J$  temos que  $f(x^k) - f(x^*) = 0$ . Disto, da relação (1.39) e o fato de  $\nu < 1$  segue a afirmação (1.32) para todo  $k$  suficientemente grande.

Para mostrarmos a afirmação (1.33), note que por (1.36) e (1.37), temos

$$\text{dist}(x^k, X) L \geq f(x^*) - f(x^k) \geq c_k \left( \frac{\text{dist}(x^k, X)}{M} \right)^{1/\nu},$$

portanto para  $\nu < 1$  segue que

$$M \left( \frac{LM}{c_k} \right)^{\nu/(1-\nu)} = \left( \frac{LM^{1/\nu}}{c_k} \right)^{\nu/(1-\nu)} \geq \text{dist}(x^k, X).$$

Finalmente, se  $\nu \geq 1$  então  $(1-\nu)/\nu \leq 0$  e como (1.39) também é mantida para  $\nu \geq 1$ , temos que  $f(x^*) - f(x^k)$  não converge para zero. Conseqüentemente, o lado esquerdo de (1.39) também não converge para zero. Mas note que se  $J$  fosse um conjunto infinito a sequência  $(1/c_k)$  convergiria para zero devido a (1.31), logo teríamos uma contradição

com (1.39). Portanto  $J$  é finito. Desta maneira, para  $k$  suficientemente grande,

$$f(x^k) = f(x^*) \quad \text{e} \quad x^k \in X.$$

Além disso, se  $x^*$  é uma solução local estrita de (1), então a última relação fornece que  $x^k = x^*$ . ■

Em particular o item (ii) do teorema acima implica na existência de um parâmetro de penalidade em que uma solução do problema de penalidade é também uma solução local exata do problema com restrições, como estabelece o resultado a seguir.

**Corolário 1.15.** [15, Cor. 4.3.13]. *Sob as hipóteses do Teorema 1.35, suponha que  $x^*$  é uma solução local estrita de (1) e  $\nu \geq 1$ . Então existe  $c^* > 0$  tal que para todo  $c > c^*$  o ponto  $x^*$  é um minimizador local estrito do problema de penalidade  $(P_c)$ .*

*Prova.* Suponha por contradição que para cada  $k \in \mathbb{N}$ , existam  $c_k > k$  e  $x^k \neq x^*$  de modo que

$$x^k \rightarrow x^* \quad \text{e} \quad \varphi_{c_k}(x^k) \leq \varphi_{c_k}(x^*) \tag{1.40}$$

visto que  $x^*$  não é minimizador local estrito de  $\varphi_{c_k}(\cdot)$ .

Como  $c_k \rightarrow \infty$  e vale (1.40), o item (ii) do Teorema 1.14 implica que  $x^k = x^*$  para  $k$  suficientemente grande, o que gera uma contradição, e donde segue o resultado. ■

Dada a convergência global do Algoritmo 1.1, qualquer sequência gerada por esse algoritmo que possui um ponto de acumulação, satisfaz a hipótese (1.31) do teorema acima. Além disto, vemos que a diminuição de valores de  $\nu < 1$  leva a uma taxa de convergência mais lenta do método de penalidade. Enquanto que, para valores de  $\nu \geq 1$  a convergência é finita.

**Exemplo 1.3.** *Vimos no Corolário 1.13 que sob a condição de MFCQ, as funções de penalidade  $\ell_p$  satisfazem uma estimativa local da violação de viabilidade como a dada em (1.30). Logo, as seguintes estimativas de taxa de convergência dos métodos de penalidade  $\ell_p$  são válidas de acordo com o Teorema 1.13:*

(i) *Para  $p > 1$  temos  $\nu = 1/p < 1$ , neste caso*

$$0 \leq f(x^*) - f(x^k) \leq \left( \frac{(LM_p)^p}{c_k} \right)^{1/(p-1)} \quad \text{e} \quad \text{dist}(x^k, X) \leq M_p \left( \frac{LM_p}{c_k} \right)^{1/(p-1)}.$$

*Note que quanto mais próximo  $p$  está de 1 melhor é a taxa de convergência para o correspondente método de penalidade  $\ell_p$ , mas com a diminuição de  $p$  a diferenciabilidade da função  $\ell_p$  torna-se pior.*

(ii) *Já para  $p \leq 1$  temos  $\nu = 1$ . Portanto, a convergência é finita para qualquer método de penalidade  $\ell_p$  em que  $p \leq 1$ . Em particular, se  $x^*$  é um minimizador local*

estrito do problema (1), então  $x^*$  é um minimizador local estrito de  $\varphi_c(\cdot)$  para  $c > c^* \in \mathbb{R}_+$  suficientemente grande, porém finito, pelo Corolário 1.15. Uma função de penalidade com essa propriedade é dita exata [1, 15, 22].

Note que para a função de penalidade ser chamada de exata, condições sobre o conjunto viável  $X$ , como MFCQ, e sobre a função objetivo de (1), como a condição de Lipschitz contínua, são requeridas.

**Definição 1.16.** [12, Def. 3.1] Uma função de penalidade  $P$  é dita exata para o problema (1) se existe um  $c^* > 0$  tal que para todo  $c > c^*$ , um minimizador global (local) do problema de penalidade  $(P_c)$  é também um minimizador global (local) de (1).

Fazemos agora uma modesta comparação entre o método de penalidade  $\ell_2$  e o método de penalidade exata  $\ell_1$ . Para isso, consideramos novamente o problema de otimização dado no Exemplo 1.1 em que foi utilizada a função de penalidade quadrática.

**Exemplo 1.4.** Considere o problema de otimização em  $\mathbb{R}$

$$\begin{aligned} \text{minimizar } f(x) &= x^2 \\ \text{sujeito a } g(x) &= x \leq -1. \end{aligned} \tag{1.41}$$

Lembremos que  $x^* = -1$  é o único minimizador de (1.41) e este ponto satisfaz a condição LICQ, visto que  $\nabla g(x^*) = 1$ , portanto satisfaz MFCQ. Em particular  $x^*$  é um ponto KKT com um único multiplicador de Lagrange associado  $\mu^* = 2$ .

Considere a função de penalidade exata  $\ell_1$ . Vamos mostrar que se  $c > c^* = 2$  então o ponto  $x^*$  é minimizador global do problema

$$\begin{aligned} \text{minimizar } \varphi_c(x) &:= x^2 + c[x + 1]_+ \\ \text{sujeito a } x &\in \mathbb{R}. \end{aligned} \tag{1.42}$$

Reescrevendo a função penalizada  $\varphi_c$ , temos

$$\varphi_c(x) = \begin{cases} x^2 + c(x + 1) & \text{se } x \geq -1, \\ x^2 & \text{caso contrário,} \end{cases}$$

logo  $x(c) = -c/2$  para  $c \leq 2$  e  $x(c) = -1$  para  $c > 2$ .

Observe que  $|\mu^*|$  é um limitante inferior para a penalização exata, no sentido que para  $c \geq |\mu^*|$  o problema com restrições (1.42) é transformado de forma equivalente em um problema irrestrito, ou seja, um minimizador global de (1.42) é também um minimizador global de (1.41), valendo também a recíproca. Tal equivalência entre os problemas restrito e irrestritos de penalidade pode ser observada na Figura 1.4, que entre outras coisas,

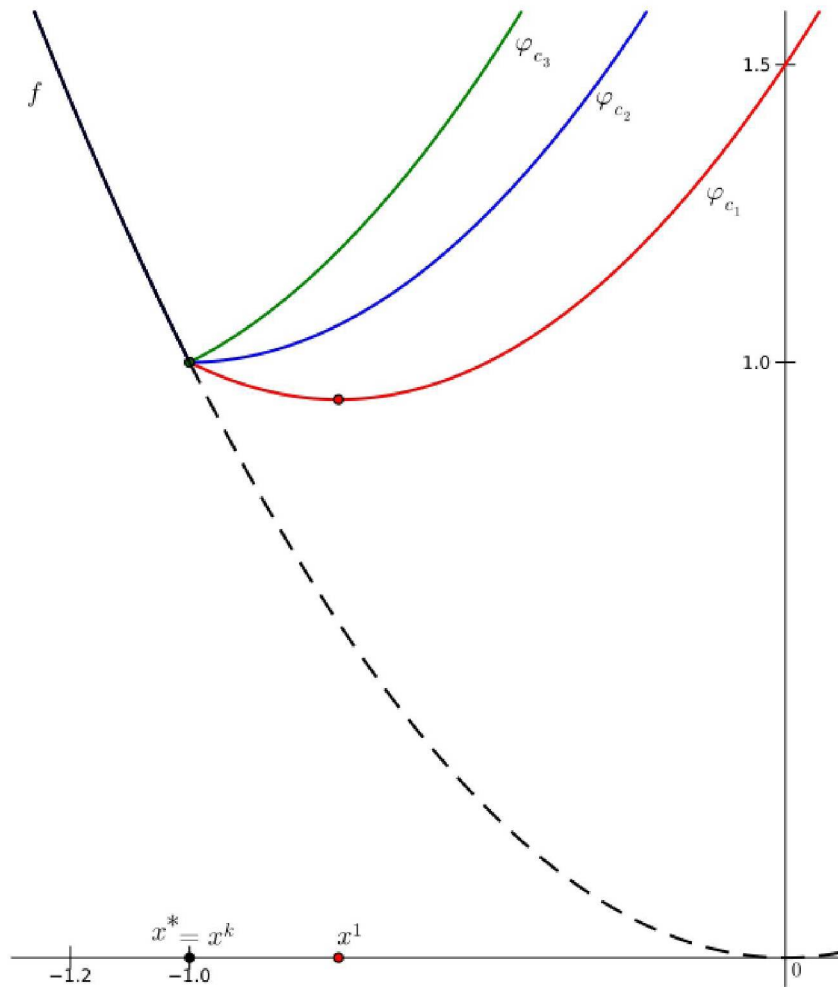


Figura 1.4: Comportamento das funções penalizadas para o problema (1.41) em que os parâmetros de penalidade são  $c_1 = 1.5$ ,  $c_2 = 2$  e  $c_3 = 2.5$ , e  $k \geq 2$ . A curva tracejada representa o gráfico de  $f$  em pontos inviáveis do problema original.

*mostra que para  $c_2, c_3 \geq 2$  os minimizadores das funções penalizadas são viáveis e portanto coincidem com o minimizador do problema restrito.*

Quando considerada a função de penalidade  $\ell_2$  obtivemos que para todo  $c > 0$  as soluções dos problemas penalizados eram inviáveis para o problema original. Ao passo que para a função de penalidade  $\ell_1$ , foi suficiente que  $c \geq 2$  para que uma solução do problema original fosse obtida minimizando a função penalizada  $\ell_1$ .

**Observação 1.17.** *Ao considerarmos um método de função de penalidade exata, o problema (1) pode ser transformado em um único problema irrestrito equivalente, sob condições razoáveis (veja a Seção 2.1) para um valor de  $c$  suficientemente grande, porém finito. Em outras palavras, não é necessário que  $c$  cresça ilimitadamente para que uma solução do problema com restrições seja obtida pelo método de penalidade.*

Por outro lado, desde que a função de penalidade seja diferenciável e o minimizador de

(1) não anule o gradiente da função objetivo  $f$  (o que não ocorre em geral), uma solução do problema de penalidade não pode resultar em uma solução do problema original, seja esta solução local ou global [34], como mostramos a seguir.

**Proposição 1.18.** *Seja  $P$  uma função de penalidade diferenciável. Suponha que  $x^*$  é uma solução local do problema (1) tal que  $\nabla f(x^*) \neq 0$ . Então, para todo  $c > 0$  o ponto  $x^*$  não é um minimizador local do problema de penalidade  $(P_c)$ .*

*Prova.* Suponha que a conclusão dessa proposição seja falsa. Então existe um  $c^* > 0$  tal que o ponto  $x^*$  é um minimizador local de  $\varphi_{c^*}(\cdot)$ . Assim, pela condição necessária de otimalidade de primeira ordem

$$0 = \nabla \varphi_{c^*}(x^*) = \nabla f(x^*) + c^* \nabla P(x^*). \quad (1.43)$$

Por outro lado, como  $x^* \in X$  segue que  $P(x^*) = 0 \leq P(x)$  para todo  $x \in \mathbb{R}^n$ . Consequentemente  $x^*$  também é um minimizador da função  $P$ , e assim  $\nabla P(x^*) = 0$ . Disto e de (1.43) segue que  $0 = \nabla f(x^*)$ , o que é uma contradição. Portanto,  $x^*$  não é uma solução local do problema de penalidade. ■

A última proposição mostra que uma função de penalidade diferenciável não alcança um minimizador de (1), dado que um minimizador de um problema com restrições não é um ponto estacionário da função objetivo desse problema em geral. Consequentemente o parâmetro de penalidade deve crescer ilimitadamente para que uma solução de (1) possa ser obtida. Em oposição, métodos de penalidade exatos podem recuperar uma solução de (1) com um número finito de iterações, ao custo da não diferenciabilidade dos subproblemas de penalidade como vemos por contra-positiva da Proposição 1.18. Diante disso, no próximo capítulo estudaremos os métodos de penalidade funções exatas e suaves separadamente .

# Capítulo 2

## Penalização exata e suave

Até o momento estudamos os métodos de penalidade externa sem especificar necessariamente a função de penalidade. Na Seção 1.3 mostramos que existem métodos de penalidade que possuem convergência finita sob hipóteses razoáveis. Em particular, vimos no Exemplo 1.3 que as funções de penalidade  $\ell_p$  com  $p \leq 1$  recuperam uma solução exata do problema original resolvendo um número finito de problemas de penalidade, essas funções são exatas, porém não são suaves. Ainda mais, a Proposição 1.18 mostra que nenhuma função de penalidade diferenciável recupera um minimizador do problema (1) se este não for um ponto estacionário da função objetivo desse problema. Logo, é necessário que  $P$  seja não diferenciável para que seja exata.

Diante disso, neste capítulo estudamos os métodos de penalidade exatos e suaves. Primeiro vamos definir uma classe de funções de penalidade que sob certas condições são exatas. Discutiremos algumas condições suficientes para penalização exata e uma condição necessária. Já na Seção 2.2, vamos mostrar que um multiplicador de Lagrange associado a uma solução do problema com restrições pode ser obtido através do método de penalidade suave. Ademais, ilustramos como o mal condicionamento do problema de penalidade aumenta à medida que o parâmetro de penalidade cresce.

### 2.1 Método de penalidade exata

O caso mais “satisfatório” de penalização exata transforma um problema com *restrições* em um problema equivalente *sem restrições* para um parâmetro de penalidade suficientemente grande. Mais especificamente, existe um limitante inferior  $c^* > 0$  tal que

$$x^* \in \operatorname{argmin}\{f(x) \mid x \in X\} \iff x^* \in \operatorname{argmin}\{\varphi_c(x) \mid x \in \mathbb{R}^n\} \cap X,$$

para todo  $c > c^*$ .

Alguns dos resultados que veremos nesta seção mostram que o limitante inferior  $c^*$ , que transforma o problema original em um problema de penalidade equivalente, é conhecido

(teoricamente). As principais referências utilizadas nesta seção são [15, 37].

### 2.1.1 O problema de penalidade exata

Trazemos nesta subseção definições e resultados necessários para discutirmos condições necessárias e suficientes para penalização exata.

**Definição 2.1.** [37] *Seja  $\sigma : \mathbb{R}^m \rightarrow \mathbb{R}$  uma função convexa tal que  $\sigma(0) = 0$ . Dizemos que  $\sigma$  é uma função de distância forte, se existe uma constante positiva  $\delta > 0$  tal que*

$$\sigma(y) \geq \delta \|y\|_1 \quad (2.1)$$

para todo  $y \in \mathbb{R}^m$ .

Dada uma função de distância forte, definimos a função de penalidade

$$P(x) = \sigma([g(x)]_+), \quad (2.2)$$

em que  $[g(x)]_+$  denota o vetor em  $\mathbb{R}_+^m$  cuja  $i$ -ésima entrada é  $[g_i(x)]_+$ . Note que  $P$  é de fato uma função de penalidade para o problema (1), pois da convexidade de  $\sigma$  em  $\mathbb{R}^m$  segue sua continuidade e, além disso, se  $x \in X$  então  $[g(x)]_+ = 0$ , conseqüentemente

$$P(x) = \sigma([g(x)]_+) = \sigma(0) = 0.$$

Por outro lado, se  $x \notin X$  então da condição (2.1), obtemos

$$P(x) = \sigma([g(x)]_+) \geq \delta \|[g(x)]_+\|_1 > 0.$$

A classe de funções de penalidade definida em (2.2) é chamada de *exata* em [37]. Veremos na Subseção 2.1.2 que estas funções recuperaram um minimizador do problema (1) se esse problema satisfaz certas condições, ou seja, mostraremos essas funções de fato são exatas conforme a Definição 1.16.

**Exemplo 2.1.** *A função de penalidade  $\ell_2$  não satisfaz a condição (2.1). Para vermos isto, observe que a função*

$$\mathbb{R}^m \ni y \mapsto \psi(y) = \frac{1}{2} \|y\|_2^2 = \frac{1}{2} \sum_{i=1}^m y_i^2$$

é tal que  $\psi(0) = 0$  e  $\ell_2(x) = \psi([g(x)]_+)$ . Entretanto, para qualquer  $\delta > 0$  dado, temos que o vetor  $y_\delta = (\delta, 0, \dots, 0) \in \mathbb{R}^m$  satisfaz

$$\psi(y_\delta) = \frac{1}{2} \delta^2 < \delta^2 = \delta \|y_\delta\|_1.$$

Por outro lado,  $\|\cdot\|_1$  satisfaz (2.1) com  $\delta = 1$  e, portanto, a função

$$\ell_1(x) = \|[g(x)]_+\|_1$$

pertence à classe de funções de penalidade exatas (2.2).

Dada uma função de penalidade definida em (2.2), associamos o problema de penalidade

$$(P_1) \quad \begin{aligned} &\text{minimizar } \varphi_c(x) := f(x) + cP(x) \\ &\text{sujeito a } x \in \mathbb{R}^n, \end{aligned}$$

em que  $P(x) = \sigma([g(x)]_+)$  para alguma função de distância forte sigma.

Uma vez que a função de penalidade é exata, a Proposição 1.18 garante que essa função não é suave, e conseqüentemente a função objetivo do problema de penalidade também não o é. Em [34, Teo. 6.9] é mostrado que para  $P = \ell_1$  vale que

$$\lim_{t \rightarrow 0^+} \frac{\ell_1([g(z+td)]_+) - \ell_1([g(z)]_+)}{t} = \sum_{i \in I(z)} [\nabla g_i(z)^T d]_+ = \|[\nabla g_i(z)^T d]_+\|_+,$$

para todo  $z \in X$ . Mais geralmente podemos estabelecer seguinte resultado.

**Lema 2.2.** *Seja  $\sigma$  uma função de distância forte. Dados  $z \in X$  e  $d \in \mathbb{R}^n$  considere a função  $\Phi : \mathbb{R}_{++} \rightarrow \mathbb{R}$  definida como*

$$\Phi(t) := \frac{\sigma([g(z+td)]_+) - \sigma([g(z)]_+)}{t},$$

e o vetor  $\nabla g_I(z)^T d \in \mathbb{R}^m$  em que

$$(\nabla g_I(z)^T d)_i := \begin{cases} \nabla g_i(z)^T d & \text{se } i \in I(z), \\ 0 & \text{se } i \notin I(z). \end{cases}$$

Então, para qualquer sequência  $(t_k) \subset \mathbb{R}_{++}$  tal que  $t_k \rightarrow 0$ , a sequência  $(\Phi(t_k))$  é limitada. Em particular, se  $\mathcal{K} \subset \mathbb{N}$  é um conjunto infinito tal que  $(\Phi(t_k))_{\mathcal{K}}$  é convergente, então

$$0 \leq \lim_{k \rightarrow \infty, k \in \mathcal{K}} \Phi(t_k) \leq \sigma([\nabla g_I(z)^T d]_+). \quad (2.3)$$

*Prova.* Seja  $z \in X$  e  $(t_k) \subset \mathbb{R}_+$  uma sequência satisfazendo as hipóteses do lema. Suponha que  $i \notin I(z)$ . Pela continuidade de  $g_i$  existe  $\delta_d > 0$  tal que  $g_i(z+td) \leq 0$  se  $|t| < \delta_d$ , em particular, existe um  $k_0(d)$  tal que se  $k \geq k_0(d)$  então

$$g_i(z+t_k d) \leq 0,$$

o que implica

$$[g_i(z + t_k d)]_+ = 0. \quad (2.4)$$

Por outro lado, se  $i \in I(z)$  então  $g_i(z) = 0$ , e como  $z \in X$  e as funções  $g_i$  são diferenciáveis, existem  $\varepsilon > 0$  e uma função  $o_i : B(0, \varepsilon) \rightarrow \mathbb{R}$  definida por

$$o_i(t_k) = g_i(z + t_k d) - t_k \nabla g_i(z)^T d$$

tal que  $o_i(t_k)/t_k \rightarrow 0$  quando  $t_k \rightarrow 0$ . Portanto

$$[g_i(z + t_k d)]_+ = [t_k \nabla g_i(z)^T d + o_i(t_k)]_+. \quad (2.5)$$

Juntando (2.4) e (2.5), obtemos

$$[g(z + t_k d)]_+ = [t_k \nabla g_I(z)^T d + o_I(t_k)]_+ \quad (2.6)$$

em que  $o_I(t_k) \in \mathbb{R}^m$  é definido como

$$(o_I(t_k))_i := \begin{cases} o_i(t_k) & \text{se } i \in I(z), \\ 0 & \text{se } i \notin I(z). \end{cases}$$

Agora, utilizando a viabilidade de  $z \in X$ , (2.6) e a convexidade de  $\sigma$ , obtemos

$$\begin{aligned} \sigma([g(z + t_k d)]_+) - \sigma([g(z)]_+) &= \sigma([g(z + t_k d)]_+) \\ &= \sigma([t_k \nabla g_I(z)^T d + o_I(t_k)]_+) \\ &\leq t_k \sigma \left( \left[ \nabla g_I(z)^T d + \frac{o_I(t_k)}{t_k} \right]_+ \right) + (1 - t_k) \sigma(0), \end{aligned}$$

para todo  $k$  suficientemente grande. Dividindo ambos os lados da desigualdade acima por  $t_k$  e usando que  $\sigma(0) = 0$ , segue que

$$\Phi(t_k) = \frac{\sigma([g(z + t_k d)]_+) - \sigma([g(z)]_+)}{t_k} \leq \sigma \left( \left[ \nabla g_I(z)^T d + \frac{o_I(t_k)}{t_k} \right]_+ \right). \quad (2.7)$$

Finalmente, como a função  $\sigma([\cdot]_+)$  é contínua obtemos

$$\lim_{k \rightarrow \infty} \sigma \left( \left[ \nabla g_I(z)^T d + \frac{o_I(t_k)}{t_k} \right]_+ \right) = \sigma([\nabla g_I(z)^T d]_+),$$

donde segue que  $(\Phi(t_k))$  é limitada superiormente. A limitação inferior decorre da viabi-

lidade de  $z$  já que

$$0 \leq \frac{\sigma([g(z + t_k d)]_+)}{t_k} = \Phi(t_k).$$

Portanto  $(\Phi(t_k))$  é limitada.

Seja  $\mathcal{K} \subset \mathbb{N}$  um conjunto infinito tal que  $(\Phi(t_k))$  é convergente. Assim, passando o limite em (2.7) sob a subsequência convergente, obtemos

$$0 \leq \lim_{k \rightarrow \infty, k \in \mathcal{K}} \Phi(t_k) \leq \sigma([\nabla g_I(z)^T d]_+),$$

o que completa a demonstração. ■

O lema anterior será importante para derivarmos uma condição necessária para penalização exata, como veremos na próxima subseção.

### 2.1.2 Condições necessárias e condições suficientes

De acordo com a Definição 1.16, uma função de penalidade é exata para o problema com restrições, se uma solução do problema de penalidade é também uma solução do problema (1). Para isto, o minimizador da função penalizada deve ser viável para o problema original. Assim, para mostrarmos que uma função é exata para o problema com restrições, verificamos que um minimizador global (local) desse problema é também um minimizador global (local) de  $(P_1)$  para algum  $c^* > 0$ .

O primeiro resultado que apresentamos nesta seção, revela uma importante condição para a existência de penalização exata. Tal resultado é apresentado em [15, Teo. 4.3.15] para a função de penalidade exata  $\ell_1$ . Aqui, estendemos parte deste resultado para a classe de funções de penalidade definida em (2.2) seguindo a mesma linha de demonstração utilizada em [15].

**Teorema 2.3.** *Seja  $P$  uma função de penalidade exata definida em (2.2). Suponha que existam  $x^* \in X$  e  $c^* > 0$  tais que  $x^*$  é um minimizador local de  $\varphi_{c^*}$  em  $\mathbb{R}^n$ . Então  $x^*$  é uma solução local de (1) e satisfaz as condições KKT. Em particular, se  $\sigma(\cdot) = \|\cdot\|_1$  e  $x^*$  satisfaz LICQ, então  $c^* \geq \|\mu^*\|_\infty$ , com  $\mu^*$  o multiplicador de Lagrange correspondente a  $x^*$ .*

*Prova.* Seja  $d \in \mathbb{R}^n$ . Como  $x^*$  é minimizador local de  $\varphi_{c^*}$  e viável para (1), existe  $\varepsilon > 0$  tal que, para todo  $x \in B(x^*, \varepsilon) \cap X$ ,

$$f(x^*) = \varphi_{c^*}(x^*) \leq \varphi_{c^*}(x) = f(x),$$

isto é,  $x^*$  é um minimizador local de (1). Em particular, existe uma sequência  $(t_k) \subset \mathbb{R}_+$  tal que  $t_k \rightarrow 0$  e

$$0 \leq \varphi_{c^*}(x^* + t_k d) - \varphi_{c^*}(x^*)$$

para todo  $k \in \mathbb{N}$ , donde segue que

$$\begin{aligned} 0 &\leq \frac{f(x^* + t_k d) - f(x^*)}{t_k} + c^* \left[ \frac{\sigma([g(x^* + t_k d)]_+) - \sigma([g(x^*)]_+)}{t_k} \right] \\ &= \frac{f(x^* + t_k d) - f(x^*)}{t_k} + c^* [\Phi(t_k)]. \end{aligned} \quad (2.8)$$

Do Lema 2.2, a sequência  $(\Phi(t_k))$  é limitada, assim, passando o limite em (2.8) utilizando uma subsequência se necessário, obtemos do limitante superior dado em (2.3) e o fato que  $f$  é diferenciável,

$$0 \leq \nabla f(x^*)^T d + c^* \sigma([\nabla g_I(x^*)^T d]_+). \quad (2.9)$$

Em particular, se  $d \in L_X(x^*) = \{d \in \mathbb{R}^m \mid \nabla g_i(x^*)^T d \leq 0, i \in I(x^*)\}$ , temos

$$\nabla f(x^*)^T d \geq 0.$$

Portanto, o vetor  $d^* = 0 \in \mathbb{R}^n$  é uma solução do problema de programação linear

$$\begin{aligned} &\text{minimizar } \nabla f(x^*)^T d \\ &\text{sujeito a } \nabla g_i(x^*)^T d \leq 0, \quad i \in I(x^*). \end{aligned} \quad (2.10)$$

Visto que as restrições do problema (2.10) são lineares, as condições de KKT são satisfeitas em  $d^*$ . Logo, existe  $\bar{\mu} \in \mathbb{R}_+^r$ , com  $r = |I(x^*)|$ , tal que

$$\nabla f(x^*) + \sum_{i \in I(x^*)} \bar{\mu}_i \nabla g_i(x^*) = 0, \quad (2.11)$$

pois  $I(x^*) = I(d^*)$ . Definindo  $\mu^* \in \mathbb{R}_+^m$  como  $\mu_i^* = 0$  se  $i \notin I(x^*)$  e  $\mu_i^* = \bar{\mu}_i$  se  $i \in I(x^*)$ , segue que

$$\mu_i^* g_i(x^*) = 0, \quad i = 1, \dots, m. \quad (2.12)$$

Portanto (2.11) e (2.12) implicam que  $(x^*, \mu^*)$  é um par KKT para (1).

Suponha agora que  $\sigma(\cdot) = \|\cdot\|_1$  e que  $x^*$  satisfaz LICQ e que  $P = \ell_1$ . Então

$$\begin{aligned} \sum_{i \in I(x^*)} \mu_i^* \nabla g_i(x^*)^T d &= -\nabla f(x^*)^T d \\ &\leq c^* \sigma([\nabla g_I(x^*)^T d]_+) \\ &= c^* \sum_{i \in I(x^*)} [\nabla g_i(x^*)^T d]_+, \end{aligned} \quad (2.13)$$

onde na primeira igualdade utilizamos (2.11), enquanto que a desigualdade é devida a

relação (2.9). Assim, de LICQ temos que para todo  $v_I = (v_1, \dots, v_r) \in \mathbb{R}^r$ , com  $r = |I(x^*)|$  o sistema linear

$$\nabla g_i(x^*)^T d = v_i, \quad i \in I(x^*)$$

tem uma solução  $d \in \mathbb{R}^n$ . Disto e de (2.13), segue que

$$\begin{aligned} v_I^T \mu^* &= \sum_{i \in I(x^*)} \mu_i^* v_i = \sum_{i \in I(x^*)} \mu_i^* \nabla g_i(x^*)^T d \\ &\leq c^* \sum_{i \in I(x^*)} [\nabla g_i(x^*)^T d]_+ = c^* \sum_{i \in I(x^*)} [v_i]_+ \\ &\leq c^* \sum_{i \in I(x^*)} |v_i| = c^* \|v_I\|_1. \end{aligned}$$

Da relação acima e do fato que  $\|\mu^*\|_\infty = \|\bar{\mu}\|_\infty$ , concluímos

$$\|\mu^*\|_\infty = \sup\{\bar{\mu}^T v_I \mid \|v_I\|_1 = 1\} = \sup\left\{\bar{\mu}^T \frac{v_I}{\|v_I\|_1} \mid v_I \in \mathbb{R}^r\right\} \leq c^*,$$

onde na primeira igualdade utilizamos a definição de norma dual. ■

Devemos observar que sem a exigência de uma condição de qualificação das restrições, um minimizador de um problema de otimização com restrições pode não ser um ponto KKT, veja por exemplo [33, Exem. 7.32]. Entretanto, o resultado anterior mostra que as condições de KKT são satisfeitas no caso de penalização exata sem que uma CQ sob o conjunto  $X$  seja exigida.

Vamos obter agora condições para que um minimizador do problema original seja também um minimizador do problema de penalidade. O primeiro teorema que apresentamos exige que um minimizador local do problema com restrições satisfaz condições suficientes de otimalidade de segunda ordem. Por completude, apresentamos a demonstração dada em [37]. Outra hipótese de segunda ordem utilizada para estabelecer penalização exata é dada por  $\nabla_{xx}^2 L(x^*, \mu^*)$  ser definida positiva no cone de direções críticas,  $L_X(x^*) \cap \{\nabla f(x^*)\}^\perp$ , a qual pode ser encontrada em [15, Teo. 4.3.18].

**Teorema 2.4.** [37, Teo. 10.6.1] *Seja  $P$  definida em (2.2). Suponha que  $x^*$  é um minimizador local do problema (1) tal que o par KKT  $(x^*, \mu^*)$  satisfaz a condição suficiente de otimalidade de segunda ordem*

$$d^T \nabla_{xx}^2 L(x^*, \mu^*) d > 0, \quad \forall 0 \neq d \in L_X(x^*) = \{d \in \mathbb{R}^m \mid d^T \nabla g_i(x^*) \leq 0, i \in I(x^*)\},$$

(neste caso  $x^*$  é um minimizador local estrito do problema com restrições). Então se

$c > 0$  é tal que

$$\delta c > \|\mu^*\|_\infty, \quad (2.14)$$

o ponto  $x^*$  é um minimizador local estrito do problema  $(P_1)$ .

*Prova.* Suponha que a conclusão do teorema seja falsa. Então, para cada  $k \in \mathbb{N}$  existem seqüências  $(x^k) \subset \mathbb{R}^n$ ,  $(c_k) \subset \mathbb{R}_+$  tais que  $c_k > k$ ,  $x^k \neq x^*$ ,  $x^k \rightarrow x^*$  e

$$\varphi_{c_k}(x^k) \leq \varphi_{c_k}(x^*) = f(x^*), \quad (2.15)$$

onde a igualdade decorre de  $x^* \in X$ . Assim, usando a definição de  $\sigma$  e (2.15), temos

$$f(x^k) + c_k(\delta \ell_1(x^k)) \leq f(x^k) + c_k\sigma([g(x^k)]_+) = \varphi_{c_k}(x^k) \leq f(x^*). \quad (2.16)$$

Escrevendo

$$x^k = x^* + t_k d^k$$

em que

$$d^k = \frac{(x^k - x^*)}{\|x^k - x^*\|} \quad \text{e} \quad t_k = \|x^k - x^*\| > 0, \quad (2.17)$$

podemos assumir sem perda de generalidade que  $d^k \rightarrow d \neq 0$ . Vamos mostrar que  $d \in L_X(x^*)$ .

