
Universidade Estadual de Campinas
Instituto de Matemática Estatística e Computação Científica
Departamento de Matemática Aplicada

**Otimização com restrições LOVO,
Restauração Inexata e o
Equilíbrio Inverso de Nash**

Luís Felipe Cesar da Rocha Bueno[†]

Doutorado em Matemática Aplicada - Campinas - SP

Orientador: Prof. Dr. José Mario Martínez Pérez

[†] Este trabalho teve apoio financeiro da FAPESP-(2007/06663-0).

Otimização com restrições LOVO, Restauração Inexata e o Equilíbrio Inverso de Nash

Este exemplar corresponde à redação final da tese devidamente corrigida e defendida por **Luís Felipe Cesar da Rocha Bueno** e aprovada pela comissão julgadora.

Campinas, 05 de novembro de 2011.



Prof. Dr. José Mario Martínez Pérez
Orientador

Banca examinadora:

Prof. Dr. José Mario Martínez Pérez.
Prof. Dr. Ernesto Julián Goldberg Birgin.
Prof. Dr. Gabriel Hacser.
Prof. Dr. Roberto Andreani.
Profa. Dra. Sandra Augusta Santos.

Tese apresentada ao Instituto de Matemática, Estatística e Computação Científica, UNICAMP, como requisito parcial para obtenção do Título de **Doutor em MATEMÁTICA APLICADA**.

FICHA CATALOGRÁFICA ELABORADA POR
MARIA FABIANA BEZERRA MÜLLER - CRB8/6162
BIBLIOTECA DO INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E
COMPUTAÇÃO CIENTÍFICA - UNICAMP

B862o	<p>Bueno, Luís Felipe Cesar da Rocha, 1983- Otimização com restrições LOVO, restauração inexata e o equilíbrio inverso de Nash / Luís Felipe Cesar da Rocha. - Campinas, SP : [s.n.], 2011.</p> <p>Orientador: José Mario Martínez Perez. Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.</p> <p>1. Programação não-linear. 2. Otimização matemática. 3. Programação (Matemática). 4. Métodos sem derivadas. 5. Otimização com restrições. I. Martínez Perez, José Mario, 1948-. II. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. III. Título.</p>
-------	--

Informações para Biblioteca Digital

Título em inglês: Optimization with LOVO constraints, inexact restoration and the
inverse Nash equilibrium

Palavras-chave em inglês:

Nonlinear programming
Mathematical optimization
Programming (Mathematics)
Derivative-free methods
Constrained optimization

Área de concentração: Matemática Aplicada

Titulação: Doutor em Matemática Aplicada

Banca examinadora:

José Mario Martínez Perez [Orientador]
Roberto Andreani
Sandra Augusta Santos
Ernesto Julián Goldberg Birgin
Gabriel Haeser

Data da defesa: 27-09-2011

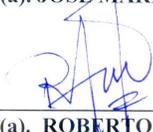
Programa de Pós-Graduação: Matemática Aplicada

Tese de Doutorado defendida em 27 de setembro de 2011 e aprovada

Pela Banca Examinadora composta pelos Profs. Drs.



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



Prof(a). Dr(a). ROBERTO ANDREANI



Prof(a). Dr(a). SANDRA AUGUSTA SANTOS



Prof(a). Dr(a). ERNESTO JULIÁN GOLDBERG BIRGIN



Prof(a). Dr(a). GABRIEL HAESER

Agradecimentos

Depois de quase dez anos na Unicamp a lista de pessoas a agradecer é imensa. Para não me alongar demasiadamente agradeço novamente a todos aqueles que já citei na minha dissertação de mestrado e vou agradecer explicitamente aqui somente aos que certamente poderiam me acusar de um crime caso esquecidos.

Primeiramente agradeço à minha família, em especial aos meus pais, por todo o apoio que me deram nessa jornada. Agradeço também à minha namorada, Fernanda, pela convivência, paciência e incentivo durante esse período.

Ao meu orientador, J. M. Martínez, pelas excelentes ideias e conselhos que me transmitiu ao longo desses anos e, mais ainda, pela grande amizade que tivemos e temos, fazendo com que esse trabalho fosse muito prazeroso.

Ao Nino também pela amizade e por ter deixado esse caminho mais divertido, além de ter sido quem me chamou para participar do grupo de otimização do IMECC. Talvez sem ele essa história sequer tivesse começado.

À Cheti por todo o apoio desde a época de graduação.

Aos coautores de partes desse trabalho, Ernesto, Natasa, Ana, Francisco e claro, Martínez novamente.

À Sandra por vários motivos, dentre eles por ter lido cuidadosamente a versão preliminar desse texto e feito inúmeras correções e comentários que melhoraram infinitamente a qualidade desse trabalho.

Ao Paulo por também ter dado sugestões que contribuíram bastante para melhorar a qualidade desse texto.

Aos membros da banca, titulares e suplentes, que prontamente aceitaram ao convite e que por diversas vezes ao longo desses anos me deram sugestões e explicações que serão usadas aqui.

Aos meus amigos de curso, em especial Francisco, Leandro e Feodor, que me ajudaram várias vezes, sobretudo com aspectos computacionais.

Finalmente, à FAPESP, pelo apoio financeiro fundamental para a realização

deste trabalho.

Resumo

Nesse trabalho serão propostos métodos de Lagrangiano Aumentado para tratar problemas com restrições do tipo LOVO, serão propostos novos métodos de Restauração Inexata e será introduzido o conceito de Equilíbrio Inverso de Nash.

Teoremas sobre condições de otimalidade para problemas do tipo LOVO serão apresentados. Um algoritmo do tipo Lagrangiano Aumentado será proposto para abordar esse problema e teoremas de convergência global serão demonstrados. Resultados computacionais serão realizados para uma aplicação em otimização de carteiras em investimentos de grande impacto.

Um método híbrido de Restauração Inexata será proposto combinando uma modificação, que usa o Lagrangiano Afiado como função de mérito, do método global de Fischer e Friedlander e o método local de Birgin e Martínez. Teoremas de convergência global e local serão apresentados.

Um método de Restauração Inexata para problemas em que as derivadas da função objetivo não estejam disponíveis será introduzido. Nesse método todas as ferramentas da otimização tradicional serão usadas na fase de restauração e uma regularização será feita na fase de otimização. Teoremas de convergência global serão demonstrados e resultados numéricos apresentados.

O conceito de Equilíbrio Inverso de Nash será introduzido e um método de Restauração Inexata será proposto para abordar esse problema. Esse método será uma extensão de um novo método de Restauração Inexata para problemas em dois níveis que também será proposto neste trabalho. Exemplos ilustrativos para uma aplicação para o problema de equilíbrio de Arrow-Debreu serão exibidos.

Palavras-chave: Programação não-linear, Otimização, Otimização com restrições do tipo LOVO, Restauração Inexata, Métodos sem derivadas, Equilíbrio Inverso de Nash .

Abstract

In this work an Augmented Lagrangian method will be proposed to deal with LOVO constraints, also some new Inexact Restoration methods will be presented and the Inverse Nash Equilibrium concept will be introduced.

Theorems about optimality conditions for LOVO-like problems will be presented. Three Augmented Lagrangian algorithms will be proposed to approach this problem and global convergence theorems will be proved. Computational results will be performed for an application in portfolio optimization with impact.

A modification of the Fischer-Friedlander global method using the Sharp Lagrangian as a merit function will be proposed. A hybrid Inexact Restoration method combining this modification and the Birgin-Martínez local method will be introduced. Global and local convergence theorems will be presented.

An Inexact Restoration method for problems in which the derivatives of the objective function are not available will be introduced. In this method it will be used all the optimization traditional tools in the restoration process as well as a regularization strategy in the optimization phase. Global convergence theorems will be demonstrated and numerical results will be presented.

The concept of Inverse Nash Equilibrium will be introduced and an Inexact Restoration method will be proposed to deal with this problem. This method is an extension of a new Inexact Restoration method for bilevel programming that will also be proposed in this work. Some illustrative examples for an application for the Arrow-Debreu equilibrium problem will be given.

Key words: Nonlinear programming, Optimization, Optimization with LOVO constraints, Inexact Restoration, Derivative-free methods, Inverse Nash Equilibrium.



Sumário

Introdução	1
1 Resultados conhecidos sobre Lagrangiano Aumentado e LOVO	7
1.1 Revisão sobre Lagrangiano Aumentado	7
1.2 Revisão sobre LOVO	11
1.2.1 Caracterização de minimizadores de problemas do tipo LOVO .	13
1.2.2 Algoritmos para problemas do tipo LOVO	16
2 Abordagem LOVO para problemas com restrições do tipo VaR	19
2.1 Condições de otimalidade para problemas com restrições do tipo LOVO	22
2.2 Algoritmo com subproblemas do tipo LOVO	24
2.2.1 Solubilidade dos subproblemas	28
2.2.2 Convergência do Algoritmo 2.2.1	28
2.2.3 Otimização Global com subproblemas do tipo LOVO	30
2.3 Algoritmo com subproblemas suaves	31
2.3.1 Solubilidade dos subproblemas	32
2.3.2 Convergência do Algoritmo 2.3.1	33
2.3.3 Propriedades de otimização global	37
2.4 Algoritmo LOVO de ponto fixo	38
3 Problema de seleção de carteiras com restrição de risco e experimentos numéricos	43

3.1	Testes preliminares	46
3.2	Modelo com custos de transação	48
3.3	Problemas de grande porte	53
3.4	Análise dos resultados computacionais	56
4	Introdução a métodos de Restauração Inexata	59
5	Um método híbrido de Restauração Inexata	69
5.1	Algoritmo Fischer-Friedlander	70
5.2	Algoritmo Birgin-Martínez	90
5.3	Estratégia híbrida combinando Fischer-Friedlander e Birgin-Martínez .	116
6	Restauração Inexata sem busca linear e otimização sem derivadas	121
6.1	Método GSS-KLT	122
6.2	Restauração Inexata sem busca linear	127
6.3	Experimentos numéricos	136
7	Introdução ao problema de Equilíbrio de Nash	145
7.1	Exemplos de problemas de equilíbrio com decisões contínuas	150
7.2	Introdução às desigualdades variacionais	163
7.3	Métodos do tipo Jacobi	172
8	Restauração Inexata para problemas com restrições de equilíbrio	181
8.1	Um novo método de RI para programação em dois níveis	182
8.2	Algoritmo para restrições de equilíbrio	204
8.3	LOVO aplicada ao Equilíbrio Inverso de Nash	226
9	Conclusões e perspectivas futuras	229
	Referências Bibliográficas	233

Introdução

Nesse trabalho serão propostos métodos de Lagrangiano Aumentado para tratar problemas com restrições do tipo LOVO, serão propostos novos métodos de Restauração Inexata e será introduzido o conceito de Equilíbrio Inverso de Nash.

Problemas do tipo OVO, em sua forma mais geral, são problemas de otimização que envolvem uma sequência de q funções, as quais, para cada x , são ordenadas de modo que $f_{i_1(x)}(x) \leq f_{i_2(x)}(x) \leq \dots \leq f_{i_q(x)}(x)$, de forma que a estrutura do problema dependa de $f_{i_j}(x)$, com j pertencente a um conjunto não vazio $J \subset \{i_1(x), i_2(x), \dots, i_q(x)\}$ [79]. As funções de valor ordenado podem estar associadas tanto com a função objetivo quanto com as restrições. O caso especial dos menores valores ordenados, ou seja, $J = \{1, 2, \dots, r\}$, onde $1 \leq r < q$, é chamado de LOVO (*Low Order-Value Optimization*) [16, 17]. Nossa contribuição aqui será propor novos métodos para tratar restrições do tipo LOVO. Teoremas de convergência, testes computacionais e uma aplicação dessa teoria em otimização de carteiras em investimentos com impacto serão apresentados.

O conceito de Equilíbrio Nash aparece como um dos pilares da Teoria de Jogos, fundamentada e difundida, principalmente, por J. von Neumann e O. Morgenstern no final dos anos quarenta em [83], e no início dos anos cinquenta por J. Nash em [84, 85, 86, 87]. Os trabalhos de Nash lhe renderam o Prêmio Nobel de Economia de 1994 juntamente com J. Harsanyi e R. Selten. Esta teoria é eficiente para modelar situações onde agentes (jogadores) têm que tomar decisões interagindo entre si, e por isso tem aplicações em diversos ramos da ciência. Na visão da otimização atual, resol-

ver um problema de Equilíbrio de Nash consiste em encontrar um ponto de equilíbrio $x^* \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_q}$ tal que para cada $i \in \{1, 2, \dots, q\}$ as coordenadas de x^* correspondentes a \mathbb{R}^{n_i} sejam solução de um problema de otimização parametrizado pelas demais coordenadas de x^* [47]. Este é um ramo que recentemente vem sendo bastante estudado e talvez seja um dos mais promissores da otimização contínua.

O conceito de Equilíbrio Inverso de Nash será introduzido nesse trabalho. O problema associado a essa formulação é um caso particular dos, também muito estudados hoje em dia, problemas com restrições de equilíbrio. Este tipo de problema é difícil de ser tratado, por várias questões teóricas que envolvem sua formulação [49, 50, 88]. No nosso caso do Equilíbrio Inverso de Nash, gostaríamos de descobrir os parâmetros usados nos problemas de otimização para que o sistema esteja de fato em equilíbrio. Uma aplicação importante desse conceito é o caso em que uma observação é feita a priori e queremos descobrir os parâmetros para que ela seja de fato um Equilíbrio de Nash do sistema. Olhado de forma geral, trata-se de um problema de quadrados mínimos com restrições de equilíbrio. Outra aplicação importante é o caso normativo em que o objetivo é determinar os parâmetros para que o equilíbrio do sistema seja o melhor possível para uma outra função objetivo dada. Esse é o caso por exemplo da situação em que o governo deve determinar parâmetros para a economia de forma que o equilíbrio atingido pelo mercado com esses parâmetros seja bom para a sociedade como um todo.

Métodos de Restauração Inexata são métodos de otimização para tratar problemas com restrições [80, 78, 29, 53]. A grande característica destes métodos é tratar a viabilidade e a otimalidade em fases diferentes. Isso se adequa muito bem quando um dos objetivos, viabilidade ou otimalidade, é bem mais fácil de ser tratado do que o outro [10]. Isso acontece claramente no tipo de problema de Equilíbrio Inverso de Nash e por isso tais métodos foram nossos escolhidos para abordar estes problemas. Ao estudar estes métodos, desenvolvemos teoria que pode ser útil independentemente do problema resolvido ser um problema de equilíbrio.

Nos métodos de Restauração Inexata, após a fase de otimização, é necessário decidir se a direção encontrada é aceita ou não. Geralmente isso é feito verificando se o

critério do decréscimo suficiente é satisfeito para alguma função de mérito. No caso da direção ser rejeitada é procurado um ponto mais próximo ao ponto restaurado até que o critério seja satisfeito. Neste trabalho vamos apresentar uma modificação na função de mérito usada em [53] e maneiras diferentes de satisfazer o decréscimo suficiente desta modificação. Vamos relatar neste texto nossas propostas para evitar o efeito Maratos em métodos globais de Restauração Inexata. Será introduzido também um novo método deste tipo para problemas em que as derivadas da função objetivo não estejam disponíveis. Além disso, vamos propor um método específico para o problema de Equilíbrio Inverso de Nash, o qual será uma extensão de um novo método para programação em dois níveis que também será descrito nesse texto.

Por fim, vamos conciliar a teoria de restrições do tipo LOVO com problemas de Equilíbrio Inverso de Nash em um exemplo ilustrativo. Nosso objetivo é modelar a situação em que os problemas envolvidos no sistema de equilíbrio estejam sujeitos a fatores externos. Dessa forma, os agentes terão de impor em seus problemas de otimização restrições de risco referentes às condições exteriores ao sistema.

A organização do texto é descrita a seguir.

Os três primeiros capítulos são baseados no artigo [26], aceito para publicação na revista *Journal of Global Optimization*, em trabalho conjunto com E. Birgin, N. Krejić e J. M. Martínez. No Capítulo 1 vamos fazer uma revisão das propriedades básicas de métodos do tipo Lagrangiano Aumentado, sobretudo as relativas às abordagens descritas em [8], e dos métodos que tratam a função do tipo LOVO como função objetivo [16, 17]. No Capítulo 2 vamos apresentar três diferentes algoritmos para tratar o problema de restrições do tipo LOVO e demonstrar suas propriedades de convergência. Testes numéricos referentes a estes algoritmos serão descritos no Capítulo 3 onde é apresentada uma aplicação em otimização de carteiras de grandes investidores, considerando custos de impacto de transações [5, 6], e restrição de risco do tipo *Value at Risk* [63].

No Capítulo 4 faremos uma introdução aos métodos de Restauração Inexata e à filosofia por trás deles. Vamos fazer isso citando as propriedades básicas deste

tipo de método e relacionando-as com outros métodos consagrados para problemas de programação não linear. No Capítulo 5 mostraremos adaptações do algoritmo de Fischer-Friedlander [53] para usar o Lagrangiano Afiado como função de mérito e proporemos uma estratégia híbrida com o algoritmo de Birgin-Martínez [29] para garantir convergência superlinear para métodos globais de Restauração Inexata. Relacionaremos o passo completo do método global com as condições exigidas pelo algoritmo local de Birgin-Martínez, evitando de certo modo o efeito Maratos.

Um método de Restauração Inexata sem busca linear, resultante de um trabalho conjunto com A. Friedlander, J. M. Martínez e F. Sobral [33], será introduzido no Capítulo 6. Sua aplicabilidade em problemas em que as derivadas da função objetivo não estejam disponíveis e testes computacionais referentes a essa situação também serão apresentados. O ponto chave desse capítulo é que se não tivermos uma direção de descida, o que é de se esperar quando as derivadas da função não estejam disponíveis, não é possível garantir que a condição do decréscimo suficiente para a função de mérito seja satisfeita apenas fazendo o *backtracking* na direção obtida na fase de otimização. Para contornar isso introduziremos uma estratégia de regularização na fase de otimização.

No Capítulo 7, faremos uma revisão sobre problemas de equilíbrio e discutiremos algumas alternativas para resolvê-los. No Capítulo 8 proporemos um novo método de Restauração Inexata para problemas de programação em dois níveis combinando a restauração sugerida em [10] com nossas adaptações do algoritmo de Fischer e Friedlander [53]. Além disso, uma extensão desse algoritmo será proposta para resolver o problema de Equilíbrio Inverso de Nash. Conclusões e perspectivas futuras serão expostas no Capítulo 9.

Notação. Vamos escrever $K_1 \subset_{\infty} K_2$ para indicar que K_1 é uma subsequência infinita de índices contidos em K_2 . O símbolo $\|\cdot\|$ denota a norma Euclidiana, embora muitas vezes ele pode ser substituído por uma norma arbitrária em \mathbb{R}^n . Para todo $v \in \mathbb{R}^n$, vamos escrever $v_+ = (\max\{0, v_1\}, \dots, \max\{0, v_n\})^T$. Para todo $a \in \mathbb{R}$ denotamos $a_+^2 = (a_+)^2$. Vamos denotar por $\#I$ o número de elementos do conjunto I . Ao longo de todo o texto diremos que uma função f é suave quando ela for de ordem C^k , onde k é o número de

derivadas necessárias para a análise envolvendo f . A constante de Lipschitz de f será denotada por L_f .

Capítulo 1

Resultados conhecidos sobre Lagrangiano Aumentado e LOVO

Por conveniência, neste capítulo, vamos recordar alguns resultados conhecidos sobre métodos do tipo Lagrangiano Aumentado para otimização suave e sobre Otimização de Valor Ordenado, em especial a do tipo LOVO como função objetivo. O motivo desta revisão é familiarizar o leitor com os conceitos recorrentes de [8], [28] e [17] que serão utilizados como ferramentas auxiliares no próximo capítulo deste trabalho.