Como  $(x^*, \mu^*)$  é um par KKT para o problema (1), temos que  $\nabla_x L(x^*, \mu^*) = 0$  e  $\mu_i^* g_i(x^*) = 0$ ,  $i = 1, \dots, m$ . Disto e da expansão de Taylor segue que

$$\begin{aligned} (\delta c_k - \|\mu^*\|_\infty) \ell_1(x^k) &= (\delta c_k - \|\mu^*\|_\infty) \ell_1(x^k) + \nabla_x L(x^*, \mu^*)^T (x^k - x^*) \\ &= (\delta c_k - \|\mu^*\|_\infty) \ell_1(x^k) + f(x^k) + \sum_{i=1}^m \mu_i^* g_i(x^k) - f(x^*) \\ &\quad - \frac{1}{2} t_k^2 (d^k)^T \nabla_{xx}^2 L(x^*, \mu^*) d^k - o(t_k^2). \end{aligned} \quad (2.18)$$

Reescrevendo (2.18) e utilizando (2.16) obtemos

$$\begin{aligned} (\delta c_k - \|\mu^*\|_\infty) \ell_1(x^k) &= f(x^k) - f(x^*) + \delta c_k \ell_1(x^k) - \sum_{i=1}^m (\|\mu^*\|_\infty [g_i(x^k)]_+ - \mu_i^* g_i(x^k)) \\ &\quad - \frac{1}{2} t_k^2 (d^k)^T \nabla_{xx}^2 L(x^*, \mu^*) d^k - o(t_k^2) \\ &\leq -\frac{1}{2} t_k^2 (d^k)^T \nabla_{xx}^2 L(x^*, \mu^*) d^k - o(t_k^2). \end{aligned} \quad (2.19)$$

Como  $(\delta c_k - \|\mu^*\|_\infty) > 0$ , a relação (2.19) resulta em

$$0 \leq \ell_1(x^k) \leq -(\delta c_k - \|\mu^*\|_\infty)^{-1} \left( \frac{1}{2} t_k^2 (d^k)^T \nabla_{xx}^2 L(x^*, \mu^*) d^k + o(t_k^2) \right). \quad (2.20)$$

Dividindo ambos os lados de (2.20) por  $t_k$ , temos

$$0 \leq \frac{\ell_1(x^k)}{t_k} \leq -(\delta c_k - \|\mu^*\|_\infty)^{-1} \left( \frac{1}{2} (d^k)^T \nabla_{xx}^2 L(x^*, \mu^*) d^k + \frac{o(t_k^2)}{t_k} \right) t_k. \quad (2.21)$$

As sequências  $d^k \rightarrow d$ ,  $t_k \rightarrow 0$  e  $(\delta c_k - \|\mu^*\|_\infty)^{-1} \rightarrow 0$ , logo o termo do lado direito de (2.21) converge para zero quanto  $k \rightarrow \infty$ . Portanto,

$$0 = \lim_{k \rightarrow \infty} \frac{\ell_1(x^k)}{t_k} = \lim_{k \rightarrow \infty} \sum_{i=1}^m \frac{[g_i(x^k)]_+}{t_k}.$$

Do fato que  $[g_i(x^k)]_+ \geq 0$  para todo  $i = 1, \dots, m$ , segue que

$$\lim_{k \rightarrow \infty} \frac{[g_i(x^k)]_+}{t_k} = 0, \quad \forall i = 1, \dots, m. \quad (2.22)$$

Da diferenciabilidade de  $g_i$ , para  $i = 1, \dots, m$ , e da viabilidade de  $x^*$  temos

$$\nabla g_i(x^*)^T d = \lim_{k \rightarrow \infty} \nabla g_i(x^*)^T d^k = \lim_{k \rightarrow \infty} \frac{g_i(x^k) - g_i(x^*)}{\|x^k - x^*\|} = \lim_{k \rightarrow \infty} \frac{g_i(x^k)}{t_k}. \quad (2.23)$$

Usando que  $g_i(x^k) \leq [g_i(x^k)]_+$ , para todo  $i = 1, \dots, m$ , (2.22) e (2.23), obtemos

$$\nabla g_i(x^*)^T d \leq 0$$

para todo  $i = 1, \dots, m$ , o que implica que  $d \in L_X(x^*)$ . Assim, pela hipótese de condição suficiente de segunda ordem, obtemos que

$$d^T \nabla_{xx}^2 L(x^*, \mu^*) d > 0. \quad (2.24)$$

Por outro lado, da relação (2.19) decorre

$$0 \geq \lim_{k \rightarrow \infty} \left[ \frac{1}{2} (d^k)^T \nabla_{xx}^2 L(x^*, \mu^*) d^k + \frac{o(t_k^2)}{t_k} \right] = \frac{1}{2} d^T \nabla_{xx}^2 L(x^*, \mu^*) d,$$

o que é um contradição com (2.24). ■

A seguinte proposição é consequência do Teorema 1.14 e do Corolário 1.15 e estabelece penalização exata sob a suposição de MFCQ.

**Proposição 2.5.** *Seja  $x^* \in X$  um minimizador local estrito do problema (1). Suponha que  $f$  é Lipschitz contínua em uma vizinhança  $U$  de  $x^*$  com constante Lipschitz  $L > 0$ .*

Suponha também que  $x^*$  cumpre MFCQ. Então qualquer função de penalidade definida em (2.2) é exata.

*Prova.* Seja  $P$  uma função de penalidade dada em (2.2). Vamos mostrar que essa função satisfaz a estimativa local de inviabilidade (1.30) para algum  $\nu \geq 1$ . De fato, do Corolário 1.13, existe  $\varepsilon_1 > 0$  tal que

$$\text{dist}(x, X) \leq M_1 \ell_1(x)$$

para todo  $x \in U := B(x, \varepsilon_1)$ . Assim a relação (2.1) implica que

$$\text{dist}(x, X) \leq M_1 \ell_1(x) \leq M_1 \delta^{-1} \sigma([g(x)]_+). \quad (2.25)$$

Tomando  $\nu = 1$ , decorre de (2.25) que  $P$  satisfaz as hipóteses do Corolário 1.15., donde segue o resultado enunciado. ■

Para estabelecermos o próximo resultado precisamos de um lema auxiliar apresentando em [15, Prop. 4.3.16] para a função de penalidade  $\ell_1$ .

**Lema 2.6.** *Seja  $P$  uma função de penalidade definida em (2.2). Então*

$$L(x, \mu) \leq f(x) + \delta^{-1} \|\mu\|_\infty P(x) \quad (2.26)$$

para quaisquer  $x \in \mathbb{R}^n$  e  $\mu \in \mathbb{R}_+^m$ .

*Prova.* Sejam  $x \in \mathbb{R}^n$  e  $\mu \in \mathbb{R}_+^m$ . Então,

$$\begin{aligned} \sum_{i=1}^m \mu_i g_i(x) &\leq \sum_{i=1}^m \mu_i [g_i(x)]_+ \leq \|\mu\|_\infty \sum_{i=1}^m [g_i(x)]_+ \\ &= \|\mu\|_\infty \|[g(x)]_+\|_1 \\ &\leq \delta^{-1} \|\mu\|_\infty P(x), \end{aligned}$$

em que na primeira desigualdade utilizamos que  $\mu_i \geq 0$  e na última desigualdade a Definição 2.1. Portanto,

$$L(x, \mu) = f(x) + \sum_{i=1}^m \mu_i g_i(x) \leq f(x) + \delta^{-1} \|\mu\|_\infty P(x)$$

para todo  $(x, \mu) \in \mathbb{R}^n \times \mathbb{R}_+^m$ . ■

**Observação 2.7** (penalidade e dualidade). *Uma das consequências do Lema 2.6, relaciona os problema de penalidade com o problema dual de Lagrange dado por*

$$\begin{aligned} (D) \quad &\text{maximizar } \theta(\mu) \\ &\text{sujeito a } \mu \in \mathbb{R}_+^m \end{aligned}$$

em que  $\theta : \mathbb{R}_+^m \rightarrow \mathbb{R} \cup \{-\infty\}$  é a função dual, definida por

$$\theta(\mu) = \inf_{x \in \mathbb{R}^n} L(x, \mu).$$

O problema dual possui interessantes propriedades, que entre elas, podemos citar que a função dual é côncava, e que para qualquer par  $(x, \mu) \in X \times \mathbb{R}_+^m$ , sempre temos que  $\theta(\mu) \leq f(x)$ . Ainda mais, se conseguimos calcular  $x_{\mu^*} \in \mathbb{R}^n$  de modo que  $\theta(\mu^*) = L(x_{\mu^*}, \mu^*)$ , então este ponto resolve a seguinte perturbação do problema original (1):

$$\begin{aligned} & \text{minimizar } f(x) \\ & \text{sujeito a } g_i(x) \leq g_i(x_{\mu^*}), \quad i \in \{1, \dots, m \mid \mu_i^* > 0\}. \end{aligned}$$

Em particular, se  $x_{\mu^*} \in X$  e  $g_i(x_{\mu^*})\mu_i^* = 0$ , para  $i = 1, \dots, m$ , então  $x_{\mu^*}$  resolve o problema original e  $f(x_{\mu^*}) = \theta(\mu^*)$ . Além disto,  $(x_{\mu^*}, \mu^*)$  é um par KKT para o problema (1).

Mais geralmente, não podemos garantir que uma solução do problema (1) seja obtida resolvendo o problema dual, sem condições adequadas de convexidade sob esse problema. Por outro lado, o Teorema 1.5 mostra que, se a sequência  $(x^k)$  é tal que  $x^k = x(c_k)$  e  $c_k \rightarrow +\infty$ , uma solução ótima do problema original é recuperada.

Vamos relacionar os valores ótimos do problema dual e do problema de penalidade. Para isto, suponha que os problemas (1),  $(P_1)$  e  $(D)$  possuem valores ótimos  $f_*$ ,  $\varphi_*$ ,  $\theta_*$ , respectivamente. Para cada  $\mu \in \mathbb{R}_+^m$  considere  $\bar{c} = \delta^{-1}\|\mu\|_\infty$ . Então pelo Lema 2.6, segue que para todo  $x \in \mathbb{R}^n$ ,

$$L(x, \mu) \leq \varphi_{\bar{c}}(x)$$

assim,

$$\theta(\mu) = \inf_{x \in \mathbb{R}^n} L(x, \mu) \leq \inf_{x \in \mathbb{R}^n} \varphi_{\bar{c}}(x) = \varphi_{\bar{c}}(x(\bar{c})).$$

Como  $\delta$  é fixo, decorre da última relação que

$$\theta_* = \sup_{\mu \in \mathbb{R}_+^m} \theta(\mu) \leq \sup_{\mu \in \mathbb{R}_+^m} \varphi_{\bar{c}}(x(\bar{c})) = \sup_{\bar{c} \geq 0} \varphi_{\bar{c}}(x(\bar{c})) = \varphi_*,$$

portanto

$$f_* - \theta_* \geq f_* - \varphi_* \geq 0.$$

O lado esquerdo da desigualdade acima é chamado de gap de dualidade. Ela mostra que a diferença entre o valor ótimo de (1) e do problema  $(P_1)$  é sempre menor ou igual ao valor de gap de dualidade. Consequentemente, se o gap é zero então  $\varphi_* = f_*$ .

Para mais detalhes e propriedades sobre dualidade lagrangiana veja [6, 16, 17], e para um estudo geométrico sobre o gap de dualidade [18].

Mostramos agora o caso mais satisfatório de penalização exata, no sentido que sob as

hipóteses desse resultado uma solução *global* do problema com restrições pode ser obtida resolvendo problemas irrestritos. A demonstração que apresentamos é novamente uma extensão da apresentada em [15, Teo. 4.3.17].

**Teorema 2.8.** *[Penalização exata no caso convexo] Suponha que  $f$  e  $g_i$ ,  $i = 1, \dots, m$  são funções convexas, e que  $x^*$  é um minimizador do problema (1) e  $\mu^*$  um vetor de multiplicador de Lagrange. Se  $P$  é uma função de penalidade como definida em (2.2), então para todo  $c > 0$  tal que  $c\delta \geq \|\mu^*\|_\infty$ , o ponto  $x^*$  é um minimizador de  $(P_1)$ .*

*Prova.* Como  $(x^*, \mu^*)$  é um par KKT, temos que

$$\nabla_x L(x^*, \mu^*) = \nabla f(x^*) + (\mu^*)^T g'(x^*) = 0.$$

A função  $L(\cdot, \mu^*)$  é convexa, visto que  $f$  e  $g_i$ ,  $i = 1, \dots, m$  são convexas, assim a relação acima mostra que  $x^*$  também é uma solução do problema

$$\text{minimizar } L(x, \mu^*), \quad \text{sujeito a } x \in \mathbb{R}^n,$$

ou seja,

$$L(x^*, \mu^*) \leq L(x, \mu^*), \quad \forall x \in \mathbb{R}^n. \quad (2.27)$$

Por outro lado,  $x^* \in X$  satisfaz junto com  $\mu^*$  a condição de complementariedade, o que implica que para qualquer  $c \geq 0$

$$\begin{aligned} \varphi_c(x^*) &= f(x^*) + cP(x^*) = f(x^*) \\ &= f(x^*) + \sum_{i=1}^m \mu_i^* g_i(x^*) \\ &= L(x^*, \mu^*). \end{aligned} \quad (2.28)$$

Juntando (2.28) e (2.27) obtemos

$$\varphi_c(x^*) = L(x^*, \mu^*) \leq L(x, \mu^*)$$

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

Seja  $c \geq c^* := \delta^{-1} \|\mu^*\|_\infty$ . Da relação acima e do Lema 2.6 temos

$$\varphi_c(x^*) \leq L(x, \mu^*) \leq f(x) + c^* P(x) \leq f(x) + cP(x) = \varphi_c(x),$$

para todo  $x \in \mathbb{R}^n$ . Portanto,  $x^*$  é um minimizador global de  $\varphi_c$  para todo  $c \geq c^*$ . ■

O diagrama na Figura 2.1 mostra uma síntese dos resultados apresentados nesta seção com as seguintes abreviações:

- CQ: Condição de qualificação MFCQ.
- Min. local estrito: Minimizador estrito do problema (1).
- Min. local: Minimizador local do problema (1).
- Penalidade Exata Min. local: Minimizador local da função de penalização exata.
- Penalidade Exata Min. local estrito: Minimizador estrito da função de penalização exata.
- KKT2: Condição suficiente de segunda ordem do Teorema 2.4.

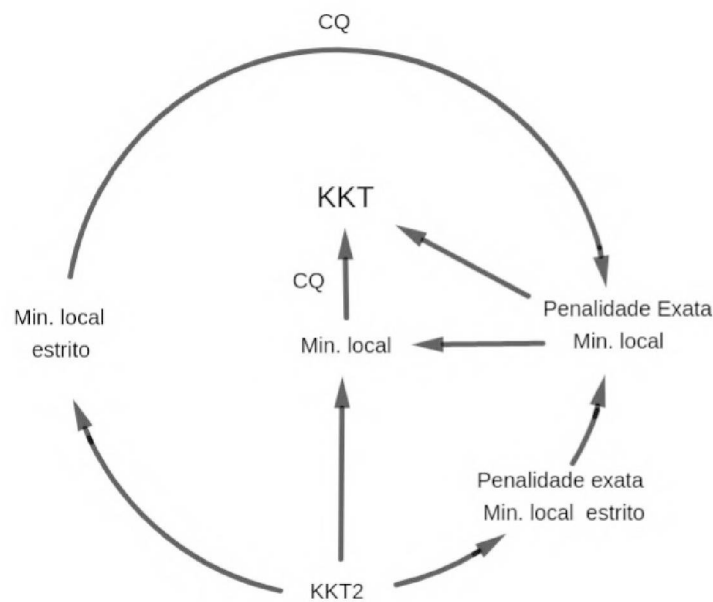


Figura 2.1: Relações entre funções de penalização exata e o problema (1).

Por fim, observamos que os resultados desta seção explicitam um limitante inferior para o parâmetro de penalidade exata. Este é obtido por meio dos multiplicadores de Lagrange, os quais infelizmente, não são conhecidos em geral. Se  $c$  for muito pequeno, a função penalizada pode ser ilimitada inferiormente. Por outro lado, se  $c$  for muito grande, surgem problemas de mal-condicionamento [23]. Além disto, mesmo que as funções  $f$  e  $g_i$ ,  $i = 1, \dots, m$ , sejam suaves, a função de penalização exata é não diferenciável em pontos em que a restrição é ativa, o que demanda métodos de otimização não suaves para resolver os subproblemas de penalidade ( $P_1$ ).

## 2.2 Método de penalidade suave

Diferentemente do método de penalidade exata em que a função penalizada não é diferenciável, o método de penalidade suave possibilita o uso de algoritmos continuamente

diferenciáveis na minimização de  $\varphi_c(\cdot)$ , como os métodos quase-Newton. Um resultado interessante que veremos nesta seção, mostra que os multiplicadores de Lagrange para o problema (1) são muito bem aproximados por este método. As principais referências para a construção desta seção são [1, 22, 34].

### 2.2.1 Construindo funções suaves

Mesmo quando  $g$  é uma função suave, a função  $[g(\cdot)]_+$  não é suave se o problema (1) é viável. Apesar disto, ao considerarmos a composição de  $\|\cdot\|_2^2$  com  $[g(\cdot)]_+$ , obtemos que a função

$$\|[g(\cdot)]_+\|_2^2$$

é suave. Mais geralmente, Luenberger e Ye [22] afirmam que sob determinadas condições acerca de  $\psi : \mathbb{R}^m \rightarrow \mathbb{R}$ , a composição  $\psi([g(\cdot)]_+)$  é uma função de penalidade suave e fornecem uma fórmula para o cálculo do gradiente. O que vamos fazer nesta subseção, mais especificamente no Teorema 2.9, é formalizar esse resultado.

**Teorema 2.9.** *Seja  $\psi : \mathbb{R}^m \rightarrow \mathbb{R}$  uma função diferenciável satisfazendo*

$$\psi(y) \geq 0 \text{ para todo } y \in \mathbb{R}^m; \quad (2.29)$$

$$\psi(y) = 0 \text{ se, e somente se, } y = 0; \quad (2.30)$$

$$\text{Se } b \in \mathbb{R}^m \text{ é tal que } b_i = 0, \text{ então } \frac{\partial \psi}{\partial y_i}(b) = 0. \quad (2.31)$$

Então a função  $P : \mathbb{R}^n \rightarrow \mathbb{R}_+$  dada por

$$P(x) = \psi([g(x)]_+)$$

é uma função de penalidade diferenciável. Além disso,

$$\nabla P(x) = g'(x)^T \nabla \psi([g(x)]_+) = \sum_{i=1}^m \frac{\partial \psi}{\partial y_i}([g(x)]_+) \nabla g_i(x). \quad (2.32)$$

Em particular, se  $\psi$  é suave então  $P$  é suave.

*Prova.* Vemos imediatamente que as condições (2.29) e (2.30) garantem que  $P$  é uma função de penalidade. Considere agora  $a \in \mathbb{R}^n$  arbitrário e defina

$$b = [g(a)]_+ \in \mathbb{R}^m, \quad A = g'(a) \in \mathbb{R}^{m \times n} \quad \text{e} \quad B = \psi'(b) = \nabla \psi(b)^T \in \mathbb{R}^{1 \times m}.$$

Dadas  $(u^k) \subset \mathbb{R}^n \setminus \{0\}$ ,  $(v^k) \subset \mathbb{R}^m \setminus \{0\}$ , com  $u^k \rightarrow 0$  e  $v^k \rightarrow 0$ , a diferenciabilidade de  $g$

e  $\psi$  nos fornece

$$\frac{g(a + u^k) - g(a) - Au^k}{\|u^k\|} \rightarrow 0 \quad \text{e} \quad \frac{\psi(b + v^k) - \psi(b) - Bv^k}{\|v^k\|} \rightarrow 0.$$

Tome uma sequência arbitrária  $(u^k) \subset \mathbb{R}^n \setminus \{0\}$  com  $u^k \rightarrow 0$  e defina

$$v^k = [g(a + u^k)]_+ - [g(a)]_+ \quad \text{e} \quad \gamma^k = \frac{v^k - Au^k}{\|u^k\|}. \quad (2.33)$$

Denotando

$$I_+ = \{i \mid g_i(a) > 0\}, \quad I_- = \{i \mid g_i(a) < 0\} \quad \text{e} \quad I_0 = \{i \mid g_i(a) = 0\},$$

afirmamos que  $\gamma_i^k \rightarrow 0$  para  $i \in I_+$  e que a sequência  $(\gamma_i^k)$  é limitada se  $i \in I_- \cup I_0$ . De fato, para  $i \in I_+$  e  $k \in \mathbb{N}$  suficientemente grande temos

$$\gamma_i^k = \frac{g_i(a + u^k) - g_i(a) - \nabla g_i(a)^T u^k}{\|u^k\|} \rightarrow 0.$$

Por outro lado, se  $i \in I_-$ , então

$$\gamma_i^k = \frac{-\nabla g_i(a)^T u^k}{\|u^k\|}$$

para todo  $k \in \mathbb{N}$  suficientemente grande. Finalmente, para  $i \in I_0$  temos

$$\gamma_i^k = \begin{cases} \frac{-\nabla g_i(a)^T u^k}{\|u^k\|}, & \text{se } k \in \{k \in \mathbb{N} \mid g_i(a + u^k) \leq 0\}, \\ \frac{g_i(a + u^k) - g_i(a) - \nabla g_i(a)^T u^k}{\|u^k\|}, & \text{se } k \in \{k \in \mathbb{N} \mid g_i(a + u^k) > 0\}. \end{cases}$$

Uma consequência imediata da afirmação que acabamos de provar é que a sequência  $\left(\frac{v^k}{\|u^k\|}\right)$  é limitada. Provaremos agora a diferenciabilidade de  $P$  e a relação (2.32). Para isto vamos considerar dois casos.

Caso 1:  $\mathcal{K}_1 = \{k \in \mathbb{N} \mid v^k = 0\}$  é infinito. Neste caso, em vista de (2.33), temos

$$\begin{aligned}
\frac{P(a + u^k) - P(a) - BAu^k}{\|u^k\|} &= \frac{\psi([g(a + u^k)]_+) - \psi([g(a)]_+) - BAu^k}{\|u^k\|} \\
&= \frac{\psi(b + v^k) - \psi(b) - BAu^k}{\|u^k\|} \\
&= B\gamma^k \\
&= \sum_{i=1}^m \frac{\partial\psi}{\partial y_i}(b)\gamma_i^k \\
&= \sum_{i \in I_+} \frac{\partial\psi}{\partial y_i}(b)\gamma_i^k \xrightarrow{\mathcal{K}_1} 0,
\end{aligned} \tag{2.34}$$

note que utilizamos a condição (2.31) para obter (2.34).

Caso 2:  $\mathcal{K}_2 = \{k \in \mathbb{N} \mid v^k \neq 0\}$  é infinito. Novamente usando (2.33) temos

$$\begin{aligned}
\frac{P(a + u^k) - P(a) - BAu^k}{\|u^k\|} &= \frac{\psi([g(a + u^k)]_+) - \psi([g(a)]_+) - BAu^k}{\|u^k\|} \\
&= \frac{\psi(b + v^k) - \psi(b) - BAu^k}{\|u^k\|} \\
&= \frac{\psi(b + v^k) - \psi(b) - Bv^k + B(v^k - Au^k)}{\|u^k\|}.
\end{aligned}$$

Assim,

$$\begin{aligned}
\frac{P(a + u^k) - P(a) - BAu^k}{\|u^k\|} &= \frac{\psi(b + v^k) - \psi(b) - Bv^k}{\|v^k\|} \frac{\|v^k\|}{\|u^k\|} + B\gamma^k \\
&= \frac{\psi(b + v^k) - \psi(b) - Bv^k}{\|v^k\|} \frac{\|v^k\|}{\|u^k\|} + \sum_{i=1}^m \frac{\partial\psi}{\partial y_i}(b)\gamma_i^k,
\end{aligned} \tag{2.35}$$

utilizando (2.31) em (2.35) obtemos

$$\begin{aligned}
\frac{P(a + u^k) - P(a) - BAu^k}{\|u^k\|} &= \frac{\psi([g(a + u^k)]_+) - \psi([g(a)]_+) - BAu^k}{\|u^k\|} \\
&= \frac{\psi(b + v^k) - \psi(b) - Bv^k}{\|v^k\|} \frac{\|v^k\|}{\|u^k\|} + \sum_{i=1}^m \frac{\partial\psi}{\partial y_i}(b)\gamma_i^k \\
&= \frac{\psi(b + v^k) - \psi(b) - Bv^k}{\|v^k\|} \frac{\|v^k\|}{\|u^k\|} + \sum_{i \in I_+} \frac{\partial\psi}{\partial y_i}(b)\gamma_i^k \xrightarrow{\mathcal{K}_2} 0.
\end{aligned}$$

Portanto,  $P$  é diferenciável e o seu gradiente é dado por (2.32). Além disso, se  $\psi$  é suave, decorre imediatamente da suavidade de  $g$  e da continuidade de  $[\cdot]_+$  que a derivada da função de penalidade é contínua. ■

Do Teorema 2.9 obtemos uma forma de construir funções de penalidade suaves como

definido a seguir.

**Definição 2.10.** *Dada  $\psi : \mathbb{R}^m \rightarrow \mathbb{R}$  uma função suave satisfazendo (2.29)-(2.31) definimos a função de penalidade suave*

$$P(x) := \psi([g(x)]_+). \quad (2.36)$$

Para uma função de penalidade definida como acima, considere o problema de penalidade

$$(P_2) \quad \begin{aligned} &\text{minimizar } \varphi_c(x) := f(x) + c\psi([g(x)]_+) \\ &\text{sujeito a } x \in \mathbb{R}^n. \end{aligned}$$

**Exemplo 2.2.** *Sabemos que a função de penalidade  $\ell_2$  definida em (1.20) é suave. Vejamos que a função  $\psi : \mathbb{R}^m \rightarrow \mathbb{R}_+$  dada por*

$$\psi(y) = \frac{1}{2} \|y\|_2^2 = \frac{1}{2} \sum_{i=1}^m y_i^2,$$

*satisfaz a condição (2.31). De fato, temos que*

$$\frac{\partial \psi}{\partial y_i}(b) = b_i = 0 \text{ se, e somente se, } b_i = 0.$$

*Como (2.29)-(2.30) são imediatas, segue que  $\ell_2$  pertence a classe de funções dada em (2.36), e pela fórmula (2.32) para o vetor gradiente, este é expresso por*

$$\nabla \ell_2(x) = \sum_{i=1}^m ([g_i(x)]_+) \nabla g_i(x),$$

*para qualquer  $x \in \mathbb{R}^n$ .*

## 2.2.2 Multiplicadores de Lagrange e o mal condicionamento

O resultado que apresentamos a seguir, mostra que uma sequência convergente para um multiplicador de Lagrange pode ser construída quando utilizado um método de penalidade suave. Em [34, Teo.6.7], tal resultado é demonstrado com a hipótese de condição de qualificação de Robinson, enquanto que [23, Teo.10.2.5] utiliza a hipótese de LICQ, ambos para o caso particular  $P = \ell_2$ . A prova que apresentamos se baseia nessas duas referências. Cabe destacar, que esse resultado considerando  $P$  como definida em (2.36) é enunciado sem demonstração em [22].

**Teorema 2.11.** *Seja  $P$  dada pela Definição 2.10. Considere  $(x^k)$  uma sequência gerada*

pele método de penalidade e a sequência

$$\mu^k := c_k \nabla \psi([g(x^k)]_+) \in \mathbb{R}^m. \quad (2.37)$$

Suponha que  $(x^k)$  converge para uma solução  $x^*$  de (1). Suponha também que  $x^*$  satisfaz LICQ. Então  $(\mu^k)$  é limitada e qualquer ponto de acumulação  $\mu^*$  é um multiplicador de Lagrange associado a  $x^*$ .

*Prova.* Considere  $k \in \mathbb{N}$  fixado. Como  $x^k$  é um minimizador global de  $(P_2)$ , temos da condição necessária de otimalidade de primeira ordem que

$$0 = \nabla \varphi_{c_k}(x^k) = \nabla f(x^k) + c_k \nabla P(x^k).$$

Juntando a relação acima, (2.32) e a definição de  $\mu^k$ , obtemos

$$0 = \nabla f(x^k) + c_k g'(x^k)^T \nabla \psi([g(x^k)]_+) = \nabla f(x^k) + g'(x^k)^T \mu^k, \quad (2.38)$$

Vamos mostrar que a sequência  $(\mu^k)$  é limitada. Por contradição, suponha que existe um conjunto infinito de índices  $\bar{\mathcal{K}} \subset \mathbb{N}$ , tal que  $\|\mu^k\| \xrightarrow{\bar{\mathcal{K}}} \infty$  com  $\|\mu^k\| \neq 0$ . Para cada  $k \in \bar{\mathcal{K}}$  defina

$$\bar{\mu}^k := \frac{\mu^k}{\|\mu^k\|}.$$

Por construção, a sequência  $(\bar{\mu}^k)$  é limitada. Logo, existe uma subsequência convergente cujo limite denotamos por  $\bar{\mu} \neq 0$ .

Agora, dividindo ambos os lados de (2.38) por  $\|\mu^k\|$ , observando que  $(x^k)$  converge para  $x^*$  e, pois  $f$  e  $g$  são suaves, passando o limite sobre a subsequência convergente, obtemos

$$0 = \lim_{k \rightarrow \infty, k \in \bar{\mathcal{K}}} \left[ \frac{\nabla f(x^k)}{\|\mu^k\|} + g'(x^k)^T \bar{\mu}^k \right] = g'(x^*)^T \bar{\mu}. \quad (2.39)$$

Note que, se  $i \notin I(x^*) \subset \{1, \dots, m\}$  então  $g_i(x^k) \rightarrow g_i(x^*) < 0$ , logo para  $k$  suficientemente grande temos  $g_i(x^k) < 0$ , para todo  $i \notin I(x^*)$ . Portanto  $[g_i(x^k)]_+ = 0$ , o que junto com a condição (2.31) resulta em

$$\frac{\partial \psi}{\partial y_i}([g(x^k)]_+) = 0,$$

consequentemente,  $\bar{\mu}_i = 0$  para todo  $i \notin I(x^*)$ . Assim, podemos reescrever (2.39) como

$$0 = g'(x^*)^T \bar{\mu} = \sum_{i \in I(x^*)} \bar{\mu}_i \nabla g_i(x^*). \quad (2.40)$$

Donde segue por LICQ que  $\bar{\mu}_i = 0$  para  $i \in I(x^*)$ . Logo  $\bar{\mu} = 0$ , o que é uma contradição. Portanto  $(\mu^k)$  é limitada.

Suponha agora que  $\mu^k \xrightarrow{\mathcal{K}} \mu^* = (\mu_1, \dots, \mu_m)$  para algum conjunto infinito de índices  $\mathcal{K} \subset \mathbb{N}$ . Então, de (2.38) e a suavidades das funções  $f$  e  $g$ , temos

$$0 = \lim_{k \rightarrow \infty, k \in \mathcal{K}} [\nabla f(x^k) + g'(x^k)^T \mu^k] = \nabla f(x^*) + g'(x^*)^T \mu^*. \quad (2.41)$$

Usando novamente que se  $i \notin I(x^*)$  então  $g_i(x^k) < 0$ , para  $k$  suficientemente grande, obtemos

$$g_i(x^k) \mu_i^k = c_k g_i(x^k) \frac{\partial \psi}{\partial y_i}([g(x^k)]_+) = 0.$$

Passando o limite sobre a subsequência convergente, obtemos

$$g_i(x^*) \mu_i = 0,$$

para todo  $i \notin I(x^*)$ . Portanto  $\mu_i = 0$  para  $i \notin I(x^*)$ . Levando isso em conta na relação (2.41), obtemos

$$0 = \nabla f(x^*) + \sum_{i \in I(x^*)} \mu_i \nabla g_i(x^*). \quad (2.42)$$

Por outro lado, o ponto  $x^*$  é um minimizador de (1) e satisfaz LICQ, então existe um multiplicador  $\lambda = (\lambda_1, \dots, \lambda_m) \in \mathbb{R}_+^m$  tal que

$$0 = \nabla f(x^*) + \sum_{i \in I(x^*)} \lambda_i \nabla g_i(x^*), \quad (2.43)$$

e  $\lambda_i g_i(x^*) = 0$  para todo  $i = 1, \dots, m$ . Usando (2.42) e (2.43), temos

$$\sum_{i \in I(x^*)} \mu_i^* \nabla g_i(x^*) = \sum_{i \in I(x^*)} \lambda_i \nabla g_i(x^*),$$

mas, como o conjunto  $\{\nabla g_i(x^*)\}_{i \in I(x^*)}$  é linearmente independente decorre que  $\mu_i^* = \lambda_i$  para  $i \in I(x^*)$ , donde segue o resultado desejado. ■

A resolução dos problema  $(P_2)$  tornam-se cada vez mais complicadas para parâmetros de penalidade muito grandes, visto que as matrizes Hessianas das funções penalizadas tornam-se cada vez mais mal condicionadas [22, Lema 1], como ilustramos no próximo exemplo.

**Exemplo 2.3.** Considere o problema de minimização em  $\mathbb{R}^2$  dado por

$$\begin{aligned} \text{minimizar } f(x) &= \frac{1}{2}x_1^2 + \frac{1}{2}(x_2 - 1)^2 \\ \text{sujeito a } g(x) &= 1 - x_1 \leq 0. \end{aligned} \quad (2.44)$$

O ponto  $x^* = (1, 1)$  é a única solução de (2.44). Note que  $x^*$  é um ponto KKT para o

problema acima com  $\mu^* = 1$ . De fato, temos que

$$\nabla f(x) = \begin{bmatrix} x_1 \\ x_2 - 1 \end{bmatrix} \quad e \quad \nabla g(x) = \begin{bmatrix} -1 \\ 0 \end{bmatrix}.$$

Assim,

$$\nabla f(x^*) + \mu^* \nabla g(x^*) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 1 \begin{bmatrix} -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

e portanto  $(x^*, \mu)$  é um par KKT para o problema (2.44).

Considere a função de penalidade  $\ell_2$ . Então, dado  $c > 0$ , o problema de penalidade  $(P_2)$  toma a forma

$$\begin{aligned} \text{minimizar } \varphi_c(x) &:= \frac{1}{2}(x_1^2 + (x_2 - 1)^2) + \frac{c}{2}([1 - x_1]_+)^2 \\ \text{sujeito a } x &\in \mathbb{R}^2. \end{aligned}$$

Como a função penalizada é convexa, para calcular um minimizador  $x(c)$  de  $\varphi_c$ , basta determinar os pontos tais que  $\nabla \varphi_c(x(c)) = 0$ , o que nos fornece

$$x(c) = \left( \frac{c}{1+c}, 1 \right).$$

Em particular, dada uma sequência  $(c_k) \subset \mathbb{R}_+$  com  $c_k \rightarrow \infty$ , segue que  $x^k = x(c_k) \rightarrow x^*$ . Além disso,  $\{\nabla g(x^*)\}$  é linearmente independente, então pelo Teorema 2.11, existe uma sequência de vetores  $(\mu^k) \subset \mathbb{R}$  convergindo para  $\mu^*$ ,

$$\mu^k = c_k [g(x^k)]_+ = \frac{c_k}{1+c_k} \rightarrow 1.$$

A matriz Hessiana de  $\varphi_c$  em  $x(c)$  é dada por

$$\nabla_{xx}^2 \varphi_c(x(c)) = \begin{bmatrix} 1+c & 0 \\ 0 & 1 \end{bmatrix},$$

com autovalores associados  $\lambda_1 = (1+c)$  e  $\lambda_2 = 1$  com respectivos autovetores  $(1, 0)$  e  $(0, 1)$ . Agora, o índice de condicionamento da matriz acima é dado por

$$\chi_c = \frac{\lambda_1}{\lambda_2} = 1+c.$$

Logo  $\chi_c \rightarrow +\infty$  quando  $c \rightarrow +\infty$ , donde segue que o problema de penalidade torna-se mais difícil de resolver, uma vez que a Hessiana torna-se arbitrariamente mal condicionada próximo de um minimizador de (1). Na Figura 2.2 ilustramos esse efeito. Aumentando  $c = 1$  para  $c = 90$  as curvas de nível das funções penalizadas tornaram-se mais íngremes na

direção do autovetor  $(1, 0)$ . É conhecido também que a taxa de convergência do algoritmo

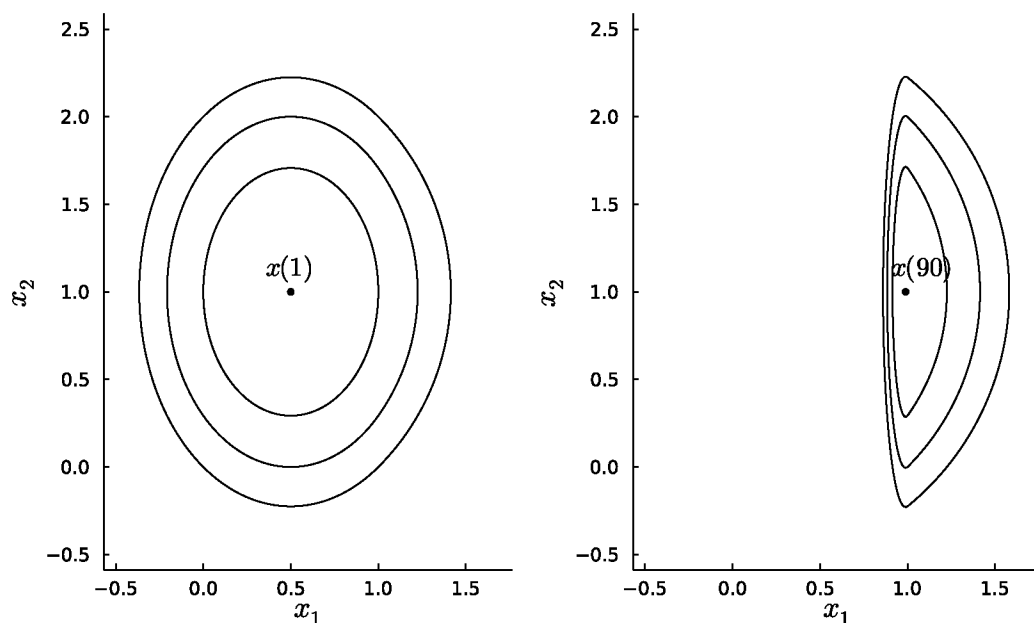


Figura 2.2: Curvas de nível da função penalizada  $\ell_2$  do Exemplo 2.2 para  $c = 1$  (esquerda) e  $c = 90$  (direita).

de máxima descida é mais lenta para valores grandes de  $\chi_c$ , veja por exemplo [34].

Além das dificuldades de mal condicionamento ilustradas pelo exemplo acima, Birgin e Martínez [5] argumentam que se o parâmetro de penalidade é muito grande, um ponto próximo a uma solução do problema original com pequeno valor de inviabilidade, pode ser considerado pior do que um ponto quase viável com um valor pobre de funções objetivo que pode estar longe da solução ótima do problema com restrições.

Uma estratégia para amenizar as dificuldades relacionadas ao mal condicionamento e à não suavidade da função de penalizada, é considerar uma família de funções suaves que aproxima uma função de penalidade exata. Esse é o assunto que discutiremos no próximo capítulo.

# Capítulo 3

## Funções de aproximação suave

Na Seção 2.1 mostramos que sob certas hipóteses (por exemplo a Proposição 2.5) uma solução do problema original é recuperada pelo método de penalidade exata. Entretanto, os subproblemas de penalidade são não suaves. Por outro lado, apesar da função de penalidade  $\ell_2$  suavizar a função  $\ell_1$ , a Proposição 1.18 garante que essa função não é exata e, como ilustrado no Exemplo 2.3 da Seção 2.2, a matriz Hessiana da função penalizada torna-se mal condicionada para parâmetros de penalidade muito grandes.

Diante dessas dificuldades, neste capítulo estudamos métodos de minimização cujos subproblemas são suaves e aproximam o problema de penalidade exata, o que é alcançado considerando uma família de funções suaves que aproxima a função máximo. Exemplos de tais métodos podem ser encontrados em [20, 35, 38, 39]. O que fazemos neste capítulo é apresentar de uma forma unificada esses métodos e resultados que relacionam minimizadores dos problemas suavizados com minimizadores dos problemas de penalidade exata e do problema com restrições.

### 3.1 Definições e propriedades

Neste capítulo utilizamos a notação de [20, 35, 38, 39] para a função  $[\cdot]_+ =: p(\cdot)$  e consideramos a família de problemas de penalidade obtidos através da função penalidade  $\ell_1$ , isto é,

$$(P_1) \quad \text{minimizar } \varphi_c(x) = f(x) + c \sum_{i=1}^m p(g_i(x))$$

sujeito a  $x \in \mathbb{R}^n$ .

**Observação 3.1.** *Lembremos que  $\ell_1$  é exata visto que pertence a classe de funções de penalidade (2.2).*

Vamos considerar em nossa análise funções que além de serem suaves também aproximam a função  $p(\cdot)$ .

**Definição 3.2.** Dizemos que  $\eta : \mathbb{R} \times \mathbb{R}_+ \rightarrow \mathbb{R}$  é uma função de aproximação suave para a função  $p(\cdot)$ , se

(i) a função  $\eta(\cdot, \varepsilon)$  é suave para qualquer  $\varepsilon > 0$ .

(ii) Existem  $a > 0$  e  $b \geq 0$  tais que

$$-b\varepsilon \leq p(t) - \eta(t, \varepsilon) \leq a\varepsilon, \quad (3.1)$$

para todo  $(t, \varepsilon) \in \mathbb{R} \times \mathbb{R}_+$ .

O lema seguinte estabelece que uma função definida como acima converge uniformemente para  $p(\cdot)$  quando  $\varepsilon \rightarrow 0^+$ . Sua demonstração segue imediatamente do item (ii), uma vez que (3.1) não depende de  $t$ .

**Lema 3.3.** Seja  $\eta : \mathbb{R} \times \mathbb{R}_+ \rightarrow \mathbb{R}$  uma função como na Definição 3.2. Então

$$\lim_{\varepsilon \rightarrow 0^+} \eta(\cdot, \varepsilon) = p(\cdot).$$

Seja  $\eta$  uma função de aproximação suave. Para cada  $c > 0$  e  $\varepsilon \geq 0$  definimos a função de suavização  $\varphi_{c,\varepsilon} : \mathbb{R}^n \rightarrow \mathbb{R}$  dada por

$$\varphi_{c,\varepsilon}(x) := f(x) + c \sum_{i=1}^m \eta(g_i(x), \varepsilon). \quad (3.2)$$

Observe que devido a condição (i) da Definição 3.2, a função acima é suave para qualquer  $(c, \varepsilon) \in \mathbb{R}_{++} \times \mathbb{R}_{++}$ . Além disto, como  $p(t) = \eta(t, 0)$  para todo  $t \in \mathbb{R}$ , segue que

$$\varphi_{c,0}(x) = f(x) + c \sum_{i=1}^m p(g_i(x)) = \ell_1(x), \quad \forall x \in \mathbb{R}^n.$$

Vamos obter agora um erro pontual de aproximação entre a função  $\ell_1$  e as funções  $\varphi_{c,\varepsilon}(\cdot)$  e verificar que, de fato, a função definida em (3.2) além de suavizar  $\varphi_c(\cdot)$ , também a aproxima pontualmente.

**Lema 3.4.** Fixe  $\varepsilon > 0$ . Então para todo  $(x, c) \in \mathbb{R}^n \times \mathbb{R}_+$  a seguinte relação é satisfeita

$$-c m b \varepsilon \leq \varphi_c(x) - \varphi_{c,\varepsilon}(x) \leq c m a \varepsilon,$$

onde  $m$  é o número de funções de restrição do problema (1).

*Prova.* Dado  $\varepsilon > 0$ , para todo  $(x, c) \in \mathbb{R}^n \times \mathbb{R}_+$  temos que

$$\begin{aligned}\varphi_c(x) - \varphi_{c,\varepsilon}(x) &= \left( f(x) + c \sum_{i=1}^m p(g_i(x)) \right) - \left( f(x) + c \sum_{i=1}^m \eta(g_i(x), \varepsilon) \right) \\ &= c \sum_{i=1}^m (p(g_i(x)) - \eta(g_i(x), \varepsilon)).\end{aligned}$$

Logo, da relação (3.1) obtemos o resultado requerido.  $\blacksquare$

Como consequência do Lema 3.4, para todo  $c \in \mathbb{R}_+$  a função  $\varphi_c(\cdot, \varepsilon)$  converge uniformemente para a função  $\varphi_c(\cdot)$  quando  $\varepsilon \rightarrow 0$ , visto que a estimativa de erro fornecida pelo Lema 3.4 não depende do ponto  $x$ . Um exemplo particular desse resultado, ou seja para uma função  $\eta$  dada, pode ser encontrado em [39, Lema 2].

**Teorema 3.5.** *Seja  $\eta : \mathbb{R} \times \mathbb{R}_+ \rightarrow \mathbb{R}$  uma função de aproximação suave e  $\varphi_{c,\varepsilon} : \mathbb{R}^n \rightarrow \mathbb{R}$  a função definida em (3.2) a partir de  $\eta$ . Então,*

$$\lim_{\varepsilon \rightarrow 0^+} \varphi_{c,\varepsilon}(\cdot) = \varphi_c(\cdot). \quad (3.3)$$

*Prova.* Dado  $\varepsilon > 0$  tome  $\delta = \varepsilon / (cm \max\{a, b\})$ . Assim, pelo Lema 3.4 temos que, para todo  $\varepsilon < \delta$ ,

$$|\varphi_{c,\varepsilon}(x) - \varphi_c(x)| \leq cm\varepsilon \max\{a, b\} < \varepsilon$$

para todo  $x \in \mathbb{R}^n$ . Portanto,  $\varphi_{c,\varepsilon}$  converge uniformemente para  $\varphi_c$ .  $\blacksquare$

Motivados pelas relações estabelecidas, definimos uma suavização do problema de penalização exata  $(P_1)$ , obtido minimizando a função definida em (3.2), isto é,

$$\begin{aligned}(\text{SP}_\varepsilon) \quad &\text{minimizar} \quad \varphi_{c,\varepsilon}(x) = f(x) + c \sum_{i=1}^m \eta(g_i(x), \varepsilon) \\ &\text{sujeito a } x \in \mathbb{R}^n.\end{aligned}$$

Tendo em vista que a função  $\varphi_{c,\varepsilon}(\cdot)$  aproxima arbitrariamente a função de penalização exata  $\varphi_c(\cdot)$ , pelo Teorema 3.5, espera-se que os minimizadores da função suavizada aproximem um minimizador do problema original, sem que o parâmetro de penalidade torne-se muito grande.

## 3.2 Relações entre as soluções dos problemas

Como uma solução de  $(\text{SP}_\varepsilon)$  se relaciona com uma solução do problema  $(P_1)$  e do problema (1)? Nosso objetivo na presente seção é estabelecer tais relações.

Começamos mostrando que o erro pontual estabelecido no Lema 3.4 é preservado entre soluções do problema suavizado e do problema de penalidade exata. Tal resultado é

análogo ao presente, por exemplo, em [20, Teorema 2.4] para a função suavizada considerada pelos autores, o qual estabelecemos a seguir para qualquer função de suavização definida como em (3.2).

**Teorema 3.6.** *Sejam  $c, \varepsilon \geq 0$ . Considere  $x^*$  uma solução do problema  $(P_1)$  e  $\bar{x}_\varepsilon$  uma solução do problema  $(SP_\varepsilon)$ . Então*

$$-c m b \varepsilon \leq \varphi_c(x^*) - \varphi_{c,\varepsilon}(\bar{x}_\varepsilon) \leq c m a \varepsilon.$$

*Prova.* Como  $x^*$  é uma solução de  $(P_1)$  e  $\bar{x}_\varepsilon$  uma solução de  $(SP_\varepsilon)$ , temos que

$$\varphi_c(x^*) - \varphi_{c,\varepsilon}(x^*) \leq \varphi_c(x^*) - \varphi_{c,\varepsilon}(\bar{x}_\varepsilon) \leq \varphi_c(\bar{x}_\varepsilon) - \varphi_{c,\varepsilon}(\bar{x}_\varepsilon).$$

Aplicando o Lema 3.4 nos dois extremos da relação acima, obtemos

$$-c m b \varepsilon \leq \varphi_c(x^*) - \varphi_{c,\varepsilon}(\bar{x}_\varepsilon) \leq c m a \varepsilon,$$

donde segue o resultado desejado. ■

O próximo teorema estende para as funções de aproximação suave o resultado presente em [39, Corolário 1], o qual estabelece relações entre minimizadores do problema  $(SP_\varepsilon)$  e soluções do problema  $(P_1)$ , mais especificamente, uma solução do problema de penalidade exata pode ser aproximada por uma sequência de soluções dos problemas de penalidade suavizados.

**Teorema 3.7.** *Seja  $(\varepsilon_k) \subset \mathbb{R}_+$  uma sequência convergente para zero. Suponha que para algum  $c > 0$ ,  $x^k \in \mathbb{R}^n$  é uma solução de  $(SP_{\varepsilon_k})$  em que  $\varepsilon_k = \varepsilon$ , para todo  $k \in \mathbb{N}$ . Então qualquer ponto de acumulação  $\bar{x}$  da sequência  $(x^k)$  é uma solução de  $(P_1)$ .*

*Prova.* Como  $x^k$  é uma solução de  $(SP_{\varepsilon_k})$ , segue que

$$\varphi_{c,\varepsilon_k}(x^k) \leq \varphi_{c,\varepsilon_k}(x), \tag{3.4}$$

para quaisquer  $x \in \mathbb{R}^n$  e  $k \in \mathbb{N}$ .

Considere  $\mathcal{K} \subset \mathbb{N}$  um conjunto infinito de índices em que  $x^k \xrightarrow{\mathcal{K}} \bar{x}$ . Do Lema 3.4 temos que

$$\varphi_c(x^k) - c m a \varepsilon_k \leq \varphi_{c,\varepsilon_k}(x^k) \leq \varphi_c(x^k) + c m b \varepsilon_k.$$

Passando o limite sobre a subsequência convergente e usando que  $\varepsilon_k \xrightarrow{\mathcal{K}} 0$ , obtemos

$$\lim_{k \rightarrow \infty, k \in \mathcal{K}} \varphi_{c,\varepsilon_k}(x^k) = \varphi_c(\bar{x}).$$

Usando a última relação e (3.4), concluímos que

$$\varphi_c(\bar{x}) = \lim_{k \rightarrow \infty, k \in \mathcal{K}} \varphi_{c, \varepsilon_k}(x^k) \leq \lim_{k \rightarrow \infty, k \in \mathcal{K}} \varphi_{c, \varepsilon_k}(x) = \varphi_c(x),$$

para todo  $x \in \mathbb{R}^n$ . Portanto,  $\bar{x}$  é uma solução para  $(P_1)$ . ■

**Definição 3.8.** [39, Definição 2] Um ponto  $x \in \mathbb{R}^n$  é chamado  $\varepsilon$ -viável para o problema (1) se satisfaz  $g_i(x) \leq \varepsilon$ , para todo  $i = 1, \dots, m$ .

O último resultado desta seção generaliza [20, Teorema 2.4.] e nos mostra que um minimizador de  $(SP_\varepsilon)$  que é  $\varepsilon$ -viável para (1) está também próximo de ser uma solução para o problema (1).

**Teorema 3.9.** Para algum  $c > 0$  fixado, suponha que o ponto  $x^*$  é uma solução do problema  $(P_1)$  e  $\bar{x}_\varepsilon$  é uma solução do problema  $(SP_\varepsilon)$ . Se  $x^* \in X$  (portanto  $x^*$  é uma solução de (1)) e  $\bar{x}_\varepsilon$  é  $\varepsilon$ -viável para (1), então  $\bar{x}_\varepsilon$  é uma solução aproximada para o problema (1), no sentido de que

$$-cm\varepsilon(a+b) \leq f(x^*) - f(\bar{x}_\varepsilon) \leq cm\varepsilon(a+b+1). \quad (3.5)$$

*Prova.* Note que  $p(g_i(x^*)) = 0$  para todo  $i = 1, \dots, m$ , pois  $x^* \in X$ . Logo, do Teorema 3.6, temos que

$$-cmb\varepsilon \leq \varphi_c(x^*) - \varphi_{c, \varepsilon}(\bar{x}_\varepsilon) = f(x^*) - \left( f(\bar{x}_\varepsilon) + c \sum_{i=1}^m \eta(g_i(\bar{x}_\varepsilon), \varepsilon) \right) \leq cma\varepsilon.$$

Rearranjando a última expressão, obtemos

$$-cmb\varepsilon + c \sum_{i=1}^m \eta(g_i(\bar{x}_\varepsilon), \varepsilon) \leq f(x^*) - f(\bar{x}_\varepsilon) \leq c \sum_{i=1}^m \eta(g_i(\bar{x}_\varepsilon), \varepsilon) + cma\varepsilon. \quad (3.6)$$

Note que do item (ii) da Definição 3.2 temos que para  $i = 1, \dots, m$ ,

$$b\varepsilon + p(g_i(\bar{x}_\varepsilon)) \geq \eta(g_i(\bar{x}_\varepsilon), \varepsilon) \geq p(g_i(\bar{x}_\varepsilon)) - a\varepsilon.$$

Por outro lado,  $\bar{x}_\varepsilon$  é  $\varepsilon$ -viável para o problema (1), ou seja,  $g_i(\bar{x}_\varepsilon) \leq \varepsilon$ , conseqüentemente  $0 \leq p(g_i(\bar{x}_\varepsilon)) \leq \varepsilon$ , para todo  $i = 1, \dots, m$ . Disso e da Definição 3.2 segue que

$$-a\varepsilon \leq \eta(g_i(\bar{x}_\varepsilon), \varepsilon) \leq (b+1)\varepsilon,$$

para todo  $i = 1, \dots, m$ . Logo,

$$-cma\varepsilon \leq c \sum_{i=1}^m \eta(g_i(\bar{x}_\varepsilon), \varepsilon) \leq cm(b+1)\varepsilon. \quad (3.7)$$

Juntando (3.6) com (3.7) obtemos a relação (3.5) o que completa a demonstração. ■

**Observação 3.10.** *Algumas implicações e observações acerca dos resultados apresentados nessa seção.*

- (i) *Em muitos casos melhores limitantes inferiores e superiores para a relação (3.5) em que  $0 < t \leq \varepsilon$  podem ser obtidos utilizando a definição de  $\eta(\cdot, \varepsilon)$ , como veremos nos exemplos da Seção 3.4.*
- (ii) *Suponha que um parâmetro  $c^*$  é tal que uma solução de  $(P_1)$  seja também viável para o problema (1). Então, o Teorema 3.7 mostra que essa solução pode ser aproximada arbitrariamente por meio do cálculo de minimizadores de  $(SP_\varepsilon)$  em que  $\varepsilon_k = \varepsilon$  e  $\varepsilon_k \rightarrow 0$ .*
- (iii) *Note que o fato da função de penalidade  $\ell_1$  ser exata nos garante que as hipóteses do Teorema 3.9 podem ser cumpridas, visto que, neste caso, é possível obter  $x^* \in X$  minimizador de  $(P_1)$  o qual é também uma solução para o problema original. Além disto, se um ponto  $\bar{x}_\varepsilon$  é  $\varepsilon$ -viável para o problema com restrições, então  $\bar{x}_\varepsilon$  é também uma solução aproximada deste problema.*

Portanto, os resultados desta seção sugerem que podemos aproximar uma solução (1) por soluções de  $(SP_\varepsilon)$ , ao considerarmos sequências

$$(c_k) \subset \mathbb{R}_+ \text{ crescente e } (\varepsilon_k) \subset \mathbb{R}_+ \text{ decrescente}$$

Isto é o que da origem ao método de penalidade suavizado que discutimos na próxima seção.

### 3.3 Algoritmo de penalidade suavizado

Consideramos nessa seção um algoritmo para encontrar minimizadores de problemas com restrições de desigualdade, em que a cada iteração é calculado um minimizador da função de penalidade suavizada. Antes de formalizarmos esse algoritmo, trazemos uma hipótese comum no contexto de métodos de penalidade suavizados e fazemos alguns comentários sobre tal hipótese.

**Hipótese 4.** *Existem  $\bar{c} > 0$  e  $\bar{\varepsilon} > 0$  tais que para quaisquer  $c \in [\bar{c}, +\infty)$  e  $\varepsilon \in (0, \bar{\varepsilon}]$ ,*

$$\operatorname{argmin}_{x \in \mathbb{R}^n} \{\varphi_{c,\varepsilon}(x)\} \neq \emptyset.$$

Primeiramente, observe que a hipótese acima tem como objetivo assegurar a boa definição do algoritmo, pois esse exige o cálculo de minimizadores da função  $\varphi_{c,\varepsilon}(\cdot)$  a

cada iteração. Assim, precisamos que  $\varphi_{c,\varepsilon}(\cdot)$  seja limitada inferiormente para todo  $c \geq \bar{c}$  e  $\varepsilon \geq \bar{\varepsilon} > 0$ . O resultado seguinte mostra que para isso ocorrer é suficiente que existam  $\bar{c} > 0$  e  $\bar{\varepsilon} > 0$  tais que  $\varphi_{\bar{c},\bar{\varepsilon}}(\cdot)$  seja limitada.

**Proposição 3.11.** *Se  $\varphi_{\bar{c},\bar{\varepsilon}}(\cdot)$  é limitada inferiormente, então para quaisquer  $c \geq \bar{c}$  e  $\varepsilon \leq \bar{\varepsilon}$ ,  $\varphi_{c,\varepsilon}(\cdot)$  é limitada inferiormente.*

*Prova.* Seja  $x \in \mathbb{R}^n$ . Pelo Lema 3.4 temos que

$$\begin{aligned}\varphi_{\bar{c},\bar{\varepsilon}}(x) &\leq \bar{c} m b \bar{\varepsilon} + \varphi_{\bar{c}}(x) \\ &\leq \bar{c} m b \bar{\varepsilon} + \varphi_c(x) \\ &\leq \bar{c} m b \bar{\varepsilon} + c m a \varepsilon + \varphi_{c,\varepsilon}(x).\end{aligned}$$

Donde segue que,

$$\inf\{\varphi_{\bar{c},\bar{\varepsilon}}(x) \mid x \in \mathbb{R}^n\} - \alpha \leq \inf\{\varphi_{c,\varepsilon}(x) \mid x \in \mathbb{R}^n\},$$

onde  $\alpha = \bar{c} m b \bar{\varepsilon} + c m a \varepsilon$ . Logo, a limitação inferior de  $\varphi_{c,\varepsilon}(\cdot)$  segue da limitação de  $\varphi_{\bar{c},\bar{\varepsilon}}(\cdot)$ . ■

A seguir, apresentamos o algoritmo de penalidade suavizado.

**Algoritmo 3.1.** *Algoritmo de penalidade suavizado*

DADOS:  $\varepsilon_0 > 0, c_0 > 0, \beta > 1, 0 < \gamma < 1$

FAÇA  $k = 0$

REPITA

Calcule  $x^k \in \operatorname{argmin}\{\varphi_{c_k,\varepsilon_k}(x) \mid x \in \mathbb{R}^n\}$

SE  $g_i(x^k) \leq 0$  para todo  $i = 1, \dots, m$

pare com  $x^k$  uma solução aproximada de (1)

SENÃO

$c_{k+1} = \beta c_k, \varepsilon_{k+1} = \gamma \varepsilon_k$

$k = k + 1$

**Observação 3.12.** *Algumas observações acerca do Algoritmo 3.1.*

(i) Para calcular  $x^k$  a cada iteração, podemos utilizar algoritmos suaves de minimização irrestrita. Alguns autores [20, 35, 38, 39] sugerem o uso de métodos quase-Newton usando a cada nova iteração o iterando anterior como ponto inicial.

(ii) Nos resultados seguintes, a suposição feita em [20] será importante: a sequência  $(x^k)$  é gerada pelo Algoritmo 3.1 com o parâmetros  $\gamma$  e  $\beta$  satisfazendo

$$\gamma\beta < 1. \tag{3.8}$$

Note que, neste caso, se o algoritmo não atinge o critério de parada, então

$$\varepsilon_{k+1}c_{k+1} = c_0\varepsilon_0(\gamma\beta)^k$$

e conseqüentemente,  $\varepsilon_{k+1}c_{k+1} \rightarrow 0$  quando  $k \rightarrow \infty$ , pela suposição (3.8).

O resultado abaixo mostra que o Algoritmo 3.1 é globalmente convergente.

**Teorema 3.13.** *Suponha que o Algoritmo 3.1 com  $\gamma\beta < 1$  gere uma seqüência infinita  $(x^k) \subset \mathbb{R}^n$ . Se  $x^*$  é um ponto de acumulação de  $(x^k)$ , então  $x^*$  é uma solução do problema (1).*

Baseado em resultados da literatura, por exemplo [20, Teorema 3.1], apresentamos um resultado mais geral do que o teorema acima. Esse consiste do cálculo de minimizadores aproximados da função  $\varphi_{c,\varepsilon}(\cdot)$ , isto é,

$$\varphi_{c_k,\varepsilon_k}(x^k) \leq \varphi_{c_k,\varepsilon_k}(x) + l_k, \quad \forall x \in \mathbb{R}^n.$$

Assim, se  $l_k \rightarrow l \geq 0$  então um ponto de acumulação da seqüência  $(x^k)$  é também uma solução aproximada do problema original. Além disso, se  $l_k \rightarrow 0$ , então esse ponto de acumulação será uma solução do problema (1).

**Teorema 3.14.** *Sejam  $(c_k), (\varepsilon_k) \subset \mathbb{R}_+$  e  $(l_k) \subset \mathbb{R}_+$ ,  $(x^k) \subset \mathbb{R}^n$  seqüências satisfazendo*

- (i)  $c_{k+1} > c_k$  e  $\varepsilon_{k+1} < \varepsilon_k \quad \forall k \in \mathbb{N}$ ;
- (ii)  $c_k \varepsilon_k \rightarrow 0$  e  $l_k \rightarrow l$ ;
- (iii)  $\varphi_{c_k,\varepsilon_k}(x^k) \leq \inf\{\varphi_{c_k,\varepsilon_k}(x) \mid x \in \mathbb{R}^n\} + l_k, \quad \forall k \in \mathbb{N}$ .

Se  $x^*$  é um ponto de acumulação de  $(x^k)$ , então  $x^* \in X$  e satisfaz

$$f(x^*) \leq f(x) + l, \quad \forall x \in X. \quad (3.9)$$

Além disso, se  $l = 0$  então  $x^*$  é uma solução do problema (1).

*Prova.* Como  $x^*$  é um ponto de acumulação de  $(x^k)$ , existe um conjunto infinito  $\mathcal{K} \subset \mathbb{N}$  tal que  $x^k \xrightarrow{\mathcal{K}} x^*$ . Vamos mostrar que  $x^*$  é viável para o problema (1). Para isto, suponha por contradição que  $x^* \notin X$ . Então existem  $i_0 \in \{1, \dots, m\}$  e  $\omega > 0$  tais que

$$g_{i_0}(x^*) \geq \omega > 0.$$

Como  $\varepsilon_k \rightarrow 0$  e  $g_{i_0}$  é contínua, existe um conjunto infinito  $\mathcal{K}_1 \subset \mathcal{K}$  tal que

$$g_{i_0}(x^k) \geq \omega > \varepsilon_k \quad (3.10)$$

para todo  $k \in \mathcal{K}_1$ . Assim, da relação (3.10) e do Lema 3.4 segue que

$$\begin{aligned}
f(x^k) + c_k(\omega - \varepsilon_k m a) &\leq f(x^k) + c_k(g_{i_0}(x^k) - \varepsilon_k m a) \\
&= f(x^k) + c_k(p(g_{i_0}(x^k)) - \varepsilon_k m a) \\
&\leq f(x^k) + c_k \sum_{i=1}^m p(g_i(x^k)) - c_k \varepsilon_k m a \\
&= \varphi_{c_k}(x^k) - c_k \varepsilon_k m a \\
&\leq \varphi_{c_k, \varepsilon_k}(x^k),
\end{aligned} \tag{3.11}$$

em que a primeira igualdade e a segunda desigualdade são devidas a  $g_{i_0}(x^k) = p(g_{i_0}(x^k))$  e  $p(\cdot) \geq 0$ , respectivamente.

Por outro lado, para todo  $k \in \mathbb{N}$ ,  $x^k$  é um minimizador aproximado de  $\varphi_{c_k, \varepsilon_k}(\cdot)$ , o que juntando com o Lema 3.4 resulta em

$$\varphi_{c_k, \varepsilon_k}(x^k) - l_k \leq \varphi_{c_k, \varepsilon_k}(x) \leq \varphi_{c_k}(x) + c_k \varepsilon_k m b, \quad \forall x \in \mathbb{R}^n.$$

Em particular, para  $x \in X$  temos  $\varphi_{c_k}(x) = f(x)$ , portanto

$$\varphi_{c_k, \varepsilon_k}(x^k) - l_k \leq f(x) + c_k \varepsilon_k m b, \quad \forall x \in X, k \in \mathbb{N}. \tag{3.12}$$

Agora, juntando (3.11) e (3.12) obtemos

$$f(x^k) + c_k(\omega - \varepsilon_k m a) \leq l_k + f(x) + c_k \varepsilon_k m b, \quad \forall k \in \mathcal{K}_1,$$

consequentemente, para quaisquer  $k \in \mathcal{K}_1$  e  $x \in X$  temos

$$c_k \omega \leq l_k + f(x) - f(x^k) + c_k \varepsilon_k m(a + b),$$

Dado que  $l_k \rightarrow l$ ,  $c_k \varepsilon_k \rightarrow 0$  e  $f(x^k) \xrightarrow{\mathcal{K}_1} f(x^*)$ , temos que o lado direito da relação acima é limitado, enquanto que o lado esquerdo é ilimitado superiormente, pois  $c_k \omega \xrightarrow{\mathcal{K}_1} +\infty$ , enquanto que o lado esquerdo é ilimitado superiormente, o que gera uma contradição. Portanto,  $x^* \in X$ .

Resta mostrar que  $x^*$  satisfaz (3.9). Do Lema 3.4 temos que

$$f(x^k) - c_k \varepsilon_k m a \leq \varphi_{c_k}(x^k) - c_k \varepsilon_k m a \leq \varphi_{c_k, \varepsilon_k}(x^k), \quad \forall k \in \mathcal{K}. \tag{3.13}$$

Usando a definição de  $x^k$  e o Lema 3.4 novamente, obtemos que

$$\varphi_{c_k, \varepsilon_k}(x^k) - l_k \leq \varphi_{c_k, \varepsilon_k}(x) \leq \varphi_{c_k}(x) + c_k \varepsilon_k m b = f(x) + c_k \varepsilon_k m b, \tag{3.14}$$

para todo  $x \in X$  e  $k \in \mathcal{K}$ . Agora, juntando (3.13), (3.14) e o fato que  $l_k \rightarrow l, c_k \varepsilon_k \xrightarrow{\mathcal{K}} 0$ , segue que

$$f(x^*) = \lim_{k \rightarrow \infty, k \in \mathcal{K}} [f(x^k) - c_k \varepsilon_k m a] \leq \lim_{k \rightarrow \infty, k \in \mathcal{K}} [f(x) + c_k \varepsilon_k m b + l_k] = f(x) + l,$$

para todo  $x \in X$ . Portanto  $x^*$  é uma solução aproximada de (1). Em particular, se  $l = 0$  então  $x^*$  é uma solução do problema original. ■

Note que, no teorema acima, se  $l_k = 0$  para todo  $k \in \mathbb{N}$  e as sequências  $c_k$  e  $\varepsilon_k$  são atualizadas como no Algoritmo 3.1, então recuperamos o Teorema 3.13. Por outro lado, se  $l_k > 0$ , o cálculo de minimizadores aproximados do problema de penalidade suavizado pode resultar em uma solução aproximada do problema original. Ainda mais, se a imprecisão do cálculo dos minimizadores globais é diminuída a cada iteração, de modo que  $l_k \rightarrow 0$ , um ponto de acumulação gerado por esse algoritmo é uma solução do problema original.