1.1 Revisão sobre Lagrangiano Aumentado

Nesta seção, apresentamos os principais algoritmos expostos em [8] e [28] e seus resultados de convergência. Estes algoritmos baseiam-se no Lagrangiano Aumentado de Powell-Hestenes-Rockafellar (PHR) [60, 91, 95]. Vamos supor que $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$, $\underline{h} : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $\underline{g} : \mathbb{R}^n \rightarrow \mathbb{R}^p$ são funções contínuas. O Algoritmo 1.1.1 exigirá continuidade dos gradientes, enquanto o Algoritmo 1.1.2 não. O problema

de otimização padrão que queremos resolver é o seguinte

$$\begin{aligned}
 & \text{Minimizar } f(x) \\
 & \quad x \in \mathbb{R}^n \\
 & \text{sujeito a: } h(x) = 0, \\
 & \quad g(x) \leq 0, \\
 & \quad \underline{h}(x) = 0, \\
 & \quad \underline{g}(x) \leq 0.
 \end{aligned} \tag{1.1}$$

As restrições $\underline{h}(x) = 0$ e $\underline{g}(x) \leq 0$ são chamadas “restrições de nível inferior”, e o conjunto dos pontos que as satisfazem será denotado por Ω , isto é $\Omega \equiv \{x \in \mathbb{R}^n; \underline{h}(x) = 0 \text{ e } \underline{g}(x) \leq 0\}$. No caso em que as restrições de nível inferior são dadas por um caixa n -dimensional, o algoritmo definido nesta seção corresponde a Algencan (o método de otimização introduzido em [8] e disponível na página do projeto Tango [104]).

O Lagrangiano Aumentado adotado nos algoritmos principais de [8] e [28] é o Lagrangiano Aumentado PHR, que mostrou melhor desempenho em uma grande bateria de testes comparativos entre vários Lagrangianos Aumentados feita por E. Birgin, R. Castillo e J. M. Martínez em [27]. Dados $\rho > 0$, $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}_+^p$, $x \in \mathbb{R}^n$, vamos definir o Lagrangiano Aumentado PHR:

$$\mathcal{L}_\rho(x, \lambda, \mu) = f(x) + \frac{\rho}{2} \left[\left\| h(x) + \frac{\lambda}{\rho} \right\|^2 + \left\| \left(g(x) + \frac{\mu}{\rho} \right)_+ \right\|^2 \right].$$

Algoritmo Conceitual 1.1.0.

Seja $\epsilon_k \downarrow 0$, $\bar{\lambda}^k \in [\lambda_{\min}, \lambda_{\max}]^m$, $\bar{\mu}^k \in [0, \mu_{\max}]^p$ para todo $k \in \mathbb{N}$, $\rho_1 > 0$, $\tau \in (0, 1)$, $\eta > 1$. Para todo $k = 1, 2, \dots$ calcule $x^k \in \mathbb{R}^n$ como uma solução aproximada de

$$\begin{aligned}
 & \text{Minimizar } \mathcal{L}_{\rho_k}(x, \bar{\lambda}^k, \bar{\mu}^k) \\
 & \quad x \in \mathbb{R}^n \\
 & \text{sujeito a: } \underline{h}(x) = 0, \\
 & \quad \underline{g}(x) \leq 0.
 \end{aligned} \tag{1.2}$$

Defina, para todo $i = 1, \dots, p$,

$$V_i^k = \max \left\{ g_i(x^k), \frac{-\bar{\mu}_i^k}{\rho_k} \right\}.$$

Se $k = 1$ ou

$$\max\{\|h(x^k)\|, \|V^k\|\} \leq \tau \max\{\|h(x^{k-1})\|, \|V^{k-1}\|\} \quad (1.3)$$

defina $\rho_{k+1} \geq \rho_k$. Caso contrário, defina, $\rho_{k+1} \geq \eta\rho_k$.

Algoritmo 1.1.1.

Proceda como no Algoritmo Conceitual 1.1.0, com x^k definido de maneira que existam $v^k \in \mathbb{R}^m$ e $w^k \in \mathbb{R}_+^p$ satisfazendo:

$$\|\nabla_x \mathcal{L}_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k) + \nabla \underline{h}(x^k)v^k + \nabla \underline{g}(x^k)w^k\| \leq \epsilon_k, \quad (1.4)$$

$$\|\underline{h}(x^k)\| \leq \epsilon_k, \quad \|\underline{g}(x^k)_+\| \leq \epsilon_k,$$

e

$$w_i^k = 0 \text{ sempre que } \underline{g}_i(x^k) < -\epsilon_k.$$

Um ponto x^k que cumpre esta condição é chamado de ϵ_k -KKT de (1.2).

Observação. Em [8], é definido $\rho_{k+1} = \rho_k$ se vale (1.3) e $\rho_{k+1} = \eta\rho_k$ caso contrário. Aqui adotamos uma forma mais geral,

$$\rho_{k+1} \geq \rho_k \text{ e } \rho_{k+1} \geq \eta\rho_k, \quad (1.5)$$

respectivamente. Desta forma, será mais fácil interpretar os métodos dos próximos capítulos em termos do Algoritmo 1.1.1.

As propriedades de convergência do Algoritmo 1.1.1 são apresentadas no seguinte teorema.

Teorema 1.1.1. *Assuma que x^* é um ponto limite da sequência gerada pelo Algoritmo 1.1.1. Então, vale uma das seguintes três possibilidades:*

1. x^* é um ponto viável de (1.1).
2. x^* é um ponto Karush-Kuhn-Tucker (KKT) do problema

$$\begin{aligned} & \text{Minimizar} \quad \|h(x)\|^2 + \|g(x)_+\|^2 \\ & \quad \quad \quad x \in \mathbb{R}^n \\ & \text{sujeito a:} \quad \underline{h}(x) = 0, \\ & \quad \quad \quad \underline{g}(x) \leq 0. \end{aligned}$$

3. As restrições $\underline{h}(x) = 0$ e $\underline{g}(x) \leq 0$ não satisfazem a condição de qualificação de Dependência Linear Positiva Constante (CPLD)¹ [19, 93] em x^* .

Se x^* é um ponto viável de (1.1) então vale uma das duas seguintes possibilidades:

1. x^* é um ponto KKT do problema (1.1).
2. As restrições $h(x) = 0$, $g(x) \leq 0$, $\underline{h}(x) = 0$, $\underline{g}(x) \leq 0$ não satisfazem a condição de qualificação CPLD em x^* .

Demonstração: Veja as demonstrações dos Teoremas 4.1 e 4.2 de [8]. Observe que as modificações (1.5) não interferem na demonstração. □

A contrapartida para otimização global do Algoritmo 1.1.1 foi definida em [28]. Nós denotaremos o método global de Lagrangiano Aumentado de [28] como Algoritmo 1.1.2 abaixo. A diferença entre estes dois algoritmos é que, no Algoritmo 1.1.2 o iterando x^k é obtido como um ϵ_k -minimizador global do Lagrangiano Aumentado.

Algoritmo 1.1.2.

Proceda como no Algoritmo 1.1.0, escolhendo $x^k \in \Omega$ de tal forma que

$$\mathcal{L}_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k) \leq \mathcal{L}_{\rho_k}(x, \bar{\lambda}^k, \bar{\mu}^k) + \epsilon_k$$

¹Definição de CPLD: Para cada subconjunto dos gradientes das restrições de desigualdade ativas e dos gradientes das restrições de igualdade, se ele é positivamente linearmente dependente em x^* , então é positivamente linearmente dependente em uma vizinhança de x^* .

para todo $x \in \Omega$.

Um ponto x^k que cumpre esta condição é chamado de ϵ_k -minimizador global do problema (1.2).

O seguinte teorema foi provado em [28], onde um conjunto de experimentos numéricos usando o algoritmo α -BB [2, 3, 4, 20] para resolver os subproblemas foram feitos.

Teorema 1.1.2. *Assuma que a região factível do problema (1.1) é não vazia. Então, todo ponto limite da sequência gerada pelo Algoritmo 1.1.2 é um minimizador global de (1.1).*

Demonstração: Veja as demonstrações dos Teoremas 1 e 2 da Seção 3 de [28]. Observe novamente que as modificações (1.5) não interferem na demonstração. \square

Na implementação prática de Algencan os multiplicadores de Lagrange são definidos como $\bar{\lambda}_i^{k+1} = \max\{\lambda_{min}, \min\{\bar{\lambda}_i^k + \rho_k h_i(x^k), \lambda_{max}\}\}$, para $i = 1, \dots, m$, e $\bar{\mu}_j^{k+1} = \max\{0, \min\{\bar{\mu}_j^k + \rho_k g_j(x_k), \mu_{max}\}\}$, para $j = 1, \dots, p$. Com essas escolhas e algumas hipóteses adicionais sobre a convergência de toda a sequência $\{x^k\}$ para um ponto viável, condições de qualificação fortes, condições de não singularidade, o uso do multiplicador de Lagrange verdadeiro e sem a modificação na escolha do parâmetro de penalidade (1.5) é possível demonstrar a limitação dos parâmetros de penalidade (veja os Teoremas 5.4 e 5.5 de [8]).

1.2 Revisão sobre LOVO

Vamos agora descrever as principais características das funções envolvidas em Otimização de Valor Ordenado (*Order-Value Optimization*). Sejam $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$, $i = 1, \dots, q$, funções dadas. Para cada $x \in \mathbb{R}^n$, seja $i_1(x), \dots, i_q(x)$ uma permutação de

$1, \dots, q$ tal que as seguintes desigualdades sejam válidas:

$$f_{i_1(x)}(x) \leq f_{i_2(x)}(x) \leq \dots \leq f_{i_q(x)}(x). \quad (1.6)$$

Veja [11, 12, 16, 17, 18] para maiores detalhes. Seja J um subconjunto não vazio de $\{1, \dots, q\}$. A Função de Valor Ordenado [79] associada a J é definida por

$$f_J(x) = \sum_{j \in J} f_{i_j(x)}(x).$$

Se $J = \{q\}$, temos $f_J(x) = \max\{f_1(x), \dots, f_q(x)\}$. Se $J = \{1\}$, temos $f_J(x) = \min\{f_1(x), \dots, f_q(x)\}$. Quando $J = \{r\}$, $f_J(x)$ pode ser interpretada como uma versão discreta da medida de risco *Value-at-Risk* (VaR) [63], associada à carteira x , cenários $\{1, \dots, q\}$, funções de perda f_1, \dots, f_q e nível de confiança $\alpha = r/q$ (veja [11, 12, 18]). Analogamente, quando $J = \{r+1, \dots, q\}$, a função $f_J(x)/(q-r)$ corresponde à função de risco “coerente” *Conditional Value-at-Risk* (CVaR) [96].

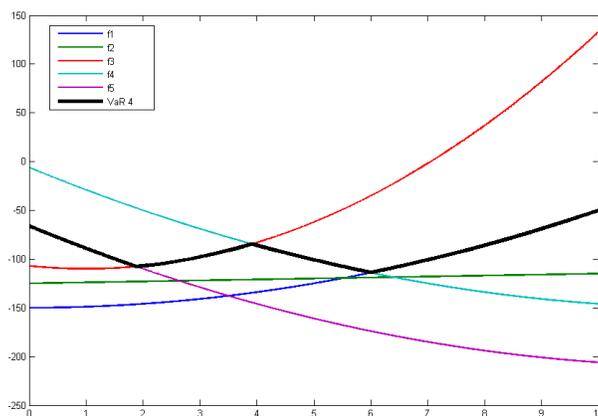


Figura 1.1: A função destacada em negrito representa $f_{i_4(x)}(x)$.

Observe que as funções $f_{i_j(x)}(x)$ podem não ser suaves e nem convexas mesmo quando todas as $f_i(x)$ são. Este fato faz com que problemas que envolvam estas funções sejam difíceis de serem abordados.

A função do tipo LOVO (*Low Order-Value Optimization*) introduzida em [16, 17] corresponde a $J = \{1, \dots, r\}$. O problema de minimizar f_J tem aplicações

interessantes em estimação de parâmetros, problemas de detecção de padrões, alinhamento de proteínas e alinhamento de estruturas em geral [16, 17, 79]. Em [17] foram introduzidos métodos para minimizar uma função tipo LOVO com e sem restrições. A seguir vamos expor com mais detalhes os resultados apresentados em [17].

1.2.1 Caracterização de minimizadores de problemas do tipo LOVO

Como dito anteriormente o problema tradicional LOVO consiste em:

$$\begin{aligned} \text{Minimizar} \quad & \sum_{j=1}^r f_{i_j(x)}(x) \\ x \in \mathbb{R}^n \quad & \\ \text{sujeito a:} \quad & x \in \Omega, \end{aligned} \tag{1.7}$$

onde Ω é um subconjunto qualquer de \mathbb{R}^n .

Um caso particular do problema (1.7) é quando $r = 1$, neste caso estamos minimizando o mínimo das q funções. Este é um caso de interesse relevante ao longo da história da otimização e, na verdade, é mais do que um caso particular do problema (1.7), pois todo problema do tipo (1.7) pode ser convertido na forma

$$\begin{aligned} \text{Minimizar} \quad & F_{i_1(x)}(x) \\ x \in \mathbb{R}^n \quad & \\ \text{sujeito a:} \quad & x \in \Omega. \end{aligned} \tag{1.8}$$

Para isto basta tomar todas as combinações de r elementos das q funções e para cada uma delas definir

$$F_i(x) = \sum_{j \in i\text{-ésima combinação}} f_j(x).$$

É evidente que os problemas são equivalentes, pois o mínimo da soma das r menores funções é o mínimo das possíveis somas de r funções. Sendo assim, vamos supor daqui em diante que, ao trabalharmos com o problema (1.7), estaremos sempre assumindo $r = 1$.

Como o problema que estamos tratando não é da forma tradicional, onde temos funções diferenciáveis, vamos fazer algumas definições e apresentar alguns resultados para analisar as condições necessárias de otimalidade do problema (1.7).

Definição: O conjunto dos índices $I_{min}(x) \equiv \{i \in \{1, \dots, m\} \mid f_i(x) = f_{i_1(x)}(x)\}$.

Definição: O ponto x^* é dito fracamente KKT se existe $i \in I_{min}(x^*)$ tal que x^* é KKT para o problema:

$$\begin{aligned} &\text{Minimizar } f_i(x) \\ &x \in \mathbb{R}^n \\ &\text{sujeito a: } x \in \Omega. \end{aligned} \tag{1.9}$$

Definição: O ponto x^* é dito fortemente KKT se, para todo $i \in I_{min}(x^*)$, x^* for KKT para o problema (1.9).

Teorema 1.2.1. *Se o ponto x^* é um minimizador global de $f_{i_1(x)}(x)$, então x^* é um minimizador global de $f_i(x)$ para todo $i \in I_{min}(x^*)$.*

Demonstração: Veja Lema 2.1 de [17]. □

A recíproca do Teorema 1.2.1 não é verdadeira conforme mostra o exemplo a seguir, apresentado em [17]. Tomando $\Omega = \mathbb{R}$, $f_1(x) = (x - 1)^2$, $f_2(x) = x$ temos que se $x^* = 1$ então $I_{min}(x^*) = \{1\}$, x^* é um minimizador global de $f_1(x)$, mas não é um minimizador global de $f_{i_1(x)}(x)$.

Teorema 1.2.2. *Se o ponto x^* é um minimizador local de $f_{i_1(x)}(x)$, então x^* é um minimizador local de $f_i(x)$ para todo $i \in I_{min}(x^*)$. Mais ainda, se as funções f_i forem contínuas em x^* para todo $i \notin I_{min}(x^*)$ então vale a recíproca.*

Demonstração: Veja Teorema 2.1 de [17]. □

Corolário 1.2.1. *Seja $x^* \in \Omega$ um minimizador local de $f_{i_1(x)}(x)$, e suponhamos que todas as funções f_i sejam diferenciáveis. Suponhamos também que o conjunto Ω seja descrito por restrições de igualdade e desigualdade diferenciáveis. Portanto, se as restrições em x^* satisfazem alguma condição de qualificação, temos que x^* é fortemente KKT.*

Demonstração: Veja Corolário 2.1 de [17]. □

Observação 1.2.1 A ideia principal tanto para as demonstrações destes resultados quanto para a elaboração e análise dos algoritmos propostos na próxima seção é que, se $f_{i_1(x)}(z) \leq f_{i_1(x)}(x)$ então vale a relação

$$f_{i_1(z)}(z) \leq f_{i_1(x)}(z) \leq f_{i_1(x)}(x).$$

Dessa forma, diminuindo alguma das funções f_i , para algum $i \in I_{min}(x)$, automaticamente diminuiríamos f_{i_1} (veja Figura 1.2).

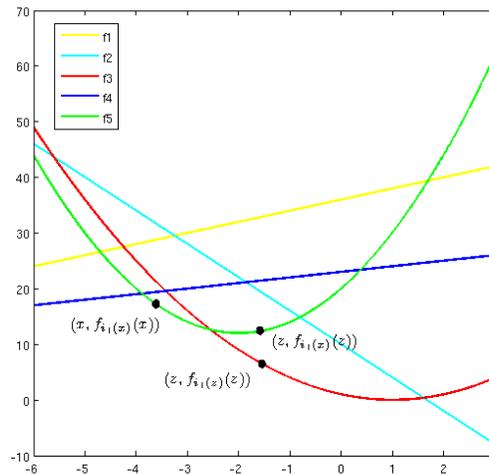


Figura 1.2: Se $5 \in I_{min}(x)$ então diminuindo $f_5(x)$ diminuimos $f_{i_1(x)}(x)$.

1.2.2 Algoritmos para problemas do tipo LOVO

Vamos expor um algoritmo proposto em [17] para minimizar uma função tipo LOVO sem restrições e outro algoritmo, que será do tipo Lagrangiano Aumentado, para o problema com restrições. A filosofia de ambos os algoritmos é agir como nos algoritmos tradicionais para otimização tradicional sendo que, sempre que requisitado “ $\nabla f_{i_1(x)}(x)$ ” consideraremos $\nabla f_i(x)$, para algum $i \in I_{min}(x)$ (e o mesmo para algoritmos que usem Hessianas para realização de algum passo). A justificativa disso vem da observação 1.2.1

Começamos pelo caso irrestrito, ou seja $\Omega = \mathbb{R}$.

Algoritmo U1.

Sejam os parâmetros $\theta \in (0, 1), \alpha \in (0, 1), \beta > 0$. Seja $x_0 \in \mathbb{R}^n$ uma aproximação inicial. Dado $x_k \in \mathbb{R}^n$, os passos para computar x_{k+1} são:

Passo 1. Escolha $i(k) \in I_{min}(x_k)$. Se $\|\nabla f_{i(k)}(x_k)\| = 0$, pare.

Passo 2. Compute $d_k \in \mathbb{R}^n$ tal que

$$\nabla f_{i(k)}(x_k)^T d_k \leq -\theta \|d_k\| \|\nabla f_{i(k)}(x_k)\| \quad \text{e} \quad \|d_k\| \geq \beta \|\nabla f_{i(k)}(x_k)\|.$$

Passo 3. Faça $t_k = 1$. Enquanto

$$f_{i_1(x_k+t_k d_k)}(x_k + t_k d_k) > f_{i_1(x)}(x_k) + \alpha t_k \nabla f_{i(k)}(x_k)^T d_k$$

faça $t_k = \frac{t_k}{2}$.

Uma vez que $f_{i_1(x_k+t_k d_k)}(x_k + t_k d_k) \leq f_{i_1(x)}(x_k) + \alpha t_k \nabla f_{i(k)}(x_k)^T d_k$ tome $x_{k+1} = z$ onde $f_{i_1(z)}(z) \leq f_{i_1(x_k+t_k d_k)}(x_k + t_k d_k)$.