Em particular, temos que se a função objetivo do problema (1) é coerciva, a existência de um ponto de acumulação pode ser garantida se  $(\varphi_{c_k, \varepsilon_k}(x^k))$  é limitada como mostra o próximo resultado.

**Teorema 3.15.** *Suponha que  $f$  é uma função coerciva. Seja  $(x^k) \subset \mathbb{R}^n$  uma sequência gerada satisfazendo as hipóteses (i) – (iii) do Teorema 3.14. Se a sequência  $(\varphi_{c_k, \varepsilon_k}(x^k))$  é limitada então  $(x^k)$  é limitada e qualquer ponto de acumulação  $x^*$  é uma solução aproximada de (1). Em particular, se  $l_k \rightarrow 0$  então  $x^*$  é uma solução de (1).*

*Prova.* Usando que  $f(x^k) \leq f(x^k) + c_k \sum_{i=1}^n p(g_i(x^k)) = \varphi_{c_k}(x^k)$  para todo  $k \in \mathbb{N}$  e o Lema 3.4, temos

$$f(x^k) - c_k m a \varepsilon_k \leq \varphi_{c_k}(x^k) - c_k m a \varepsilon_k \leq \varphi_{c_k, \varepsilon_k}(x^k), \quad (3.15)$$

para todo  $k \in \mathbb{N}$ .

Por hipótese a sequência  $(\varphi_{c_k, \varepsilon_k}(x^k))$  é limitada, assim existe  $L > 0$  tal que  $L > \varphi_{c_k, \varepsilon_k}(x^k)$  para todo  $k \in \mathbb{N}$ . Juntando isto com (3.15), obtemos

$$L > \varphi_{c_k, \varepsilon_k}(x^k) \geq f(x^k) - c_k m a \varepsilon_k,$$

para todo  $k \in \mathbb{N}$ . Donde segue que  $(x^k)$  é limitada, pois caso contrário, poderíamos obter um conjunto infinito  $\mathcal{K} \subset \mathbb{N}$  de modo que  $\|x^k\| \xrightarrow{\mathcal{K}} \infty$ , o que pela coercividade de  $f$  resultaria

$$f(x^k) - c_k m a \varepsilon_k \xrightarrow{\mathcal{K}} +\infty$$

pois  $c_k \varepsilon_k \xrightarrow{\mathcal{K}} 0$ . Portanto  $(x^k)$  é limitada e pelo Teorema 3.14 qualquer ponto de acumulação  $x^*$  da sequência  $(x^k)$  é uma solução aproximada do problema (1). Em particular, se  $l_k \rightarrow 0$  então  $x^*$  é uma solução (1). ■

Note que ao considerarmos a hipótese de coercividade da função objetivo no problema de penalidade

$$\text{minimizar}\{f(x) + cP(x) \mid x \in \mathbb{R}^n\}$$

obtemos que esse problema sempre é limitado inferiormente para qualquer  $c > 0$ . Consequentemente, as iterações do Algoritmo 1.2 estariam bem definidas para qualquer função de penalidade. Assim, um ponto de acumulação de uma sequência gerada por esse algoritmo seria, pelo menos, uma solução aproximada do problema original (Teorema 1.6).

### 3.4 Exemplos de aproximação suave

A seguir apresentamos exemplos de suavizações da literatura que satisfazem a Definição 3.2. Em especial, os resultados vistos nas Seções 3.1 – 3.3 são válidos para essas funções.

#### Função 1 ([20])

Seja  $\eta_1 : \mathbb{R} \times \mathbb{R}_+ \rightarrow \mathbb{R}$  definida como

$$\eta_1(t, \varepsilon) = \begin{cases} \frac{3}{2}\varepsilon e^{t/\varepsilon} - 2\varepsilon & t \leq 0, \\ t - \frac{1}{2}\varepsilon e^{-t/\varepsilon} & t > 0. \end{cases} \quad (3.16)$$

Note que esta função assume valores negativos para  $t < 0$  e, como pode ser visto na Figura 3.1, quanto menor o valor do parâmetro de aproximação, menor é o erro de aproximação entre essas funções. Vejamos que a função  $\eta_1$  é de fato uma função de aproximação suave.

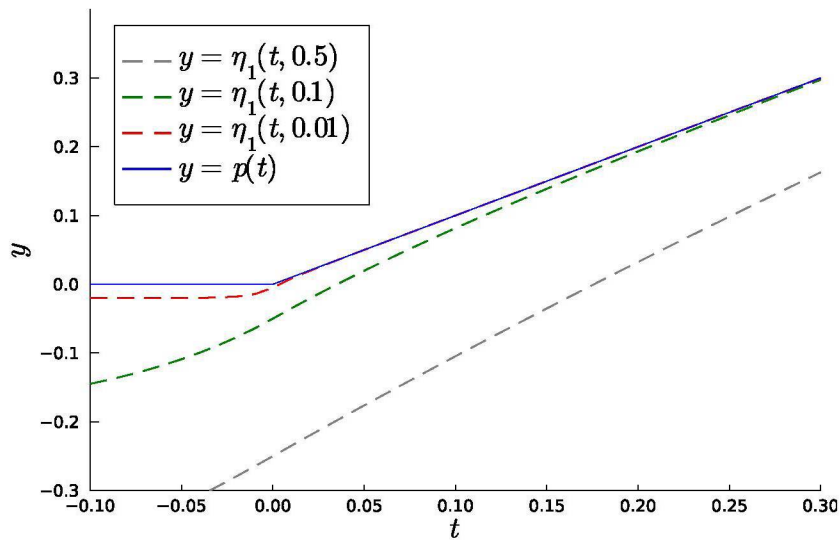


Figura 3.1: Comportamento da função  $\eta_1(\cdot, \varepsilon)$  e  $p(\cdot)$ .

Para cada  $\varepsilon > 0$  fixado, temos que

$$E_1(t) := p(t) - \eta_1(t, \varepsilon) = \begin{cases} 2\varepsilon - \frac{3}{2}\varepsilon e^{t/\varepsilon} & t \leq 0, \\ \frac{1}{2}\varepsilon e^{-t/\varepsilon} & t > 0. \end{cases}$$

A função  $E_1$  é decrescente no intervalo  $(-\infty, 0)$ , pois  $E_1(t) = -\frac{3}{2}\varepsilon e^{t/\varepsilon} < 0$ , para  $t < 0$ . Analogamente, vemos que  $E_1$  é decrescente em  $(0, +\infty)$ . Além disto,

$$\lim_{t \rightarrow 0^+} E_1(t) = \frac{1}{2}\varepsilon \quad \text{e} \quad \lim_{t \rightarrow -\infty} E_1(t) = 2\varepsilon,$$

o que resulta em

$$\frac{1}{2} \leq E_1(t) \leq 2\varepsilon, \text{ se } t < 0 \quad \text{e} \quad 0 \leq E_1(t) \leq \frac{1}{2}\varepsilon, \text{ se } t > 0.$$

Portanto, para todo  $t \in \mathbb{R}$ ,

$$0 \leq p(t) - \eta_1(t, \varepsilon) \leq 2\varepsilon.$$

Logo,  $\eta_1$  satisfaz o item (ii) da Definição 3.2 com  $b = 0$  e  $a = 2$ . Para o item (i), vamos calcular o limite fundamental dessa função pela esquerda e direita de zero. Temos que

$$\lim_{t \rightarrow 0^+} \frac{t - 1/2\varepsilon e^{-t/\varepsilon} + 1/2\varepsilon}{t} = 1 + \lim_{t \rightarrow 0^+} \frac{1}{2}\varepsilon \left( \frac{1 - \varepsilon e^{t/\varepsilon}}{t} \right) = 1 + \frac{1}{2} \ln(e) = \frac{3}{2}$$

e

$$\lim_{t \rightarrow 0^-} \frac{3/2\varepsilon e^{t/\varepsilon} - 2\varepsilon + 1/2\varepsilon}{t} = \lim_{t \rightarrow 0^-} \frac{3}{2}\varepsilon \left( \frac{e^{t/\varepsilon} - 1}{t} \right) = \frac{3}{2}.$$

Consequentemente,  $\eta_1$  é derivável no ponto zero, além disto,

$$\frac{d}{dt}(\eta_1(t, \varepsilon)) = \begin{cases} \frac{3}{2}\varepsilon e^{t/\varepsilon} & t \leq 0, \\ 1 + \frac{1}{2}\varepsilon e^{-t/\varepsilon} & t > 0, \end{cases}$$

assim  $\eta_1$  satisfaz as condições (i) e (ii). Portanto,  $\eta_1$  é uma função de aproximação suave para  $p(\cdot)$ .

## Função 2 ([7, 19])

Outra função que consideramos  $\eta_2 : \mathbb{R} \times \mathbb{R}_+ \rightarrow \mathbb{R}$  é definida como

$$\eta_2(t, \varepsilon) = \begin{cases} 0 & t \leq 0, \\ \frac{t^2}{2\varepsilon} & 0 < t \leq \varepsilon, \\ t - \frac{\varepsilon}{2} & t > \varepsilon. \end{cases}$$

Esta é utilizada mais recentemente em [7] para derivar um algoritmo de subgradiente primal-dual. Já em [19] uma variante dessa função é utilizada para desenvolver um algoritmo de penalidade suavizado, o qual aproxima o método de penalidade exata  $\ell_p$  com  $p < 1$ .

Note que, diferentemente de  $\eta_1$  que assume valores negativos, a função  $\eta_2$  é não negativa em  $\mathbb{R}$ , isto pode ser observado na Figura 3.2 que ilustra a aproximação da função  $\eta(\cdot, \varepsilon)$  de  $p(\cdot)$ , a medida que  $\varepsilon$  se aproxima de zero.

Vamos estimar analiticamente o erro de aproximação. Considere  $\varepsilon > 0$  fixado, então

$$E_2(t) := p(t) - \eta_2(t, \varepsilon) = \begin{cases} 0 & t \leq 0, \\ t - \frac{t^2}{2\varepsilon} & 0 < t \leq \varepsilon \\ \frac{\varepsilon}{2} & t > \varepsilon. \end{cases}$$

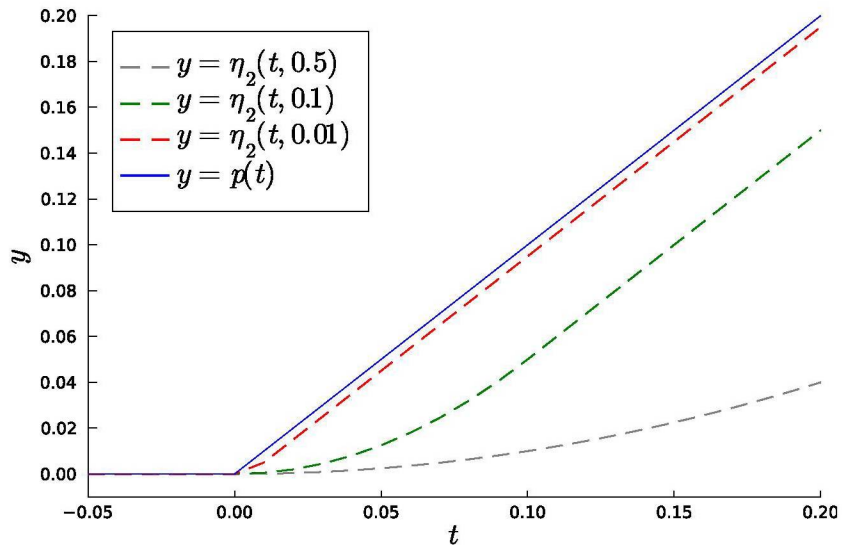


Figura 3.2: Comportamento de  $\eta_2(\cdot, \varepsilon)$  e  $p(\cdot)$ .

Temos que  $E_2$  é crescente em  $(0, \varepsilon)$ , visto que  $\frac{d}{dt}E_2(t) = 1 - t/\varepsilon \geq 0$  neste intervalo. Donde segue que

$$0 \leq E_2(t) \leq \frac{\varepsilon}{2},$$

para  $t \in (-\infty, 0) \cup (\varepsilon, +\infty)$ . Portanto

$$0 \leq p(t) - \eta_2(t, \varepsilon) \leq \frac{\varepsilon}{2}. \quad (3.17)$$

Logo  $\eta_2$  satisfaz (3.1) com  $b = 0$  e  $a = 1/2$ . De forma análoga a  $\eta_1$ , podemos mostrar que  $\eta_2(\cdot, \varepsilon)$  é suave.

### Função 3 ([39])

Seja  $\eta_3 : \mathbb{R} \times \mathbb{R}_+ \rightarrow \mathbb{R}_+$  dada por

$$\eta_3(t, \varepsilon) = \begin{cases} 0 & t < 0, \\ \frac{t^3}{6\varepsilon^2} & 0 \leq t < \varepsilon, \\ t + \frac{\varepsilon^2}{2t} - \frac{4\varepsilon}{3} & \varepsilon \leq t. \end{cases} \quad (3.18)$$

A Figura 3.3 mostra que quanto menor o valor de  $\varepsilon$ , menor é também o erro de aproximação. Dado  $\varepsilon > 0$  temos que

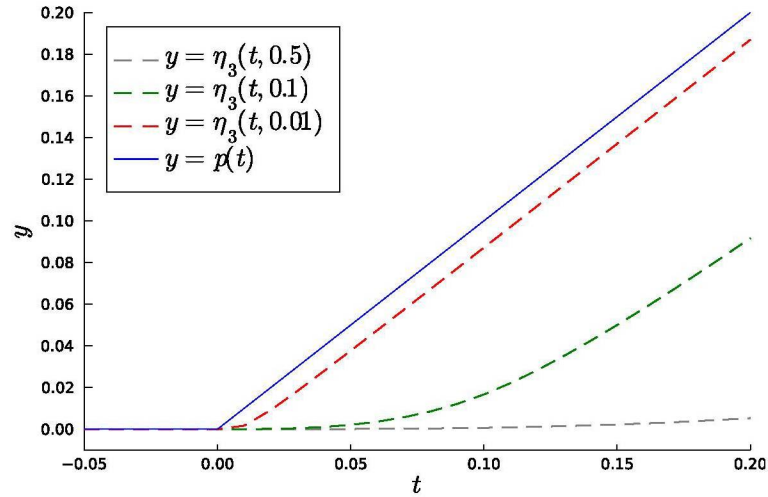


Figura 3.3: Comportamento de  $\eta_3(\cdot, \varepsilon)$  e  $p(\cdot)$ .

$$E_3(t) := p(t) - \eta_3(t, \varepsilon) = \begin{cases} 0 & t < 0, \\ t - \frac{t^3}{6\varepsilon^2} & 0 \leq t < \varepsilon, \\ \frac{4\varepsilon}{3} - \frac{\varepsilon^2}{2t} & \varepsilon \leq t. \end{cases}$$

Como a função  $E_3$  é crescente para  $t \in [0, +\infty)$ , segue que

$$0 \leq p(t) - \eta_3(t, \varepsilon) \leq \frac{4}{3}\varepsilon.$$

Logo,  $\eta_3$  satisfaz a condição (ii) da Definição 3.2. Ainda, podemos mostrar que  $\eta_3$  é suave e que a sua derivada é dada por

$$\eta_3'(t, \varepsilon) = \begin{cases} 0 & t < 0, \\ \frac{t^2}{2\varepsilon^2} & 0 \leq t < \varepsilon, \\ 1 - \frac{\varepsilon^2}{2t^2} & \varepsilon < t. \end{cases}$$

Portanto,  $\eta_3$  é uma função de aproximação suave. Ainda mais, essa função é duas vezes diferenciável com a segunda derivada dada por

$$\eta_3''(t) = \begin{cases} 0 & t < 0, \\ \frac{t}{\varepsilon^2} & 0 \leq t < \varepsilon, \\ \frac{\varepsilon^2}{t^3} & \varepsilon \leq t. \end{cases}$$

Para  $x \in \mathbb{R}^n$  tal que  $g_i(x) \leq \varepsilon$ , temos que a matriz Hessiana da função suavizada é

$$\begin{aligned} \nabla^2 \varphi_{c,\varepsilon}(x) &= \nabla^2 f(x) + c \sum_{i=1}^m \eta_3''(g_i(x), \varepsilon) \nabla g_i(x) \nabla g_i(x)^T + c \sum_{i=1}^m \eta_3'(g_i(x), \varepsilon) \nabla^2 g_i(x) \\ &= \nabla^2 f(x) + c \sum_{i \in I_+} \frac{g_i(x)}{\varepsilon^2} \nabla g_i(x) \nabla g_i(x)^T + \frac{g_i(x)^2}{2\varepsilon^2} \nabla^2 g_i(x) \\ &= L(x, \mu) + \sum_{i \in I_+} \frac{2\mu_i}{g_i(x)} \nabla g_i(x) \nabla g_i(x)^T, \end{aligned}$$

onde  $\mu_i = \frac{c([g_i(x)]_+)^2}{2\varepsilon^2}$  e  $I_+ = \{i \in \{1, \dots, m\} \mid g_i(x) > 0\}$ .

Observe que, para  $\mu_i$  e  $\mu_i/g_i(x)$ ,  $i \in I_+$  serem limitadas, digamos por  $\omega > 0$ , é necessário que  $g_i(x) \leq (2\omega\varepsilon^2)/c$ . Mas, tal condição não pode ser garantida em geral. Além disso, a ilimitação desses termos pode levar ao mal condicionamento da Hessiana. De fato, considerando o problema de otimização convexo do Exemplo 2.3 temos que a função penalizada suavizada é

$$\varphi_{c,\varepsilon}(x) = \frac{1}{2} (x_1^2 + (x_2 - 1)^2) + c\eta_3(1 - x_1, \varepsilon).$$

Em pontos  $x$  que são  $\varepsilon$ -viáveis para esse problema, temos

$$\nabla^2 \varphi_{c,\varepsilon}(x) = \begin{bmatrix} 1 + \frac{c}{\varepsilon^2}(x_1 + 1) & 0 \\ 0 & 1 \end{bmatrix}.$$

Logo, o número de condicionamento da matriz acima é

$$\chi_{c,\varepsilon}(x) = 1 + \frac{c}{\varepsilon^2}(x_1 + 1).$$

Se para algum  $\omega > 0$  temos que  $(x_1^k + 1) \leq \omega \frac{\varepsilon_k^2}{c_k}, \forall k \in \mathbb{N}$ , então a matriz acima é sempre bem condicionada. Caso contrário,  $\nabla^2 \varphi_{c_k, \varepsilon_k}(x^k)$  torna-se arbitrariamente mal condicionada, a medida que  $\varepsilon_k$  e  $c_k^{-1}$  tendem a zero.

#### Função 4 ([38])

A última função de suavização que trazemos neste trabalho é denotada por  $\eta_4 : \mathbb{R} \times \mathbb{R}_+ \rightarrow \mathbb{R}$  e definida utilizando o número de restrições  $g_i$  do problema original, dado por  $m$ , e o parâmetro de penalidade  $c$  como segue:

$$\eta_4(t, \varepsilon) = \begin{cases} 0 & t < 0, \\ \frac{m^3 c^3 t^4}{10 \varepsilon^3} & 0 \leq t < \frac{\varepsilon}{mc}, \\ t + \frac{3 \varepsilon^2}{5 m^2 c^2 t} - \frac{3 \varepsilon}{2 mc} & t \geq \frac{\varepsilon}{mc}. \end{cases}$$

A Figura 3.4 mostra que o parâmetro de penalidade  $c$  também comporta-se como um parâmetro de suavização para a função definida acima, já que conforme  $c$  aumenta, o gráfico de  $\eta_4(t, \varepsilon)$  aproxima-se mais de  $p(t)$ . O mesmo efeito pode ser verificado se o número de restrições  $m$  é muito grande, ou ainda, se diminuirmos  $\varepsilon$  como ilustrado na Figura 3.5.

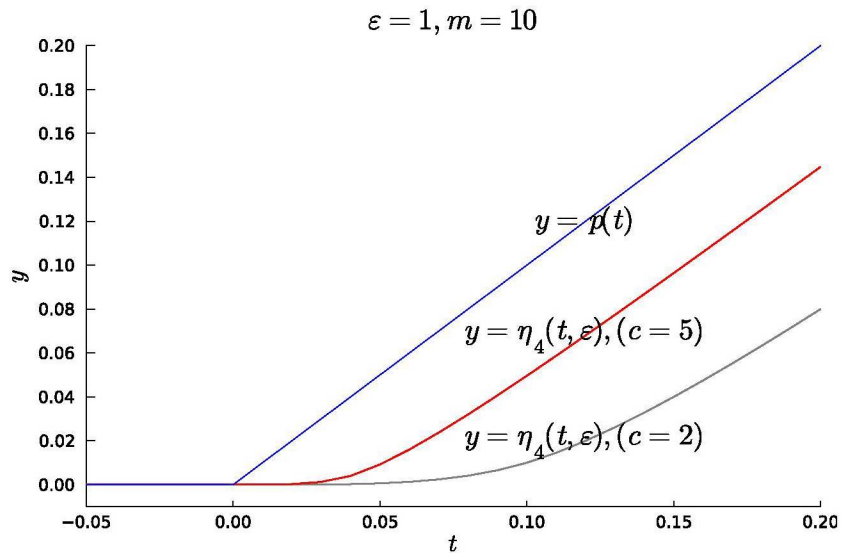


Figura 3.4: Comportamento de  $\eta_4(\cdot, \varepsilon)$  em relação a  $c$ .

Vamos estimar um erro de aproximação entre  $p\eta_4$ . Para  $\varepsilon$  e  $c$  fixados, temos

$$E_4(t) = p(t) - \eta_4(t, \varepsilon) = \begin{cases} 0 & t < 0, \\ t - \frac{m^3 c^3 t^4}{10\varepsilon^3} & 0 \leq t < \frac{\varepsilon}{mc}, \\ \frac{3\varepsilon}{2mc} - \frac{3\varepsilon^2}{5m^2 c^2 t} & t \geq \frac{\varepsilon}{mc}. \end{cases}$$

Se  $0 \leq t \leq \frac{\varepsilon}{mc}$  então a derivada de  $E_4$  é positiva, uma vez que

$$\frac{d}{dt}E_4(t) = 1 - \frac{2}{5} \left(\frac{mc}{\varepsilon}\right)^3 t^3 \geq 1 - \frac{2}{5} > 0.$$

Consequentemente,  $0 \leq E_4(t) \leq (9/10)\varepsilon mc$ , e como

$$\frac{d}{dt}E_4(t) = \frac{3\varepsilon^2}{5m^2 c^2 t^2} > 0, \quad t \geq \frac{\varepsilon}{mc},$$

segue que  $9/10 \leq E_4(t) \leq 3\varepsilon/2mc$ . Essa análise mostra que

$$0 \leq p(t) - \eta_4(t, \varepsilon) \leq \frac{3\varepsilon}{2mc}.$$

Portanto,  $\eta_4$  satisfaz a Definição 3.2 para  $a = 3/2\varepsilon mc$  e  $b = 0$ .

Ainda é possível mostrar que a primeira e segunda derivada de  $\eta_3$  são dadas, reespecivamente, por

$$\eta_4'(t, \varepsilon) = \begin{cases} 0 & t \leq 0, \\ \frac{2m^3 c^3 t^3}{5\varepsilon^3} & 0 \leq t \leq \frac{\varepsilon}{mc}, \\ 1 - \frac{3\varepsilon^2}{5m^2 c^2 t^2} & t \geq \frac{\varepsilon}{mc}, \end{cases} \quad \text{e} \quad \eta_4''(t) = \begin{cases} 0 & t < 0, \\ \frac{6m^3 c^3 t^2}{5\varepsilon^3} & 0 \leq t \leq \frac{\varepsilon}{mc}, \\ \frac{6\varepsilon^2}{5m^2 c^2 t^3} & t \geq \frac{\varepsilon}{mc}. \end{cases}$$

Assim,

$$\nabla \varphi_{c,\varepsilon}(x) = \nabla f(x) + c \sum_{i=1}^m \eta_4'(g_i(x), \varepsilon) \nabla g_i(x)$$

e

$$\nabla^2 \varphi_{c,\varepsilon}(x) = \nabla^2 f(x) + c \sum_{i=1}^m \eta_4''(g_i(x), \varepsilon) \nabla g_i(x) \nabla g_i(x)^T + c \sum_{i=1}^m \eta_4'(g_i(x), \varepsilon) \nabla^2 g_i(x).$$

Em particular, se  $g_i(x) \leq \frac{\varepsilon}{mc}$ , para todo  $i \in \{1, \dots, m\}$ , então

$$\nabla^2 \varphi_{c,\varepsilon}(x) = \nabla^2 f(x) + \frac{2c^4 m^3}{5\varepsilon^3} \sum_{i \in I_+} [3g_i(x)^2 \nabla g_i(x) \nabla g_i(x)^T + g_i(x)^3 \nabla^2 g_i(x)],$$

onde  $I_+ = \{i \in \{1, \dots, m\} \mid g_i(x) > 0\}$ . Reescrevendo a fórmula acima, temos

$$\nabla^2 \varphi_{c,\varepsilon}(x) = \nabla_{xx}^2 L(x, \mu) + \sum_{i \in I_+} \frac{3\mu_i}{g_i(x)} \nabla g_i(x) \nabla g_i(x)^T$$

onde  $\mu_i = \frac{2m^3 c^4}{5\varepsilon^3} [g_i(x)^3]_+$ .

Novamente, temos que  $\nabla^2 \varphi_{c,\varepsilon}(\cdot)$  pode tornar-se arbitrariamente mal condicionada, se  $\mu_i$  e  $\mu_i/g_i(x)$  não forem limitadas. Por exemplo, considerando novamente o problema

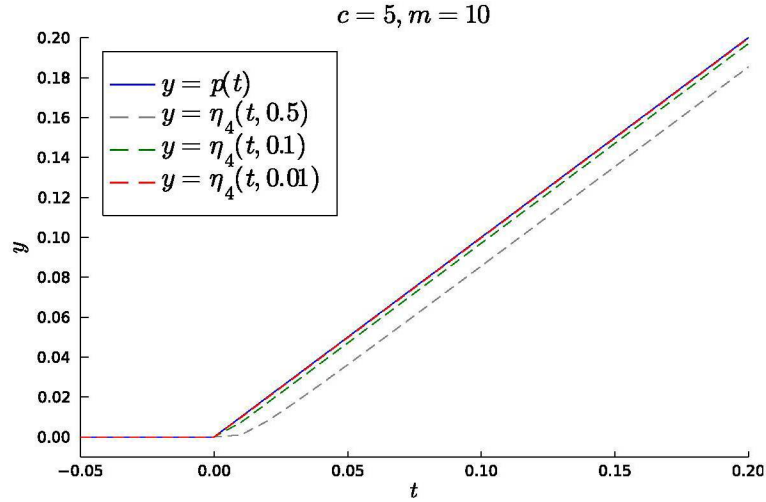


Figura 3.5: Comportamento de  $\eta_4(\cdot, \varepsilon)$  e  $p(\cdot)$ ,

convexo do Exemplo 2.3, vemos que se  $g(x) \leq \varepsilon/mc$  então

$$\nabla^2 \varphi_{c,\varepsilon}(x) = \begin{bmatrix} 1 + \frac{6c^4}{5\varepsilon^3} (1 - x_1)^2 & 0 \\ 0 & 1 \end{bmatrix}.$$

Logo, o número de condicionamento dessa matriz é

$$\chi_{c,\varepsilon} = 1 + \frac{6c^4}{5\varepsilon^3} (1 - x_1)^2.$$

Assim, se existe  $\omega > 0$  tal que  $(1 - x_1^k) \leq \omega(\varepsilon_k^{3/2}/c_k^2)$  para todo  $k \in \mathbb{N}$ , temos que  $\nabla^2 \varphi_{c_k, \varepsilon_k}(x^k)$  é sempre bem condicionada, caso contrário o mal condicionamento é crescente a medida que  $k$  cresce.

Nesta capítulo foram estabelecidas condições para obtermos uma função que aproxima e suaviza  $p(\cdot)$ . A partir disso, determinamos relações entre soluções do problema  $(SP_\varepsilon)$  com soluções de  $(P_1)$  e de (1). Em particular, vimos que uma solução do problema com restrições pode ser aproximada resolvendo o problema  $(SP_\varepsilon)$  (Teorema 3.7). Além disso, um algoritmo para minimizar (1) com uma abordagem de penalidade foi definido e sua convergência global provada, no sentido de que se a sequência gerada pelo algoritmo

possui um ponto de acumulação, esse ponto é uma solução de (1). Por fim, mostramos que diversas funções da literatura satisfazem as condições impostas para ser uma função de aproximação suave.

# Capítulo 4

## Experimentos numéricos

Em [19, 20, 38, 39] são apresentados testes numéricos realizados com os algoritmos de penalidade  $\ell_1, \ell_2$  e algoritmos de suavização, para um conjunto de problemas-teste na linguagem de programação MATLAB. A partir dos resultados obtidos é observado que o algoritmo proposto produz bons resultados para problemas de otimização com restrições. Tendo isto em vista, nesta seção comparamos o desempenho dos algoritmos estudados por meio de *testes numéricos na linguagem de programação Julia*, na resolução de todos os 51 problemas com restrições de desigualdade da coleção CUTEst [11] da forma

$$\begin{aligned} & \text{minimizar } f(x) && (4.1) \\ & \text{sujeito a } g_i(x) \leq 0, \quad i = 1, \dots, m, \\ & x \in \mathbb{R}^n. \end{aligned}$$

com  $f, g_i : \mathbb{R}^n \rightarrow \mathbb{R}$  funções duas vezes diferenciáveis em que  $0 \leq m, n \leq 100$ .

Os problemas foram coletados por meio do pacote CUTEst.jl [28]. Para acessá-los e integrá-los à linguagem de programação Julia, utilizamos o pacote NLPModelsIpopt.jl [29].

### 4.1 Algoritmos e critérios de parada

O critério para declararmos parada dos Algoritmos 1.1 e 3.1 considerados nos Capítulos 1 e 3, respectivamente, dependem unicamente da viabilidade do iterando  $x^{k+1}$ . Nos experimentos numéricos utilizamos o critério de parada de tolerância para a viabilidade igual a  $\epsilon = 10^{-6}$ , isto é,

$$\|[g(x^{k+1})]_+\|_\infty \leq 10^{-6}. \quad (4.2)$$

Se o critério (4.2) é atingido, paramos a execução do algoritmo e declaramos convergência. Neste caso, identificamos esse resultado com a FLAG de saída do algoritmo igual a “V”.

Tal critério de tolerância para a viabilidade também é utilizada nos trabalhos [19, 20, 38, 39]. Além de (4.2), utilizamos um segundo critério de parada para os algoritmos, o qual exige a redução da inviabilidade relativa ao ponto inicial, ou seja,

$$\|[g(x^{k+1})]_+\|_\infty \leq 10^{-6} \|[g(x^0)]_+\|_\infty. \quad (4.3)$$

Se o critério acima é atingido, paramos a execução e atribuímos a FLAG de saída do algoritmo igual a “V”.

Entretanto, os algoritmos podem não atingir o critério (4.2) ou (4.3) e, conseqüentemente, não parar, ou ainda o parâmetro de penalidade pode vir a se tornar muito grande e provocar erros numéricos. Assim, consideramos alguns critérios de interrupção com insucesso dos algoritmos com as respectivas FLAGS de saída: *Limitante superior para o parâmetro de penalidade*,

$$c^{k+1} > 10^{20}, \quad \text{FLAG} = \text{C}; \quad (4.4)$$

*Limitante superior para o tempo de execução em segundos:*

$$\text{max\_tempo} = 600, \quad \text{FLAG} = \text{T}. \quad (4.5)$$

O critério (4.4) tem como objetivo evitar erros numéricos, visto que na linguagem de programação Julia são retornados valores negativos para valores maiores ou iguais a  $10^{26}$ . Enquanto que (4.5) visa impedir que os algoritmos parem por insucesso na resolução dos subproblemas. Mas, se este critério de parada é atingido, isto é devido à dificuldade em calcular um minimizador para os subproblemas. A execução dos algoritmos também é interrompida se algum *erro numérico* é identificado, tais como  $\varepsilon_k < 0$  ou  $x_i^k = NaN$ , para algum  $i = 1, \dots, n$ , ou valores funcionais muito grandes como  $|f(x^{k+1})| > 10^{100}$  ou  $|f(x^{k+1})| = \infty$ . Em quaisquer desses casos é atribuído à FLAG de saída do algoritmo o valor “E”.

Levando em consideração os critérios de parada, o algoritmo de penalidade com as funções  $\ell_1(p = 1)$  e  $\ell_2(p = 2)$  assume a forma descrita abaixo.

**Algoritmo 4.1.** *Algoritmo de penalidade*

DADOS:  $x^0 \in \mathbb{R}^n$ ,  $c_0 > 0$ ,  $\beta > 1$ ,  $\epsilon > 0$ ,  $p \in \{1, 2\}$  e  $\tau \in \{1, \|[g(x^0)]_+\|_\infty\}$

$k = 0$

REPITA

$$\text{CALCULE } x^{k+1} \in \operatorname{argmin} \left\{ f(x) + \frac{c_k}{p} \sum_{i=1}^m \max\{0, g_i(x)\}^p \mid x \in \mathbb{R}^n \right\}$$

SE  $\|[g(x^{k+1})]_+\|_\infty \leq \epsilon\tau$

*pare com sucesso*

SENÃO

defina  $c_{k+1} = \beta c_k$

SE ocorre (4.4) ou (4.5) ou algum erro numérico  
pare com insucesso

SENÃO

$k = k + 1$

O Algoritmo de Suavização 3.1 com as funções  $\eta_j$ ,  $j = 1, 2, 3, 4$ , descritas na Seção 3.4 toma a forma abaixo.

**Algoritmo 4.2.** *Algoritmo de suavização*

DADOS:  $x^0 \in \mathbb{R}^n$ ,  $c_0 > 0$ ,  $\beta > 1$ ,  $\varepsilon_0 > 0$ ,  $0 < \gamma < 1$ ,  $\epsilon > 0$ ,  $j \in \{1, 2, 3, 4\}$  e  
 $\tau \in \{1, \|[g(x^0)]_+\|_\infty\}$

$k = 0$

REPITA

CALCULE  $x^{k+1} \in \operatorname{argmin} \left\{ f(x) + \sum_{i=1}^m \eta_j(g_i(x), \varepsilon_k) \mid x \in \mathbb{R}^n \right\}$

SE  $\|[g(x^{k+1})]_+\|_\infty \leq \epsilon$   
pare com sucesso

SENÃO

defina  $c_{k+1} = \beta c_k$  e  $\varepsilon_{k+1} = \gamma \varepsilon_k$

SE ocorre (4.4) ou (4.5) ou algum erro numérico  
pare com insucesso

SENÃO

$k = k + 1$

Os algoritmos acima foram implementados na linguagem de programação Julia, com base nas infraestruturas JuliaSmoothOptimizers [27] e Optim [25]. Os trabalhos [19, 20, 38, 39] sugerem o uso de algoritmos do tipo quase-Newton e do iterando anterior como ponto inicial para a resolução dos subproblemas dos algoritmos de suavização. Assim, foi seguida essa sugestão, adotando-se o método BFGS disponível no pacote Optim para os algoritmos de suavização e de penalidade  $\ell_2$ . Para a resolução dos subproblemas do algoritmo de penalidade  $\ell_1$ , utilizamos o método Nelder-Mead, também disponível no pacote Optim.

A seguir, são descritos os parâmetros iniciais para a execução de cada um dos algoritmos.

#### 4.1.1 Parâmetros iniciais

Nos experimentos numéricos apresentados para os algoritmos propostos em [19, 20, 38, 39], em cada problema testado, são utilizados um ou mais conjuntos de parâmetros de

entrada  $\{c_0, \varepsilon_0, \beta, \gamma\}$ . Em nossos experimentos, utilizamos os conjuntos de parâmetros trazidos na referência que a introduziu. Mais especificamente, fixado um dos critérios de parada (4.2) ou (4.3) para o Algoritmo 3.1 com as funções  $\eta_1, \eta_2, \eta_3, \eta_4$ , testamos todos os conjuntos de parâmetros utilizados nos experimentos numéricos de [19], [20], [38] e [39], respectivamente. Na Tabela 4.1 são mostrados esses parâmetros. A notação “Alg $j$ - $r$ ” significa que o Algoritmo 4.2 é considerado com a função  $\eta_j$  e o conjunto de parâmetros descritos na linha  $r$ .

	$\varepsilon_0$	$\gamma$	$c_0$	$\beta$		$\varepsilon_0$	$\gamma$	$c_0$	$\beta$
Alg1_1	0.1	0.1	10	3	Alg3_1	0.1	0.01	100	2
Alg1_2	1	0.1	5	2	Alg3_2	0.1	0.01	1	2
Alg1_3	1	0.01	100	2	Alg3_3	0.1	0.01	10	10
Alg1_4	0.1	0.1	1000	2	Alg3_4	0.1	0.1	100	2
Alg1_5	0.1	0.01	100	5	Alg3_5	1	0.5	1	2
Alg2_1	0.01	0.01	1	2	Alg4_1	1	0.5	1	2
Alg2_2	0.001	0.001	2	10	Alg4_2	0.1	0.01	10	10
Alg2_3	0.1	0.1	5	3	Alg4_3	0.1	0.01	1	2
Alg2_4	0.0001	0.001	2	10	Alg4_4	0.1	0.01	1000	2
-	-	-	-	-	Alg4_5	0.1	0.1	100	2

Tabela 4.1: Conjunto de parâmetros para os algoritmos de suavização.

Para os algoritmos de penalidade, utilizamos todos os pares de parâmetros  $(c_0, \beta)$  obtidos na Tabela 4.1, além do par  $(1, 10)$  sugerido em [23], como descritos na Tabela 4.2. A notação “ $\ell p$ - $r$ ” significa que o Algoritmo 4.1 é considerado com a função  $\ell_p$  e o conjunto de parâmetros descritos na linha  $r$  da tabela.

	$c_0$	$\beta$	Alg	$c_0$	$\beta$
$\ell 1$ _1	100	2	$\ell 2$ _2	100	2
$\ell 1$ _2	1	2	$\ell 2$ _2	1	2
$\ell 1$ _3	10	10	$\ell 2$ _3	10	10
$\ell 1$ _4	10	3	$\ell 2$ _4	10	3
$\ell 1$ _5	1000	2	$\ell 2$ _5	1000	2
$\ell 1$ _6	2	10	$\ell 2$ _6	2	10
$\ell 1$ _7	5	2	$\ell 2$ _7	5	2
$\ell 1$ _8	5	3	$\ell 2$ _8	5	3
$\ell 1$ _9	100	5	$\ell 2$ _9	100	5
$\ell 1$ _10	1	10	$\ell 2$ _10	1	10

Tabela 4.2: Conjunto de parâmetros dos algoritmos de penalidade.

## 4.2 Metodologia de comparação

Como para cada critério de parada (4.2) e (4.3) são testadas 19 instâncias do algoritmo de suavização e 20 instâncias do algoritmo de penalidade, os testes numéricos foram

realizados em duas etapas. A primeira etapa consiste da escolha de parâmetros entre os descritos nas Tabelas 4.1 e 4.2 que fornece o melhor desempenho para cada algoritmo. Em seguida, foi comparado o desempenho de todos os algoritmos com seu respectivo conjunto de parâmetros. Em ambas as etapas foram considerados todos os 51 problemas-testes.

Para a análise dos resultados, foram utilizados os mesmos critérios comparativos entre as soluções nas duas etapas. Dado um problema na forma (4.1), seja  $f_i^*$  o valor de função encontrado pelo algoritmo  $i$  e  $f_{\min}$  o menor valor de função encontrado por todos os algoritmos cuja solução é viável. Declaramos que o algoritmo *encontrou uma solução para este problema* se a condição sugerida em [4] é verificada, isto é, se a solução é viável e satisfaz

$$\frac{f_i^* - f_{\min}}{\max\{1, |f_{\min}|\}} \leq 0.01. \quad (4.6)$$

Em seguida, analisamos os resultados através de gráficos de perfis de desempenho [8] em relação ao tempo computacional, gerados por meio do pacote SolverBenchmark.jl [30]. Utilizamos um critério de empate para o tempo computacional, sugerido em [2]. Mais especificamente, consideramos que um algoritmo é o mais eficiente se *encontrou uma solução para o problema e o seu tempo de execução não ultrapassou em 5% do tempo do mais rápido*. Assim, um algoritmo é tido como *o mais robusto* quando resolve o maior número de problemas-teste entre os algoritmos comparados.

### 4.3 Resultados numéricos

Nesta seção trazemos os resultados numéricos obtidos com os critérios de parada (4.2) (Subseção 4.3.1) e (4.3) (Subseção 4.3.2). Em cada subseção exibimos os gráficos de performance profile e identificamos o conjunto de parâmetros que ofereceu o melhor desempenho computacional para cada um dos algoritmos. Em seguida, comparamos o desempenhos dos seis algoritmos com o conjunto de parâmetros selecionados na etapa anterior. Por fim, trazemos algumas observações gerais sobre o desempenhos dos algoritmos com os dois critérios de parada.

Adotamos a notação  $\ell 1$  e  $\ell 2$  para nos referirmos ao Algoritmo 4.1 com  $p = 1$  e  $p = 2$ , respectivamente, e Alg $j$  para o Algoritmo 4.2 com a função de suavização  $\eta_j$ , para  $j = 1, 2, 3, 4$ .

#### 4.3.1 Resultados obtidos com o critério de parada (4.2)

Os resultados que descrevemos nesta seção levam em consideração o critério de parada (4.2) e os critérios comparativos entre as soluções descritos na Seção 4.2.

## Seleção dos conjuntos de parâmetros

Nosso presente interesse é selecionar um conjunto de parâmetros iniciais para cada um dos algoritmos da Seção 4.1. Assim, apresentamos os resultados dos testes numéricos em gráficos de perfil de desempenho nas Figuras 4.1-4.4 e determinamos o conjunto de parâmetros em que cada um dos algoritmos apresentaram o melhor desempenho computacional.

A Figura 4.1 mostra que o algoritmo de penalidade com a função  $\ell_1$  teve maior robustez com o segundo par de parâmetros (1, 2), o qual é também um dos mais eficientes, como pode-se ver no gráfico da direita. Já a Figura 4.2 mostra que o algoritmo  $\ell_2$ -10, ou seja o algoritmo de penalidade  $\ell_2$  com o par de parâmetros (1, 10), foi o mais robusto e eficiente. Logo,  $\ell_1$ -2 e  $\ell_2$ -10 foram as instâncias selecionadas para os algoritmos de penalidade  $\ell_1$  e  $\ell_2$  para a segunda fase de experimentos numéricos.

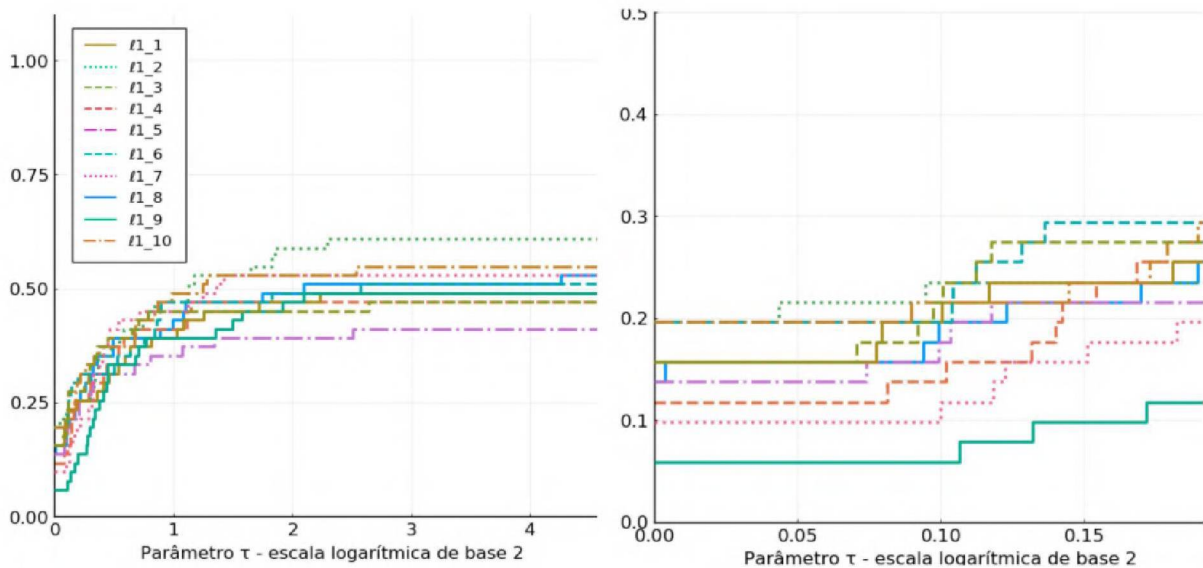


Figura 4.1: Perfil de desempenho do Algoritmo 4.1 com o critério de parada (4.2) e a função de penalidade  $\ell_1$ , em relação ao tempo computacional.

Na Figura 4.3 são apresentados os resultados para os algoritmos de suavização com a função  $\eta_1$  (à esquerda) e  $\eta_2$  (à direita). Temos que Alg1 apresentou o seu melhor desempenho com o primeiro conjunto de parâmetros  $\{0.1, 0.1, 10, 3\}$ , visto que resolveu três vezes mais problemas com essa instância do algoritmo do que com as outras instâncias testadas para a função  $\eta_1$ . Já os resultados obtidos para Alg2, mostram que esse algoritmo foi mais eficiente com o quarto conjunto de parâmetros, ou seja,  $\{0.0001, 0.001, 2, 10\}$ , porém, essa instância do algoritmo é a que apresenta menor robustez. Enquanto que o primeiro conjunto de parâmetros, dado por  $\{0.01, 0.01, 1, 2\}$ , é o qual o algoritmo de suavização com  $\eta_2$  teve sua maior robustez e também o segundo mais eficiente. Portanto, adotamos o Alg1.1 e o Alg2.1 para a segunda etapa de testes numéricos.

Na Figura 4.4 apresentamos os resultados obtidos para o algoritmo de suavização com

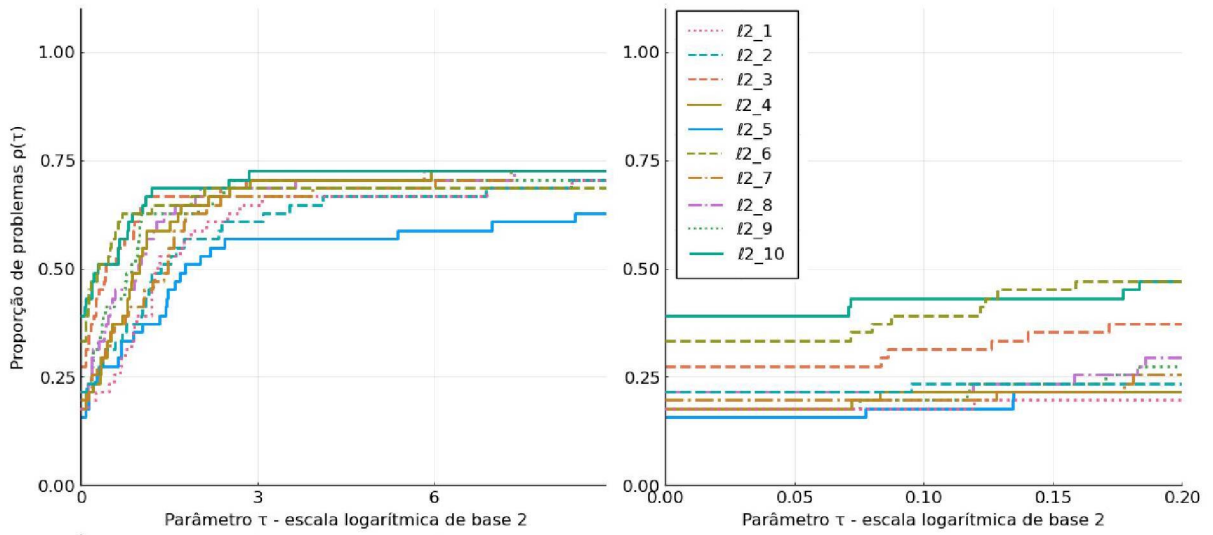


Figura 4.2: Perfil de desempenho do Algoritmo 4.1 com o critério de parada (4.2) e a função de penalidade  $\ell_2$ , em relação ao tempo computacional.

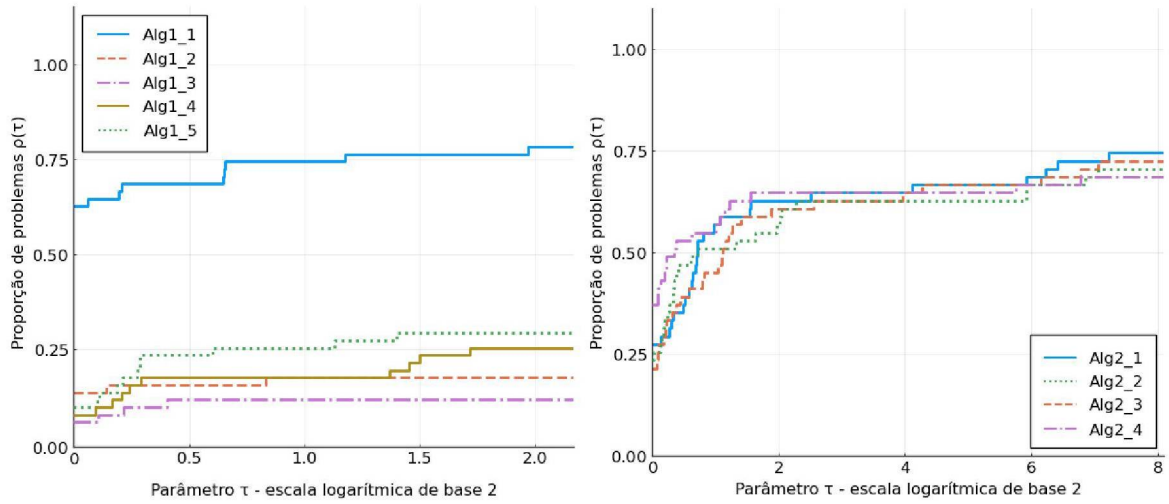


Figura 4.3: Perfil de desempenho do Algoritmo 4.2 com o critério de parada (4.2) e as funções de suavização  $\eta_1$  e  $\eta_2$ , em relação ao tempo computacional.

as funções  $\eta_3$  (gráfico da esquerda) e  $\eta_4$  (gráfico da direita). Estes evidenciam o melhor desempenho desses algoritmos com os parâmetros  $\varepsilon_0 = 0.1, \gamma = 0.01, c_0 = 1$  e  $\beta = 2$ , correspondente ao segundo e terceiro conjunto de parâmetros descritos na Tabela 4.1. Portanto, esses foram os parâmetros adotados para a segunda etapa para o Algoritmo 4.2 com as funções  $\eta_3$  e  $\eta_4$ .

Em decorrência da análise acima, a Tabela 4.3 mostra os parâmetros selecionados para a comparação entre os algoritmos.

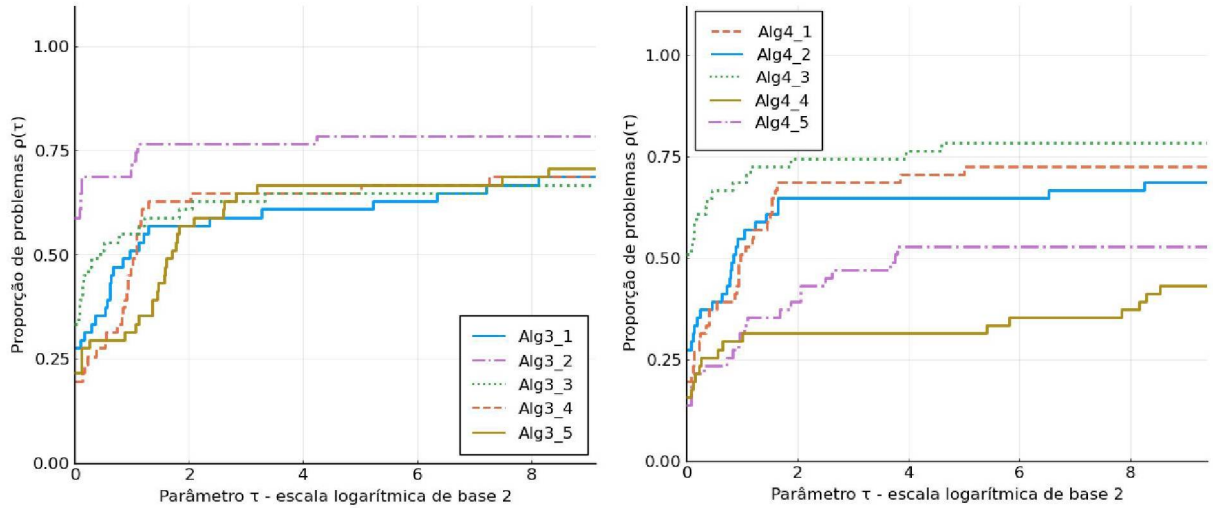


Figura 4.4: Perfil de desempenho do Algoritmo 4.2 com o critério de parada (4.2) e as funções de suavização  $\eta_3$  e  $\eta_4$ , em relação ao tempo computacional.

	$\varepsilon_0$	$\gamma$	$c_0$	$\beta$
$\ell_1$	-	-	1	2
$\ell_2$	-	-	1	10
Alg1	0.1	0.1	10	3
Alg2	0.001	0.001	1	2
Alg3	0.1	0.01	1	2
Alg4	0.1	0.01	1	2

Tabela 4.3: Parâmetros iniciais selecionados para os algoritmos com o critério de parada (4.2).

### Comparação entre os algoritmos

Apresentamos agora os resultados obtidos na comparação de desempenho computacional entre os algoritmos de penalidade e de suavização com os parâmetros descritos na Tabela 4.3 para os 51 problemas-teste da coleção CUTEst.

Na Tabela 4.4 é apresentada a taxa de robustez e eficiência atingida por cada um dos seis algoritmos. O algoritmo de penalidade  $\ell_1$  é o mais eficiente, resolvendo 35,29% dos problemas com o menor tempo, seguido pelo algoritmo de suavização com  $\eta_2$  o qual resolveu 25,49% problemas. Esse algoritmo é também o mais robusto, atingido 76,47% de problemas resolvidos, a mesma robustez apresentada pelo algoritmo de penalidade  $\ell_2$ . Já os algoritmos de suavização com as funções  $\eta_3$  e  $\eta_4$  resolveram 37 e 38 de 51 problemas, respectivamente.

Em consonância com os resultados apresentados na Tabela 4.4, o gráfico da esquerda da Figura 4.5 mostra que o algoritmo de suavização com  $\eta_2$  e o algoritmo de penalidade  $\ell_2$  são os mais robustos, sendo que Alg2 exigiu menos tempo computacional para resolver o mesmo número de problemas, além de ser o segundo mais eficiente, perdendo apenas para o algoritmo de penalidade  $\ell_1$ , que foi o mais eficiente em 18 de 51 problemas. O gráfico

	Robustez	Eficiência
$\ell_1$	56,86%	35,29%
$\ell_2$	76,47%	11,76 %
Alg1	29,41%	7,84 %
Alg2	76,47%	25,49%
Alg3	72,54%	7,84%
Alg4	74,51%	3,92%

Tabela 4.4: Robustez e eficiência dos 6 algoritmos com o critério de parada (4.2)

à direita representa o perfil de desempenho dos mesmos algoritmos desconsiderando o critério (4.6), ou seja, aceitando que um algoritmo resolveu um problema simplesmente se sua FLAG de saída é de sucesso (V). Neste gráfico, observamos que os algoritmos de suavização com  $\eta_1$  e de penalidade  $\ell_1$  foram os mais rápidos para atingir a tolerância para a viabilidade em 39 e 35 dos 51 problemas-testes, respectivamente. Em contraste, o gráfico da esquerda mostra que esses dois algoritmos foram os que apresentaram o pior desempenho quando se leva em consideração o critério (4.6). Isto mostra a importância de se adotar um critério de análise de qualidade da solução obtida pelos algoritmos como (4.6).

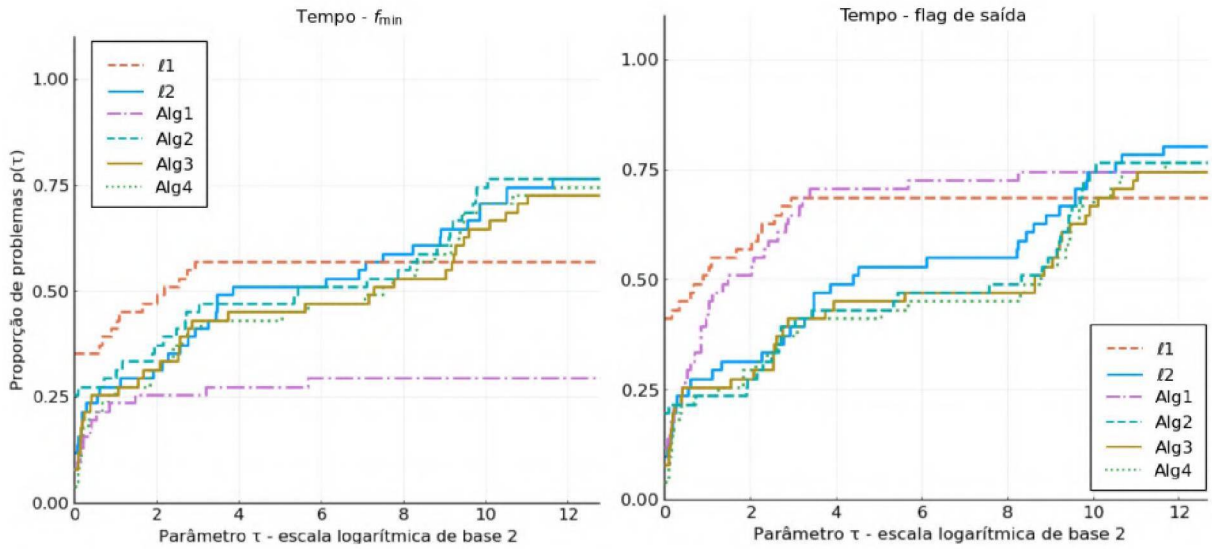


Figura 4.5: Perfil de desempenho de todos os algoritmos com o critério de parada (4.2) e os conjuntos de parâmetros selecionados, em relação ao tempo computacional. A figura da esquerda leva em consideração o critério (4.6), enquanto a da direita, não.

Na Tabela 4.5 é apresentado o número de vezes que os algoritmos atingiram cada uma das FLAGS de saída. O algoritmo de penalidade  $\ell_2$  foi o que atingiu o critério de viabilidade o maior número de vezes, 41 dos 51 problemas. Enquanto que o algoritmo com a função  $\ell_1$  foi o que menos vezes atingiu esse critério, totalizando 35 problemas, mas note que em 25 deles o algoritmo  $\ell_1$  foi o mais eficiente. Observamos que os algoritmos,  $\ell_1, \eta_2$  e  $\eta_4$  foram interrompidos por erros computacionais. Enquanto que em pelo menos 5

	V	C	T	E
$\ell_1$	35	5	0	11
$\ell_2$	41	10	0	0
Alg1	38	13	0	0
Alg2	39	11	0	1
Alg3	38	12	1	0
Alg4	39	11	0	1

Tabela 4.5: FLAGS de saída dos 6 algoritmos com o critério (4.2)

problemas, todos os algoritmos tiveram sua execução interrompida devido ao parâmetro de penalidade ser superior a  $10^{20}$ .

### 4.3.2 Resultados obtidos com o critério de parada (4.3)

Os resultados que descrevemos nessa subseção seguir levam em consideração o critério de parada (4.3) e os critérios comparativos entre as soluções descritos na Seção 4.2.

#### Seleção dos conjuntos de parâmetros

Apresentamos nesta seção os resultados dos testes numéricos em gráficos de perfil de desempenho nas Figuras 4.6-4.9 e determinamos um conjunto de parâmetros que oferece o melhor desempenho computacional para cada um dos seis algoritmos.

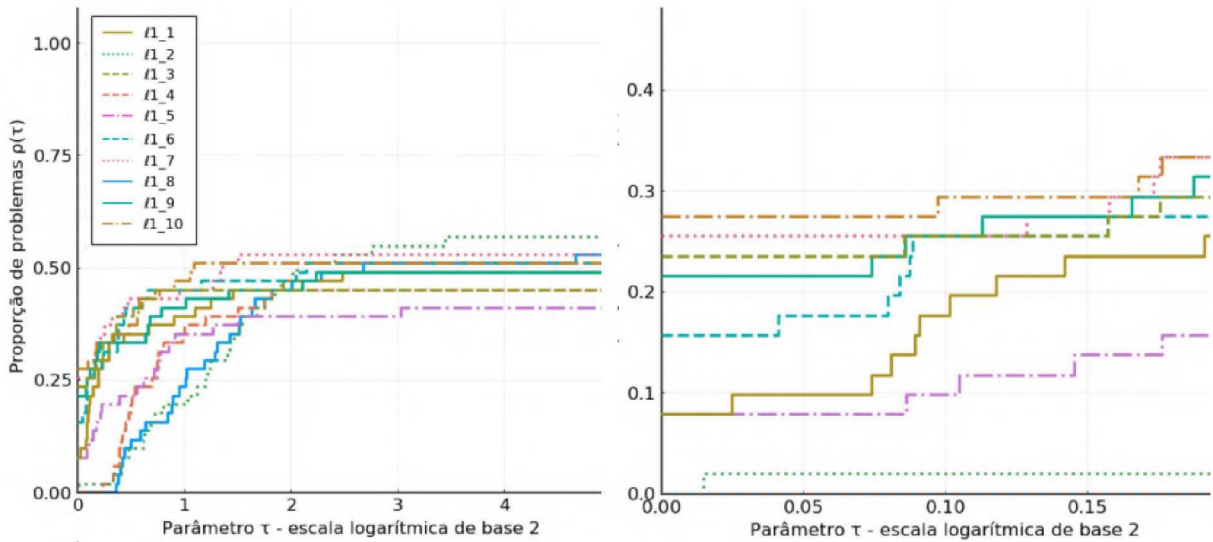


Figura 4.6: Perfil de desempenho do Algoritmo 4.1 com o critério de parada (4.3) e a função de penalidade  $\ell_1$ , em relação ao tempo computacional.

A Figura 4.6 mostra o gráfico de perfil de desempenho para o algoritmo de penalidade  $\ell_1$ . Dentre os parâmetros comparados, esse algoritmo resolveu o maior número de problemas com a instância  $\ell_1_2$ , isto é,  $c_0 = 1$  e  $\beta = 2$ . Já a Figura 4.7 mostra gráfico de perfil de desempenho com respeito ao algoritmo de penalidade  $\ell_2$ . Esse algoritmo apresentou

maior robustez com o par de parâmetros  $c_0 = 5$  e  $\beta = 2$  ( $\ell_{2.7}$ ), visto que essa instância do algoritmo foi a que resolveu o maior número de problemas dentre os parâmetros comparados. Portanto,  $\ell_{1.2}$  e  $\ell_{2.7}$  são as instâncias selecionadas para a segunda fase de testes numéricos.

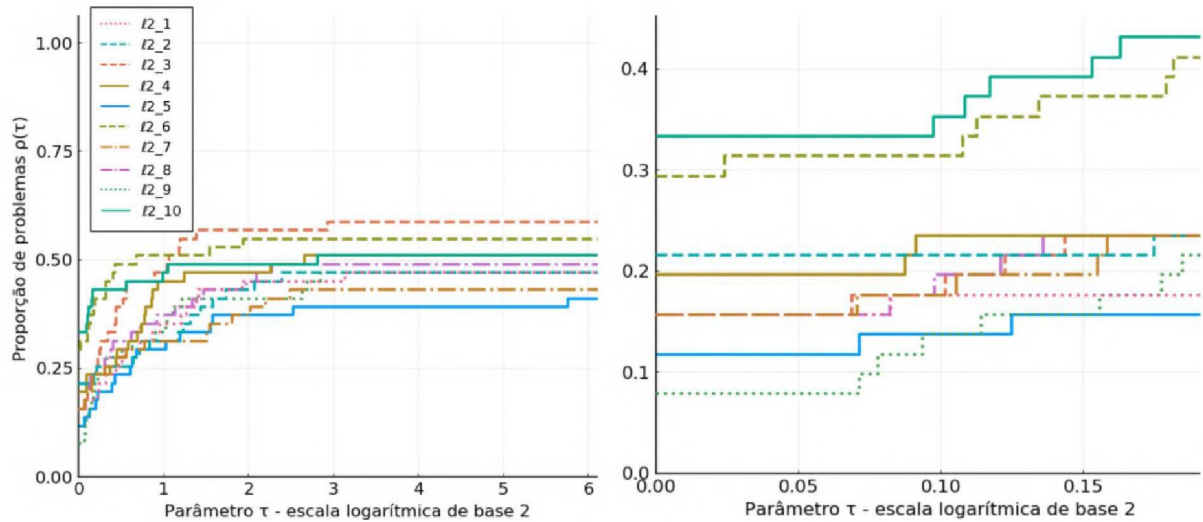


Figura 4.7: Perfil de desempenho do Algoritmo 4.1 com o critério de parada (4.3) e a função de penalidade  $\ell_2$ , em relação ao tempo computacional.

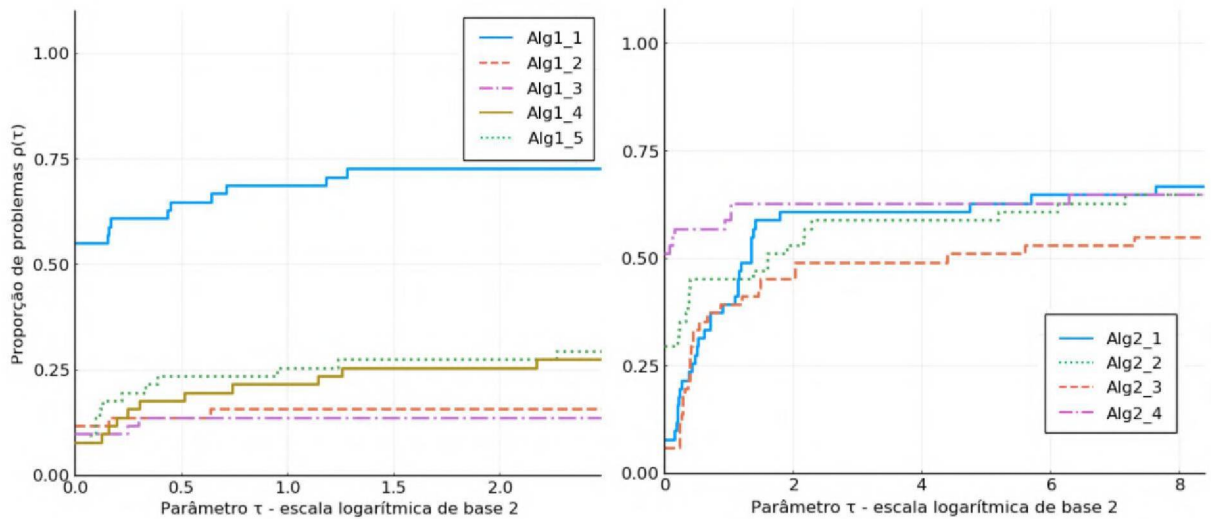


Figura 4.8: Perfil de desempenho do Algoritmo 4.2 com o critério de parada (4.3) e as funções de suavização  $\eta_1$  e  $\eta_2$ , em relação ao tempo computacional.