Teorema 1.2.3. *Se x^* é um ponto limite da sequencia gerada pelo Algoritmo U1 então x^* é fracamente KKT para o problema (1.7). Ou seja, como estamos no caso irrestrito, existe $i \in I_{min}(x^*)$ tal que $\nabla f_i(x^*) = 0$.*

Demonstração: Veja Teorema 3.2 de [17]. \square

Um comentário pertinente aqui é que o algoritmo pode não parar mesmo que encontre um ponto fracamente KKT, desde que este não seja fortemente KKT. Por exemplo tome $f_1(x) = x$, $f_2(x) = x^2$, se $x_k = 0$ e o algoritmo escolhe $i(k) = 1$ ele continuará, mesmo valendo que $\nabla f_2(0) = 0$.

Algoritmos convergentes a pontos fortemente KKT também são desenvolvidos em [17], mas, segundo os próprios autores, eles parecem não apresentar muita vantagem sobre os convergentes a pontos fracamente KKT.

O algoritmo descrito em [17] para minimizar uma função tipo LOVO com restrições suaves é do tipo Lagrangiano Aumentado e é chamado de C-LOVO. A única diferença do Algoritmo C-LOVO e o Algoritmo 1.1.1 é que na condição (1.4) substituímos $\nabla_x \mathcal{L}_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k)$ por

$$\nabla f_i(x^k) + \frac{\rho_k}{2} \nabla_x \left[\left\| h(x^k) + \frac{\lambda^k}{\rho_k} \right\|^2 + \left\| \left(g(x^k) + \frac{\mu^k}{\rho_k} \right) \right\|^2 \right],$$

para algum $i \in I_{min}(x^k)$. Portanto podemos usar o Algoritmo U1 para a resolução dos subproblemas. Essencialmente, o C-LOVO consiste em aplicar Algencan para o problema, “ignorando” o fato de que a primeira derivada pode não estar definida. Os resultados de convergência para o C-LOVO estão resumidos no teorema a seguir.

Teorema 1.2.4. *Cada ponto limite x^* da sequência gerada por C-LOVO é viável, ou é estacionário para a soma de quadrados da inviabilidade. Além disso, qualquer ponto limite viável é fracamente KKT, desde que as restrições cumpram a condição de qualificação de dependência linear positiva constante (CPLD) [19, 93].*

Demonstração: Veja os Teoremas 5.4 e 5.5 de [17]. \square

Uma última observação que justifica ainda mais o uso do Algoritmo U1 na resolução dos subproblemas de C-LOVO é que mesmo que conseguíssemos um ponto

fortemente KKT em todos os subproblemas não teríamos a garantia de que o ponto limite de C-LOVO seria fortemente KKT para o problema 1.7. Isso está descrito no exemplo da observação após o Teorema 5.5 de [17], o qual vamos reproduzir a seguir.

Considerando $f_1(x) = (x - 1)^2$, $f_2(x) = (x + 1)^2$, $g(x) = x$, $r = 1$ e os multiplicadores nulos temos que cada ponto da sequência $x_k = \frac{1}{1+\rho_k}$ é fortemente KKT para os subproblemas. Isso acontece pois $I_{min}(x_k) = \{1\}$ para todo k , já que $f_1(x_k) < f_2(x_k)$, e x_k é minimizador de $L_1(x) = (x - 1)^2 + \rho_k x_+^2$. Entretanto $\lim_{k \rightarrow \infty} x_k = 0$ mas $I_{min}(0) = \{1, 2\}$ e $x^* = 0$ não é um ponto KKT do problema de minimizar $f_2(x)$ sujeito a $g(x) \leq 0$. A Figura 1.3 a seguir ilustra esse exemplo.

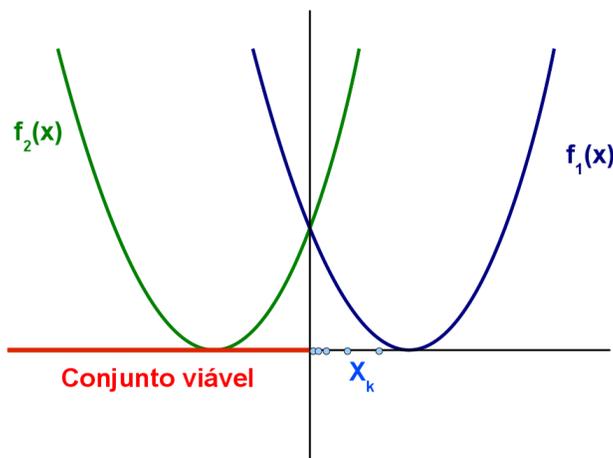


Figura 1.3: Exemplo em que pontos fortemente KKT para o subproblemas não implica que o ponto limite é fortemente KKT. Para a sequência x_k vemos que $I_{min} = \{1\}$ mas $I_{min}(0) = \{1, 2\}$ e 0 não é um ponto KKT do problema de minimizar $f_2(x)$ sujeito a $g(x) \leq 0$.

Capítulo 2

Abordagem LOVO para problemas com restrições do tipo VaR

Neste capítulo vamos tratar do problema onde as restrições possuem uma estrutura do tipo OVO. Este tipo de problema surge naturalmente da modelagem do problema de seleção de carteiras de investimento com uma restrição em relação ao risco do investimento, caso este risco seja mensurado pela popular medida de risco *Value-at-Risk* (VaR) [63]. Por este motivo vamos chamar estas restrições do tipo VaR. Como veremos mais adiante, vamos propor uma reformulação deste tipo de restrição para o que chamamos de restrições do tipo LOVO. Este tipo de problema poderá ser abordado com ideias semelhantes às utilizadas no capítulo anterior. Neste capítulo vamos estudar a estrutura deste tipo de problema e apresentar três diferentes métodos do tipo Lagrangiano Aumentado para resolvê-lo. Os resultados referentes aos métodos expostos aqui fazem parte do artigo [26], aceito para publicação na revista *Journal of Global Optimization* em trabalho conjunto com E. Birgin, N. Krejić e J. M. Martínez. Neste artigo também relatamos o desempenho computacional dos nossos algoritmos na seleção de carteiras com restrição de risco. Esta aplicação e os testes computacionais a ela referentes serão reproduzidos no próximo capítulo (Capítulo 3).

Como no Capítulo 1, consideramos $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$, $i = 1, \dots, q$, dadas e, para

cada $x \in \mathbb{R}^n$, $i_1(x), \dots, i_q(x)$ de forma que valha (1.6), ou seja:

$$f_{i_1(x)}(x) \leq f_{i_2(x)}(x) \leq \dots \leq f_{i_q(x)}(x).$$

Estamos interessados agora em problemas de otimização envolvendo “restrições do tipo VaR”, isto é, restrições da forma

$$f_{i_r(x)}(x) \leq M. \quad (2.1)$$

Sem perda de generalidade (redefinindo $f_i(x) \leftarrow f_i(x) - M$) vamos assumir que $M = 0$ em (2.1). Isso significa que, dado $r \in \{1, \dots, q\}$, nosso problema será:

$$\begin{aligned} &\text{Minimizar } f(x) \\ &x \in \Omega \subset \mathbb{R}^n \\ &\text{sujeito a: } f_{i_r(x)}(x) \leq 0, \\ &h(x) = 0, \\ &g(x) \leq 0. \end{aligned} \quad (2.2)$$

O conjunto Ω será descrito pelas *restrições do nível inferior* da forma

$$\underline{h}(x) = 0, \underline{g}(x) \leq 0.$$

Muitas vezes, Ω será uma caixa n -dimensional:

$$\Omega = \{x \in \mathbb{R}^n \mid \ell \leq x \leq u\}.$$

Vamos supor que as funções $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$, $i = 1, 2, \dots, q$, $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$, $\underline{h} : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $\underline{g} : \mathbb{R}^n \rightarrow \mathbb{R}^p$ tenham primeira derivada contínua em \mathbb{R}^n .

Por (1.6), o problema (2.2) é equivalente a:

$$\begin{aligned} &\text{Minimizar } f(x) \\ &x \in \Omega \subset \mathbb{R}^n \\ &\text{sujeito a: } f_{i_j(x)}(x) \leq 0, j = 1, \dots, r, \\ &h(x) = 0, \\ &g(x) \leq 0. \end{aligned} \quad (2.3)$$

Este tipo de restrições é o que chamamos de restrições do tipo LOVO. Neste capítulo vamos abordar e propor soluções práticas para o problema (2.3).

Como comentamos no Capítulo 1, a grande dificuldade de trabalhar com funções do tipo OVO é que, mesmo que todas as f_i sejam suaves e convexas a função OVO pode não ser nem suave nem convexa, e isso é o que geralmente acontece em situações práticas. Quando trabalhamos com restrições do tipo LOVO a não convexidade acarreta uma dificuldade ainda maior que é a possibilidade do conjunto factível ser desconexo. Com isso, mesmo no caso de funções objetivo muito bem comportadas, podem existir vários minimizadores locais não globais (ver Figura 2.1).

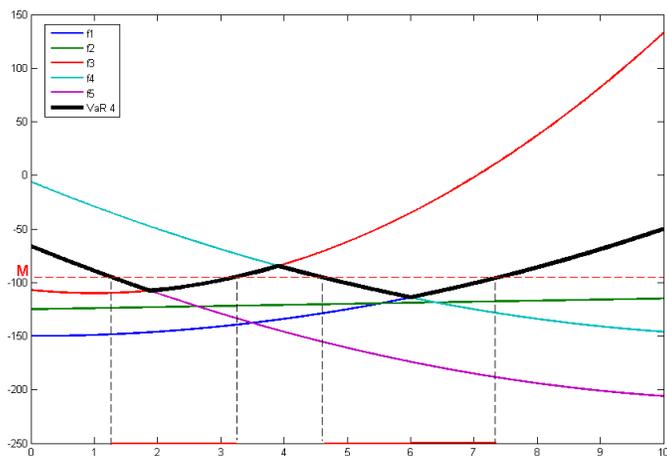


Figura 2.1: Para o nível de risco aceitável o domínio é desconexo.

Estamos especialmente interessados em casos de grande porte, em que n , que estará relacionado ao número de ativos de nossa aplicação do Capítulo 3, ou q , que estará associado ao número de cenários, ou ambos, são grandes. Em nossa aplicação descrita no Capítulo 3 as funções envolvidas na formulação do problema (2.3) são não-lineares em nosso caso de maior interesse, quando os custos de transação são incorporados. Além disso, mesmo no caso linear, a sua estrutura não é padrão, de modo que sua resolução necessita da invenção de métodos modernos de otimização. Vamos

mostrar que a abordagem de Lagrangiano Aumentado descrita em [8] é uma ferramenta adequada para tratar desse tipo de problemas. Essa abordagem lida bem com grande número de variáveis e, sobretudo, com grande número de restrições de desigualdade. Além disso, Lagrangianos Aumentados são facilmente adaptáveis ao paradigma LOVO, como mostrado no Capítulo 1.

2.1 Condições de otimalidade para problemas com restrições do tipo LOVO

Assim como no Capítulo 1 vamos definir condições de otimalidade para o nosso problema de interesse, uma vez que ele não é da forma padrão da otimização. Para isso precisamos de algumas definições similares às apresentadas para a otimização do tipo LOVO associada à função objetivo. As possíveis combinações das funções relativas aos r menores valores ordenados será denotada por $I_{min}(x)$. Ou seja, $I_{min}(x)$ é o conjunto dos conjuntos de índices referentes às r menores $f_i(x)$.

Definição: $I_{min}(x) \equiv \{I = \{i_1, i_2, \dots, i_r\}$ tais que $f_{i_j}(x) \leq f_s(x), \quad \forall s \notin I\}$.

Também como no capítulo anterior vamos definir os conceitos de fracamente e fortemente KKT.

Definição: Um ponto x^* é dito fracamente KKT se existe $I \in I_{min}(x^*)$ tal que x^* é KKT para o problema:

$$\begin{aligned} & \text{Minimizar} && f(x) \\ & && x \in \hat{\Omega} \\ & \text{Sujeito a:} && f_i(x) \leq 0; \quad \forall i \in I, \end{aligned} \tag{2.4}$$

onde $\hat{\Omega} \equiv \{x \in \mathbb{R}^n; x \in \Omega, h(x) = 0, g(x) \leq 0\}$.

Definição: Um ponto x^* é dito fortemente KKT se, para todo $I \in I_{min}(x^*)$, tem se que x^* é KKT para o problema (2.4).

Teorema 2.1.1. *Se x^* é um minimizador global do problema (2.3) então x^* é minimizador global de (2.4) para todo $I \in I_{min}(x^*)$.*

Demonstração: Suponhamos por absurdo que não valesse a tese, então existem $I \in I_{min}(x^*)$ e $y \in \hat{\Omega}$ tais que $f(y) < f(x^*)$ e $f_i(y) \leq 0 \forall i \in I$. Como $I \in I_{min}(x^*)$ a cardinalidade de I é r , e portanto existem pelo menos r funções tais que $f_i(y) \leq 0$. Conseqüentemente, com certeza as r menores ($f_{i_1(y)}(y) \leq f_{i_2(y)}(y) \leq \dots \leq f_{i_r(y)}(y)$) também serão menores do que 0, e neste caso, temos que y é admissível para o problema (2.3). Mas isso é um absurdo uma vez que $f(y) < f(x^*)$ e x^* é um minimizador global de (2.3). □

A recíproca do Teorema 2.1.1 é falsa. Como contra-exemplo tome $f_1(x) = x^2$, $f_2(x) = x + 1$, $f(x) = x$, $\hat{\Omega} = \mathbb{R}$ e $r = 1$. Se $x^* = 0$ então $I_{min}(x^*) = \{\{1\}\}$ e o problema (2.4) neste caso teria x^* como único ponto viável, sendo assim ele seria minimizador global. Entretanto é fácil ver que a região viável de (2.3) é $(-\infty, -1] \cup \{0\}$ e portanto o problema original é ilimitado inferiormente.

Teorema 2.1.2. *Um ponto x^* é um minimizador local do problema (2.3) se e somente se x^* é minimizador local de (2.4) para todo $I \in I_{min}(x^*)$.*

Demonstração: Suponhamos que x^* é um minimizador local do problema (2.3), então existe $\epsilon > 0$ tal que x^* é um minimizador global do problema (2.3) restrito à bola $B(x^*, \epsilon)$. Tomando $\Omega' = \hat{\Omega} \cap B(x^*, \epsilon)$ e aplicando o Teorema 2.1.1 temos que x^* é minimizador global em Ω' , e portanto é um minimizador local de (2.4) em Ω , para todo $I \in I_{min}(x^*)$.

Suponhamos agora que x^* é minimizador local de (2.4) para todo $I \in I_{min}(x^*)$. Pela continuidade das f_i s temos que $I_{min}(x) \subset I_{min}(x^*)$ para todo x suficientemente próximo de x^* . Sendo assim, perto de x^* , temos que x é viável para o problema (2.3) somente se é admissível para (2.4) com algum $I \in I_{min}(x) \subset I_{min}(x^*)$. Pelo fato de x^* ser minimizador local de (2.4) em I , se x está suficientemente próximo de x^* então $f(x) \geq f(x^*)$ e portanto x^* é um minimizador local do problema (2.3).

□

Como no nosso caso de interesse estamos trabalhando com funções suaves temos os seguintes corolários.

Corolário 2.1.1. *Se x^* é um minimizador local do problema (2.3) e a condição de qualificação CPLD é satisfeita em (2.4) para todo $I \in I_{min}(x^*)$ então x^* é fortemente KKT.*

Corolário 2.1.2. *Se x^* é um minimizador local do problema (2.3) e a condição de qualificação CPLD é satisfeita em (2.4) para algum $I \in I_{min}(x^*)$ então x^* é fracamente KKT.*

Nas próximas três seções vamos apresentar três métodos do tipo Lagrangiano Aumentado para resolver o problema (2.3). O primeiro deles, exposto na Seção 2.2, garante convergência a pontos fracamente KKT. Os outros dois, expostos nas Seções 2.3 e 2.4, apresentam resultados de convergência mais fracos.

2.2 Algoritmo com subproblemas do tipo LOVO

Em [17] um método do tipo Lagrangiano Aumentado (C-LOVO) foi definido para minimizar uma função dos Menores Valores-Ordenados com restrições suaves. Essencialmente, o C-Lovo consiste em aplicar Algencan para o problema, “ignorando” o fato de que a primeira derivada pode não estar definida. Cada ponto limite da sequência gerada por C-Lovo é viável ou é estacionário para a soma de quadrados da inviabilidade. Além

disso, qualquer ponto limite viável satisfaz as condições Karush-Kuhn-Tucker (KKT), desde que as restrições cumpram a condição de qualificação de dependência linear positiva constante (CPLD) [19, 93]. Estes fatos tornam desejável resolver problemas com restrições do tipo LOVO (2.3), por meio de uma sequência de problemas com restrições com função objetivo do tipo LOVO. Nesta seção apresentamos um algoritmo com essas características. O algoritmo será, como o C-LOVO, do tipo Lagrangiano Aumentado. Por outro lado, a cada iteração exterior, um subproblema será aproximadamente resolvido usando C-LOVO.

Dados $\rho > 0$, $\lambda \in \mathbb{R}^m$, $\mu \in \mathbb{R}_+^p$, $\nu \in \mathbb{R}_+$, $x \in \mathbb{R}^n$, definimos a função Lagrangiano Aumentado LOVO $L_\rho(x, \lambda, \mu, \nu)$ como:

$$L_\rho(x, \lambda, \mu, \nu) = f(x) + \frac{\rho}{2} \left[\left\| h(x) + \frac{\lambda}{\rho} \right\|^2 + \left\| \left(g(x) + \frac{\mu}{\rho} \right)_+ \right\|^2 + \sum_{j=1}^r \left(f_{i_j(x)}(x) + \frac{\nu}{\rho} \right)_+^2 \right]. \quad (2.5)$$

Em cada iteração (exterior), os algoritmos introduzidos nesta seção minimizam (aproximadamente) $L_\rho(x, \lambda, \mu, \nu)$ sujeito a $x \in \Omega$. Vamos justificar por que usar a função objetivo L_ρ , dada por (2.5), em vez de o (talvez mais intuitivo) “Lagrangiano Aumentado” \tilde{L}_ρ associado ao problema (2.2). Neste caso, teríamos:

$$\tilde{L}_\rho(x, \lambda, \mu, \nu) = f(x) + \frac{\rho}{2} \left[\left\| h(x) + \frac{\lambda}{\rho} \right\|^2 + \left\| \left(g(x) + \frac{\mu}{\rho} \right)_+ \right\|^2 + \left(f_{i_r(x)}(x) + \frac{\nu}{\rho} \right)_+^2 \right]. \quad (2.6)$$

O ponto chave é que minimizar a função L_ρ é mais fácil que minimizar \tilde{L}_ρ . De fato, dado x , uma vez encontrado um ponto tentativo z tal que

$$\Phi_x(z) \equiv f(z) + \frac{\rho}{2} \left[\left\| h(z) + \frac{\lambda}{\rho} \right\|^2 + \left\| \left(g(z) + \frac{\mu}{\rho} \right)_+ \right\|^2 + \sum_{j=1}^r \left(f_{i_j(x)}(z) + \frac{\nu}{\rho} \right)_+^2 \right] < L_\rho(x, \lambda, \mu, \nu), \quad (2.7)$$

automaticamente temos que

$$L_\rho(z, \lambda, \mu, \nu) < L_\rho(x, \lambda, \mu, \nu).$$

Note que os índices $i_j(x)$, $i = 1, \dots, r$, são fixados na definição de $\Phi_x(z)$, em contraste com os índices na definição de $L_\rho(x, \lambda, \mu, \nu)$ em (2.5), que dependem da variável x .