Na Figura 4.8 são mostrados os gráficos de perfil de desempenho para o algoritmo de suavização com as funções  $\eta_1$  (gráfico a esquerda) e  $\eta_2$  (gráfico a direita). Dentre os parâmetros comparados para  $\eta_1$ , o Alg1.1 é o mais eficiente e robusto. Enquanto que, Alg2.1 é a instância que apresentou a maior robustez para  $\eta_2$ . Portanto,  $(0.1, 0.1, 10, 3)$  e  $(0.01, 0.01, 1, 2)$ , são os conjuntos de parâmetros selecionados para o algoritmo de suavização com  $\eta_1$  e  $\eta_2$ , respectivamente.

A Figura 4.9 apresenta os gráficos de perfil de desempenho para o algoritmo de suavização com a função  $\eta_3$  (a esquerda) e  $\eta_4$  (a direita). Observamos que as instâncias Alg3.2 e Alg4.3 são as mais robustas e eficientes, dentre os parâmetros comparados. Portanto,  $\{0.1, 0.01, 1, 2\}$  é o conjunto de parâmetros selecionado para o algoritmo de suavização com as funções  $\eta_3$  e  $\eta_4$ .

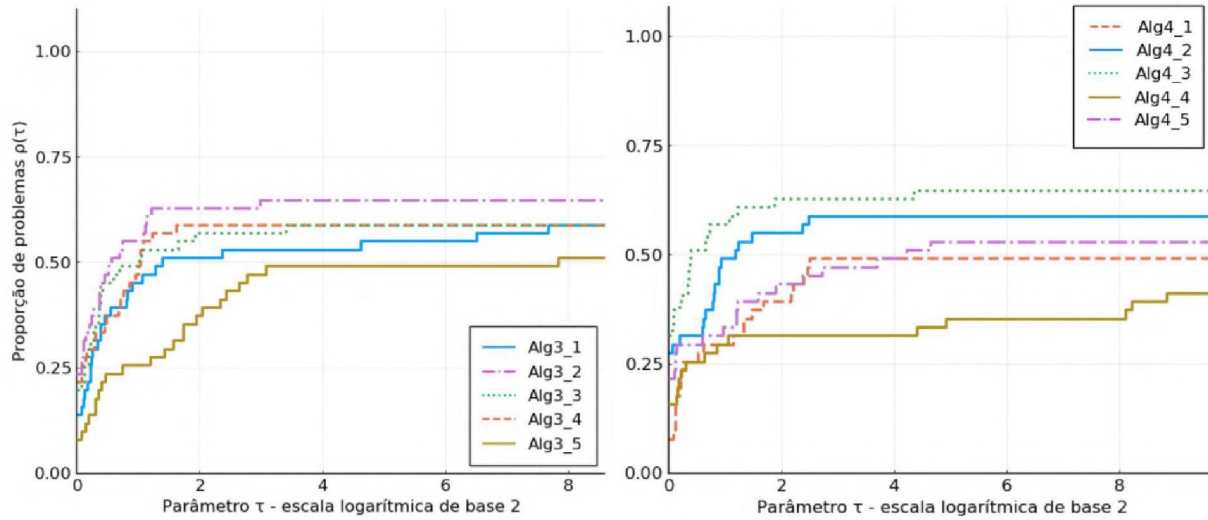


Figura 4.9: Perfil de desempenho do Algoritmo 4.2 com o critério de parada (4.3) e as funções de suavização  $\eta_3$  e  $\eta_4$ , em relação ao tempo computacional.

### Comparação entre os algoritmos

Trazemos agora os resultados computacionais obtidos para os seis algoritmos estudados, com os conjuntos de parâmetros selecionados na subseção anterior, que estão descritos na Tabela 4.6

	$\varepsilon_0$	$\gamma$	$c_0$	$\beta$
$\ell_1$	-	-	1	2
$\ell_2$	-	-	10	10
Alg1	0.1	0.1	10	3
Alg2	0.01	0.01	1	2
Alg3	0.1	0.01	1	2
Alg4	0.1	0.01	1	2

Tabela 4.6: Parâmetros iniciais selecionados para os algoritmos com o critério de parada (4.3)

A Figura 4.10 mostra os gráficos de perfil de desempenho em relação ao tempo computacional para resolver os problemas-teste.

O algoritmo de penalidade  $\ell_1$  foi o mais eficiente, resolvendo 31,37% dos problemas com o melhor tempo. Já o algoritmo de suavização com  $\eta_2$  se manteve o mais robusto, alcançando a taxa de 66,66% de problemas resolvidos. Enquanto que os algoritmos de

suavização com as funções  $\eta_3$  e  $\eta_4$  resolveram 64,70% dos problemas, ou seja, um problema a menos do que o algoritmo mais robusto, com este critério de parada, o algoritmo de penalidade  $\ell_2$  foi o quarto mais robusto, resolvendo 60,78% dos problemas.

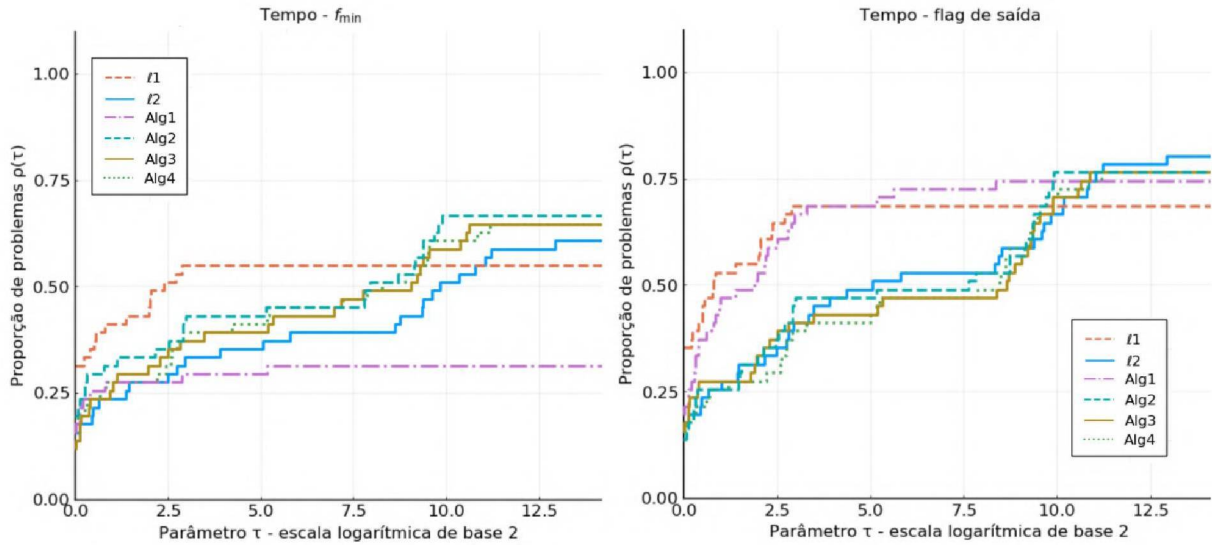


Figura 4.10: Perfil de desempenho de todos os algoritmos com o critério de parada (4.3) e os conjuntos de parâmetros selecionados, em relação ao tempo computacional. A figura da esquerda leva em consideração o critério (4.6), enquanto a da direita, não.

	Robustez	Eficiência
$\ell_1$	54,90%	31,37%
$\ell_2$	60,78%	15,68%
Alg1	31,37%	15,68 %
Alg2	66,66%	19,60%
Alg3	64,70%	11,76%
Alg4	64,70%	17,60%

Tabela 4.7: Robustez e eficiência dos 6 algoritmos mais viabilidade com o critério de parada (4.3)

A Tabela 4.8 apresenta a quantidade de vezes que os algoritmos atingiram cada uma das Flags de saída e a viabilidade. O algoritmo  $\ell_2$  foi o que mais vezes atingiu o critério de parada, totalizando 41 problemas, seguido pelos algoritmos de suavização  $\eta_2, \eta_3, \eta_4$  com 39 problemas. Note que nenhum dos algoritmos foram interrompidos por tempo computacional. Entretanto, com exceção do método de penalidade  $\ell_1$ , mais de 19% dos problemas tiveram sua execução interrompida devido ao parâmetro de penalidade se tornar superior a  $10^{20}$ .

### 4.3.3 Comentários gerais

Os resultados presentes nas subseções anteriores mostram que os algoritmos com o critério de parada (4.2) resolveram mais problemas-teste do que com o critério (4.3), com exceção

	V	C	T	E	Viab
$\ell_1$	35	5	0	11	28
$\ell_2$	41	10	0	0	31
Alg1	38	13	0	0	37
Alg2	39	11	0	1	34
Alg3	39	12	0	0	34
Alg4	39	11	0	1	34

Tabela 4.8: Número de problemas-teste que atingiram a tolerância para viabilidade e as FLAGS de saída para os 6 algoritmos com o critério de parada (4.3).

do algoritmo de suavização com  $\eta_1$ , o qual apresentou um melhor desempenho com o critério de parada (4.3). Por outro lado, o algoritmo de penalidade  $\ell_2$  foi o quarto mais robusto com (4.3), enquanto que com critério de parada de viabilidade resolveu o mesmo número de problemas de  $\eta_2$ , sendo o algoritmo mais robusto conforme podemos ver nas Tabelas 4.4 e 4.7.

Observamos nos Apêndice A que 9 dos 51 problemas não atingiram nenhum dos dois critérios de parada testados, CB2, CB3, EXPFITA, GIGOMEZ2, GIGOMEZ3, HS268, HS29, S268, SPIRAL. O que significa que nem a viabilidade, nem uma redução da inviabilidade do primeiro iterando foram obtidas para esses problemas-teste em nenhum dos algoritmos, fazendo com que o parâmetro de penalidade atingisse um valor superior a  $10^{20}$  ou apresentasse erros computacionais (como é o caso do algoritmo de penalidade  $\ell_1$ ) levando ao insucesso da resolução do problema-teste.

Em ambos os critérios de parada, o algoritmo de suavização com  $\eta_2$  foi o mais robusto e o algoritmo de penalidade exata  $\ell_1$  foi o mais eficiente. Por outro lado, o algoritmo de penalidade  $\ell_2$  está entre os mais robustos, mas é pouco eficiente. Os algoritmos de suavização com as funções  $\eta_3$  e  $\eta_4$  também apresentaram um desempenho computacional semelhante ao da função  $\eta_2$ , diferindo por no máximo dois problemas não resolvidos.

Na Tabela 4.9 são mostradas informações relativas aos problemas não resolvidos. A segunda coluna mostra o tipo de função objetivo do problema. O termo “mais geral”, significa que o tipo de função objetivo é mais geral do que funções do tipo: a soma de quadrados, quadrática, linear e constante. Na terceira coluna é apresentado o tipo da restrição mais complexa que o problema envolve. Neste caso, o termo “mais geral” significa que a restrição mais complexa é mais geral do que quaisquer um dos tipos de restrições abaixo:

- as únicas restrições do problema são variáveis fixas,
- as únicas restrições do problema são os limites das variáveis,
- as restrições do problema representam a matriz de adjacência de uma rede (linear),
- as restrições do problema são lineares,

- as restrições do problema são quadráticas.

Além das informações dos problemas-testes apresentadas na Tabela 4.9, enfatizamos que esses problemas são classificados como acadêmicos, isto é, problemas criados por pesquisadores para testar o desempenho de algoritmos.

problema	função objetivo	restrição mais geral	$n$	$m$
CB2	linear	mais geral	3	3
CB3	linear	mais geral	3	3
EXPFITA	mais geral	linear	5	22
GIGOMEZ2	linear	mais geral	3	3
GIGOMEZ3	linear	mais geral	3	3
HS268	quadrática	linear	5	5
HS29	mais geral	quadrática	3	1
S268	quadrática	linear	5	5
SPIRAL	linear	mais geral	3	2

Tabela 4.9: Descrição dos problemas não resolvidos por nenhum dos algoritmos.

Por fim, os resultados presentes neste capítulo indicam a necessidade de controlar o parâmetro de penalidade, visto que em cerca de 20% dos problemas-teste, em 5 algoritmos (com exceção possivelmente de  $\ell_1$ )  $c$  tornou-se arbitrariamente grande. Além disto, obtivemos insucesso na resolução dos subproblemas em 9 problemas-teste em todos os 6 algoritmos. Assim, uma possibilidade para trabalhos futuros seria testar outros algoritmos na resolução dos subproblemas.

# Conclusões

Neste trabalho foi estudado a teoria de métodos de penalidade para o problema geral de otimização com restrições de desigualdade. Discutimos a dicotomia entre métodos de penalidade suaves e métodos de penalidade exata e estendemos resultados já conhecidos na literatura. Apresentamos de forma unificada suavizações para o método de penalidade exata  $\ell_1$ . Também, derivamos condições suficientes para obter uma função de suavização e, como consequência, um algoritmo globalmente convergente para o problema com restrições.

Ademais, foi comparado o desempenho computacional dos algoritmos estudados na resolução dos problemas da biblioteca CUTEst [11]. Observamos que, três dos quatro algoritmos de suavização tiveram um melhor desempenho computacional do que os algoritmos de penalidade  $\ell_1$  e  $\ell_2$ . Sendo que em muitos problemas-teste o parâmetro de penalidade tornou-se muito grande, ocasionando a falha do algoritmo. Diante dessa dificuldade, uma das possibilidades para trabalhos futuros seria considerar diferentes parâmetros de penalidade para cada uma das restrições, o que pode ser conveniente pela possível diferença de escala entre as restrições. Outra possibilidade seria utilizar estratégias de restauração inexata, visando melhorar a viabilidade do ponto obtido a cada iteração e evitar que o parâmetro de penalidade torne-se muito grande.

# Bibliografia

- [1] M. S. Bazaraa, H. D. Sherali e C. M. Shetty. *Nonlinear Programming: Theory and Algorithms*. New York: John Wiley & Sons, 2006.
- [2] Y. Bello-Cruz, O. Ferreira e L. Prudente. “On the Global Convergence of the Inexact Semi-Smooth Newton Method for Absolute Value Equation”. Em: *Computational Optimization and Applications* 65 (set. de 2016). DOI: 10.1007/s10589-016-9837-x.
- [3] D. P. Bertsekas. “Nondifferentiable Optimization via Approximation.” English (US). Em: *Math Program Study* 4 (1975). Copyright: Copyright 2020 Elsevier B.V., All rights reserved., pp. 1–25. DOI: 10.1007/bfb0120696.
- [4] E. G. Birgin e J. Gentil. “Evaluating Bound-Constrained Minimization Software”. Em: *Computational Optimization and Applications* (fev. de 2012). DOI: 10.1007/s10589-012-9466-y.
- [5] E. G. Birgin e J. M. Martínez. *Practical Augmented Lagrangian Methods for Constrained Optimization*. Philadelphia, PA: Society for Industrial e Applied Mathematics, 2014. DOI: 10.1137/1.9781611973365.
- [6] J. Bonnans, J. C. Gilbert, C. Lemaréchal e C. Sagastizabal. *Numerical Optimization – Theoretical and Practical Aspects*. Jan. de 2006.
- [7] R. S. Burachik, C. Y. Kaya e C. J. Price. “A Primal–Dual Penalty Method via Rounded Weighted-1 Lagrangian Duality”. Em: *Optimization* 0.0 (2021), pp. 1–37. DOI: 10.1080/02331934.2021.1934680.
- [8] E. D. Dolan e J. J. Moré. “Benchmarking Optimization Software with Performance Profiles”. Em: *Mathematical Programming* 91 (2002), pp. 201–213.
- [9] M. V. Dolgopolik. “A unifying theory of exactness of linear penalty functions II: parametric penalty functions”. Em: *Optimization* 66.10 (2017), pp. 1577–1622. DOI: 10.1080/02331934.2017.1349125. URL: <https://doi.org/10.1080/02331934.2017.1349125>.

- [10] C. Gonzaga e R. Castillo. “A Nonlinear Programming Algorithm Based on Non-Coercive Penalty Functions”. Em: *Math. Program.* 96 (abr. de 2003), pp. 87–101. DOI: 10.1007/s10107-002-0332-z.
- [11] N. Gould, D. Orban e P. Toint. “CUTEst: a Constrained and Unconstrained Testing Environment with safe threads for Mathematical Optimization”. Em: *Computational Optimization and Applications* 60 (jul. de 2014). DOI: 10.1007/s10589-014-9687-3.
- [12] J. Griffin e T. Kolda. “Nonlinearly Constrained Optimization Using Heuristic Penalty Methods and Asynchronous Parallel Generating Set Search”. Em: *Applied Mathematics Research eXpress* 00 (abr. de 2010), pp. 1–17. DOI: 10.1093/amrx/abq003.
- [13] Z. Israel. “A Smoothing-Out Technique for Min—Max Optimization”. Em: *Mathematical Programming* 19 (1980), pp. 61–77. DOI: 10.1007/BF01581628.
- [14] A. Izmailov e M. Solodov. *Otimização: Condições de Otimalidade, Elementos de Análise Convexa e Dualidade*. terceira. Vol. 1. Rio de Janeiro: Instituto de Matemática Pura e Aplicada, 2014. ISBN: 978-85-244-0238-8.
- [15] A. Izmailov e M. Solodov. *Otimização: Métodos Computacionais*. terceira. Vol. 2. Rio de Janeiro: Instituto de Matemática Pura e Aplicada, 2018. ISBN: 978-85-244-0454-2.
- [16] C. Lemaréchal. “Lagrangian Relaxation”. Em: *Jünger, D. Naddef (Eds.), Computational combinatorial optimization* (2001), pp. 112–156.
- [17] C. Lemaréchal. “The Omnipresence of Lagrange”. Em: *Quarterly Journal of the Belgian, French and Italian Operations Research Societies* (2007), pp. 9–27.
- [18] C. Lemaréchal e A. Renaud. “A Geometric Study of Duality Gaps, With Applications.” Em: *Mathematical Programming* 90 (2001), 399–427.
- [19] S. Lian. “On the Smoothing of the Lower Order Exact Penalty Function for Inequality Constrained Optimization”. Em: *Operations Research Transactions* 16 (jul. de 2016). DOI: 10.1186/s13660-016-1126-9.
- [20] S. Lian. “Smoothing Approximation to l1 Exact Penalty Function for Inequality Constrained Optimization”. Em: *Applied Mathematics and Computation* 219.6 (2012), pp. 3113–3121. ISSN: 0096-3003. DOI: <https://doi.org/10.1016/j.amc.2012.09.042>.
- [21] E. L. Lima. “A Desigualdade do Valor Médio”. Em: *Curso de Análise*. Rio de Janeiro: Instituto de Matemática Pura e Aplicada, 1999, pp. 264–266. ISBN: 85-244-0108-7.

- [22] D. G. Luenberger e Y. Ye. “Penalty and Barrier Methods”. Em: *Linear and Nonlinear Programming*. Cham: Springer International Publishing, 2016, pp. 397–428. DOI: 10.1007/978-3-319-18842-3\_13.
- [23] J. M. Martínez e S. A. Santos. *Métodos Computacionais de Otimização*. Rio de Janeiro: Sociedade Brasileira de Matemática, 1995, p. 256.
- [24] L. Minchenko e S. Stakhovski. “On Relaxed Constant Rank Regularity Condition in Mathematical Programming”. Em: *Optimization* 60.4 (2011), pp. 429–440. DOI: 10.1080/02331930902971377.
- [25] P. K. Mogensen e A. N. Riseth. “Optim: A Mathematical Optimization Package for Julia”. Em: *Journal of Open Source Software* 3.24 (2018), p. 615. DOI: 10.21105/joss.00615.
- [26] J. Nocedal e S. J. Wright. *Numerical Optimization*. Springer Series in Operations Research. New York: Springer-Verlag, 2006.
- [27] D. Orban e A. S. Siqueira. *JuliaSmoothOptimizers: Infrastructure and Solvers for Continuous Optimization in Julia*. 2019. DOI: 10.5281/zenodo.2655082. URL: <https://juliasmoothoptimizers.github.io>.
- [28] D. Orban, A. S. Siqueira e contributors. *CUTEst.jl: Julia’s CUTEst Interface*. <https://github.com/JuliaSmoothOptimizers/CUTEst.jl>. 2020. DOI: 10.5281/zenodo.1188851.
- [29] D. Orban, A. S. Siqueira e contributors. *NLPModelsIpopt.jl: A Thin IPOPT Wrapper for NLPModels*. <https://github.com/JuliaSmoothOptimizers/NLPModelsIpopt.jl>. 2020. DOI: 10.5281/zenodo.2629034.
- [30] D. Orban, A. S. Siqueira e contributors. *SolverBenchmark.jl: Benchmark Tools for Solvers*. <https://github.com/JuliaSmoothOptimizers/SolverBenchmark.jl>. 2020. DOI: 10.5281/zenodo.3948381.
- [31] G. D. Pillo. “Exact Penalty Methods”. Em: *In I. Ciocco (Ed.), Algorithms for Continuous Optimization*. Kluwer Acad. Publ, 1994, pp. 1–45.
- [32] M. Pinar e S. Zenios. “On Smoothing Exact Penalty Functions for Convex Constrained Optimization”. Em: *SIAM Journal on Optimization* 4 (ago. de 1994), pp. 486–511. DOI: 10.1137/0804027.
- [33] A. A. Ribeiro e E. W. Karas. *Otimização Contínua: Aspectos Teóricos e Computacionais*. Cengage Learnig, 2013. ISBN: 9788522115013.
- [34] A. Ruszczyński. *Nonlinear Optimization*. Princeton University Press, 2011. DOI: 10.1515/9781400841059.

- [35] A. Sahiner, G. Kapusuz e N. Yilmaz. “A New Smoothing Approach to Exact Penalty Functions for Inequality Constrained Optimization Problems”. Em: *Numerical Algebra, Control and Optimization* 6 (jun. de 2016), pp. 161–173. DOI: 10.3934/naco.2016006.
- [36] M. V. Solodov. “Constraint Qualifications”. Em: *Wiley Encyclopedia of Operations Research and Management Science*. John Wiley & Sons, 2011. ISBN: 9780470400531. DOI: <https://doi.org/10.1002/9780470400531.eorms0978>.
- [37] S. Wenyu e Y. Ya-Xiang. *Optimization Theory and Methods*. Springer US, 2006, p. 688. ISBN: 978-0-387-24976-6. DOI: 10.1007/b106451.
- [38] X. Xu, C. Dang, F. Chan e Y. Wang. “On Smoothing ll Exact Penalty Function for Constrained Optimization Problems”. Em: *Numerical Functional Analysis and Optimization* 40 (dez. de 2018), pp. 1–18. DOI: 10.1080/01630563.2018.1483948.
- [39] X. Xu, Z. Meng, J. Sun, L. Huang e R. Shen. “A Second-Order Smooth Penalty Function Algorithm for Constrained Optimization Problems”. Em: *Computational Optimization and Applications* 55 (mai. de 2013). DOI: 10.1007/s10589-012-9504-9.
- [40] W. I. Zangwill. “Non-Linear Programming Via Penalty Functions”. Em: *Management Science* 13.5 (1967), pp. 344–358. URL: <https://EconPapers.repec.org/RePEc:inm:ormnsc:v:13:y:1967:i:5:p:344-358>.

# Apêndice A

## Resultados numéricos: comparação entre os algoritmos

Neste apêndice apresentamos em tabelas os resultados numéricos obtidos com os critérios de parada (4.2) e (4.3) que foram utilizados na geração dos gráficos de perfil de desempenho mostrados nas Figuras 4.5 e 4.10, respectivamente. As colunas “s” e “tempo” indicam a *flag* de saída e o tempo gasto na execução de cada algoritmo, respectivamente. O símbolo  $\infty$  (ou  $-\infty$ ), nas referidas tabelas, significam que  $f(x^k)$  ou  $\|[g(x^k)]_+\|_\infty$  é superior a  $e^{710}$  (inferior a  $-e^{710}$ ). Se a condição (4.6) é satisfeita, os resultados são destacados em negrito.

### A.1 Resultados com o critério de parada (4.2)

problema	alg	n	m	s	$f(x^k)$	$\ [g(x^k)]_+\ _\infty$	tempo	k
CB2	Alg1	3	3	C	1.00e+00	1.00e+00	4.26e-02	39
	Alg2	3	3	C	1.00e+00	1.00e+00	7.46e-02	64
	Alg3	3	3	C	1.00e+00	1.00e+00	6.71e-02	64
	Alg4	3	3	C	1.00e+00	1.00e+00	2.33e-02	64
	$\ell_1$	3	3	E	-4.89e+133	$\infty$	3.36e-03	1
	$\ell_2$	3	3	C	1.00e+00	1.00e+00	1.08e-02	14
CB3	Alg1	3	3	C	1.00e+00	1.00e+00	4.04e-02	39
	Alg2	3	3	C	1.00e+00	1.00e+00	7.64e-02	64
	Alg3	3	3	C	1.00e+00	1.00e+00	6.10e-02	64
	Alg4	3	3	C	1.00e+00	1.00e+00	3.78e-02	64
	$\ell_1$	3	3	E	-4.89e+133	$\infty$	3.76e-03	1
	$\ell_2$	3	3	C	1.00e+00	1.00e+00	1.27e-02	14
CHACONN1	Alg1	3	3	V	2.29e+00	0.00e+00	2.59e-03	1
	<b>Alg2</b>	<b>3</b>	<b>3</b>	<b>V</b>	<b>1.95e+00</b>	<b>0.00e+00</b>	<b>3.06e-01</b>	<b>3</b>
	<b>Alg3</b>	<b>3</b>	<b>3</b>	<b>V</b>	<b>1.95e+00</b>	<b>0.00e+00</b>	<b>1.26e+00</b>	<b>4</b>
	<b>Alg4</b>	<b>3</b>	<b>3</b>	<b>V</b>	<b>1.95e+00</b>	<b>0.00e+00</b>	<b>3.06e-01</b>	<b>3</b>
	$\ell_1$	<b>3</b>	<b>3</b>	<b>V</b>	<b>1.95e+00</b>	<b>0.00e+00</b>	<b>8.18e-04</b>	<b>1</b>

Continua na próxima página

problema	alg	n	m	s	$f(x^k)$	$\ g(x^k)\ _\infty$	tempo	$k$
	$\ell 2$	<b>3</b>	<b>3</b>	<b>V</b>	<b>1.95e+00</b>	<b>0.00e+00</b>	<b>9.00e-01</b>	<b>5</b>
CHACONN2	Alg1	3	3	V	2.37e+00	0.00e+00	1.68e-03	1
	<b>Alg2</b>	<b>3</b>	<b>3</b>	<b>V</b>	<b>2.00e+00</b>	<b>0.00e+00</b>	<b>7.10e-03</b>	<b>3</b>
	<b>Alg3</b>	<b>3</b>	<b>3</b>	<b>V</b>	<b>2.00e+00</b>	<b>1.83e-07</b>	<b>1.18e-02</b>	<b>3</b>
	<b>Alg4</b>	<b>3</b>	<b>3</b>	<b>V</b>	<b>2.00e+00</b>	<b>0.00e+00</b>	<b>5.24e-03</b>	<b>3</b>
	$\ell 1$	3	3	V	2.00e+00	4.62e-10	9.12e-04	1
	$\ell 2$	3	3	V	2.00e+00	0.00e+00	1.11e-02	5
CONGIGMZ	Alg1	3	5	C	-6.54e+26	0.00e+00	4.63e-02	39
	<b>Alg2</b>	<b>3</b>	<b>5</b>	<b>V</b>	<b>-2.61e+08</b>	<b>0.00e+00</b>	<b>1.54e-02</b>	<b>5</b>
	Alg3	3	5	V	-4.00e+01	0.00e+00	2.84e-03	2
	Alg4	3	5	V	-1.08e+02	0.00e+00	8.02e-03	4
	$\ell 1$	3	5	E	-4.62e+155	$\infty$	1.56e-02	2
	$\ell 2$	3	5	C	1.18e+00	9.27e+00	1.27e-02	14
DEMYMALO	Alg1	3	3	V	-2.62e+00	0.00e+00	1.49e-03	1
	<b>Alg2</b>	<b>3</b>	<b>3</b>	<b>V</b>	<b>-3.00e+00</b>	<b>0.00e+00</b>	<b>5.47e-03</b>	<b>3</b>
	<b>Alg3</b>	<b>3</b>	<b>3</b>	<b>V</b>	<b>-3.00e+00</b>	<b>0.00e+00</b>	<b>5.11e-03</b>	<b>3</b>
	<b>Alg4</b>	<b>3</b>	<b>3</b>	<b>V</b>	<b>-3.00e+00</b>	<b>0.00e+00</b>	<b>3.44e-03</b>	<b>3</b>
	$\ell 1$	3	3	V	-3.00e+00	0.00e+00	6.69e-04	1
	$\ell 2$	3	3	V	-3.00e+00	1.06e-07	3.06e-03	4
DIPIGRI	Alg1	7	4	V	6.81e+02	0.00e+00	6.33e-03	1
	Alg2	7	4	V	6.81e+02	0.00e+00	4.38e+00	3
	Alg3	7	4	V	6.81e+02	0.00e+00	3.72e+00	3
	Alg4	7	4	V	6.81e+02	0.00e+00	3.24e+00	3
	$\ell 1$	7	4	V	6.81e+02	0.00e+00	1.49e-02	3
	$\ell 2$	7	4	V	6.81e+02	0.00e+00	2.29e+00	4
EXPFITA	Alg1	5	22	C	1.08e+04	1.00e+00	8.72e+01	39
	Alg2	5	22	C	1.67e+03	1.00e+00	2.21e+02	64
	Alg3	5	22	C	2.60e+00	1.00e+00	1.58e+02	64
	Alg4	5	22	C	5.54e+00	1.00e+00	1.44e+02	64
	$\ell 1$	5	22	C	1.30e+43	5.34e+23	2.62e+00	64
	$\ell 2$	5	22	C	3.67e+02	1.00e+00	4.08e+01	14
GIGOMEZ1	Alg1	3	3	V	1.35e+00	0.00e+00	4.53e-03	2
	<b>Alg2</b>	<b>3</b>	<b>3</b>	<b>V</b>	<b>-7.54e+00</b>	<b>0.00e+00</b>	<b>1.98e-02</b>	<b>2</b>
	<b>Alg3</b>	<b>3</b>	<b>3</b>	<b>V</b>	<b>-7.54e+00</b>	<b>0.00e+00</b>	<b>2.49e-03</b>	<b>2</b>
	<b>Alg4</b>	<b>3</b>	<b>3</b>	<b>V</b>	<b>-7.54e+00</b>	<b>0.00e+00</b>	<b>1.53e-03</b>	<b>2</b>
	$\ell 1$	3	3	E	-1.46e+267	$\infty$	6.36e-03	2
	$\ell 2$	3	3	V	1.39e+00	0.00e+00	1.34e-03	3
GIGOMEZ2	Alg1	3	3	C	2.00e+00	2.00e+00	5.82e-02	39
	Alg2	3	3	C	2.00e+00	2.00e+00	5.55e-02	64
	Alg3	3	3	C	2.00e+00	2.00e+00	7.29e-02	64
	Alg4	3	3	C	2.00e+00	2.00e+00	2.60e-02	64
	$\ell 1$	3	3	E	-9.55e+133	$\infty$	3.45e-03	1
	$\ell 2$	3	3	C	2.00e+00	2.00e+00	8.66e-03	14
GIGOMEZ3	Alg1	3	3	C	2.00e+00	2.00e+00	4.60e-02	39
	Alg2	3	3	C	2.00e+00	2.00e+00	5.60e-02	64
	Alg3	3	3	C	2.00e+00	2.00e+00	7.23e-02	64
	Alg4	3	3	C	2.00e+00	2.00e+00	2.73e-02	64
	$\ell 1$	3	3	E	-9.55e+133	$\infty$	3.69e-03	1
	$\ell 2$	3	3	C	2.00e+00	2.00e+00	8.31e-03	14
GOFFIN	Alg1	51	50	V	6.83e-01	0.00e+00	6.00e+02	1
	Alg2	51	50	T	5.50e-04	1.98e-05	1.20e+03	3
	<b>Alg3</b>	<b>51</b>	<b>50</b>	<b>V</b>	<b>1.85e-03</b>	<b>0.00e+00</b>	<b>1.20e+03</b>	<b>3</b>
	<b>Alg4</b>	<b>51</b>	<b>50</b>	<b>V</b>	<b>3.49e-04</b>	<b>0.00e+00</b>	<b>1.20e+03</b>	<b>2</b>
	$\ell 1$	51	50	V	6.74e+01	0.00e+00	1.94e+00	8
	$\ell 2$	<b>51</b>	<b>50</b>	<b>V</b>	<b>8.76e-03</b>	<b>0.00e+00</b>	<b>6.00e+02</b>	<b>4</b>
HAIFAS	Alg1	13	9	V	-6.78e-02	0.00e+00	7.89e-02	1
	<b>Alg2</b>	<b>13</b>	<b>9</b>	<b>V</b>	<b>-4.50e-01</b>	<b>0.00e+00</b>	<b>1.28e+01</b>	<b>3</b>
	<b>Alg3</b>	<b>13</b>	<b>9</b>	<b>V</b>	<b>-4.50e-01</b>	<b>0.00e+00</b>	<b>2.73e+01</b>	<b>4</b>