Assim, obter (2.7) não é difícil, porque isso equivale a encontrar uma direção de descida para a função suave $\Phi_x(z)$. Como consequência, resultados adequados de convergência global são obtidos [17]. Por outro lado,

$$f(z) + \frac{\rho}{2} \left[\left\| h(z) + \frac{\lambda}{\rho} \right\|^2 + \left\| \left(g(z) + \frac{\mu}{\rho} \right)_+ \right\|^2 + \left(f_{i_r(x)}(z) + \frac{\nu}{\rho} \right)_+^2 \right] < \tilde{L}_\rho(x, \lambda, \mu, \nu)$$

não implica que

$$\tilde{L}_\rho(z, \lambda, \mu, \nu) < \tilde{L}_\rho(x, \lambda, \mu, \nu). \quad (2.8)$$

Neste caso, o fato de que $i_r(z)$ é, em geral, diferente de $i_r(x)$ inibe que propriedade desejável (2.8) seja satisfeita. Estas observações também justificam o uso de (2.3) em vez de (2.2). O fato de que os problemas do tipo LOVO são mais fáceis de resolver do que problemas do tipo OVO [11, 12] é explorado em [17].

As derivadas de L_ρ em relação a x podem não existir nos pontos x em que o conjunto de índices que definem os r menores valores de $f_i(x)$ não é univocamente definido. Entanto, para simplificar a notação, vamos escrever:

$$\begin{aligned} \nabla L_\rho(x, \lambda, \mu, \nu) = \\ \nabla f(x) + \frac{\rho}{2} \left\{ \nabla \left(\left\| h(x) + \frac{\lambda}{\rho} \right\|^2 + \left\| \left(g(x) + \frac{\mu}{\rho} \right)_+ \right\|^2 \right) + \sum_{j=1}^r \nabla \left[\left(f_{i_j(x)}(x) + \frac{\nu}{\rho} \right)_+^2 \right] \right\}. \end{aligned}$$

Algoritmo Conceitual 2.2.0.

Os parâmetros que definem o algoritmo são: $\tau \in [0, 1)$, $\eta > 1$, $\lambda_{\min} < \lambda_{\max}$, $\mu_{\max} > 0$. Na primeira iteração externa usamos um parâmetro de penalidade $\rho_1 > 0$ e estimativas com salvaguardas dos multiplicadores de Lagrange $\bar{\lambda}^1 \in \mathbb{R}^m$, $\bar{\mu}^1 \in \mathbb{R}_+^p$, $\bar{\nu}_1 \in \mathbb{R}_+$ tais que $\bar{\lambda}_i^1 \in [\lambda_{\min}, \lambda_{\max}]$, $i = 1, \dots, m$, $\|\bar{\mu}^1\|_\infty \leq \mu_{\max}$ e $\bar{\nu}_1 \leq \nu_{\max}$. Assumimos que $x^0 \in \mathbb{R}^n$ é um ponto inicial arbitrário e $\{\epsilon_k\}$ é uma sequência de números positivos que satisfaz $\lim_{k \rightarrow \infty} \epsilon_k = 0$.

Passo 1. Inicialização.

Faça $k \leftarrow 1$.

Passo 2. *Resolva o subproblema.*

Calcule $x^k \in \mathbb{R}^n$ como uma solução aproximada de

$$\begin{aligned} & \underset{x \in \mathbb{R}^n}{\text{Minimizar}} && L_{\rho_k}(x, \bar{\lambda}^k, \bar{\mu}^k, \bar{\nu}_k) \\ & \text{sujeito a:} && \underline{h}(x) = 0, \\ & && \underline{g}(x) \leq 0. \end{aligned} \tag{2.9}$$

Passo 3. *Atualize o parâmetro de penalidade.*

Para todo $i = 1, \dots, p$, calcule $V_i^k = \max \left\{ g_i(x^k), -\frac{\bar{\mu}_i^k}{\rho_k} \right\}$.

Para todo $i = 1, \dots, q$, calcule $W_i^k = \begin{cases} \max \left\{ f_i(x^k), -\frac{\bar{\nu}_i^k}{\rho_k} \right\}, & \text{se } i \in \{i_1(x^k), \dots, i_r(x^k)\}, \\ 0, & \text{caso contrário.} \end{cases}$

Calcule

$$R_k = \max\{\|h(x^k)\|_\infty, \|V^k\|_\infty, \|W^k\|_\infty\}.$$

Se $k > 1$ e $R_k > \tau R_{k-1}$, defina $\rho_{k+1} = \eta \rho_k$. Caso contrário, defina $\rho_{k+1} = \rho_k$.

Passo 4. *Estime os multiplicadores.*

Calcule $\bar{\lambda}_i^{k+1} \in [\lambda_{\min}, \lambda_{\max}]$ para todo $i = 1, \dots, m$, $\bar{\mu}_i^{k+1} \in [0, \mu_{\max}]$ para todo $i = 1, \dots, p$, e $\bar{\nu}_{k+1} \in [0, \nu_{\max}]$. Faça $k \leftarrow k + 1$ e vá ao Passo 2.

Algoritmo 2.2.1. Proceda como no Algoritmo Conceitual 2.2.0, encontrando $x^k \in \mathbb{R}^n$ tal que existam $v^k \in \mathbb{R}^m$ e $w^k \in \mathbb{R}^p$ satisfazendo

$$\|\nabla L_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k, \bar{\nu}_k) + \sum_{i=1}^m v_i^k \nabla \underline{h}_i(x^k) + \sum_{i=1}^p w_i^k \nabla \underline{g}_i(x^k)\| \leq \epsilon_k, \tag{2.10}$$

$$w^k \geq 0, \underline{g}(x^k) \leq \epsilon_k, \tag{2.11}$$

$$\underline{g}_i(x^k) < -\epsilon_k \Rightarrow w_i^k = 0 \text{ para todo } i = 1, \dots, p, \tag{2.12}$$

$$\|\underline{h}(x^k)\| \leq \epsilon_k. \tag{2.13}$$

2.2.1 Solubilidade dos subproblemas

Em cada iteração do Algoritmo 2.2.1 minimizamos, aproximadamente, $L_{\rho_k}(x, \bar{\lambda}^k, \bar{\mu}^k, \bar{\nu}_k)$ em relação a x no conjunto viável definido por $\{x; \underline{h}(x) = 0 \text{ e } \underline{g}(x) \leq 0\}$. O critério de parada para o correspondente processo iterativo se reflete nas condições (2.10)-(2.13). Queremos mostrar que a obtenção de (2.10)-(2.13) é possível, em tempo finito, usando um algoritmo bem estabelecido.

Vamos definir:

$$F_{\min}(x) = \min_{I \mid \#I=r} \left\{ f(x) + \frac{\rho}{2} \left[\left\| h(x) + \frac{\lambda}{\rho} \right\|^2 + \left\| \left(g(x) + \frac{\mu}{\rho} \right)_+ \right\|^2 + \sum_{i \in I} \left(f_i(x) + \frac{\nu}{\rho} \right)_+^2 \right] \right\}.$$

O subproblema

$$\begin{aligned} \text{Minimizar} \quad & L_{\rho_k}(x, \bar{\lambda}^k, \bar{\mu}^k, \bar{\nu}_k) \\ & x \in \mathbb{R}^n \end{aligned}$$

$$\begin{aligned} \text{sujeito a:} \quad & \underline{h}(x) = 0, \\ & \underline{g}(x) \leq 0. \end{aligned}$$

é equivalente a

$$\begin{aligned} \text{Minimizar} \quad & F_{\min}(x) \\ & x \in \mathbb{R}^n \end{aligned}$$

$$\begin{aligned} \text{sujeito a:} \quad & \underline{h}(x) = 0, \\ & \underline{g}(x) \leq 0. \end{aligned}$$

Este é um problema do tipo LOVO com restrições como os definidos em [17]. Portanto, pode-se empregar Algoritmo C-LOVO de [17] para a sua resolução. Assim, assumindo que o conjunto viável $\{x; \underline{h}(x) = 0, \underline{g}(x) \leq 0\}$ é não vazio e que o conjunto definido por $\{x; \underline{g}(x) \leq \epsilon\}$ é limitado para algum $\epsilon > 0$, temos que as condições (2.10)-(2.13) são satisfeitas por algum iterando de C-LOVO em tempo finito.

2.2.2 Convergência do Algoritmo 2.2.1

Na seção anterior vimos que Algoritmo 2.2.1 está bem definido. Aqui queremos analisar suas propriedades de convergência. De agora em diante vamos supor que $K \subseteq \mathbb{N}$ é tal

que

$$\lim_{k \in K} x^k = x^*. \quad (2.14)$$

Como o número de subconjuntos de $\{1, \dots, q\}$ é finito, existem $K_1 \subseteq K$, $I \subset \{1, \dots, q\}$, $\#I = r$, tais que, para todo $k \in K_1$,

$$\{i_1(x^k), \dots, i_r(x^k)\} = I. \quad (2.15)$$

Para todo $k \in K_1$, $i \in I, j \notin I$ temos que:

$$f_i(x^k) \leq f_j(x^k). \quad (2.16)$$

Tomando limites em (2.16) podemos ver que para todo $i \in I, j \notin I$,

$$f_i(x^*) \leq f_j(x^*). \quad (2.17)$$

Com essas definições, graças a (1.5), a sequência gerada pelo Algoritmo 2.2.1 pode ser pensada como se tivesse sido gerada pelo Algoritmo 1.1.1, aplicado ao problema:

$$\begin{aligned} &\text{Minimizar} && f(x) \\ &&& x \in \mathbb{R}^n \\ &\text{sujeito a:} && h(x) = 0, \\ &&& g(x) \leq 0, \\ &&& f_i(x) \leq 0 \forall i \in I, \\ &&& \underline{h}(x) = 0, \\ &&& \underline{g}(x) \leq 0. \end{aligned} \quad (2.18)$$

Portanto, o resultado de convergência dado pelo Teorema 1.1.1 é válido para essa sequência. O resultado de convergência global está condensado no teorema abaixo.

Teorema 2.2.1. *Assuma que x^* , K , K_1 e I satisfazem (2.14), (2.15). Então, uma das seguintes três possibilidades vale:*

1. x^* é um ponto viável de (2.3).

2. x^* é um ponto KKT de

$$\begin{aligned} & \underset{x \in \mathbb{R}^n}{\text{Minimizar}} \quad \|h(x)\|^2 + \|g(x)_+\|^2 + \sum_{i \in I} f_i(x)_+^2 \\ & \text{sujeito a:} \quad \underline{h}(x) = 0, \\ & \quad \underline{g}(x) \leq 0. \end{aligned}$$

3. As restrições $\underline{h}(x) = 0$, $\underline{g}(x) \leq 0$ não satisfazem a condição de qualificação de dependência linear positiva constante (CPLD) [19, 93] em x^* .

Se x^* é um ponto viável de (2.3) então uma das seguintes duas possibilidades vale:

1. x^* é um ponto KKT do problema (2.18).
2. As restrições $h(x) = 0$, $g(x) \leq 0$, $f_i(x) \leq 0 \forall i \in I$, $\underline{h}(x) = 0$, $\underline{g}(x) \leq 0$ não satisfazem a condição de qualificação CPLD em x^* .

Observação. Por (2.17), o conjunto I associado às restrições do tipo LOVO $f_i(x) \leq 0$, $\forall i \in I$ no problema (2.18) corresponde aos índices dos r menores valores de $f_1(x^*), \dots, f_q(x^*)$. Essa relação entre I e x^* é boa consequência prática do Teorema 2.2.1. A importância desta propriedade se tornará mais aparente quando apresentarmos os algoritmos das seções seguintes.

2.2.3 Otimização Global com subproblemas do tipo LOVO

Como no caso do Algoritmo 1.1.0, vamos ver que uma versão do Algoritmo 2.2.0 converge para minimizadores globais de (2.3). Como na Seção 1.1, só precisamos encontrar minimizadores globais aproximados dos subproblemas.

Algoritmo 2.2.2.

Proceda como no Algoritmo Conceitual 2.2.0, com $x^k \in \Omega$ tal que

$$L_{\rho_k}(x^k, \bar{\lambda}^k, \bar{\mu}^k, \nu_k) \leq L_{\rho_k}(x, \bar{\lambda}^k, \bar{\mu}^k, \nu_k) + \epsilon_k$$

para todo $x \in \Omega$.

Observe que as funções \bar{f}_j definidas por $\bar{f}_j(x) = f_{i_j(x)}(x)$ são contínuas e que as restrições do tipo LOVO do problema são $\bar{f}_j(x) \leq 0$, $j = 1, \dots, r$. Portanto, o Algoritmo 2.2.2 é um caso particular do algoritmo principal de [28]. Como consequência, pelo Teorema 2 de [28], podemos provar o seguinte teorema.

Teorema 2.2.2. *Assuma que a região viável do problema (2.3) é não vazia. Então, cada ponto limite da sequência gerada pelo Algoritmo 2.2.2 é um minimizador global de (2.3).*

2.3 Algoritmo com subproblemas suaves

Em cada iteração do Algoritmo 2.2.0 resolvemos um problema de otimização com uma função objetivo do tipo LOVO. As condições (2.10)-(2.13) são as condições de otimalidade aproximadas para este subproblema. Este subproblema não suave pode ser resolvido usando C-LOVO [17], mas a alternativa de usar subproblemas suaves merece uma consideração cuidadosa. A ideia consiste em definir, no início de cada iteração externa k ,

$$I(x^{k-1}) = \{i_1(x^{k-1}), \dots, i_r(x^{k-1})\} \quad (2.19)$$

e fixar esse conjunto de índices na definição do Lagrangiano Aumentado. Ou seja, ao invés de (2.5), definimos:

$$L_\rho^k(x, \lambda, \mu, \nu) = f(x) + \frac{\rho}{2} \left[\left\| h(x) + \frac{\lambda}{\rho} \right\|^2 + \left\| \left(g(x) + \frac{\mu}{\rho} \right)_+ \right\|^2 + \sum_{j \in I(x^{k-1})} \left(f_j(x) + \frac{\nu}{\rho} \right)_+^2 \right]. \quad (2.20)$$

Novamente, é pertinente perguntar por que usar o Lagrangiano Aumentado L_ρ^k em vez de \tilde{L}_ρ^k , associado ao problema (2.2) e definido por:

$$\tilde{L}_\rho^k(x, \lambda, \mu, \nu) = f(x) + \frac{\rho}{2} \left[\left\| h(x) + \frac{\lambda}{\rho} \right\|^2 + \left\| \left(g(x) + \frac{\mu}{\rho} \right)_+ \right\|^2 + \left(f_{i_r(x^{k-1})}(x) + \frac{\nu}{\rho} \right)_+^2 \right].$$

A razão é a seguinte: Depois de minimizar L_{ρ}^k , provavelmente obtém-se um ponto de x tal que $f_j(x) \leq 0$, para pelo menos, r índices j (aqueles pertencentes a $I(x^{k-1})$). Isto implica que $f_{i_r(x)}(x) \leq 0$. Por outro lado, o melhor que podemos esperar de minimizar \tilde{L}_{ρ}^k é um ponto x tal que $f_{i_r(x^{k-1})}(x) \leq 0$. Uma vez que, muito provavelmente, a ordem das f'_i s muda de uma iteração para outra, esta propriedade não implica que $f_{i_r(x)}(x) \leq 0$. A função (2.20) tem a primeira derivada contínua, de modo que seu gradiente não precisa de uma definição especial. O Algoritmo Conceitual 2.3.0 é idêntico ao Algoritmo 2.2.0 exceto pelo fato de L_{ρ_k} ser substituído por $L_{\rho_k}^k$ em (2.10).

Algoritmo Conceitual 2.3.0.

Definir $\tau, \eta, \lambda_{\min}, \lambda_{\max}, \mu_{\max}, \rho_1, \bar{\lambda}^1, \bar{\mu}^1, \bar{\nu}_1$ e ϵ_k como no Algoritmo Conceitual 2.2.0. Os Passos 1, 3 e 4 são os mesmos do Algoritmo Conceitual 2.2.0. No Passo 2, o ponto $x^k \in \mathbb{R}^n$ é obtido como uma solução aproximada de

$$\begin{aligned} & \underset{x \in \mathbb{R}^n}{\text{Minimizar}} && L_{\rho_k}^k(x, \bar{\lambda}^k, \bar{\mu}^k, \bar{\nu}_k) \\ & \text{sujeito a:} && \underline{h}(x) = 0, \\ & && \underline{g}(x) \leq 0. \end{aligned} \tag{2.21}$$

Algoritmo 2.3.1.

Proceda como no Algoritmo Conceitual 2.3.0. Para o cálculo de uma solução aproximada de (2.21), proceda como no Algoritmo 2.2.1, substituindo a condição (2.10) por:

$$\|\nabla L_{\rho_k}^k(x^k, \bar{\lambda}^k, \bar{\mu}^k, \bar{\nu}_k) + \sum_{i=1}^m v_i^k \nabla \underline{h}_i(x^k) + \sum_{i=1}^p w_i^k \nabla \underline{g}_i(x^k)\| \leq \epsilon_k. \tag{2.22}$$

As condições (2.11)-(2.13) continuam as mesmas, como no Algoritmo 2.2.1.

2.3.1 Solubilidade dos subproblemas

Em cada iteração do Algoritmo 2.3.1 minimizamos, aproximadamente, $L_{\rho_k}^k(x, \bar{\lambda}^k, \bar{\mu}^k, \bar{\nu}_k)$ em relação a x no conjunto viável definido por $\{x; \underline{h}(x) = 0, \underline{g}(x) \leq 0\}$. O critério

de parada para o correspondente processo iterativo é dado pelas condições (2.22) e (2.11)-(2.13).

Como no caso do Algoritmo 2.2.1, obter o critério de parada é possível, em tempo finito, usando um algoritmo bem estabelecido. Neste caso, definimos:

$$F_{\min}^k(x) = f(x) + \frac{\rho}{2} \left[\left\| h(x) + \frac{\lambda}{\rho} \right\|^2 + \left\| \left(g(x) + \frac{\mu}{\rho} \right)_+ \right\|^2 + \sum_{j \in I(x^{k-1})} \left(f_j(x) + \frac{\nu}{\rho} \right)_+^2 \right],$$

o subproblema

$$\begin{aligned} & \text{Minimizar} && L_{\rho_k}^k(x, \bar{\lambda}^k, \bar{\mu}^k, \bar{\nu}_k) \\ & && x \in \mathbb{R}^n \\ & \text{sujeito a:} && \underline{h}(x) = 0, \\ & && \underline{g}(x) \leq 0. \end{aligned}$$

é equivalente a

$$\begin{aligned} & \text{Minimizar} && F_{\min}^k(x) \\ & && x \in \mathbb{R}^n \\ & \text{sujeito a:} && \underline{h}(x) = 0, \\ & && \underline{g}(x) \leq 0. \end{aligned}$$

Este é um problema suave de otimização com restrições. Assim, pode-se utilizar o Algoritmo 1.1.1 para a sua resolução. Supondo que o conjunto viável $\{x; \underline{h}(x) = 0, \underline{g}(x) \leq 0\}$ é não vazio e que o conjunto definido por $\{x; \underline{g}(x) \leq \epsilon\}$ é limitado para algum $\epsilon > 0$, sabemos que as condições (2.20) e (2.11)-(2.13) são satisfeitas por alguma iteração do Lagrangiano Aumentado [8].

2.3.2 Convergência do Algoritmo 2.3.1

Nesta seção, analisamos as propriedades de convergência do Algoritmo 2.3.1. Supomos aqui, como na Seção 2.2.2, que $K \subset_{\infty} \mathbb{N}$ é tal que

$$\lim_{k \in K} x^k = x^*. \tag{2.23}$$

Se necessário é possível redefinir K , tomando uma subsequência apropriada, de forma a assegurar que:

$$\lim_{k \in K} x^{k-1} = x^{**}. \quad (2.24)$$

Uma vez que o número de subconjuntos de $\{1, \dots, q\}$ é finito, existem $K_1 \subset K$, $I \subset \{1, \dots, q\}$, $\#I = r$, tais que, para todo $k \in K_1$,

$$\{i_1(x^{k-1}), \dots, i_r(x^{k-1})\} = I. \quad (2.25)$$

Para todo $k \in K_1$, $i \in I, j \notin I$ temos que:

$$f_i(x^{k-1}) \leq f_j(x^{k-1}). \quad (2.26)$$

Tomando o limite em (2.26) vemos que para todo $i \in I, j \notin I$,

$$f_i(x^{**}) \leq f_j(x^{**}). \quad (2.27)$$

Com essas definições, a sequência gerada pelo Algoritmo 2.3.1 pode ser considerada como sendo gerada pelo Algoritmo 1.1.1 aplicado ao problema:

$$\begin{aligned} & \text{Minimizar} && f(x) \\ & && x \in \mathbb{R}^n \\ & \text{sujeito a:} && h(x) = 0, \\ & && g(x) \leq 0, \\ & && f_i(x) \leq 0 \forall i \in I, \\ & && \underline{h}(x) = 0, \\ & && \underline{g}(x) \leq 0. \end{aligned} \quad (2.28)$$

Portanto, o Teorema 1.1.1 pode ser aplicado. O resultado de convergência global é declarado no teorema abaixo.

Teorema 2.3.1. *Assuma que x^* , x^{**} , K , K_1 e I satisfaz (2.23)-(2.25). Então, uma das seguintes três possibilidades vale:*

1. x^* é um ponto viável de (2.3).

2. x^* é um ponto KKT de

$$\begin{aligned} \text{Minimizar} \quad & \|h(x)\|^2 + \|g(x)_+\|^2 + \sum_{i \in I} f_i(x)_+^2 \\ & x \in \mathbb{R}^n \\ \text{sujeito a:} \quad & \underline{h}(x) = 0, \\ & \underline{g}(x) \leq 0. \end{aligned}$$

3. As restrições $\underline{h}(x) = 0$, $\underline{g}(x) \leq 0$ não satisfazem a condição de qualificação de dependência linear positiva constante (CPLD) [19, 93] em x^* .

Se x^* é um ponto viável de (2.28) Então, uma das duas seguintes possibilidades vale:

1. x^* é um ponto KKT do problema (2.28).
2. As restrições $h(x) = 0$, $g(x) \leq 0$, $f_i(x) \leq 0 \forall i \in I$, $\underline{h}(x) = 0$, $\underline{g}(x) \leq 0$ não satisfazem a condição de qualificação CPLD em x^* .

Observação. O fato de um ponto ser viável para (2.28) implica que ele é viável para (2.3). Entretanto um ponto viável para (2.3) pode não ser viável para (2.28). No caso em que

$$x^{**} = \lim_{k \in K_1} x^{k-1} = \lim_{k \in K_1} x^k = x^*, \quad (2.29)$$

o resultado do Teorema 2.3.1 é o mesmo que o de Teorema 2.2.1. No entanto, (2.29) é uma propriedade da sequência gerada pelo algoritmo (não uma propriedade do problema) e, portanto, pode não ser verificada pela sequência gerada pelo Algoritmo 2.3.1. Em todo caso, o ponto limite x^* geralmente satisfaz todas as restrições $f_i(x) \leq 0$, $\forall i \in I$. Assim, $f_i(x^*) \leq 0$, para pelo menos, r índices i . Isto significa que a restrição $f_{i_r(x)}(x) \leq 0$ é certamente satisfeita por x^* mas existe a possibilidade de $f_{i_{r+s}}(x^*) \leq 0$ também para algum $s > 0$. Desde modo é possível que x^* satisfaça mais restrições do que necessário, e portanto o ponto x^* seja talvez, apenas sub-ótimo.

Para ver a consequência prática da observação acima, temos que discutir primeiro o critério de parada adequada para o Algoritmo 2.3.1. Pelo Teorema 2.3.1,

podemos esperar que os pontos limite do algoritmo resolvam o problema (2.28). Portanto, o algoritmo deve parar no iterando x^k quando x^k satisfaz aproximadamente a condição KKT para esse problema. (Para uma melhor discussão sobre as condições aproximadamente KKT, consulte [15]) Em outras palavras, o algoritmo deve parar em x^k quando x^k é, provavelmente, uma solução aproximada de

$$\begin{aligned}
 & \text{Minimizar} && f(x) \\
 & && x \in \mathbb{R}^n \\
 & \text{sujeito a:} && h(x) = 0, \\
 & && g(x) \leq 0, \\
 & && f_i(x) \leq 0 \quad \forall i \in I(x^{k-1}), \\
 & && \underline{h}(x) = 0, \\
 & && \underline{g}(x) \leq 0.
 \end{aligned} \tag{2.30}$$

Uma vez que (com alguma pequena tolerância ϵ) $f_i(x^k) \leq 0$ para todo $i \in I(x^{k-1})$ e $I(x^{k-1})$ contendo r índices, verifica-se que $f_i(x^k) \leq 0$ para todo $i \in I(x^k)$ (com tolerância ϵ). Portanto, x^k é um provável minimizador de $f(x)$ num conjunto que, desconsiderando a tolerância, está (talvez de forma estrita) contido no conjunto viável do seguinte problema:

$$\begin{aligned}
 & \text{Minimizar} && f(x) \\
 & && x \in \mathbb{R}^n \\
 & \text{sujeito a:} && h(x) = 0, \\
 & && g(x) \leq 0, \\
 & && f_i(x) \leq 0 \quad \forall i \in I(x^k), \\
 & && \underline{h}(x) = 0, \\
 & && \underline{g}(x) \leq 0.
 \end{aligned} \tag{2.31}$$

Isto significa que o mínimo de (2.30) pode ser maior do que o mínimo de (2.31). Então, embora o Algoritmo 2.3.1 geralmente calcule pontos (quase) viáveis para o problema (2.3), seus iterandos podem ser piores do que os gerados pelo Algoritmo 2.2.1.

2.3.3 Propriedades de otimização global

Seguindo as ideias de seções anteriores, é natural perguntar sobre as propriedades do Algoritmo 2.3.1 quando, em vez de (2.22), (2.11), (2.12) e (2.13), exigimos que x^k seja um minimizador global aproximado de $L_{\rho_k}^k(x, \bar{\lambda}^k, \bar{\mu}^k, \bar{\nu}_k)$ em relação a $x \in \Omega$.

Isso corresponde ao seguinte algoritmo.

Algoritmo 2.3.2.

Proceda como no Algoritmo 2.3.1 exceto que $x^k \in \Omega$ é tal que

$$L_{\rho_k}^k(x^k, \bar{\lambda}^k, \bar{\mu}^k, \bar{\nu}_k) \leq L_{\rho_k}^k(x, \bar{\lambda}^k, \bar{\mu}^k, \bar{\nu}_k) + \epsilon_k$$

para todo $x \in \Omega$.

As propriedades de otimização global do Algoritmo 2.3.2 são dadas no seguinte teorema.

Teorema 2.3.2. *Assuma que a região viável do problema (2.3) é não vazia. Seja $\{x^k\}$ uma sequência gerada pelo Algoritmo 2.3.2 e x^* um ponto limite, ao passo que x^{**} , K, K_1 e I satisfazem (2.23)-(2.25). Então, x^* é uma solução global do problema (2.28).*

Demonstração: Como na Seção 2.3.2, podemos assumir que a subsequência que converge para x^* é gerada pelo Algoritmo 1.1.2 aplicado ao problema (2.28). Então, pelo Teorema 1.1.2, x^* é uma solução global de (2.28), como queríamos demonstrar. \square

Note que, mesmo no caso de $x^* = x^{**}$ este resultado é mais fraco do que o obtido no Teorema 2.2.2 pelo Algoritmo 2.2.2. De fato, se $x^* = x^{**}$, segue-se que x^* é um minimizador global de $f(x)$ sujeito a: $h(x) = 0$, $g(x) \leq 0$, $x \in \Omega$ e $f_i(x) \leq 0$ para todo $i \in I$, onde $I = \{i_1(x^*), \dots, i_r(x^*)\}$, mas isso não implica que x^* é uma solução de (2.3). Por exemplo, considere o problema

$$\begin{aligned} &\text{Minimizar } x \\ &\text{sujeito a: } \min\{(x-1)^2 - 1, (x+1)^2 - 1\} \leq 0. \end{aligned} \tag{2.32}$$

Se tomarmos $x^k = x^* = 0$, para todo k teremos que $x^{**} = 0$, $K = K_1 = \mathbb{N}$ e $I = \{\{1\}\}$ é uma escolha possível que satisfaz (2.23)-(2.25). Entretanto, x^* é um minimizador global da função objetivo sujeito a $f_{i_1(x^*)}(x) \leq 0$ mas o minimizador global do problema (2.32) é -2 .

2.4 Algoritmo LOVO de ponto fixo

O último algoritmo apresentado neste trabalho para resolver (2.3) é do tipo ponto fixo. A ideia é resolver o problema por meio de um pequeno número de problemas (talvez um único) de otimização restrita suave. Para todo $k = 1, 2, \dots$ definimos $I(x^{k-1})$ como em (2.19). A definição do algoritmo é a seguinte.

Algoritmo 2.4.1.

Tomemos $x^0 \in \mathbb{R}^n$ como um ponto inicial arbitrário.

Passo 1. *Inicialização.*

Faça $k \leftarrow 1$.

Passo 2. *Resolva o subproblema.*

Calcule x^k como (possível) solução de

$$\begin{aligned}
 & \text{Minimizar} && f(x) \\
 & && x \in \mathbb{R}^n \\
 & \text{sujeito a:} && h(x) = 0, \\
 & && g(x) \leq 0, \\
 & && f_i(x) \leq 0 \quad \forall i \in I(x^{k-1}), \\
 & && \underline{h}(x) = 0, \\
 & && \underline{g}(x) \leq 0.
 \end{aligned} \tag{2.33}$$

Passo 3. Se $I(x^{k-1}) = I(x^k)$, defina $x^{\text{final}} = x^k$ e pare. Caso contrário, faça $k \leftarrow k + 1$ e vá ao Passo 1.

Uma alternativa ao problema (2.33) pode ser:

$$\begin{aligned}
 &\text{Minimizar} && f(x) \\
 &&& x \in \mathbb{R}^n \\
 &\text{sujeito a:} && h(x) = 0, \\
 &&& g(x) \leq 0, \\
 &&& f_{i_r(x^{k-1})}(x) \leq 0, \\
 &&& \underline{h}(x) = 0, \\
 &&& \underline{g}(x) \leq 0.
 \end{aligned} \tag{2.34}$$

No entanto, no caso de resolver com sucesso (2.34), obteríamos apenas $f_{i_r(x^{k-1})}(x^k) \leq 0$. Ou seja, o cumprimento de $f_j(x) \leq 0$ estaria garantido para apenas um índice j . Por outro lado, resolvendo com sucesso (2.33) conseguimos x^k que satisfaz $f_j(x) \leq 0$, para pelo menos, r índices. Portanto, $f_{i_r(x^k)}(x^k) \leq 0$.

Como “possível” solução do problema (2.33) entendemos um ponto limite de uma sequência gerada por um algoritmo para otimização restrita suave. Podemos usar o Algoritmo 1.1.1 ou o Algoritmo 1.1.2 para esse fim. Se

$$I(x^k) = I(x^{k-1}) \tag{2.35}$$

então, muito provavelmente, x^{k+s} seria idêntico a x^k para todo $s = 1, 2, \dots$. Portanto, a identidade entre $I(x^{k-1})$ e $I(x^k)$ é um critério de parada prático sensato para o Algoritmo 2.4.1.

O Algoritmo 2.4.1 lembra o terceiro algoritmo proposto no artigo de Gaivronski e Pflug [56]. Estes autores consideraram o caso de função objetivo, retornos e restrições adicionais, lineares, observando que, neste caso, (2.33) se reduz a um problema de Programação Linear. Em vez de usar sempre $I(x^{k-1})$, para as restrições de risco, eles sugerem como selecionar um conjunto adequado de r cenários usando heurísticas convenientes.

Teorema 2.4.1. *Assuma que usamos o Algoritmo 1.1.1 no Passo 2 do Algoritmo*

2.4.1 e que a sequência $\{x^k\}$, gerada pelo Algoritmo 2.4.1, pára no ponto viável x^{final} . Então, pelo menos uma de duas possibilidades a seguir é válida:

1. x^{final} é um ponto KKT do problema

$$\begin{aligned}
 & \text{Minimizar} && f(x) \\
 & && x \in \mathbb{R}^n \\
 & \text{sujeito a:} && h(x) = 0, \\
 & && g(x) \leq 0, \\
 & && f_i(x) \leq 0 \quad i \in I(x^{\text{final}}), \\
 & && \underline{h}(x) = 0, \\
 & && \underline{g}(x) \leq 0.
 \end{aligned} \tag{2.36}$$

2. As restrições do problema (2.36) não satisfazem a condição de qualificação CPLD em x^{final} .

Demonstração: A prova decorre do Teorema 1.1.1 usando a identidade (2.35). \square

A prova do teorema final desta seção segue diretamente da definição do Algoritmo 2.4.1 e do fato de que $I(x^k) = I(x^{k-1})$ na iteração final.

Teorema 2.4.2. *Assuma que usamos o Algoritmo 1.1.2 no Passo 2 do Algoritmo 2.4.1 e que a sequência $\{x^k\}$, gerada pelo Algoritmo 2.4.1, pára no ponto viável x^{final} . Então, x^{final} é uma solução global de*

$$\begin{aligned}
 & \text{Minimizar} && f(x) \\
 & && x \in \mathbb{R}^n \\
 & \text{sujeito a:} && h(x) = 0, \\
 & && g(x) \leq 0, \\
 & && f_i(x) \leq 0 \quad i \in I(x^{\text{final}}), \\
 & && \underline{h}(x) = 0, \\
 & && \underline{g}(x) \leq 0.
 \end{aligned}$$

Note que a tese do Teorema 2.4.2 não implica que x^{final} é uma solução do problema original (2.3). Na verdade, o exemplo no final de Seção 2.3 mostra que $x^* = 0$ pode ser um ponto fixo, que não é uma solução de (2.3).

Uma última observação é que em todos os resultados apresentados nos Capítulos 1 e 2 é possível substituir a CPLD por condições de qualificação mais fracas tais como a R-CPLD [13], a CRSC [14] ou a CPG [14] que foram recentemente introduzidas.

Capítulo 3

Problema de seleção de carteiras com restrição de risco e experimentos numéricos

Neste capítulo vamos apresentar uma aplicação natural da formulação de problemas com restrições do tipo VaR que vem da seleção de carteiras com restrição de risco. Os três algoritmos propostos no Capítulo 2 serão utilizados na resolução numérica de variações de nossa aplicação, e um estudo comparativo entre eles é apresentado.

Problemas de otimização de portfólio lidam com a distribuição do capital entre ativos diferentes, em geral, ativos de risco ou da combinação de um ativo livre de risco e outros arriscados. O objetivo é selecionar uma combinação que maximiza o ganho esperado com um futuro nível de risco tolerável ou encontrar uma carteira com o menor risco entre todas as carteiras que têm um ganho futuro esperado de ao menos algum valor especificado. Modelagem e medição de risco, bem como estimar os ganhos futuros, não são tarefas triviais e uma quantidade considerável de pesquisa tem sido dedicada a estes temas (ver, por exemplo, [97]). Três medidas de risco tradicionais estão atualmente em uso: a variância da carteira, desde o trabalho pioneiro de Markowitz [76], *Value-at-Risk* (VaR) [63] e *Conditional Value-at-Risk* (CVaR) [96]. Suas relações

mútuas são analisados em [57] e todos as três apresentam vantagens e desvantagens. Medidas de risco mais gerais são discutidas em [24].

Neste capítulo trabalharemos com a VaR e com carteiras compostas por ações e um ativo livre de risco. As atuais normas para a indústria financeira regulam em termos da VaR o risco dos investimentos de uma instituição e, portanto a VaR é uma ferramenta padrão para a gestão de risco. Por definição, a VaR é o percentil da função de distribuição da perda futura dado um nível de confiança α . A medida α -VaR de um instrumento financeiro é a menor quantidade tal que a perda é menor do que ou igual a ela, com probabilidade α . Agências reguladoras geralmente exigem que o capital disponível deva ser um múltiplo da VaR. Uma visão detalhada das propriedades da VaR e de suas aplicações em gestão de risco é dada em [63].

O problema de otimização, que produz uma carteira com o ganho máximo e satisfaça as restrições de VaR é considerado em [39, 56], onde alguns algoritmos são sugeridos. O problema é difícil, devido à geometria complicada do conjunto viável. Neste trabalho vamos incluir os custos de transação no problema de otimização de carteiras. Estes custos são inevitavelmente presentes na vida real e podem diminuir significativamente o rendimento de uma carteira. Vários trabalhos tratam da otimização de carteiras com custos de transação usando diferentes medidas de risco. Em [25, 71, 74, 90] os custos de transação são modelados como funções lineares ou lineares por partes.

Os custos de transação podem ser divididos em dois tipos, as despesas fixas e os custos do impacto. Os custos fixos são as taxas e impostos, sendo em geral proporcionais ao valor da transação. Devido ao rápido desenvolvimento de pregão eletrônico nas últimas duas décadas e o grande número de participantes, os custos fixos não podem ser uma parte dominante do total de custos, especialmente para grandes investidores institucionais. Por outro lado, eles ainda são significativos para pequenos investidores e, portanto, precisam ser incluídos em um modelo realista. Descrever os custos do impacto é uma questão muito complicada e não há na literatura um acordo geral sobre o modelo adequado. Um custo de impacto é um desvio do preço de equilíbrio causado por

uma atividade comercial própria. Se nós formos comprar uma grande quantidade de uma determinada ação, então estamos obviamente aumentando a demanda e aumentando o preço do ativo. É difícil distinguir se as mudanças de preços são causadas por uma atividade comercial ou por ruídos. Grandes instituições financeiras utilizam modelos próprios para medir o impacto. Embora com base em trabalhos acadêmicos, por razões comerciais, esses modelos não são disponíveis ao público. Diferentes abordagens para essa modelagem foram publicadas em [5, 6, 31, 70]. Neste trabalho adotaremos o modelo de impacto de mercado proposto em [5, 6].

Como já foi dito anteriormente, dadas as funções $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$, $i = 1, \dots, q$, a função $f_{i_r(x)}(x)$, onde $f_{i_1(x)}(x) \leq f_{i_2(x)}(x) \leq \dots \leq f_{i_q(x)}(x)$ pode ser interpretada como uma versão discreta da função VaR associada à carteira x , cenários $\{1, \dots, q\}$, funções de perda f_1, \dots, f_q e nível de confiança $\alpha = r/q$ (veja [11, 12, 18]). Em geral $r < q$ e $r \approx q$. Na situação mais típica deseja-se minimizar a perda média associada a algum investimento financeiro sujeito à condição de que a VaR que corresponde à decisão ótima não deva exceder a tolerância c . Dessa forma temos um problema com a restrição:

$$f_{i_r(x)}(x) \leq c,$$

portanto temos um problema do tipo (2.2) e conseqüentemente temos um problema do tipo (2.3). Dessa forma os algoritmos propostos no Capítulo 2 são aplicáveis.

Para os testes computacionais foram implementados os Algoritmos 2.2.1, 2.3.1 e 2.4.1, fazendo as modificações adequadas na versão 2.2.1 de Algencan e utilizando todos os parâmetros padrões. Isso implica que as propriedades de otimização global mencionadas nos Teoremas 2.2.2, 2.3.2 e 2.4.2 não podem ser garantidas para essas implementações. No entanto, as implementações Algencan são projetadas de tal forma que minimizadores globais dos subproblemas são buscados, independentemente do cumprimento das condições de estacionariedade aproximada dos subproblemas. Em outras palavras, os *solvers* de nossos subproblemas tentam sempre encontrar o menor valor possível de função, mesmo que isso não seja necessário para a obtenção de mi-

minimizadores locais aproximados. Como consequência, o comportamento prático dos métodos do tipo Algencan costumam ser bem explicados pelas suas propriedades de otimização global. A “preferência por minimizadores globais” do método suave original Algencan tem sido discutida em [8].