Continua na próxima página

problema	alg	n	m	s	$f(x^k)$	$\ g(x^k)\ _\infty$	tempo	k
	<b>Alg4</b>	<b>13</b>	<b>9</b>	<b>V</b>	<b>-4.49e-01</b>	<b>0.00e+00</b>	<b>3.19e+01</b>	<b>4</b>
	$\ell_1$	13	9	V	-4.50e-01	1.57e-07	9.58e-03	1
	$\ell_2$	13	9	V	-4.49e-01	7.59e-07	3.04e+01	6
HALDMADS	Alg1	6	42	V	6.47e-01	0.00e+00	3.07e-01	1
	<b>Alg2</b>	<b>6</b>	<b>42</b>	<b>V</b>	<b>2.18e-04</b>	<b>0.00e+00</b>	<b>5.32e+01</b>	<b>3</b>
	<b>Alg3</b>	<b>6</b>	<b>42</b>	<b>V</b>	<b>2.52e-04</b>	<b>8.29e-08</b>	<b>1.26e+02</b>	<b>4</b>
	<b>Alg4</b>	<b>6</b>	<b>42</b>	<b>V</b>	<b>1.35e-04</b>	<b>0.00e+00</b>	<b>8.05e+01</b>	<b>3</b>
	$\ell_1$	6	42	V	1.93e-02	0.00e+00	2.62e-01	3
	$\ell_2$	6	42	V	2.29e+00	0.00e+00	5.89e-01	5
<b>HS10</b>	<b>Alg1</b>	<b>2</b>	<b>1</b>	<b>V</b>	<b>-2.00e+01</b>	<b>0.00e+00</b>	<b>4.57e-02</b>	<b>1</b>
	<b>Alg2</b>	<b>2</b>	<b>1</b>	<b>V</b>	<b>-2.00e+01</b>	<b>0.00e+00</b>	<b>3.72e-02</b>	<b>1</b>
	<b>Alg3</b>	<b>2</b>	<b>1</b>	<b>V</b>	<b>-2.00e+01</b>	<b>0.00e+00</b>	<b>4.43e-02</b>	<b>1</b>
	<b>Alg4</b>	<b>2</b>	<b>1</b>	<b>V</b>	<b>-2.00e+01</b>	<b>0.00e+00</b>	<b>3.02e-02</b>	<b>1</b>
	$\ell_1$	2	1	E	-6.72e+227	$\infty$	2.08e-03	1
	$\ell_2$	2	1	V	-2.00e+01	0.00e+00	5.26e-02	2
HS100	Alg1	7	4	V	1.64e+02	0.00e+00	1.34e-02	1
	Alg2	7	4	V	1.63e+02	0.00e+00	6.24e+00	4
	Alg3	7	4	V	1.63e+02	0.00e+00	7.20e+00	4
	Alg4	7	4	V	1.63e+02	0.00e+00	5.88e+00	4
	$\ell_1$	7	4	V	1.63e+02	6.62e-07	9.31e-03	3
	$\ell_2$	7	4	V	1.63e+02	0.00e+00	2.63e+00	4
HS100MOD	Alg1	7	4	V	2.40e+02	0.00e+00	2.60e-02	1
	Alg2	7	4	V	2.41e+02	0.00e+00	8.81e+00	4
	Alg3	7	4	V	2.40e+02	0.00e+00	1.10e+01	3
	Alg4	7	4	V	2.41e+02	0.00e+00	1.03e+01	4
	$\ell_1$	7	4	V	2.40e+02	0.00e+00	1.54e-02	3
	$\ell_2$	7	4	V	2.42e+02	7.89e-08	2.06e+01	8
HS11	Alg1	2	1	V	-2.50e+01	0.00e+00	9.70e-05	1
	Alg2	2	1	V	-2.50e+01	0.00e+00	8.49e-05	1
	Alg3	2	1	V	-2.50e+01	0.00e+00	9.89e-05	1
	Alg4	2	1	V	-2.50e+01	0.00e+00	1.04e-04	1
	$\ell_1$	2	1	V	-2.50e+01	0.00e+00	1.05e-04	1
	$\ell_2$	2	1	V	-2.50e+01	0.00e+00	1.03e-04	2
HS113	Alg1	10	8	V	-3.10e+01	0.00e+00	7.62e-02	1
	Alg2	10	8	V	-3.10e+01	0.00e+00	5.31e-02	1
	Alg3	10	8	V	-3.10e+01	0.00e+00	6.88e-02	1
	Alg4	10	8	V	-3.10e+01	0.00e+00	4.79e-02	1
	$\ell_1$	10	8	V	-3.10e+01	0.00e+00	8.16e-03	1
	$\ell_2$	10	8	V	-3.10e+01	0.00e+00	1.42e-01	2
HS12	Alg1	2	1	V	-1.22e+02	0.00e+00	1.17e-04	1
	Alg2	2	1	V	-1.22e+02	0.00e+00	1.08e-04	1
	Alg3	2	1	V	-1.22e+02	0.00e+00	1.14e-04	1
	Alg4	2	1	V	-1.22e+02	0.00e+00	1.36e-04	1
	$\ell_1$	2	1	V	-1.22e+02	0.00e+00	1.44e-04	1
	$\ell_2$	2	1	V	-1.22e+02	0.00e+00	8.30e-05	2
HS22	Alg1	2	2	V	2.29e-07	0.00e+00	1.90e-04	1
	Alg2	2	2	V	1.97e-31	0.00e+00	1.20e-04	1
	Alg3	2	2	V	6.52e-21	0.00e+00	9.11e-05	1
	Alg4	2	2	V	7.85e-24	0.00e+00	1.14e-04	1
	$\ell_1$	2	2	V	4.97e-10	0.00e+00	1.45e-04	1
	$\ell_2$	2	2	V	1.97e-31	0.00e+00	8.11e-05	2
HS268	Alg1	5	5	C	7.88e+05	1.85e+01	3.01e+01	39
	Alg2	5	5	C	3.05e+05	2.20e+01	4.06e+01	64
	Alg3	5	5	C	1.49e+05	2.70e+01	4.31e+01	64
	Alg4	5	5	C	6.85e+04	2.64e+01	5.22e+01	64
	$\ell_1$	5	5	C	1.86e+37	3.66e+18	1.18e+00	64
	$\ell_2$	5	5	C	7.15e+04	1.45e+01	4.37e-01	14
HS29	Alg1	3	1	C	4.56e-02	4.70e+01	2.25e-02	39

Continua na próxima página

problema	alg	n	m	s	$f(x^k)$	$\ g(x^k)\ _\infty$	tempo	$k$
	Alg2	3	1	C	4.56e-02	4.70e+01	3.47e-02	64
	Alg3	3	1	C	4.56e-02	4.70e+01	3.66e-02	64
	Alg4	3	1	C	-1.00e+00	4.10e+01	2.18e-02	64
	$\ell_1$	3	1	C	$-\infty$	0.00e+00	3.52e-03	64
	$\ell_2$	<b>3</b>	<b>1</b>	<b>V</b>	<b>-2.03e+01</b>	<b>0.00e+00</b>	<b>5.71e-03</b>	<b>12</b>
<b>HS43</b>	<b>Alg1</b>	<b>4</b>	<b>3</b>	<b>V</b>	<b>-7.99e+01</b>	<b>0.00e+00</b>	<b>4.57e-02</b>	<b>1</b>
	<b>Alg2</b>	<b>4</b>	<b>3</b>	<b>V</b>	<b>-7.99e+01</b>	<b>0.00e+00</b>	<b>3.80e-02</b>	<b>1</b>
	<b>Alg3</b>	<b>4</b>	<b>3</b>	<b>V</b>	<b>-7.99e+01</b>	<b>0.00e+00</b>	<b>4.41e-02</b>	<b>1</b>
	<b>Alg4</b>	<b>4</b>	<b>3</b>	<b>V</b>	<b>-7.99e+01</b>	<b>0.00e+00</b>	<b>3.08e-02</b>	<b>1</b>
	$\ell_1$	<b>4</b>	<b>3</b>	<b>V</b>	<b>-7.99e+01</b>	<b>0.00e+00</b>	<b>9.90e-04</b>	<b>1</b>
	$\ell_2$	<b>4</b>	<b>3</b>	<b>V</b>	<b>-7.99e+01</b>	<b>0.00e+00</b>	<b>4.92e-02</b>	<b>2</b>
HS88	Alg1	2	1	V	1.76e-01	0.00e+00	1.70e-02	1
	<b>Alg2</b>	<b>2</b>	<b>1</b>	<b>V</b>	<b>1.36e-23</b>	<b>0.00e+00</b>	<b>2.80e-03</b>	<b>1</b>
	<b>Alg3</b>	<b>2</b>	<b>1</b>	<b>V</b>	<b>1.36e-23</b>	<b>0.00e+00</b>	<b>3.47e-03</b>	<b>1</b>
	<b>Alg4</b>	<b>2</b>	<b>1</b>	<b>V</b>	<b>1.36e-23</b>	<b>0.00e+00</b>	<b>2.30e-03</b>	<b>1</b>
	$\ell_1$	<b>2</b>	<b>1</b>	<b>V</b>	<b>7.75e-09</b>	<b>0.00e+00</b>	<b>7.15e-03</b>	<b>1</b>
	$\ell_2$	<b>2</b>	<b>1</b>	<b>V</b>	<b>1.36e-23</b>	<b>0.00e+00</b>	<b>2.60e-03</b>	<b>2</b>
HS89	Alg1	3	1	V	1.76e-01	0.00e+00	4.19e-02	1
	<b>Alg2</b>	<b>3</b>	<b>1</b>	<b>V</b>	<b>2.04e-23</b>	<b>0.00e+00</b>	<b>5.90e-03</b>	<b>1</b>
	<b>Alg3</b>	<b>3</b>	<b>1</b>	<b>V</b>	<b>2.04e-23</b>	<b>0.00e+00</b>	<b>5.92e-03</b>	<b>1</b>
	<b>Alg4</b>	<b>3</b>	<b>1</b>	<b>V</b>	<b>2.04e-23</b>	<b>0.00e+00</b>	<b>7.14e-03</b>	<b>1</b>
	$\ell_1$	<b>3</b>	<b>1</b>	<b>V</b>	<b>3.08e-09</b>	<b>0.00e+00</b>	<b>2.08e-02</b>	<b>1</b>
	$\ell_2$	<b>3</b>	<b>1</b>	<b>V</b>	<b>2.04e-23</b>	<b>0.00e+00</b>	<b>5.34e-03</b>	<b>2</b>
HS90	Alg1	4	1	V	1.76e-01	0.00e+00	8.74e-02	1
	<b>Alg2</b>	<b>4</b>	<b>1</b>	<b>V</b>	<b>2.72e-23</b>	<b>0.00e+00</b>	<b>1.03e-02</b>	<b>1</b>
	<b>Alg3</b>	<b>4</b>	<b>1</b>	<b>V</b>	<b>2.72e-23</b>	<b>0.00e+00</b>	<b>1.28e-02</b>	<b>1</b>
	<b>Alg4</b>	<b>4</b>	<b>1</b>	<b>V</b>	<b>2.72e-23</b>	<b>0.00e+00</b>	<b>1.25e-02</b>	<b>1</b>
	$\ell_1$	<b>4</b>	<b>1</b>	<b>V</b>	<b>4.17e-09</b>	<b>0.00e+00</b>	<b>4.71e-02</b>	<b>1</b>
	$\ell_2$	<b>4</b>	<b>1</b>	<b>V</b>	<b>2.72e-23</b>	<b>0.00e+00</b>	<b>9.76e-03</b>	<b>2</b>
HS91	Alg1	5	1	V	1.76e-01	0.00e+00	1.56e-01	1
	<b>Alg2</b>	<b>5</b>	<b>1</b>	<b>V</b>	<b>3.41e-23</b>	<b>0.00e+00</b>	<b>1.72e-02</b>	<b>1</b>
	<b>Alg3</b>	<b>5</b>	<b>1</b>	<b>V</b>	<b>3.41e-23</b>	<b>0.00e+00</b>	<b>2.03e-02</b>	<b>1</b>
	<b>Alg4</b>	<b>5</b>	<b>1</b>	<b>V</b>	<b>3.41e-23</b>	<b>0.00e+00</b>	<b>1.27e-02</b>	<b>1</b>
	$\ell_1$	<b>5</b>	<b>1</b>	<b>V</b>	<b>8.04e-09</b>	<b>0.00e+00</b>	<b>1.12e-01</b>	<b>1</b>
	$\ell_2$	<b>5</b>	<b>1</b>	<b>V</b>	<b>3.41e-23</b>	<b>0.00e+00</b>	<b>1.32e-02</b>	<b>2</b>
HS92	Alg1	6	1	V	1.76e-01	0.00e+00	3.17e-01	1
	<b>Alg2</b>	<b>6</b>	<b>1</b>	<b>V</b>	<b>4.09e-23</b>	<b>0.00e+00</b>	<b>3.76e-02</b>	<b>1</b>
	<b>Alg3</b>	<b>6</b>	<b>1</b>	<b>V</b>	<b>4.09e-23</b>	<b>0.00e+00</b>	<b>2.80e-02</b>	<b>1</b>
	<b>Alg4</b>	<b>6</b>	<b>1</b>	<b>V</b>	<b>4.09e-23</b>	<b>0.00e+00</b>	<b>4.43e-02</b>	<b>1</b>
	$\ell_1$	<b>6</b>	<b>1</b>	<b>V</b>	<b>1.13e-08</b>	<b>0.00e+00</b>	<b>1.89e-01</b>	<b>1</b>
	$\ell_2$	<b>6</b>	<b>1</b>	<b>V</b>	<b>4.09e-23</b>	<b>0.00e+00</b>	<b>2.26e-02</b>	<b>2</b>
KIWCRESC	Alg1	3	2	V	3.27e-01	0.00e+00	1.29e-03	1
	<b>Alg2</b>	<b>3</b>	<b>2</b>	<b>V</b>	<b>5.26e-06</b>	<b>0.00e+00</b>	<b>7.78e-01</b>	<b>3</b>
	<b>Alg3</b>	<b>3</b>	<b>2</b>	<b>V</b>	<b>4.71e-06</b>	<b>0.00e+00</b>	<b>1.81e+00</b>	<b>4</b>
	<b>Alg4</b>	<b>3</b>	<b>2</b>	<b>V</b>	<b>3.87e-06</b>	<b>0.00e+00</b>	<b>7.18e-01</b>	<b>3</b>
	$\ell_1$	<b>3</b>	<b>2</b>	<b>V</b>	<b>-6.39e-08</b>	<b>1.38e-07</b>	<b>9.60e-04</b>	<b>1</b>
	$\ell_2$	<b>3</b>	<b>2</b>	<b>V</b>	<b>5.94e-06</b>	<b>0.00e+00</b>	<b>1.27e+00</b>	<b>5</b>
MADSEN	Alg1	3	6	C	-1.08e+98	0.00e+00	3.12e-01	39
	Alg2	3	6	C	-1.27e+157	0.00e+00	3.35e-02	64
	Alg3	3	6	C	-2.20e+154	0.00e+00	3.50e-02	64
	Alg4	3	6	C	-1.21e+164	0.00e+00	3.07e-02	64
	$\ell_1$	3	6	C	1.65e+287	2.26e+289	7.48e-01	64
	$\ell_2$	<b>3</b>	<b>6</b>	<b>V</b>	<b>-5.59e+19</b>	<b>0.00e+00</b>	<b>1.83e-02</b>	<b>2</b>
MAKELA1	Alg1	3	2	V	-1.08e+00	0.00e+00	7.66e-04	1
	<b>Alg2</b>	<b>3</b>	<b>2</b>	<b>V</b>	<b>-1.41e+00</b>	<b>0.00e+00</b>	<b>1.83e-03</b>	<b>3</b>
	<b>Alg3</b>	<b>3</b>	<b>2</b>	<b>V</b>	<b>-1.41e+00</b>	<b>0.00e+00</b>	<b>2.56e-03</b>	<b>4</b>
	<b>Alg4</b>	<b>3</b>	<b>2</b>	<b>V</b>	<b>-1.41e+00</b>	<b>0.00e+00</b>	<b>2.47e-03</b>	<b>3</b>
	$\ell_1$	<b>3</b>	<b>2</b>	<b>V</b>	<b>-1.41e+00</b>	<b>3.56e-08</b>	<b>7.12e-04</b>	<b>1</b>

Continua na próxima página

problema	alg	n	m	s	$f(x^k)$	$\ g(x^k)\ _\infty$	tempo	$k$
	$\ell 2$	<b>3</b>	<b>2</b>	<b>V</b>	<b>-1.41e+00</b>	<b>0.00e+00</b>	<b>2.53e-03</b>	<b>5</b>
MAKELA2	Alg1	3	3	V	7.53e+00	0.00e+00	1.69e-03	1
	<b>Alg2</b>	<b>3</b>	<b>3</b>	<b>V</b>	<b>7.20e+00</b>	<b>0.00e+00</b>	<b>1.06e+00</b>	<b>3</b>
	<b>Alg3</b>	<b>3</b>	<b>3</b>	<b>V</b>	<b>7.21e+00</b>	<b>0.00e+00</b>	<b>1.60e+00</b>	<b>4</b>
	<b>Alg4</b>	<b>3</b>	<b>3</b>	<b>V</b>	<b>7.21e+00</b>	<b>0.00e+00</b>	<b>1.01e+00</b>	<b>3</b>
	$\ell 1$	3	3	V	7.20e+00	0.00e+00	2.58e-03	2
	$\ell 2$	3	3	V	7.21e+00	0.00e+00	7.27e-01	5
MAKELA3	Alg1	21	20	V	5.70e-01	0.00e+00	3.67e-01	1
	<b>Alg2</b>	<b>21</b>	<b>20</b>	<b>V</b>	<b>4.25e-06</b>	<b>0.00e+00</b>	<b>4.83e+01</b>	<b>3</b>
	<b>Alg3</b>	<b>21</b>	<b>20</b>	<b>V</b>	<b>1.56e-06</b>	<b>0.00e+00</b>	<b>4.03e-01</b>	<b>3</b>
	<b>Alg4</b>	<b>21</b>	<b>20</b>	<b>V</b>	<b>5.72e-06</b>	<b>0.00e+00</b>	<b>3.91e+01</b>	<b>3</b>
	$\ell 1$	21	20	V	3.33e-01	0.00e+00	4.00e-02	2
	$\ell 2$	21	20	V	2.58e-06	0.00e+00	1.14e+00	4
MAKELA4	Alg1	21	40	V	6.40e-01	0.00e+00	6.35e-01	1
	<b>Alg2</b>	<b>21</b>	<b>40</b>	<b>V</b>	<b>5.53e-07</b>	<b>0.00e+00</b>	<b>7.63e-01</b>	<b>2</b>
	<b>Alg3</b>	<b>21</b>	<b>40</b>	<b>V</b>	<b>-6.66e-07</b>	<b>6.67e-07</b>	<b>1.14e+02</b>	<b>3</b>
	<b>Alg4</b>	<b>21</b>	<b>40</b>	<b>V</b>	<b>5.57e-06</b>	<b>0.00e+00</b>	<b>1.90e+02</b>	<b>3</b>
	$\ell 1$	21	40	V	2.47e-01	0.00e+00	1.03e-01	2
	$\ell 2$	21	40	V	3.59e-06	0.00e+00	8.30e+01	4
MIFFLIN1	Alg1	3	2	V	-6.60e-01	0.00e+00	6.63e-04	1
	<b>Alg2</b>	<b>3</b>	<b>2</b>	<b>V</b>	<b>-1.00e+00</b>	<b>0.00e+00</b>	<b>2.03e-03</b>	<b>3</b>
	<b>Alg3</b>	<b>3</b>	<b>2</b>	<b>V</b>	<b>-1.00e+00</b>	<b>0.00e+00</b>	<b>4.81e-03</b>	<b>4</b>
	<b>Alg4</b>	<b>3</b>	<b>2</b>	<b>V</b>	<b>-1.00e+00</b>	<b>0.00e+00</b>	<b>1.80e-03</b>	<b>3</b>
	$\ell 1$	3	2	V	-1.00e+00	1.10e-07	5.18e-04	1
	$\ell 2$	3	2	V	-1.00e+00	0.00e+00	2.23e-03	5
MIFFLIN2	Alg1	3	2	V	-7.03e-01	0.00e+00	8.87e-04	1
	<b>Alg2</b>	<b>3</b>	<b>2</b>	<b>V</b>	<b>-1.00e+00</b>	<b>0.00e+00</b>	<b>6.75e-03</b>	<b>3</b>
	<b>Alg3</b>	<b>3</b>	<b>2</b>	<b>V</b>	<b>-1.00e+00</b>	<b>0.00e+00</b>	<b>4.03e-03</b>	<b>4</b>
	<b>Alg4</b>	<b>3</b>	<b>2</b>	<b>V</b>	<b>-1.00e+00</b>	<b>0.00e+00</b>	<b>5.01e-03</b>	<b>3</b>
	$\ell 1$	3	2	V	-1.00e+00	1.56e-07	7.36e-04	1
	$\ell 2$	3	2	V	-1.00e+00	0.00e+00	3.78e-03	5
MINMAXBD	Alg1	5	20	V	1.16e+02	0.00e+00	1.61e-01	1
	<b>Alg2</b>	<b>5</b>	<b>20</b>	<b>V</b>	<b>1.16e+02</b>	<b>0.00e+00</b>	<b>3.33e+01</b>	<b>3</b>
	<b>Alg3</b>	<b>5</b>	<b>20</b>	<b>V</b>	<b>1.16e+02</b>	<b>0.00e+00</b>	<b>1.96e+01</b>	<b>3</b>
	<b>Alg4</b>	<b>5</b>	<b>20</b>	<b>V</b>	<b>1.16e+02</b>	<b>0.00e+00</b>	<b>1.74e+01</b>	<b>2</b>
	$\ell 1$	5	20	V	1.19e+02	0.00e+00	3.70e-02	2
	$\ell 2$	5	20	V	1.16e+02	0.00e+00	1.33e+01	4
MINMAXRB	Alg1	3	4	C	-7.93e+163	0.00e+00	8.31e-02	39
	Alg2	3	4	E	NaN	NaN	5.75e-03	1
	Alg3	3	4	C	-3.96e+155	0.00e+00	1.29e-02	64
	Alg4	3	4	E	NaN	NaN	7.38e-03	1
	$\ell 1$	3	4	E	NaN	NaN	1.92e-01	34
	$\ell 2$	<b>3</b>	<b>4</b>	<b>V</b>	<b>-5.74e+04</b>	<b>0.00e+00</b>	<b>7.69e-01</b>	<b>2</b>
PENTAGON	Alg1	6	15	V	8.36e-02	0.00e+00	1.52e-02	1
	<b>Alg2</b>	<b>6</b>	<b>15</b>	<b>V</b>	<b>1.37e-04</b>	<b>0.00e+00</b>	<b>9.67e+00</b>	<b>2</b>
	<b>Alg3</b>	<b>6</b>	<b>15</b>	<b>V</b>	<b>1.38e-04</b>	<b>0.00e+00</b>	<b>1.56e+01</b>	<b>4</b>
	<b>Alg4</b>	<b>6</b>	<b>15</b>	<b>V</b>	<b>1.44e-04</b>	<b>0.00e+00</b>	<b>2.14e+01</b>	<b>3</b>
	$\ell 1$	6	15	V	2.76e-04	0.00e+00	9.95e-03	1
	$\ell 2$	6	15	V	1.37e-04	0.00e+00	7.68e+00	3
POLAK1	Alg1	3	2	V	3.06e+00	0.00e+00	1.53e-03	1
	<b>Alg2</b>	<b>3</b>	<b>2</b>	<b>V</b>	<b>2.72e+00</b>	<b>0.00e+00</b>	<b>3.57e-03</b>	<b>3</b>
	<b>Alg3</b>	<b>3</b>	<b>2</b>	<b>V</b>	<b>2.72e+00</b>	<b>0.00e+00</b>	<b>3.30e-03</b>	<b>3</b>
	<b>Alg4</b>	<b>3</b>	<b>2</b>	<b>V</b>	<b>2.72e+00</b>	<b>0.00e+00</b>	<b>5.48e-03</b>	<b>3</b>
	$\ell 1$	3	2	V	2.72e+00	3.20e-08	6.27e-04	1
	$\ell 2$	3	2	V	2.72e+00	0.00e+00	7.75e-03	5
POLAK2	Alg1	11	2	V	5.49e+01	0.00e+00	1.49e-02	1
	<b>Alg2</b>	<b>11</b>	<b>2</b>	<b>V</b>	<b>5.46e+01</b>	<b>0.00e+00</b>	<b>6.53e-03</b>	<b>2</b>
	<b>Alg3</b>	<b>11</b>	<b>2</b>	<b>V</b>	<b>5.46e+01</b>	<b>0.00e+00</b>	<b>8.66e-03</b>	<b>3</b>

Continua na próxima página

problema	alg	n	m	s	$f(x^k)$	$\ g(x^k)\ _\infty$	tempo	k
	<b>Alg4</b>	11	2	V	5.46e+01	0.00e+00	2.18e-02	3
	$\ell_1$	11	2	V	5.46e+01	0.00e+00	9.57e-03	3
	$\ell_2$	11	2	V	5.46e+01	0.00e+00	1.09e-02	4
POLAK3	Alg1	12	10	V	6.30e+00	0.00e+00	7.27e-01	1
	<b>Alg2</b>	12	10	V	5.94e+00	0.00e+00	6.80e+01	3
	<b>Alg3</b>	12	10	V	5.94e+00	0.00e+00	9.20e+01	4
	<b>Alg4</b>	12	10	V	5.93e+00	0.00e+00	5.56e+01	3
	$\ell_1$	12	10	V	5.95e+00	0.00e+00	1.25e-01	2
	$\ell_2$	12	10	V	5.93e+00	0.00e+00	1.51e+00	4
POLAK4	Alg1	3	3	V	3.40e-01	0.00e+00	1.84e-01	1
	<b>Alg2</b>	3	3	V	2.07e-05	0.00e+00	1.65e+00	3
	<b>Alg3</b>	3	3	V	2.05e-05	0.00e+00	2.58e+00	4
	<b>Alg4</b>	3	3	V	1.98e-05	0.00e+00	1.29e+00	3
	$\ell_1$	3	3	V	1.53e-07	8.88e-08	6.08e-01	2
	$\ell_2$	3	3	V	2.22e-05	0.00e+00	2.81e+00	5
POLAK5	Alg1	3	2	V	5.03e+01	0.00e+00	7.44e-03	1
	<b>Alg2</b>	3	2	V	5.00e+01	0.00e+00	1.61e-02	2
	<b>Alg3</b>	3	2	V	5.00e+01	0.00e+00	1.06e-02	3
	<b>Alg4</b>	3	2	V	5.00e+01	0.00e+00	1.75e-02	3
	$\ell_1$	3	2	V	5.00e+01	1.99e-08	3.02e-03	2
	$\ell_2$	3	2	V	5.00e+01	0.00e+00	1.54e-02	4
POLAK6	Alg1	5	4	V	-4.36e+01	0.00e+00	1.28e-02	1
	<b>Alg2</b>	5	4	V	-4.40e+01	0.00e+00	2.21e+00	3
	<b>Alg3</b>	5	4	V	-4.40e+01	0.00e+00	2.62e+00	3
	<b>Alg4</b>	5	4	V	-4.40e+01	0.00e+00	3.70e+00	3
	$\ell_1$	5	4	V	-1.78e+01	0.00e+00	4.14e-03	1
	$\ell_2$	5	4	V	-4.40e+01	0.00e+00	1.79e+00	4
ROSENMMX	Alg1	5	4	V	-4.36e+01	0.00e+00	9.95e-03	1
	<b>Alg2</b>	5	4	V	-4.40e+01	0.00e+00	1.96e+00	3
	<b>Alg3</b>	5	4	V	-4.40e+01	0.00e+00	2.33e+00	3
	<b>Alg4</b>	5	4	V	-4.40e+01	0.00e+00	1.99e+00	3
	$\ell_1$	5	4	V	-4.40e+01	7.71e-08	1.22e-02	2
	$\ell_2$	5	4	V	-4.39e+01	0.00e+00	2.85e+00	5
S268	Alg1	5	5	C	7.88e+05	1.85e+01	3.06e+01	39
	Alg2	5	5	C	3.05e+05	2.20e+01	4.14e+01	64
	Alg3	5	5	C	1.49e+05	2.70e+01	5.13e+01	64
	Alg4	5	5	C	6.85e+04	2.64e+01	3.92e+01	64
	$\ell_1$	5	5	C	1.86e+37	3.66e+18	9.97e-01	64
	$\ell_2$	5	5	C	7.15e+04	1.45e+01	3.73e-01	14
SNAKE	<b>Alg1</b>	2	2	V	1.00e+00	0.00e+00	9.12e-04	1
	<b>Alg2</b>	2	2	V	1.00e+00	0.00e+00	7.70e-04	1
	<b>Alg3</b>	2	2	V	1.00e+00	0.00e+00	7.58e-04	1
	<b>Alg4</b>	2	2	V	1.00e+00	0.00e+00	1.24e-03	1
	$\ell_1$	2	2	E	$\infty$	NaN	1.87e-02	4
	$\ell_2$	2	2	V	1.00e+00	0.00e+00	5.84e-04	2
SPIRAL	Alg1	3	2	C	1.00e+00	8.75e-01	8.57e-02	39
	Alg2	3	2	C	1.00e+00	8.75e-01	7.39e-02	64
	Alg3	3	2	C	1.00e+00	8.75e-01	8.12e-02	64
	Alg4	3	2	C	1.00e+00	8.75e-01	2.66e-02	64
	$\ell_1$	3	2	E	-2.18e+154	NaN	4.01e-02	3
	$\ell_2$	3	2	C	1.00e+00	8.75e-01	1.18e-02	14
WOMFLET	Alg1	3	3	C	-5.05e+114	0.00e+00	6.23e-02	39
	Alg2	3	3	C	-3.61e+160	0.00e+00	1.16e-02	64
	Alg3	3	3	C	-2.22e+158	0.00e+00	2.46e-02	64
	Alg4	3	3	C	-3.76e+174	0.00e+00	9.73e-03	64
	$\ell_1$	3	3	E	NaN	NaN	1.12e-01	30
	$\ell_2$	3	3	C	-2.28e+23	0.00e+00	1.37e-02	14

## A.2 Resultados com o critério de parada (4.3)

problema	alg	n	m	s	$f(x^k)$	$\ g(x^k)_+\ _\infty$	tempo	$k$
CB2	Alg1	3	3	C	1.00e+00	1.00e+00	3.06e-02	39
	Alg2	3	3	C	1.00e+00	1.00e+00	5.04e-02	64
	Alg3	3	3	C	1.00e+00	1.00e+00	4.95e-02	64
	Alg4	3	3	C	1.00e+00	1.00e+00	2.80e-02	64
	$\ell_1$	3	3	E	-4.89e+133	$\infty$	3.86e-03	1
	$\ell_2$	3	3	C	1.00e+00	1.00e+00	1.23e-02	14
CB3	Alg1	3	3	C	1.00e+00	1.00e+00	3.10e-02	39
	Alg2	3	3	C	1.00e+00	1.00e+00	6.43e-02	64
	Alg3	3	3	C	1.00e+00	1.00e+00	5.49e-02	64
	Alg4	3	3	C	1.00e+00	1.00e+00	2.60e-02	64
	$\ell_1$	3	3	E	-4.89e+133	$\infty$	4.10e-03	1
	$\ell_2$	3	3	C	1.00e+00	1.00e+00	1.19e-02	14
CHACONN1	Alg1	3	3	V	2.29e+00	0.00e+00	1.49e-03	1
	<b>Alg2</b>	<b>3</b>	<b>3</b>	<b>V</b>	<b>1.95e+00</b>	<b>0.00e+00</b>	<b>2.61e-01</b>	<b>3</b>
	<b>Alg3</b>	<b>3</b>	<b>3</b>	<b>V</b>	<b>1.95e+00</b>	<b>0.00e+00</b>	<b>8.27e-01</b>	<b>4</b>
	<b>Alg4</b>	<b>3</b>	<b>3</b>	<b>V</b>	<b>1.95e+00</b>	<b>0.00e+00</b>	<b>2.60e-01</b>	<b>3</b>
	$\ell_1$	3	3	V	1.95e+00	0.00e+00	9.91e-04	1
	$\ell_2$	3	3	V	1.95e+00	0.00e+00	1.14e+00	5
CHACONN2	Alg1	3	3	V	2.37e+00	0.00e+00	1.05e-03	1
	<b>Alg2</b>	<b>3</b>	<b>3</b>	<b>V</b>	<b>2.00e+00</b>	<b>0.00e+00</b>	<b>5.71e-03</b>	<b>3</b>
	<b>Alg3</b>	<b>3</b>	<b>3</b>	<b>V</b>	<b>2.00e+00</b>	<b>1.83e-07</b>	<b>8.13e-03</b>	<b>3</b>
	<b>Alg4</b>	<b>3</b>	<b>3</b>	<b>V</b>	<b>2.00e+00</b>	<b>0.00e+00</b>	<b>4.67e-03</b>	<b>3</b>
	$\ell_1$	3	3	V	2.00e+00	4.62e-10	7.14e-04	1
	$\ell_2$	3	3	V	2.00e+00	0.00e+00	1.31e-02	5
CONGIGMZ	Alg1	3	5	C	-6.54e+26	0.00e+00	4.87e-02	39
	<b>Alg2</b>	<b>3</b>	<b>5</b>	<b>V</b>	<b>-2.61e+08</b>	<b>0.00e+00</b>	<b>1.19e-02</b>	<b>5</b>
	Alg3	3	5	V	-4.00e+01	0.00e+00	3.19e-03	2
	Alg4	3	5	V	-1.08e+02	0.00e+00	6.30e-03	4
	$\ell_1$	3	5	E	-4.62e+155	$\infty$	1.36e-02	2
	$\ell_2$	3	5	C	1.18e+00	9.27e+00	3.14e-02	14
DEMYMALO	Alg1	3	3	V	-2.62e+00	0.00e+00	9.04e-04	1
	<b>Alg2</b>	<b>3</b>	<b>3</b>	<b>V</b>	<b>-3.00e+00</b>	<b>0.00e+00</b>	<b>4.38e-03</b>	<b>3</b>
	<b>Alg3</b>	<b>3</b>	<b>3</b>	<b>V</b>	<b>-3.00e+00</b>	<b>0.00e+00</b>	<b>3.16e-03</b>	<b>3</b>
	<b>Alg4</b>	<b>3</b>	<b>3</b>	<b>V</b>	<b>-3.00e+00</b>	<b>0.00e+00</b>	<b>2.97e-03</b>	<b>3</b>
	$\ell_1$	3	3	V	-3.00e+00	0.00e+00	6.26e-04	1
	$\ell_2$	3	3	V	-3.00e+00	1.06e-07	4.11e-03	4
DIPIGRI	Alg1	7	4	V	6.81e+02	0.00e+00	3.99e-03	1
	Alg2	7	4	V	6.81e+02	0.00e+00	3.52e+00	3
	Alg3	7	4	V	6.81e+02	0.00e+00	2.40e+00	3
	Alg4	7	4	V	6.81e+02	0.00e+00	2.96e+00	3
	$\ell_1$	7	4	V	6.81e+02	0.00e+00	1.73e-02	3
	$\ell_2$	7	4	V	6.81e+02	0.00e+00	2.21e+00	4
EXPFITA	Alg1	5	22	C	1.08e+04	1.00e+00	6.08e+01	39
	Alg2	5	22	C	1.67e+03	1.00e+00	1.97e+02	64
	Alg3	5	22	C	2.60e+00	1.00e+00	1.12e+02	64
	Alg4	5	22	C	5.54e+00	1.00e+00	1.55e+02	64
	$\ell_1$	5	22	C	1.30e+43	5.34e+23	2.97e+00	64
	$\ell_2$	5	22	C	3.67e+02	1.00e+00	4.08e+01	14
GIGOMEZ1	Alg1	3	3	V	1.35e+00	0.00e+00	2.29e-03	2
	<b>Alg2</b>	<b>3</b>	<b>3</b>	<b>V</b>	<b>-7.54e+00</b>	<b>0.00e+00</b>	<b>2.43e-03</b>	<b>2</b>
	<b>Alg3</b>	<b>3</b>	<b>3</b>	<b>V</b>	<b>-7.54e+00</b>	<b>0.00e+00</b>	<b>2.38e-03</b>	<b>2</b>
	<b>Alg4</b>	<b>3</b>	<b>3</b>	<b>V</b>	<b>-7.54e+00</b>	<b>0.00e+00</b>	<b>2.19e-03</b>	<b>2</b>
	$\ell_1$	3	3	E	-1.46e+267	$\infty$	6.94e-03	2
	$\ell_2$	3	3	V	1.39e+00	0.00e+00	3.70e-03	3
GIGOMEZ2	Alg1	3	3	C	2.00e+00	2.00e+00	3.07e-02	39