Para a resolução do problema (2.3) com os Algoritmos 2.2.1, 2.3.1 e 2.4.1, as avaliações de função objetivo, restrições e suas derivadas são as tarefas mais demoradas computacionalmente. Em particular, entre elas, a tarefa mais demorada é a avaliação de restrições $f_{i_j(x)}(x) \leq 0, j = 1, \dots, r$, que envolve, para um dado x , a seleção dos r menores valores entre $f_i(x), i = 1, \dots, q$. Foi implementado um algoritmo do tipo “*divide-and-conquer*” chamado *Randomized-Select* (veja [38] pp. 185-192), cuja complexidade esperada é linear sobre o número de cenários q (para qualquer valor de r).

Os códigos das implementações estão em Fortran 77 e, juntamente com alguns dados dos problemas que tornam possível a reprodução de experimentos numéricos, estão disponíveis para *download* em [105]. Os códigos foram compilados com gfortran (GNU Fortran versão 4.2.1) e a opção de compilador adotada foi -O4. Todos os experimentos foram executados em um processador 2.4GHz Intel Core 2 Quad Q6600, com 4 GB de memória RAM e sistema operacional Linux.

3.1 Testes preliminares

Antes de detalhar a nossa aplicação principal, nessa seção vamos comparar os Algoritmos 2.2.1, 2.3.1 e 2.4.1 usando um problema mais simples de seleção de carteiras com restrições de risco. Assuma que temos um histórico de retornos percentuais para n ativos e, utilizando estas informações, vamos simular uma lista de q n -uplas igualmente prováveis de preços no final de um período de tempo (igual a 10 dias úteis nos experimentos abaixo). Cada n -upla representa um cenário diferente e define uma função perda. Queremos minimizar a perda média sujeito a $f_{i_r(x)}(x) \leq 0$, onde f_j é a diferença entre as perdas no cenário j e a perda tolerada, para $j = 1, \dots, q$. Assumimos que x_j é o valor investido no ativo j , $j = 1, \dots, n$. Seja θ_{ij} o quociente entre o preço final

do ativo j no cenário i e o preço inicial desse ativo. Portanto, no final do período temos $\sum_{j=1}^n \theta_{ij} x_j$ unidades monetárias. Em média, o nosso valor final de dinheiro será: $\sum_{j=1}^n \hat{\theta}_j x_j$, onde $\hat{\theta}_j = \frac{1}{q} \sum_{i=1}^q \theta_{ij}$ para $j = 1, \dots, n$. A função natural a ser minimizada é, portanto,

$$\text{Perda média esperada} \equiv f(x) = - \sum_{j=1}^n \hat{\theta}_j x_j.$$

A restrição orçamentária natural é dada por

$$\sum_{j=1}^n x_j = M \tag{3.1}$$

e, para evitar investimentos negativos, devemos incluir as restrições de não-negatividade

$$x_j \geq 0, \quad j = 1, \dots, n. \tag{3.2}$$

O problema de minimizar $f(x)$ sujeito a (3.1) e (3.2) tem pelo menos uma solução trivial x^* , dada por

$$x_j^* = M \text{ se } \hat{\theta}_j = \max\{\hat{\theta}_1, \dots, \hat{\theta}_n\}, \quad x_j^* = 0, \text{ caso contrário.}$$

A restrição do tipo VaR adicional impõe que, em pelo menos r dos cenários, a perda não deva exceder a tolerância T_{loss} . Isso significa que:

$$\text{Perda} \equiv M - \sum_{j=1}^n \theta_{ij} x_j \leq T_{\text{loss}} \tag{3.3}$$

para pelo menos r índices i .

Em nossos experimentos tomamos $T_{\text{loss}} = 0.05M$ e $r = 0.99q$. Isto significa que a perda máxima permitida com probabilidade $\alpha = r/q = 0.99$ deve ser menor do que $(T_{\text{loss}}/M)100\% = 5\%$ do capital investido.

Supomos que um dos ativos (o que corresponde a $j = n$) é isento de riscos e tem taxa de retorno igual a zero. Este fato é expresso afirmando que $\theta_{in} = 1$ para todo $i = 1, \dots, q$. Como consequência, a carteira dada por:

$$x_n = M, x_j = 0 \text{ for all } j = 1, \dots, n - 1, \tag{3.4}$$

é atraente como uma aproximação inicial, mas tem a desvantagem de que $f_i(x)$ são as mesmas para todo cenário i . Isto faz com que a escolha de $i_1(x), \dots, i_r(x)$ seja bastante ambígua. Se uma má escolha é feita essa poderia levar a minimizadores locais ruins para o problema. Portanto, a escolha inicial

$$x_n = 0.5M, x_j = 0.5M/(n-1) \text{ para todo } j = 1, \dots, n-1 \quad (3.5)$$

é mais promissora e foi a adotada. Observe que a carteira (3.5) pode não satisfazer a restrição (3.3), mas este não é um grave inconveniente para abordagens do tipo Lagrangiano Aumentado.

Resumindo, nesta seção, queremos resolver o problema (2.3) enunciado como:

$$\text{Minimizar} \quad - \sum_{j=1}^n \hat{\theta}_j x_j \quad (3.6)$$

$$\text{sujeito a: } f_{i_k(x)}(x) \leq 0, k = 1, \dots, r, \quad (3.7)$$

$$\sum_{j=1}^n x_j = M, \quad (3.8)$$

$$x \in \Omega, \quad (3.9)$$

onde $f_i(x) = M - \sum_{j=1}^n \theta_{ij} x_j - T_{\text{loss}}$, $i = 1, \dots, q$, e $\Omega = \{x \in \mathbb{R}^n \mid x \geq 0\}$.

Usamos $n = 8$ e $q = 1000$ nos nossos experimentos. Para a reprodutibilidade, a matriz dos cenários $\theta \in \mathbb{R}^{q \times n}$ pode ser encontrada em [105]. Uma informação que pode ser útil para analisar os resultados obtidos é que, como consequência de ter

$$\hat{\theta} \approx (0.9994, 1.0005, 0.9958, 1.0003, 1.0063, 1.0059, 1.0023, 1.0000)^T,$$

a solução do problema (3.6)-(3.9) ignorando a restrição do tipo VaR (3.7) é dada por $(100/M) \times x_5^* = 100$ e $x_i^* = 0$, $\forall i \neq 5$, com $(-100/M) \times f(x^*) \approx 100.63$. As soluções obtidas pelos Algoritmos 2.2.1, 2.3.1 e 2.4.1 são apresentadas na Tabela 3.1.

3.2 Modelo com custos de transação

Vamos agora descrever os custos de transação para o problema (3.6)-(3.9). De acordo com o modelo descrito em [5, 6], os custos do impacto podem ser temporários ou

Método	$(-100/M) \times f(x^*)$	$(100/M) \times x^*$
Algoritmo 2.2.1	100.40	$x_1 = x_2 = x_3 = x_4 = 0$ $x_5 \approx 1.58$ $x_6 \approx 65.79$ $x_7 \approx 1.62$ $x_8 \approx 31.01$
Algoritmo 2.3.1	100.41	$x_1 = x_2 = x_3 = x_4 = x_7 = 0$ $x_5 \approx 27.54$ $x_6 \approx 39.92$ $x_8 \approx 32.54$
Algoritmo 2.4.1	100.42	$x_1 = x_2 = x_3 = x_4 = 0$ $x_5 \approx 20.98$ $x_6 \approx 46.60$ $x_7 \approx 4.96$ $x_8 \approx 27.46$

Tabela 3.1: Desempenho dos Algoritmos 2.2.1, 2.3.1 e 2.4.1 aplicados ao problema simples, sem custos de transação.

permanentes. Um impacto temporário é um desvio de curta duração a partir do preço de equilíbrio causado por nossa própria negociação, e isso pode afetar nossas operações em andamento. Sua duração é regida principalmente pelo padrão de liquidez das ações que estamos negociando. Um impacto permanente pode continuar bem depois que a transação seja executada, mas é significativamente menor (por uma ordem de grandeza) do que o impacto temporário. Ambos os impactos, temporários e permanentes, são funções côncavas. Ao contrário dos custos fixos, custos de impacto são importantes para grandes investidores institucionais uma vez que suas operações tendem a ser de grande volume, enquanto a sua importância para os pequenos investidores não é grande. Devido a todas essas considerações, vamos incluir tanto os custos fixos quanto os custos de impacto em nosso modelo e considerar separadamente pequenos e grandes investidores nos testes numéricos. Devido à presença de custos de impacto, teremos uma função objetivo não-linear em nosso problema de otimização. Os custos de transação também são divididos em custos fixos e de impacto em [65], onde o problema é formulado como minimização da VaR com restrições de custos e de produtividade.

Vamos descrever o modelo que vamos considerar a partir de agora. Suponha que o conjunto de n diferentes ativos $\{1, 2, \dots, n\}$ está disponível. Denotando por x_j o montante investido no ativo j , temos:

$$\sum_{j=1}^n x_j = M, \quad x_j \geq 0, \quad j = 1, \dots, n. \quad (3.10)$$

Segundo [5, 6], o impacto é uma função de intensidade de comércio, que depende de vários parâmetros das ações específicas: o *spread* ϵ_j , o preço inicial do ativo π_j , o volume médio diário ADV_j e um parâmetro adicional $\beta \in (0, 1]$ que determina a não-linearidade da função de impacto.

Vamos definir a função do impacto temporário [5], em relação ao ativo j , da seguinte forma:

$$H(x_j) = \frac{\epsilon_j}{2\pi_j} + \frac{\epsilon_j}{2\pi_j} \left(\frac{100x_j}{ADV_j\pi_j} \right)^\beta. \quad (3.11)$$

A função de impacto permanente [5] será dada por:

$$G(x_j) = \frac{\epsilon_j}{\pi_j} \left(\frac{10x_j}{ADV_j \pi_j} \right)^\beta. \quad (3.12)$$

Supondo que $c_j x_j$ é o custo fixo de negociar x_j unidades monetárias do ativo j , o custo total da operação que inclui o impacto e os custos fixos serão os seguintes:

$$t_j(x_j) = x_j(H(x_j) + G(x_j) + c_j).$$

Desta forma, a regra dada em [5] é levada em conta. Esta regra estabelece que uma ordem de compra de 1% do ADV_j do ativo j move temporariamente o preço de um spread ϵ_j enquanto o mesmo impacto permanente é atingido com volume igual a 10% do ADV_j . Portanto, o impacto temporário é maior que o permanente na magnitude de uma ordem de grandeza.

Os custos fixos, compostos de todos os impostos e taxas, são em geral, proporcionais ao valor da operação, mas com diferentes fatores para os investidores pequenos e grandes. Vamos supor que um pequeno investidor tem um custo fixo igual a 1% do valor da transação para qualquer ativo. Por outro lado, um grande investidor institucional vai pagar 0,1 % para qualquer ativo.

Vamos supor que a sequência histórica (ou simulada) de retornos dos ativos esteja disponível. Isto significa que a matriz $\theta = (\theta_{ij})$, que nos dá o quociente entre o preço final e o preço inicial do ativo j no cenário i , está disponível. O valor final de nossa carteira, no cenário i , será dado por $\sum_{j=1}^n \theta_{ij} x_j$. Portanto, definindo, como na Seção 3.1, $\hat{\theta}_j = \sum_{i=1}^q \theta_{ij} / q$, o valor médio da carteira no final do período é $\sum_{j=1}^n \hat{\theta}_j x_j$. Queremos maximizar esse valor, descontando os custos de transação e considerando a restrição do tipo VaR. Isto significa que a nossa função objetivo (2.3) deve ser:

$$f(x) = - \sum_{j=1}^n \hat{\theta}_j x_j + \sum_{j=1}^n t_j(x_j).$$

A restrição do tipo VaR é determinada pela definição de $f_i(x)$ para cada cenário i . No caso considerado temos:

$$f_i(x) = M - \sum_{j=1}^n \theta_{ij} x_j + \sum_{j=1}^n t_j(x_j) - T_{\text{loss}},$$

onde T_{loss} denota a perda tolerada.

Como na Seção 3.1, tomamos $T_{\text{loss}} = 0.05M$, $r = 0.99q$ e $q = 1000$ em nossos experimentos. Além disso, assumimos também que um dos ativos (o que corresponde a $j = n$) é isento de riscos afirmando que $\theta_{in} = 1$ para todo $i = 1, \dots, q$, $\epsilon_n = 0$ e $c_n = 0$. Finalmente, nós também usamos $x_n = 0.5M$, $x_j = 0.5M/(n - 1)$ para $j = 1, \dots, n - 1$ como escolha inicial.

Resumindo, nesta seção, queremos resolver o problema (2.3) enunciado como:

$$\text{Minimizar} \quad - \sum_{j=1}^n \widehat{\theta}_j x_j + \sum_{j=1}^n t_j(x_j) \quad (3.13)$$

$$\text{sujeito a:} \quad f_{i_k(x)} \leq 0, k = 1, \dots, r, \quad (3.14)$$

$$\sum_{j=1}^n x_j = M, \quad (3.15)$$

$$x \in \Omega, \quad (3.16)$$

onde

$$f_i(x) = M - \sum_{j=1}^n \theta_{ij} x_j + \sum_{j=1}^n t_j(x_j) - T_{\text{loss}}, i = 1, \dots, q,$$

e

$$\Omega = \{x \in \mathbb{R}^n \mid x \geq 0\}.$$

Os algoritmos foram testados utilizando os problemas descritos abaixo. O conjunto de $n = 8$ ativos disponíveis consiste de sete ativos do *FTSE*: AZN.L, BARC.L, KGF.L, LLOY.L, MKS.L, TSCO.L e VOD.L, além de um ativo livre de risco. Consideramos $q = 1000$ cenários. As matrizes de cenários de θ foram geradas com os dados históricos de retornos diários de 19 de março 2004 a 12 de fevereiro 2008. Para o caso de pequenos investidores consideramos $M = 10,000$ (libras) e carteiras com um tempo de vida de 120 e 240 dias úteis. Para o caso de grande investidor consideramos $M = 100,000,000$ e carteiras com 10 dias de tempo de vida. As matrizes θ , assim como os custos de transação relacionadas, constantes ϵ_j , π_j e $ADV_j, j = 1, \dots, n$ podem ser encontradas em [105]. A Tabela 3.2 mostra as soluções encontradas pelos Algoritmos 2.2.1, 2.3.1 e 2.4.1 no caso dos grandes investidores usando $\beta = 1$ e $\beta = 0.6$ nas fórmulas

dos custos de transação, respectivamente. As soluções, no caso dos pequenos investidores para o caso em que $\beta = 1$ são apresentadas na Tabela 3.3. Como era esperado, uma vez que fatores de impacto são insignificantes para os pequenos investidores, os testes feitos com $\beta = 0.6$ produzem resultados quase idênticos.

3.3 Problemas de grande porte

Nesta seção, vamos considerar casos de grande escala do problema (3.13)-(3.16) com $n \in \{500, 1000\}$ ativos e $q = 1000$ cenários. Os dados $\theta \in \mathbb{R}^{q \times n}$ e $\epsilon, \pi, ADV \in \mathbb{R}^n$ podem ser encontrados em [105]. Usamos $\beta = 0.6$ para lidar com os custos de transação não-linear. Vamos lidar com o caso do investidor de grande porte, com $M = 100,000,000$, $c_j = 0.001$ para todo j . A matriz θ simula cenários para 120 dias futuros. Consideramos também que o ativo n é livre de riscos e usamos $x_n = 0.5M$, $x_j = 0.5M/(n-1)$ para $j = 1, \dots, n-1$ como ponto inicial.

A fim de avaliar a influência da restrição do tipo VaR na qualidade do valor da função objetivo na solução, resolvemos o problema para diferentes valores de parâmetros para a restrição do tipo VaR

$$\text{nível de confiança} = r/q \in \{0.90, 0.91, \dots, 0.99\}$$

e

$$\text{Perda máxima percentual} = (100/M) \times T_{\text{loss}} \in \{1, 2, \dots, 10\}.$$

As Tabelas 3.4 e 3.5 mostram o valor ótimo obtido através dos Algoritmos 2.2.1, 2.3.1 e 2.4.1 para cada combinação de r e T_{loss} . Note que quanto mais rigorosa a restrição do tipo VaR (inferior direito), o ganho esperado é menor.

Os tempos médio de CPU de cada método são 64.82, 60.10 e 10.06 segundos, para o problema com $n = 500$, e 324.65, 451.02 e 50.27 segundos, para o problema com $n = 1000$, respectivamente. Em média, os valores ótimos encontrados pelo Algoritmo 2.2.1 são um pouco melhores do que os obtidos pelo Algoritmo 2.3.1, e estes são ligeiramente melhores do que os obtidos pelo Algoritmo 2.4.1

	Método	$(-100/M) \times f(x^*)$	$(100/M) \times x^*$
$\beta = 1.0$	Algoritmo 2.2.1	100.31	$x_1 = x_2 = x_3 = x_4 = x_7 = 0$ $x_5 \approx 16.20$ $x_6 \approx 53.03$ $x_8 \approx 30.77$
	Algoritmo 2.3.1	100.31	$x_1 = x_2 = x_3 = x_4 = x_7 = 0$ $x_5 \approx 16.20$ $x_6 \approx 53.03$ $x_8 \approx 30.77$
	Algoritmo 2.4.1	100.31	$x_1 = x_2 = x_3 = x_4 = x_7 = 0$ $x_5 \approx 10.17$ $x_6 \approx 58.82$ $x_8 \approx 31.01$
$\beta = 0.6$	Algoritmo 2.2.1	100.30	$x_1 = x_2 = x_3 = x_4 = x_7 = 0$ $x_5 \approx 10.15$ $x_6 \approx 58.71$ $x_8 \approx 31.13$
	Algoritmo 2.3.1	100.29	$x_1 = x_2 = x_3 = x_4 = 0$ $x_5 \approx 1.55$ $x_6 \approx 64.36$ $x_7 \approx 1.58$ $x_8 \approx 32.51$
	Algoritmo 2.4.1	100.30	$x_1 = x_2 = x_3 = x_4 = x_7 = 0$ $x_5 \approx 10.15$ $x_6 \approx 58.71$ $x_8 \approx 31.13$

Tabela 3.2: Desempenho dos Algoritmos 2.2.1, 2.3.1 e 2.4.1 usando $\beta \in \{0.6, 1.0\}$, no caso do grande investidor, considerando carteiras com 10 dias de tempo de vida.

Método	$(-100/M) \times f(x^*)$	$(100/M) \times x^*$	
$\beta = 1.0$	Algoritmo 2.2.1	101.77	$x_2 = x_3 = x_4 = 0$ $x_1 \approx 1.02$ $x_5 \approx 6.04$ $x_6 \approx 18.53$ $x_7 \approx 1.31$ $x_8 \approx 73.11$
	Algoritmo 2.3.1	101.82	$x_2 = x_3 = x_4 = x_7 = 0$ $x_1 \approx 0.13$ $x_5 \approx 6.19$ $x_6 \approx 19.37$ $x_8 \approx 74.31$
	Algoritmo 2.4.1	101.83	$x_2 = x_3 = x_4 = x_7 = 0$ $x_1 \approx 0.19$ $x_5 \approx 6.17$ $x_6 \approx 19.41$ $x_8 \approx 74.23$
$\beta = 0.6$	Algoritmo 2.2.1	103.00	$x_1 = x_2 = x_3 = x_4 = x_7 = 0$ $x_5 \approx 2.63$ $x_6 \approx 17.23$ $x_8 \approx 80.14$
	Algoritmo 2.3.1	103.09	$x_3 = x_4 = x_5 = x_7 = 0$ $x_1 \approx 0.15$ $x_2 \approx 0.54$ $x_6 \approx 19.49$ $x_8 \approx 79.82$
	Algoritmo 2.4.1	103.09	$x_3 = x_4 = x_5 = x_7 = 0$ $x_1 \approx 0.15$ $x_2 \approx 0.55$ $x_6 \approx 19.49$ $x_8 \approx 79.81$

Tabela 3.3: Desempenho dos Algoritmos 2.2.1, 2.3.1 e 2.4.1 usando $\beta = 1.0$ no caso do pequeno investidor, considerando carteiras com 120 e 240 dias de tempo de vida.

3.4 Análise dos resultados computacionais

Os resultados numéricos parecem mostrar que existem pequenas diferenças de desempenho entre os algoritmos apresentados neste trabalho. No entanto, estes resultados não são definitivos e a experiência prática com esta abordagem ainda é incipiente. Os resultados teóricos são bastante enfáticos no sentido de mostrar que um melhor comportamento deve ser esperado dos algoritmos da Seção 2.2. Como mencionado anteriormente, os algoritmos efetivamente implementados aqui não garantem a minimização global dos problemas. A complexidade do problema torna muito improvável que existam métodos práticos que garantam a convergência para minimizadores globais, pelo menos com os atuais recursos computacionais e na presença de situações de grande escala. No entanto, pode-se esperar que em situações reais existam boas aproximações iniciais para minimizadores global, fornecidas por investidores treinados. Além disso, quando os algoritmos do tipo Algencan são utilizados, os procedimentos heurísticos de otimização global podem ser utilizados na resolução dos subproblemas para aumentar a probabilidade de convergência para minimizadores globais.

Outras vezes, os investidores estão dispostos a fazer pequenas mudanças em suas carteiras, mas não grandes mudanças que poderiam ser sugeridas por um algoritmo de minimização global. A vida real contém critérios subjetivos, que são difíceis de capturar em modelos. Esses critérios são normalmente implícitos na aproximação inicial. Assim, por um lado, a aproximação inicial motivada pela experiência dos investidores é, provavelmente, um bom palpite para a solução global de problema. Por outro lado, no caso em que minimizadores globais estão muito longe das suposições iniciais, é muito provável que alguma restrição implicitamente considerada pelo usuário não esteja contemplada no modelo matemático. Por esta razão, acreditamos que os algoritmos de minimização local podem ser bastante úteis.

Entre as melhorias de algoritmos que temos em mente, relatamos que no futuro possa ser interessante adaptar algumas técnicas de otimização global empregadas nos subproblemas de [28] para o problema introduzido neste trabalho.

3.4. Análise dos resultados computacionais

		Algoritmo 2.2.1										
		Nível de confiança $\alpha = r/q$										\bar{f}
		0.90	0.91	0.92	0.93	0.94	0.95	0.96	0.97	0.98	0.99	
% Perda $(100/M) \times T_{\text{loss}}$	10%	101.51	101.48	101.42	101.31	101.25	101.16	101.10	101.06	100.99	100.87	101.22
	9%	101.37	101.33	101.27	101.18	101.11	101.07	101.00	100.96	100.86	100.79	101.09
	8%	101.21	101.18	101.14	101.05	101.00	100.92	100.87	100.84	100.80	100.70	100.97
	7%	101.06	101.02	100.94	100.92	100.88	100.81	100.76	100.75	100.67	100.63	100.84
	6%	100.92	100.72	100.85	100.77	100.73	100.69	100.66	100.65	100.58	100.52	100.71
	5%	100.77	100.74	100.70	100.66	100.63	100.59	100.56	100.54	100.50	100.45	100.61
	4%	100.61	100.59	100.56	100.53	100.50	100.48	100.46	100.43	100.41	100.35	100.49
	3%	100.46	100.45	100.42	100.40	100.36	100.36	100.34	100.33	100.29	100.27	100.37
	2%	100.31	100.30	100.29	100.27	100.25	100.25	100.23	100.22	100.20	100.18	100.25
	1%	100.16	100.16	100.15	100.14	100.13	100.13	100.12	100.11	100.11	100.10	100.13
\bar{f}		100.84	100.80	100.77	100.72	100.68	100.65	100.61	100.59	100.54	100.49	100.67

		Algoritmo 2.3.1										
		Nível de confiança $\alpha = r/q$										\bar{f}
		0.90	0.91	0.92	0.93	0.94	0.95	0.96	0.97	0.98	0.99	
% Perda $(100/M) \times T_{\text{loss}}$	10%	101.51	101.48	101.42	101.31	101.25	101.19	101.11	101.06	100.98	100.89	101.22
	9%	101.37	101.33	101.27	101.18	101.11	101.08	101.00	100.96	100.84	100.80	101.09
	8%	101.22	101.18	101.12	100.91	100.97	100.87	100.89	100.82	100.77	100.70	100.95
	7%	101.03	101.04	100.99	100.90	100.83	100.78	100.78	100.73	100.69	100.59	100.84
	6%	100.91	100.89	100.85	100.76	100.74	100.69	100.67	100.62	100.58	100.54	100.73
	5%	100.77	100.72	100.71	100.64	100.61	100.60	100.55	100.53	100.50	100.45	100.61
	4%	100.55	100.55	100.51	100.51	100.49	100.49	100.46	100.42	100.39	100.36	100.47
	3%	100.44	100.43	100.40	100.38	100.38	100.35	100.34	100.32	100.30	100.27	100.36
	2%	100.31	100.31	100.28	100.26	100.25	100.23	100.23	100.22	100.21	100.18	100.25
	1%	100.15	100.14	100.14	100.13	100.13	100.12	100.11	100.10	100.10	100.10	100.12
\bar{f}		100.83	100.81	100.77	100.70	100.68	100.64	100.61	100.58	100.54	100.49	100.66

		Algoritmo 2.4.1										
		Nível de confiança $\alpha = r/q$										\bar{f}
		0.90	0.91	0.92	0.93	0.94	0.95	0.96	0.97	0.98	0.99	
% Perda $(100/M) \times T_{\text{loss}}$	10%	101.40	101.33	101.13	101.12	101.09	101.01	101.09	101.03	100.93	100.82	101.10
	9%	101.26	101.20	101.02	100.99	100.95	100.91	100.98	100.93	100.84	100.74	100.98
	8%	101.12	101.07	100.91	100.87	100.88	100.81	100.87	100.83	100.75	100.66	100.88
	7%	100.98	100.93	100.80	100.77	100.77	100.71	100.76	100.72	100.65	100.58	100.77
	6%	100.84	100.80	100.68	100.66	100.66	100.61	100.65	100.62	100.56	100.49	100.66
	5%	100.70	100.67	100.57	100.55	100.55	100.51	100.55	100.52	100.47	100.41	100.55
	4%	100.56	100.54	100.46	100.44	100.44	100.41	100.44	100.42	100.37	100.33	100.44
	3%	100.42	100.40	100.34	100.33	100.33	100.30	100.33	100.31	100.28	100.25	100.33
	2%	100.28	100.27	100.23	100.22	100.22	100.20	100.22	100.21	100.20	100.17	100.22
	1%	100.14	100.14	100.12	100.11	100.11	100.10	100.11	100.11	100.10	100.08	100.11
\bar{f}		100.77	100.74	100.63	100.61	100.60	100.56	100.60	100.57	100.52	100.45	100.60

Tabela 3.4: Desempenho dos Algoritmos 2.2.1, 2.3.1 e 2.4.1 em um problema com $n = 500$ ativos, variando o nível de confiança e a perda máxima. A restrição do tipo VaR impõe que a perda máxima deva ser menor do que $(100/M) \times T_{\text{loss}}\%$ do capital investido, com uma probabilidade $\alpha = r/q$. Cada entrada da tabela mostra o valor de $-(100/M) \times f(x^*)$. Colunas e linhas em negrito representam os valores médios.

Capítulo 3. Problema de seleção de carteiras com restrição de risco e experimentos numéricos

		Algoritmo 2.2.1										\bar{f}
		Nível de confiança $\alpha = r/q$										
		0.90	0.91	0.92	0.93	0.94	0.95	0.96	0.97	0.98	0.99	
% Perda $(100/M) \times T_{\text{loss}}$	10%	101.78	101.75	101.72	101.63	101.54	101.41	101.28	101.17	101.12	101.04	101.44
	9%	101.63	101.59	101.51	101.47	101.39	101.27	101.16	101.06	101.02	100.94	101.30
	8%	101.45	101.39	101.38	101.31	101.24	101.11	100.98	100.94	100.92	100.81	101.15
	7%	101.26	101.24	101.19	101.15	101.08	100.97	100.86	100.83	100.81	100.73	101.01
	6%	101.09	101.02	101.01	100.97	100.93	100.86	100.78	100.71	100.69	100.63	100.87
	5%	100.86	100.83	100.88	100.83	100.78	100.71	100.64	100.60	100.58	100.52	100.72
	4%	100.73	100.69	100.65	100.67	100.63	100.57	100.51	100.49	100.46	100.41	100.58
	3%	100.56	100.53	100.53	100.49	100.47	100.43	100.39	100.35	100.35	100.31	100.44
	2%	100.37	100.36	100.33	100.33	100.32	100.28	100.27	100.24	100.24	100.21	100.30
	1%	100.19	100.19	100.18	100.17	100.16	100.14	100.14	100.12	100.12	100.10	100.15
\bar{f}		100.99	100.96	100.94	100.90	100.85	100.78	100.70	100.65	100.63	100.57	100.80

		Algoritmo 2.3.1										\bar{f}
		Nível de confiança $\alpha = r/q$										
		0.90	0.91	0.92	0.93	0.94	0.95	0.96	0.97	0.98	0.99	
% Perda $(100/M) \times T_{\text{loss}}$	10%	101.80	101.74	101.72	101.63	101.55	101.41	101.29	101.19	101.08	101.04	101.45
	9%	101.66	101.59	101.53	101.47	101.35	101.27	101.11	101.07	101.02	100.93	101.30
	8%	101.45	101.40	101.33	101.30	101.24	101.11	100.97	100.94	100.89	100.84	101.15
	7%	101.28	101.26	101.14	101.14	101.07	100.99	100.85	100.82	100.77	100.72	101.00
	6%	101.09	101.02	100.98	100.98	100.88	100.85	100.73	100.71	100.69	100.63	100.86
	5%	100.90	100.87	100.85	100.82	100.76	100.71	100.60	100.59	100.57	100.51	100.72
	4%	100.74	100.67	100.67	100.58	100.63	100.57	100.53	100.46	100.46	100.41	100.57
	3%	100.55	100.53	100.49	100.46	100.45	100.43	100.40	100.36	100.34	100.31	100.43
	2%	100.36	100.35	100.35	100.33	100.31	100.30	100.26	100.24	100.23	100.20	100.29
	1%	100.20	100.18	100.18	100.17	100.16	100.15	100.13	100.12	100.12	100.10	100.15
\bar{f}		101.00	100.96	100.92	100.89	100.84	100.78	100.69	100.65	100.62	100.57	100.79

		Algoritmo 2.4.1										\bar{f}
		Nível de confiança $\alpha = r/q$										
		0.90	0.91	0.92	0.93	0.94	0.95	0.96	0.97	0.98	0.99	
% Perda $(100/M) \times T_{\text{loss}}$	10%	101.68	101.64	101.46	101.52	101.31	101.26	101.16	101.08	101.04	100.99	101.31
	9%	101.64	101.41	101.31	101.31	101.27	101.14	101.05	100.97	100.93	100.89	101.19
	8%	101.35	101.26	101.17	101.16	101.13	101.01	100.93	100.86	100.82	100.80	101.05
	7%	101.28	101.10	101.02	101.01	100.92	100.89	100.82	100.75	100.73	100.70	100.92
	6%	101.08	100.95	100.87	100.87	100.79	100.76	100.70	100.65	100.65	100.60	100.79
	5%	100.91	100.79	100.73	100.73	100.66	100.64	100.58	100.54	100.53	100.50	100.66
	4%	100.74	100.66	100.58	100.58	100.53	100.51	100.47	100.43	100.42	100.40	100.53
	3%	100.55	100.47	100.44	100.44	100.40	100.38	100.35	100.35	100.32	100.31	100.40
	2%	100.37	100.36	100.29	100.29	100.27	100.25	100.23	100.24	100.21	100.20	100.27
	1%	100.18	100.18	100.15	100.15	100.13	100.13	100.12	100.11	100.11	100.10	100.14
\bar{f}		100.98	100.88	100.80	100.81	100.74	100.70	100.64	100.60	100.58	100.55	100.73

Tabela 3.5: Desempenho dos Algoritmos 2.2.1, 2.3.1 e 2.4.1 em um problema com $n = 1000$ ativos, variando o nível de confiança e a perda máxima. A restrição do tipo VaR impõe que a perda máxima deva ser menor do que $(100/M) \times T_{\text{loss}}\%$ do capital investido, com uma probabilidade $\alpha = r/q$. Cada entrada da tabela mostra o valor de $-(100/M) \times f(x^*)$. Colunas e linhas em negrito representam os valores médios.

Capítulo 4

Introdução a métodos de Restauração Inexata

Nos próximos capítulos proporemos métodos de Restauração Inexata para a resolução de problemas de Programação Não Linear com restrições. Nosso objetivo neste capítulo é descrever as principais características deste tipo de método. Vamos fazer também comparações com as ideias centrais de outros métodos bem estabelecidos para tratar esse tipo de problema. Nosso objetivo aqui é levantar vantagens e desvantagens dos métodos de Restauração Inexata para deixar mais clara e intuitiva a apresentação dos métodos propostos nos capítulos seguintes.

O problema de Programação Não Linear padrão que consideraremos será

$$\begin{aligned} &\text{Minimizar } f(x) \\ &\quad x \in \mathbb{R}^n \\ &\text{sujeito a: } h(x) = 0, \\ &\quad \Omega \subset \mathbb{R}^n, \end{aligned} \tag{4.1}$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ são funções duas vezes continuamente diferenciáveis e o conjunto Ω é uma caixa n -dimensional limitada: $\Omega = \{x \in \mathbb{R}^n \mid \ell \leq x \leq u\}$. Em muitas situações poderíamos considerar Ω como sendo um polítopo ou, mais geralmente ainda, um conjunto compacto convexo associado a restrições fáceis de serem tratadas, para as

quais exista uma maneira eficiente de calcular a projeção de um vetor em Ω . Entretanto, por simplicidade, vamos nos limitar ao caso de caixas. Outra generalização que poderíamos considerar em muitos dos nossos resultados seria a inclusão de restrições de desigualdade, mas outra vez por simplicidade vamos considerar apenas igualdades. Na verdade, como estamos considerando restrições de caixa, toda restrição de desigualdade pode ser transformada em uma restrição de igualdade acrescentando uma variável de folga, com o único inconveniente de aumentarmos a dimensão do problema.

Inúmeros métodos computacionais foram introduzidos para resolver o problema (4.1). Nosso objetivo aqui é situar a filosofia principal de métodos de Restauração Inexata em relação aos métodos consagrados de Penalidade Externa [103], Lagrangianos Aumentados [95], métodos de direções no espaço reduzido [1] e Programação Quadrática Sequencial [99]. Estes métodos podem ter convergência global, no sentido de terem boas propriedades de convergência a partir de qualquer ponto inicial, ou local, no sentido das propriedades de convergência valerem apenas numa vizinhança de uma solução do problema. Geralmente a convergência local está associada a taxas de convergência quadrática ou superlinear. Estratégias de globalização também costumam ser propostas para os métodos com convergência local. As principais estratégias de globalização são o uso de uma função de mérito com busca linear [22], regiões de confiança [36] ou regularizações [44], e o uso de filtros [54]. Uma questão importante quando utilizamos uma estratégia de globalização é verificar se, pelo menos em situações ideais, perto de uma solução ela permite que o passo eficiente dado pelo método local seja aceito pelo algoritmo, evitando assim o chamado efeito Maratos [75].

A maior característica de métodos de Restauração Inexata é que eles tratam otimalidade e viabilidade em fases diferentes. Na primeira fase melhoramos a viabilidade mantendo uma deterioração limitada da otimalidade em relação ao ponto corrente. Na segunda nos preocupamos com a otimalidade sem deteriorar muito a viabilidade. Essa característica pode ser uma vantagem desse tipo de método quando o problema abordado apresenta uma estrutura particular em que um dos objetivos, viabilidade ou otimalidade, seja mais fácil de ser tratado do que o outro. Isso acontece porque não

perdemos tempo avaliando o objetivo difícil de ser tratado quando o foco é o objetivo fácil.

Métodos do tipo Penalidade Externa e Lagrangiano Aumentado baseiam-se em funções do tipo

$$\phi(x, \rho) = \rho f(x) + (1 - \rho)H(\|h(x)\|) \text{ e} \quad (4.2)$$

$$\phi(x, \lambda, \rho) = \rho(f(x) + \lambda^T h(x)) + (1 - \rho)H(\|h(x)\|), \quad (4.3)$$

respectivamente, onde $H : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ é uma função contínua tal que $H(0) = 0$ e $H(x) > 0$ se $x \neq 0$. Usualmente $H(x) = x$ ou $H(x) = x^2$; no primeiro caso, chamado de penalidade exata, as funções ϕ podem não ser diferenciáveis. Nesses métodos as funções ϕ são minimizadas em relação a x em cada iteração e são atualizados $\rho \in (0, 1)$ e λ . Por adotarem uma combinação convexa entre a função objetivo e a medida de inviabilidade, estes métodos trabalham com ambas ao mesmo tempo. Desse modo eles não aproveitam a estrutura de um problema que tenha ou função objetivo, ou restrições, fáceis. Os dois métodos tem convergência global. Para penalidade externa com $H(x) = x$ é possível provar, sob hipóteses bem fracas, que o parâmetro ρ fica afastado de zero. Já para o Lagrangiano Aumentado conseguimos este resultado inclusive utilizando a função ϕ suave correspondente ao uso de $H(x) = x^2$. Entretanto, hipóteses bem mais fortes são necessárias. A importância do parâmetro ρ não tender a zero é que assim garantimos um peso suficiente para a função objetivo, isso se reflete na estabilidade numérica na resolução dos subproblemas. Outra boa propriedade desses métodos é que podem ser utilizadas estratégias simples de *backtracking* em direções de busca para os subproblemas.

Métodos do tipo direção no espaço reduzido, por exemplo RG [100] e GRG [1], agem, como nos métodos do tipo penalidade, resolvendo uma sequência de problemas irrestritos. Entretanto, nos métodos de direção no espaço reduzido o espírito é completamente oposto a métodos do tipo penalidade. Enquanto nos métodos de penalidade não estamos preocupados em manter a viabilidade dos iterandos os métodos de direções no espaço reduzido tentam manter a viabilidade em todas as iterações. A maneira de fazer isso é usar as restrições para definir algumas variáveis implicitamente

em função das outras, pelo menos localmente, e trabalhar no espaço reduzido \mathbb{R}^{n-m} das variáveis livres. Este tipo de método utiliza fortemente o fato de termos restrições cujos gradientes são linearmente independentes para conseguirmos escrever algumas variáveis em função das outras, o que é uma de suas desvantagens. Outra grande dificuldade deste tipo de método é que, no caso de restrições não lineares, é praticamente impossível computacionalmente se manter no espaço reduzido, sendo assim é necessário restaurar a viabilidade dos iterandos, o que nem sempre é uma tarefa fácil. No caso de exigirmos a cada passo do método que a restrição $h(x) = 0$ seja sempre satisfeita chamá-riamos a restauração de exata. Infelizmente, não é possível garantir que seja computável um ponto exatamente viável a cada iteração. Sendo assim, os métodos do tipo GRG para restrições não lineares fazem na prática um passo de otimização e um passo de restauração inexata.

Os métodos de Restauração Inexata que vamos analisar nos próximos capítulos têm a vantagem de já ter a teoria desenvolvida para que a restauração não seja exata. Nesses métodos é estipulado qual o grau de exatidão que vamos exigir em cada iteração. Além disso, o passo de otimização é dado na linearização da superfície de nível das restrições no ponto restaurado, que é a aproximação mais simples do espaço reduzido referente ao conjunto de interesse, mas não é de fato usada a redução do espaço.