Continua na próxima página

problema	alg	n	m	s	$f(x^k)$	$\ g(x^k)\ _\infty$	tempo	$k$
	Alg2	3	3	C	2.00e+00	2.00e+00	5.07e-02	64
	Alg3	3	3	C	2.00e+00	2.00e+00	4.89e-02	64
	Alg4	3	3	C	2.00e+00	2.00e+00	2.58e-02	64
	$\ell_1$	3	3	E	-9.55e+133	$\infty$	3.87e-03	1
	$\ell_2$	3	3	C	2.00e+00	2.00e+00	1.88e-02	14
GIGOMEZ3	Alg1	3	3	C	2.00e+00	2.00e+00	3.08e-02	39
	Alg2	3	3	C	2.00e+00	2.00e+00	5.00e-02	64
	Alg3	3	3	C	2.00e+00	2.00e+00	4.88e-02	64
	Alg4	3	3	C	2.00e+00	2.00e+00	3.10e-02	64
	$\ell_1$	3	3	E	-9.55e+133	$\infty$	4.01e-03	1
	$\ell_2$	3	3	C	2.00e+00	2.00e+00	1.30e-02	14
GOFFIN	Alg1	51	50	V	6.86e-01	0.00e+00	6.00e+02	1
	<b>Alg2</b>	<b>51</b>	<b>50</b>	<b>V</b>	<b>4.06e-04</b>	<b>0.00e+00</b>	<b>1.20e+03</b>	<b>2</b>
	Alg3	51	50	T	2.09e-03	9.07e-04	1.20e+03	3
	<b>Alg4</b>	<b>51</b>	<b>50</b>	<b>V</b>	<b>4.27e-04</b>	<b>0.00e+00</b>	<b>1.20e+03</b>	<b>2</b>
	$\ell_1$	51	50	V	6.74e+01	0.00e+00	1.93e+00	8
	$\ell_2$	<b>51</b>	<b>50</b>	<b>V</b>	<b>8.76e-03</b>	<b>0.00e+00</b>	<b>6.00e+02</b>	<b>4</b>
HAIFAS	Alg1	13	9	V	-6.78e-02	0.00e+00	5.02e-02	1
	<b>Alg2</b>	<b>13</b>	<b>9</b>	<b>V</b>	<b>-4.50e-01</b>	<b>0.00e+00</b>	<b>1.05e+01</b>	<b>3</b>
	<b>Alg3</b>	<b>13</b>	<b>9</b>	<b>V</b>	<b>-4.50e-01</b>	<b>0.00e+00</b>	<b>1.86e+01</b>	<b>4</b>
	<b>Alg4</b>	<b>13</b>	<b>9</b>	<b>V</b>	<b>-4.49e-01</b>	<b>0.00e+00</b>	<b>2.91e+01</b>	<b>4</b>
	$\ell_1$	13	9	V	-4.50e-01	1.57e-07	1.01e-02	1
	$\ell_2$	13	9	V	-4.49e-01	7.59e-07	3.21e+01	6
HALDMADS	Alg1	6	42	V	6.47e-01	0.00e+00	2.15e-01	1
	<b>Alg2</b>	<b>6</b>	<b>42</b>	<b>V</b>	<b>2.18e-04</b>	<b>0.00e+00</b>	<b>4.85e+01</b>	<b>3</b>
	<b>Alg3</b>	<b>6</b>	<b>42</b>	<b>V</b>	<b>2.52e-04</b>	<b>8.29e-08</b>	<b>8.48e+01</b>	<b>4</b>
	<b>Alg4</b>	<b>6</b>	<b>42</b>	<b>V</b>	<b>1.35e-04</b>	<b>0.00e+00</b>	<b>8.33e+01</b>	<b>3</b>
	$\ell_1$	6	42	V	1.93e-02	0.00e+00	2.55e-01	3
	$\ell_2$	6	42	V	2.29e+00	0.00e+00	5.61e-01	5
HS10	Alg1	2	1	V	-2.00e+01	0.00e+00	3.00e-02	1
	Alg2	2	1	V	-2.00e+01	0.00e+00	3.06e-02	1
	Alg3	2	1	V	-2.00e+01	0.00e+00	3.02e-02	1
	Alg4	2	1	V	-2.00e+01	0.00e+00	4.87e-02	1
	$\ell_1$	2	1	E	-6.72e+227	$\infty$	1.87e-03	1
	$\ell_2$	2	1	V	-2.00e+01	0.00e+00	3.14e-02	2
HS100	Alg1	7	4	V	1.64e+02	0.00e+00	8.31e-03	1
	Alg2	7	4	V	1.63e+02	0.00e+00	5.27e+00	4
	Alg3	7	4	V	1.63e+02	0.00e+00	4.83e+00	4
	Alg4	7	4	V	1.63e+02	0.00e+00	5.22e+00	4
	$\ell_1$	7	4	V	1.63e+02	6.62e-07	1.00e-02	3
	$\ell_2$	7	4	V	1.63e+02	0.00e+00	2.89e+00	4
HS100MOD	Alg1	7	4	V	2.40e+02	0.00e+00	1.73e-02	1
	Alg2	7	4	V	2.41e+02	0.00e+00	7.48e+00	4
	Alg3	7	4	V	2.40e+02	0.00e+00	7.42e+00	3
	Alg4	7	4	V	2.41e+02	0.00e+00	6.74e+00	4
	$\ell_1$	7	4	V	2.40e+02	0.00e+00	1.35e-02	3
	$\ell_2$	7	4	V	2.42e+02	7.89e-08	1.84e+01	8
HS11	Alg1	2	1	V	-2.50e+01	0.00e+00	7.39e-05	1
	Alg2	2	1	V	-2.50e+01	0.00e+00	6.39e-05	1
	Alg3	2	1	V	-2.50e+01	0.00e+00	6.79e-05	1
	Alg4	2	1	V	-2.50e+01	0.00e+00	8.30e-05	1
	$\ell_1$	2	1	V	-2.50e+01	0.00e+00	1.72e-04	1
	$\ell_2$	2	1	V	-2.50e+01	0.00e+00	7.58e-05	2
HS113	Alg1	10	8	V	-3.10e+01	0.00e+00	5.26e-02	1
	Alg2	10	8	V	-3.10e+01	0.00e+00	4.14e-02	1
	Alg3	10	8	V	-3.10e+01	0.00e+00	4.13e-02	1
	Alg4	10	8	V	-3.10e+01	0.00e+00	4.94e-02	1
	$\ell_1$	10	8	V	-3.10e+01	0.00e+00	7.30e-03	1

Continua na próxima página

problema	alg	n	m	s	$f(x^k)$	$\ g(x^k)\ _\infty$	tempo	$k$
	$\ell 2$	10	8	V	-3.10e+01	0.00e+00	9.32e-02	2
HS12	Alg1	2	1	V	-1.22e+02	0.00e+00	9.01e-05	1
	Alg2	2	1	V	-1.22e+02	0.00e+00	8.32e-05	1
	Alg3	2	1	V	-1.22e+02	0.00e+00	9.51e-05	1
	Alg4	2	1	V	-1.22e+02	0.00e+00	1.21e-04	1
	$\ell 1$	2	1	V	-1.22e+02	0.00e+00	1.48e-04	1
	$\ell 2$	2	1	V	-1.22e+02	0.00e+00	1.03e-04	2
HS22	Alg1	2	2	V	2.29e-07	0.00e+00	1.30e-04	1
	Alg2	2	2	V	1.97e-31	0.00e+00	1.06e-04	1
	Alg3	2	2	V	6.52e-21	0.00e+00	8.80e-05	1
	Alg4	2	2	V	7.85e-24	0.00e+00	8.49e-05	1
	$\ell 1$	2	2	V	4.97e-10	0.00e+00	1.30e-04	1
	$\ell 2$	2	2	V	1.97e-31	0.00e+00	8.39e-05	2
HS268	Alg1	5	5	C	7.88e+05	1.85e+01	2.01e+01	39
	Alg2	5	5	C	3.05e+05	2.20e+01	3.46e+01	64
	Alg3	5	5	C	1.49e+05	2.70e+01	3.53e+01	64
	Alg4	5	5	C	6.85e+04	2.64e+01	3.39e+01	64
	$\ell 1$	5	5	C	1.86e+37	3.66e+18	1.13e+00	64
	$\ell 2$	5	5	C	7.15e+04	1.45e+01	4.06e-01	14
HS29	Alg1	3	1	C	4.56e-02	4.70e+01	1.93e-02	39
	Alg2	3	1	C	4.56e-02	4.70e+01	3.09e-02	64
	Alg3	3	1	C	4.56e-02	4.70e+01	6.08e-02	64
	Alg4	3	1	C	-1.00e+00	4.10e+01	1.75e-02	64
	$\ell 1$	3	1	C	$-\infty$	0.00e+00	2.86e-03	64
	$\ell 2$	3	1	V	-2.03e+01	0.00e+00	5.55e-03	12
HS43	Alg1	4	3	V	-7.99e+01	0.00e+00	3.09e-02	1
	Alg2	4	3	V	-7.99e+01	0.00e+00	3.12e-02	1
	Alg3	4	3	V	-7.99e+01	0.00e+00	3.08e-02	1
	Alg4	4	3	V	-7.99e+01	0.00e+00	3.09e-02	1
	$\ell 1$	4	3	V	-7.99e+01	0.00e+00	7.22e-04	1
	$\ell 2$	4	3	V	-7.99e+01	0.00e+00	3.48e-02	2
HS88	Alg1	2	1	V	1.76e-01	0.00e+00	1.12e-02	1
	Alg2	2	1	V	1.36e-23	0.00e+00	2.31e-03	1
	Alg3	2	1	V	1.36e-23	0.00e+00	2.31e-03	1
	Alg4	2	1	V	1.36e-23	0.00e+00	2.23e-03	1
	$\ell 1$	2	1	V	7.75e-09	0.00e+00	6.84e-03	1
	$\ell 2$	2	1	V	1.36e-23	0.00e+00	2.67e-03	2
HS89	Alg1	3	1	V	1.76e-01	0.00e+00	2.77e-02	1
	Alg2	3	1	V	2.04e-23	0.00e+00	4.71e-03	1
	Alg3	3	1	V	2.04e-23	0.00e+00	4.88e-03	1
	Alg4	3	1	V	2.04e-23	0.00e+00	4.69e-03	1
	$\ell 1$	3	1	V	3.08e-09	0.00e+00	2.09e-02	1
	$\ell 2$	3	1	V	2.04e-23	0.00e+00	5.34e-03	2
HS90	Alg1	4	1	V	1.76e-01	0.00e+00	5.74e-02	1
	Alg2	4	1	V	2.72e-23	0.00e+00	8.13e-03	1
	Alg3	4	1	V	2.72e-23	0.00e+00	8.17e-03	1
	Alg4	4	1	V	2.72e-23	0.00e+00	9.49e-03	1
	$\ell 1$	4	1	V	4.17e-09	0.00e+00	4.50e-02	1
	$\ell 2$	4	1	V	2.72e-23	0.00e+00	1.03e-02	2
HS91	Alg1	5	1	V	1.76e-01	0.00e+00	1.03e-01	1
	Alg2	5	1	V	3.41e-23	0.00e+00	1.39e-02	1
	Alg3	5	1	V	3.41e-23	0.00e+00	1.27e-02	1
	Alg4	5	1	V	3.41e-23	0.00e+00	1.61e-02	1
	$\ell 1$	5	1	V	8.04e-09	0.00e+00	8.73e-02	1
	$\ell 2$	5	1	V	3.41e-23	0.00e+00	1.57e-02	2
HS92	Alg1	6	1	V	1.76e-01	0.00e+00	2.10e-01	1
	Alg2	6	1	V	4.09e-23	0.00e+00	2.15e-02	1
	Alg3	6	1	V	4.09e-23	0.00e+00	2.14e-02	1

Continua na próxima página

problema	alg	n	m	s	$f(x^k)$	$\ g(x^k)\ _\infty$	tempo	$k$
	<b>Alg4</b>	<b>6</b>	<b>1</b>	<b>V</b>	<b>4.09e-23</b>	<b>0.00e+00</b>	<b>2.13e-02</b>	<b>1</b>
	$\ell_1$	<b>6</b>	<b>1</b>	<b>V</b>	<b>1.13e-08</b>	<b>0.00e+00</b>	<b>1.57e-01</b>	<b>1</b>
	$\ell_2$	<b>6</b>	<b>1</b>	<b>V</b>	<b>4.09e-23</b>	<b>0.00e+00</b>	<b>2.49e-02</b>	<b>2</b>
KIWCRES	Alg1	3	2	V	3.27e-01	0.00e+00	6.90e-04	1
	<b>Alg2</b>	<b>3</b>	<b>2</b>	<b>V</b>	<b>5.26e-06</b>	<b>0.00e+00</b>	<b>6.14e-01</b>	<b>3</b>
	<b>Alg3</b>	<b>3</b>	<b>2</b>	<b>V</b>	<b>4.71e-06</b>	<b>0.00e+00</b>	<b>1.17e+00</b>	<b>4</b>
	<b>Alg4</b>	<b>3</b>	<b>2</b>	<b>V</b>	<b>3.87e-06</b>	<b>0.00e+00</b>	<b>6.42e-01</b>	<b>3</b>
	$\ell_1$	<b>3</b>	<b>2</b>	<b>V</b>	<b>-6.39e-08</b>	<b>1.38e-07</b>	<b>9.79e-04</b>	<b>1</b>
	$\ell_2$	<b>3</b>	<b>2</b>	<b>V</b>	<b>5.94e-06</b>	<b>0.00e+00</b>	<b>1.34e+00</b>	<b>5</b>
MADSEN	Alg1	3	6	C	-1.08e+98	0.00e+00	2.06e-01	39
	Alg2	3	6	C	-1.27e+157	0.00e+00	5.82e-02	64
	Alg3	3	6	C	-2.20e+154	0.00e+00	2.78e-02	64
	Alg4	3	6	C	-1.21e+164	0.00e+00	2.79e-02	64
	$\ell_1$	3	6	C	1.65e+287	2.26e+289	7.02e-01	64
	$\ell_2$	<b>3</b>	<b>6</b>	<b>V</b>	<b>-5.59e+19</b>	<b>0.00e+00</b>	<b>1.45e-02</b>	<b>2</b>
MAKELA1	Alg1	3	2	V	-1.08e+00	0.00e+00	5.26e-04	1
	<b>Alg2</b>	<b>3</b>	<b>2</b>	<b>V</b>	<b>-1.41e+00</b>	<b>0.00e+00</b>	<b>1.50e-03</b>	<b>3</b>
	<b>Alg3</b>	<b>3</b>	<b>2</b>	<b>V</b>	<b>-1.41e+00</b>	<b>0.00e+00</b>	<b>1.96e-03</b>	<b>4</b>
	<b>Alg4</b>	<b>3</b>	<b>2</b>	<b>V</b>	<b>-1.41e+00</b>	<b>0.00e+00</b>	<b>1.75e-03</b>	<b>3</b>
	$\ell_1$	<b>3</b>	<b>2</b>	<b>V</b>	<b>-1.41e+00</b>	<b>3.56e-08</b>	<b>3.54e-04</b>	<b>1</b>
	$\ell_2$	<b>3</b>	<b>2</b>	<b>V</b>	<b>-1.41e+00</b>	<b>0.00e+00</b>	<b>2.53e-03</b>	<b>5</b>
MAKELA2	Alg1	3	3	V	7.53e+00	0.00e+00	1.10e-03	1
	<b>Alg2</b>	<b>3</b>	<b>3</b>	<b>V</b>	<b>7.20e+00</b>	<b>0.00e+00</b>	<b>8.41e-01</b>	<b>3</b>
	<b>Alg3</b>	<b>3</b>	<b>3</b>	<b>V</b>	<b>7.21e+00</b>	<b>0.00e+00</b>	<b>1.02e+00</b>	<b>4</b>
	<b>Alg4</b>	<b>3</b>	<b>3</b>	<b>V</b>	<b>7.21e+00</b>	<b>0.00e+00</b>	<b>9.16e-01</b>	<b>3</b>
	$\ell_1$	<b>3</b>	<b>3</b>	<b>V</b>	<b>7.20e+00</b>	<b>0.00e+00</b>	<b>1.66e-03</b>	<b>2</b>
	$\ell_2$	<b>3</b>	<b>3</b>	<b>V</b>	<b>7.21e+00</b>	<b>0.00e+00</b>	<b>7.03e-01</b>	<b>5</b>
MAKELA3	Alg1	21	20	V	5.70e-01	0.00e+00	2.31e-01	1
	<b>Alg2</b>	<b>21</b>	<b>20</b>	<b>V</b>	<b>4.25e-06</b>	<b>0.00e+00</b>	<b>4.25e+01</b>	<b>3</b>
	<b>Alg3</b>	<b>21</b>	<b>20</b>	<b>V</b>	<b>1.56e-06</b>	<b>0.00e+00</b>	<b>2.72e-01</b>	<b>3</b>
	<b>Alg4</b>	<b>21</b>	<b>20</b>	<b>V</b>	<b>5.72e-06</b>	<b>0.00e+00</b>	<b>3.84e+01</b>	<b>3</b>
	$\ell_1$	21	20	V	3.33e-01	0.00e+00	4.22e-02	2
	$\ell_2$	<b>21</b>	<b>20</b>	<b>V</b>	<b>2.58e-06</b>	<b>0.00e+00</b>	<b>1.07e+00</b>	<b>4</b>
MAKELA4	Alg1	21	40	V	6.40e-01	0.00e+00	4.30e-01	1
	<b>Alg2</b>	<b>21</b>	<b>40</b>	<b>V</b>	<b>5.53e-07</b>	<b>0.00e+00</b>	<b>6.40e-01</b>	<b>2</b>
	<b>Alg3</b>	<b>21</b>	<b>40</b>	<b>V</b>	<b>-6.66e-07</b>	<b>6.67e-07</b>	<b>9.61e+01</b>	<b>3</b>
	<b>Alg4</b>	<b>21</b>	<b>40</b>	<b>V</b>	<b>5.57e-06</b>	<b>0.00e+00</b>	<b>1.29e+02</b>	<b>3</b>
	$\ell_1$	21	40	V	2.47e-01	0.00e+00	1.01e-01	2
	$\ell_2$	<b>21</b>	<b>40</b>	<b>V</b>	<b>3.59e-06</b>	<b>0.00e+00</b>	<b>8.38e+01</b>	<b>4</b>
MIFFLIN1	Alg1	3	2	V	-6.60e-01	0.00e+00	4.00e-04	1
	<b>Alg2</b>	<b>3</b>	<b>2</b>	<b>V</b>	<b>-1.00e+00</b>	<b>0.00e+00</b>	<b>1.59e-03</b>	<b>3</b>
	<b>Alg3</b>	<b>3</b>	<b>2</b>	<b>V</b>	<b>-1.00e+00</b>	<b>0.00e+00</b>	<b>1.68e-03</b>	<b>4</b>
	<b>Alg4</b>	<b>3</b>	<b>2</b>	<b>V</b>	<b>-1.00e+00</b>	<b>0.00e+00</b>	<b>1.40e-03</b>	<b>3</b>
	$\ell_1$	<b>3</b>	<b>2</b>	<b>V</b>	<b>-1.00e+00</b>	<b>1.10e-07</b>	<b>5.12e-04</b>	<b>1</b>
	$\ell_2$	<b>3</b>	<b>2</b>	<b>V</b>	<b>-1.00e+00</b>	<b>0.00e+00</b>	<b>1.78e-03</b>	<b>5</b>
MIFFLIN2	Alg1	3	2	V	-7.03e-01	0.00e+00	6.14e-04	1
	<b>Alg2</b>	<b>3</b>	<b>2</b>	<b>V</b>	<b>-1.00e+00</b>	<b>0.00e+00</b>	<b>5.00e-03</b>	<b>3</b>
	<b>Alg3</b>	<b>3</b>	<b>2</b>	<b>V</b>	<b>-1.00e+00</b>	<b>0.00e+00</b>	<b>2.57e-03</b>	<b>4</b>
	<b>Alg4</b>	<b>3</b>	<b>2</b>	<b>V</b>	<b>-1.00e+00</b>	<b>0.00e+00</b>	<b>4.42e-03</b>	<b>3</b>
	$\ell_1$	<b>3</b>	<b>2</b>	<b>V</b>	<b>-1.00e+00</b>	<b>1.56e-07</b>	<b>7.28e-04</b>	<b>1</b>
	$\ell_2$	<b>3</b>	<b>2</b>	<b>V</b>	<b>-1.00e+00</b>	<b>0.00e+00</b>	<b>4.07e-03</b>	<b>5</b>
MINMAXBD	Alg1	5	20	V	1.16e+02	0.00e+00	1.03e-01	1
	Alg2	5	20	V	1.16e+02	0.00e+00	2.79e+01	3
	Alg3	5	20	V	1.16e+02	0.00e+00	1.33e+01	3
	Alg4	5	20	V	1.16e+02	0.00e+00	1.72e+01	2
	$\ell_1$	5	20	V	1.19e+02	0.00e+00	5.50e-02	2
	$\ell_2$	<b>5</b>	<b>20</b>	<b>V</b>	<b>1.16e+02</b>	<b>0.00e+00</b>	<b>1.42e+01</b>	<b>4</b>
MINMAXRB	Alg1	3	4	C	-7.93e+163	0.00e+00	4.25e-02	39

Continua na próxima página

problema	alg	n	m	s	$f(x^k)$	$\ g(x^k)\ _\infty$	tempo	$k$
	Alg2	3	4	E	NaN	NaN	5.15e-03	1
	Alg3	3	4	C	-3.96e+155	0.00e+00	1.68e-02	64
	Alg4	3	4	E	NaN	NaN	5.37e-03	1
	$\ell 1$	3	4	E	NaN	NaN	1.46e-01	34
	$\ell 2$	<b>3</b>	<b>4</b>	<b>V</b>	<b>-5.74e+04</b>	<b>0.00e+00</b>	<b>8.17e-01</b>	<b>2</b>
PENTAGON	Alg1	6	15	V	8.36e-02	0.00e+00	9.76e-03	1
	<b>Alg2</b>	<b>6</b>	<b>15</b>	<b>V</b>	<b>1.37e-04</b>	<b>0.00e+00</b>	<b>8.06e+00</b>	<b>2</b>
	<b>Alg3</b>	<b>6</b>	<b>15</b>	<b>V</b>	<b>1.38e-04</b>	<b>0.00e+00</b>	<b>1.34e+01</b>	<b>4</b>
	<b>Alg4</b>	<b>6</b>	<b>15</b>	<b>V</b>	<b>1.44e-04</b>	<b>0.00e+00</b>	<b>1.42e+01</b>	<b>3</b>
	$\ell 1$	6	15	V	2.76e-04	0.00e+00	8.24e-03	1
	$\ell 2$	6	15	V	1.37e-04	0.00e+00	7.52e+00	3
POLAK1	Alg1	3	2	V	3.06e+00	0.00e+00	1.00e-03	1
	<b>Alg2</b>	<b>3</b>	<b>2</b>	<b>V</b>	<b>2.72e+00</b>	<b>0.00e+00</b>	<b>2.46e-03</b>	<b>3</b>
	<b>Alg3</b>	<b>3</b>	<b>2</b>	<b>V</b>	<b>2.72e+00</b>	<b>0.00e+00</b>	<b>5.08e-03</b>	<b>3</b>
	<b>Alg4</b>	<b>3</b>	<b>2</b>	<b>V</b>	<b>2.72e+00</b>	<b>0.00e+00</b>	<b>1.98e-03</b>	<b>3</b>
	$\ell 1$	3	2	V	2.72e+00	3.20e-08	5.31e-04	1
	$\ell 2$	3	2	V	2.72e+00	0.00e+00	4.23e-03	5
POLAK2	Alg1	11	2	V	5.49e+01	0.00e+00	6.99e-03	1
	<b>Alg2</b>	<b>11</b>	<b>2</b>	<b>V</b>	<b>5.46e+01</b>	<b>0.00e+00</b>	<b>5.16e-03</b>	<b>2</b>
	<b>Alg3</b>	<b>11</b>	<b>2</b>	<b>V</b>	<b>5.46e+01</b>	<b>0.00e+00</b>	<b>7.91e-03</b>	<b>3</b>
	<b>Alg4</b>	<b>11</b>	<b>2</b>	<b>V</b>	<b>5.46e+01</b>	<b>0.00e+00</b>	<b>7.80e-03</b>	<b>3</b>
	$\ell 1$	11	2	V	5.46e+01	0.00e+00	7.31e-03	3
	$\ell 2$	11	2	V	5.46e+01	0.00e+00	1.20e-02	4
POLAK3	Alg1	12	10	V	6.30e+00	0.00e+00	4.92e-01	1
	<b>Alg2</b>	<b>12</b>	<b>10</b>	<b>V</b>	<b>5.94e+00</b>	<b>0.00e+00</b>	<b>6.08e+01</b>	<b>3</b>
	<b>Alg3</b>	<b>12</b>	<b>10</b>	<b>V</b>	<b>5.94e+00</b>	<b>0.00e+00</b>	<b>6.26e+01</b>	<b>4</b>
	<b>Alg4</b>	<b>12</b>	<b>10</b>	<b>V</b>	<b>5.93e+00</b>	<b>0.00e+00</b>	<b>5.29e+01</b>	<b>3</b>
	$\ell 1$	12	10	V	5.95e+00	0.00e+00	1.43e-01	2
	$\ell 2$	12	10	V	5.93e+00	0.00e+00	1.63e+00	4
POLAK4	Alg1	3	3	V	3.40e-01	0.00e+00	1.33e-01	1
	<b>Alg2</b>	<b>3</b>	<b>3</b>	<b>V</b>	<b>2.07e-05</b>	<b>0.00e+00</b>	<b>1.32e+00</b>	<b>3</b>
	<b>Alg3</b>	<b>3</b>	<b>3</b>	<b>V</b>	<b>2.05e-05</b>	<b>0.00e+00</b>	<b>1.72e+00</b>	<b>4</b>
	<b>Alg4</b>	<b>3</b>	<b>3</b>	<b>V</b>	<b>1.98e-05</b>	<b>0.00e+00</b>	<b>1.17e+00</b>	<b>3</b>
	$\ell 1$	3	3	V	1.53e-07	8.88e-08	5.11e-02	2
	$\ell 2$	3	3	V	2.22e-05	0.00e+00	1.74e+00	5
POLAK5	Alg1	3	2	V	5.03e+01	0.00e+00	4.87e-03	1
	<b>Alg2</b>	<b>3</b>	<b>2</b>	<b>V</b>	<b>5.00e+01</b>	<b>0.00e+00</b>	<b>1.08e-02</b>	<b>2</b>
	<b>Alg3</b>	<b>3</b>	<b>2</b>	<b>V</b>	<b>5.00e+01</b>	<b>0.00e+00</b>	<b>8.50e-03</b>	<b>3</b>
	<b>Alg4</b>	<b>3</b>	<b>2</b>	<b>V</b>	<b>5.00e+01</b>	<b>0.00e+00</b>	<b>2.44e-02</b>	<b>3</b>
	$\ell 1$	3	2	V	5.00e+01	1.99e-08	4.92e-03	2
	$\ell 2$	3	2	V	5.00e+01	0.00e+00	1.53e-02	4
POLAK6	Alg1	5	4	V	-4.36e+01	0.00e+00	8.59e-03	1
	<b>Alg2</b>	<b>5</b>	<b>4</b>	<b>V</b>	<b>-4.40e+01</b>	<b>0.00e+00</b>	<b>1.86e+00</b>	<b>3</b>
	<b>Alg3</b>	<b>5</b>	<b>4</b>	<b>V</b>	<b>-4.40e+01</b>	<b>0.00e+00</b>	<b>1.85e+00</b>	<b>3</b>
	<b>Alg4</b>	<b>5</b>	<b>4</b>	<b>V</b>	<b>-4.40e+01</b>	<b>0.00e+00</b>	<b>2.43e+00</b>	<b>3</b>
	$\ell 1$	5	4	V	-1.78e+01	0.00e+00	4.04e-03	1
	$\ell 2$	5	4	V	-4.40e+01	0.00e+00	1.52e+00	4
ROSENMMX	Alg1	5	4	V	-4.36e+01	0.00e+00	3.08e-03	1
	<b>Alg2</b>	<b>5</b>	<b>4</b>	<b>V</b>	<b>-4.40e+01</b>	<b>0.00e+00</b>	<b>1.82e+00</b>	<b>3</b>
	<b>Alg3</b>	<b>5</b>	<b>4</b>	<b>V</b>	<b>-4.40e+01</b>	<b>0.00e+00</b>	<b>1.49e+00</b>	<b>3</b>
	<b>Alg4</b>	<b>5</b>	<b>4</b>	<b>V</b>	<b>-4.40e+01</b>	<b>0.00e+00</b>	<b>1.78e+00</b>	<b>3</b>
	$\ell 1$	5	4	V	-4.40e+01	7.71e-08	6.47e-03	2
	$\ell 2$	5	4	V	-4.39e+01	0.00e+00	2.58e+00	5
S268	Alg1	5	5	C	7.88e+05	1.85e+01	2.01e+01	39
	Alg2	5	5	C	3.05e+05	2.20e+01	3.48e+01	64
	Alg3	5	5	C	1.49e+05	2.70e+01	3.43e+01	64
	Alg4	5	5	C	6.85e+04	2.64e+01	3.51e+01	64
	$\ell 1$	5	5	C	1.86e+37	3.66e+18	1.11e+00	64

Continua na próxima página

problema	alg	n	m	s	$f(x^k)$	$\ [g(x^k)]_+\ _\infty$	tempo	$k$
	$\ell_2$	5	5	C	7.15e+04	1.45e+01	4.94e-01	14
SNAKE	Alg1	2	2	V	1.00e+00	0.00e+00	1.09e-03	1
	Alg2	2	2	V	1.00e+00	0.00e+00	1.11e-03	1
	Alg3	2	2	V	1.00e+00	0.00e+00	1.01e-03	1
	Alg4	2	2	V	1.00e+00	0.00e+00	1.16e-03	1
	$\ell_1$	2	2	E	$\infty$	NaN	1.67e-02	4
	$\ell_2$	2	2	V	1.00e+00	0.00e+00	9.81e-04	2
SPIRAL	Alg1	3	2	C	1.00e+00	8.75e-01	3.69e-02	39
	Alg2	3	2	C	1.00e+00	8.75e-01	8.31e-02	64
	Alg3	3	2	C	1.00e+00	8.75e-01	5.91e-02	64
	Alg4	3	2	C	1.00e+00	8.75e-01	2.96e-02	64
	$\ell_1$	3	2	E	-2.18e+154	NaN	2.68e-02	3
	$\ell_2$	3	2	C	1.00e+00	8.75e-01	1.33e-02	14
WOMFLET	Alg1	3	3	C	-5.05e+114	0.00e+00	2.61e-02	39
	Alg2	3	3	C	-3.61e+160	0.00e+00	1.72e-02	64
	Alg3	3	3	C	-2.22e+158	0.00e+00	2.00e-02	64
	Alg4	3	3	C	-3.76e+174	0.00e+00	1.54e-02	64
	$\ell_1$	3	3	E	NaN	NaN	1.18e-01	30
	$\ell_2$	3	3	C	-2.28e+23	0.00e+00	2.54e-02	14