Métodos locais mais eficientes, que apresentam boas taxas de convergência, quadráticas ou superlineares, estão geralmente baseados em métodos do tipo Newton. Isso acontece pois os métodos de Newton são referenciais de excelência tanto para problemas de otimização irrestrita quanto para resolução de sistemas não lineares. O primeiro caso consiste em fazer modelos quadráticos da função objetivo, já o segundo em fazer uma linearização das equações. Para otimização com restrições outra ideia local semelhante muito eficiente seria resolver o sistema KKT pelo método de Newton. Inconvenientes desse tipo de abordagem é que o método não diferencia minimizadores de maximizadores e que acrescenta uma variável para cada restrição, aumentando a dimensão do problema. Talvez a alternativa que mais se aproxime dos métodos de

Newton para resolver o problema é a Programação Quadrática Sequencial (PQS), que pode ser globalizada utilizando uma função de mérito e estratégia de região de confiança.

Vamos agora descrever um pouco melhor esse tipo de método. Como sabemos, a curvatura das restrições de um problema de otimização é essencial para identificarmos se um ponto é um minimizador ou um maximizador. (Isso se reflete no fato de que um minimizador x^* de f no conjunto $h(x) = 0$ pode ser um maximizador de f no espaço tangente $J_h(x^*)(x - x^*) = 0$. Este tipo de situação não acontece se tivermos trabalhando com o Lagrangiano $L(x, \lambda) = f(x) + \lambda^T h(x)$, com o multiplicador de Lagrange exato da solução, no lugar de simplesmente a função objetivo $f(x)$.) Por esse motivo, minimizar o Lagrangiano é muito mais interessante quando estamos trabalhando com aproximações lineares das restrições. A grosso modo, o método de PQS consiste primeiramente em calcular uma direção de busca de acordo com um modelo quadrático do Lagrangiano numa aproximação linear das restrições dentro de uma região de confiança. E depois em dar um passo nessa direção, preferencialmente o passo 1, de modo a diminuir alguma função de mérito.

A direção de PQS a partir de um ponto x , com estimativa de multiplicador de Lagrange λ , ignorando a região de confiança, é definida como solução do problema:

$$\begin{aligned} &\text{Minimizar} && d^T H(x, \lambda) d + \nabla_x L(x, \lambda)^T d + L(x, \lambda) \\ &&& d \in \mathbb{R}^n \\ &\text{sujeito a:} && J_h(x) d = -h(x) \\ &&& x + d \in \Omega. \end{aligned} \tag{4.4}$$

onde $H(x, \lambda) \approx \nabla_{xx}^2 L(x, \lambda)$ é uma aproximação da Hessiana do Lagrangiano com respeito a x . Vamos chamar a direção solução desse problema de PQS Pura d_{pq} . Essa direção pode ser obtida também por uma iteração do método de Newton para anular o sistema KKT:

$$F(x, \lambda) = \begin{pmatrix} \nabla f(x) + J_h(x)^T \lambda \\ h(x) \end{pmatrix}.$$

Nesse caso teríamos que

$$\begin{pmatrix} \nabla_{xx}^2 L(x, \lambda) & J_h(x)^T \\ J_h(x) & 0 \end{pmatrix} \begin{pmatrix} d_{pq} \\ d_\lambda \end{pmatrix} = \begin{pmatrix} -\nabla f(x) - J_h(x)^T \lambda \\ -h(x) \end{pmatrix}, \quad (4.5)$$

que são exatamente as condições para anular o sistema KKT do problema (4.4), para $H(x, \lambda) = \nabla_x^2 L(x, \lambda)$.

É sabido que se d_{pq} for sempre aceito no método de Programação Quadrática Sequencial então o método tem, sob hipóteses adequadas, convergência local superlinear se supusermos boas estimativas para os multiplicadores de Lagrange e para a hessiana do Lagrangiano, e quadrática no caso em que, além das boas estimativas para os multiplicadores, tivermos $H(x, \lambda) = \nabla_{xx}^2 L(x, \lambda)$. Isso é consequência imediata dos resultados do método de Newton e a formulação (4.5). Como d_{pq} é obtida minimizando um modelo quadrático do Lagrangiano no espaço tangente das restrições é interessante que uma função de mérito envolvida na globalização esteja relacionada com $L(x, \lambda)$ para que haja uma maior possibilidade de que d_{pq} seja aceita. De fato, em [43] foi apresentado para o problema com apenas igualdades um critério para aceitar um novo iterando baseado no Lagrangiano Aumentado de Powell, Hestenes e Rockafellar [60, 91, 95]

$$\phi(x, \lambda, \rho) = f(x) + \lambda^T h(x) + \frac{\rho}{2} \|h(x)\|^2.$$

Também em [43] foi demonstrado que, sob hipóteses bem razoáveis, d_{pq} é aceita quando o ponto corrente está numa vizinhança da solução, e portanto este método não tem efeito Maratos.

Um inconveniente dos métodos PQS é a possibilidade dos subproblemas serem inviáveis. Uma observação é que eles também tratam simultaneamente otimalidade e viabilidade. Embora não sejam feitas avaliações de restrições na resolução do subproblema, o fato de tentar resolver os dois objetivos de uma só vez pode fazer com que o método gaste esforço desnecessário ao tentar resolver um objetivo difícil e outro fácil simultaneamente. Outra observação importante é que estamos utilizando um modelo do Lagrangiano no ponto x mas o espaço afim $J_h(x)d = -h(x)$, resultante da linearização

das restrições pode estar longe do ponto x . A Figura 4.1 ilustra essa situação. Se isso ocorrer o modelo quadrático pode não ser parecido com o Lagrangiano nos pontos de interesse e conseqüentemente a direção de PQS pode ser rejeitada pela função de mérito. Nesse caso, um novo problema de programação quadrática deve ser resolvido para encontrar uma nova direção. Visto isto é coerente pensar se não seria interessante fazer o modelo quadrático do Lagrangiano em algum dos pontos de interesse na linearização das restrições.

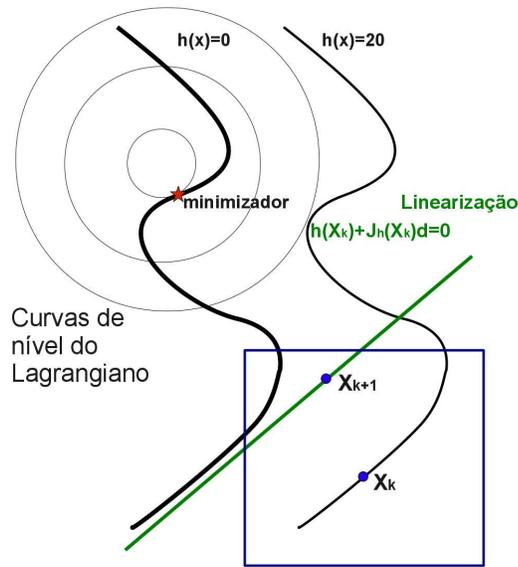


Figura 4.1: No método de PQS o modelo quadrático é construído com informações do ponto x_k o qual pode estar afastado da linearização. Além disso se a região de confiança for pequena ela pode não interseccionar a linearização.

Os métodos de Restauração Inexata (RI) tentam aproveitar todas as observações anteriores. Como já foi dito, por trabalhar em fases diferentes a otimalidade e a viabilidade, o método pode utilizar melhor a estrutura do problema. Vamos descrever agora, de forma geral, as ideias principais de como executar cada uma dessas fases. Na fase de restauração, dado um ponto x , primeiro encontramos um ponto mais viável y sem piorar muito a otimalidade. Geralmente esse passo é feito resolvendo inexatamente a linearização das restrições em x , possivelmente dando um passo do

tipo Newton. Depois disso, na fase de otimalidade, construímos um modelo quadrático do Lagrangiano em y . Computamos a direção de otimização minimizando o modelo quadrático no espaço tangente da superfície de nível das restrições em y . Por estarmos no espaço tangente $J_h(y)d = 0$ a deterioração da viabilidade é limitada, e além disso, os subproblemas sempre tem conjunto viável não vazio, pois $d = 0$ é factível. Após a obtenção de d testamos se uma função de mérito aceita o passo. Caso ele não seja aceito testamos outros pontos mais próximos de y_k até que algum deles seja aceito. A Figura 4.2 ilustra esse processo.

Observe que a obtenção da direção de otimalidade vem de um problema de programação quadrática e, como em PQS, no caso de não usarmos regiões de confiança, também vamos denotar a direção de otimalidade por d_{pq} . Observe também que caso os iterandos da fase de otimalidade da RI forem sempre factíveis não é necessário fazer a restauração e portanto os métodos de PQS e RI coincidem um com o outro.

Para Martínez e Pilota em [80] o subproblema é resolvido em uma região de confiança e com uma função de mérito envolvendo f e não o Lagrangiano como medida de otimalidade. Em [78] o subproblema proposto por Martínez ainda é resolvido em uma região de confiança mas a função de mérito envolve o Lagrangiano, o que melhora as chances do passo d_{pq} ser aceito. Outro método global, e de mais fácil compreensão, é proposto por Fischer e Friedlander em [53]. Nesse artigo é usada busca linear como estratégia de globalização, a função de penalidade exata $(\rho f(x) + (1 - \rho)\|h(x)\|)$ e é demonstrado que ρ fica afastado do zero. Em [59] Haeser propõe uma gama de direções Newtonianas, mais promissora que a direção do tipo gradiente sugerida em [53], que se enquadra na teoria de [53]. Um dos resultados deste trabalho, que será apresentado no Capítulo 5, é mostrar uma extensão do Algoritmo de Fischer-Friedlander na qual valem todas as propriedades demonstradas em [53]. Para isso vamos usar as direções propostas em [59] e considerar o Lagrangiano Afiado $(\rho(f(x) + \lambda^T h(x)) + (1 - \rho)\|h(x)\|)$ como função de mérito.

Em [29] Birgin e Martínez propõem um método local onde a restauração e a otimalidade tem apelo Newtoniano e provam convergência superlinear e quadrática,

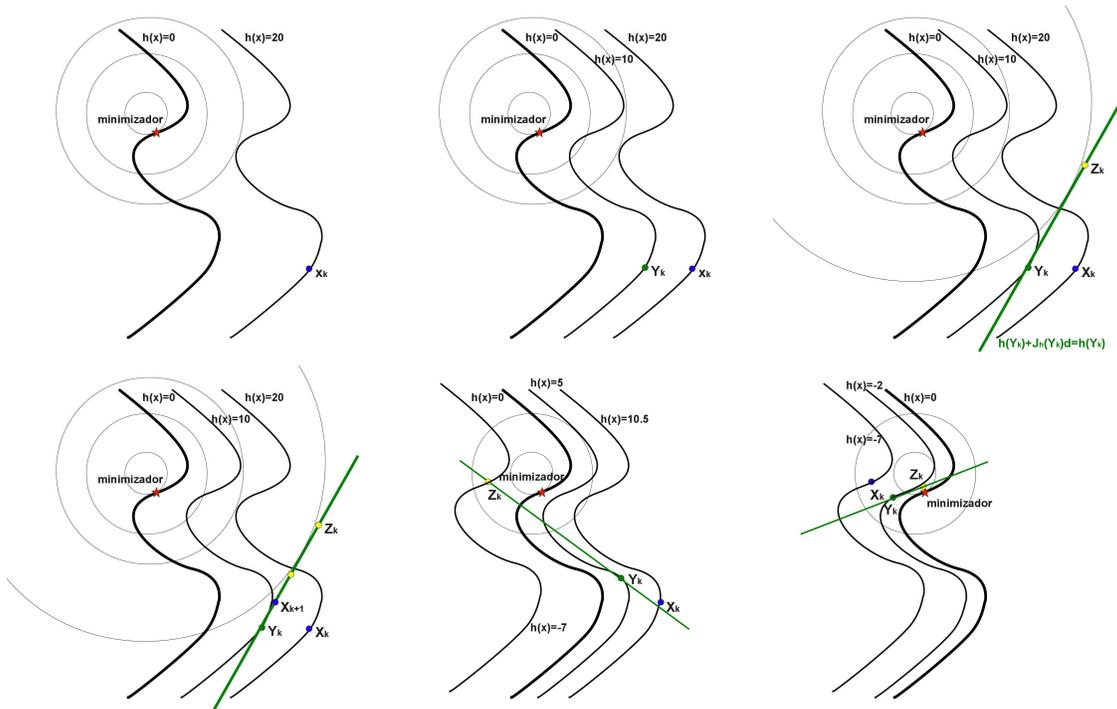


Figura 4.2: Sequência de passos realizados em iterações de Restauração Inexata. Primeiro um ponto mais viável y_k é obtido. Depois é construída a linearização da curva de nível das restrições nesse ponto. É construído também em y_k um modelo quadrático, o qual é minimizado inexatamente sujeito as restrições lineares encontrando z_k . É feito o teste se uma função de mérito aceita esse ponto. Caso ele não seja aceito testamos outros pontos mais próximos de y_k até que algum deles seja aceito. Uma vez que o ponto é aceito o processo continua fazendo mais iterações até que um critério de parada é satisfeito.

dependendo das escolhas feitas nos subproblemas. Também no Capítulo 5 vamos refazer os resultados de [29] medindo o resíduo de forma mais natural e fazendo com mais rigor a demonstração de convergência quadrática, deixando mais claros tais resultados. Ainda no Capítulo 5 vamos propor uma estratégia híbrida entre nossa modificação do Algoritmo global de Fischer-Friedlander e o Algoritmo local Birgin-Martínez, obtendo assim um método global com convergência quadrática e superlinear. No Capítulo 6 propomos um algoritmo parecido com o de Fischer-Friedlander mas com uma regularização

como estratégia de globalização e mostramos sua utilidade em problemas de otimização em que as derivadas da função objetivo não estejam disponíveis. No Capítulo 8 propomos um novo método de Restauração inexata para problemas de programação em dois níveis seguindo as ideias centrais de [10] e uma extensão dessa ideia para problemas com restrições de equilíbrio.

Capítulo 5

Um método híbrido de Restauração Inexata

O objetivo central deste capítulo é propor um método global de Restauração Inexata que tenha convergência local quadrática ou superlinear. Como vimos no Capítulo 4, estas taxas de convergência estão associadas com escolhas Newtonianas e com o fato da função de mérito não rejeitar esta escolha numa vizinhança da solução. Para o problema apenas com restrições de igualdade foi mostrado em [43] que, sob hipóteses adequadas, uma escolha particular de Lagrangiano Aumentado como função de mérito evita o efeito Maratos, ou seja, faz com que o passo Newtoniano seja sempre aceito numa vizinhança de uma solução. Para o caso de desigualdades um método de identificação das restrições ativas, tal como o proposto em [101], pode garantir o mesmo resultado. Nesses casos obtemos as boas taxas de convergência para esses métodos. Pensando nisso, estendemos os resultados do método global de Fischer e Friedlander em [53] substituindo, na função de mérito, a função objetivo $f(x)$ pelo Lagrangiano $L(x, \lambda)$.

Um algoritmo local de Restauração Inexata com convergência superlinear ou quadrática foi proposto por Birgin e Martínez em [29]. Neste capítulo vamos apresentar alguns dos resultados referentes ao Algoritmo Birgin-Martínez modificando ligeiramente os enunciados do artigo original. Estas modificações tem o propósito de deixar mais

natural a maneira de medir o resíduo referente aos iterandos do algoritmo. Dessa forma as demonstrações ficam mais claras, e nos possibilitam demonstrar com maior rigor a convergência quadrática. Além disso vamos mostrar que em casos especiais o passo 1 do algoritmo de Fischer-Friedlander satisfaz as condições do algoritmo de Birgin-Martínez.

Mesmo considerando o Lagrangiano na função de mérito do Algoritmo global Fischer-Friedlander não foi possível demonstrar que nossa modificação garante que estamos livres do Efeito Maratos. Dessa forma construímos uma estratégia híbrida global-local combinando os Algoritmos Fischer-Friedlander e Birgin-Martínez. Nossa proposta é uma estratégia *watchdog* em que sempre que possível tentamos o método local mas salvaguardamos com o método global de forma a obtermos um método global com convergência quadrática ou superlinear.

O problema alvo deste capítulo continua a ser o problema com restrições de igualdade e caixa (4.1).

5.1 Algoritmo Fischer-Friedlander

A seguir vamos apresentar nossa extensão do Algoritmo Fischer-Friedlander em que será usado o Lagrangiano Afiado como função de mérito. Vamos utilizar também a proposta de direção de otimização exibida por Haeser em [59] que se encaixa na teoria do algoritmo original e que é mais promissora do que a sugerida em [53]. Vamos demonstrar a boa definição do algoritmo e provar suas propriedades de convergência. Para nossa análise vamos definir o Lagrangiano Afiado $\Phi(x, \lambda, \rho) = \rho L(x, \lambda) + (1 - \rho)\|h(x)\|$.

Nosso algoritmo global modificado, está descrito a seguir.

Algoritmo 5.1

Sejam $r \in [0, 1)$, $\beta, \gamma > 0$ constantes algorítmicas e $\Omega_\lambda \subset \mathbb{R}^m$ um compacto.

Passo 0: Escolher $x_0 \in \Omega$, $\lambda_0 \in \Omega_\lambda$, $\rho_0 \in (0, 1)$ e fazer $k = 0$.

Passo 1: Escolher $y_k \in \Omega$ e $\lambda_{k+1} \in \Omega_\lambda$ tais que

$$\|h(y_k)\| \leq r\|h(x_k)\|, \text{ ou seja, } \|h(y_k)\| - \|h(x_k)\| \leq -(1-r)\|h(x_k)\|,$$

$$L(y_k, \lambda_{k+1}) \leq L(x_k, \lambda_k) + \beta\|h(x_k)\|.$$

Passo 2: Escolher $\rho_{k+1} \in \{2^{-i}\rho_k : i \in \mathbb{N}\}$ o maior possível tal que

$$\Phi(y_k, \lambda_{k+1}, \rho_{k+1}) - \Phi(x_k, \lambda_k, \rho_{k+1}) \leq \frac{1}{2}(1-r)(\|h(y_k)\| - \|h(x_k)\|).$$

Passo 3: Computar a direção de busca para otimalidade d_k como solução do modelo quadrático:

$$\begin{aligned} \text{Minimizar } & d^T H(y_k, \lambda_{k+1})d + \nabla_x L(y_k, \lambda_{k+1})^T d + L(y_k, \lambda_{k+1}) \\ & d \in \mathbb{R}^n \\ \text{sujeito a: } & J_h(y_k)d = 0 \\ & y_k + d \in \Omega. \end{aligned} \tag{5.1}$$

Passo 4: Determinar $t_k \in \{2^{-i} : i \in \mathbb{N}\}$ o maior possível tal que

$$L(y_k + t_k d_k, \lambda_{k+1}) \leq L(y_k, \lambda_{k+1}) - \gamma t_k \|d_k\|^2 \tag{5.2}$$

$$\Phi(y_k + t_k d_k, \lambda_{k+1}, \rho_{k+1}) - \Phi(x_k, \lambda_k, \rho_{k+1}) \leq \frac{1}{2}(1-r)(\|h(y_k)\| - \|h(x_k)\|) \tag{5.3}$$

Passo 5: Fazer $x_{k+1} = y_k + t_k d_k$, $k = k + 1$ e voltar ao passo 1.

Além do uso da direção de otimização sugerida em [59], do Lagrangiano Afiado como função de mérito e dos multiplicadores de Lagrange desde a fase de restauração, o Algoritmo 5.1 difere do original apresentado em [53] pois exige o teste da condição 5.2. Esse teste é necessário para garantir a convergência do método e já foi incluído por Fischer e Friedlander nas novas publicações sobre o algoritmo.

Vamos agora provar que o algoritmo global está bem definido. Para demonstrar que sempre é possível executar o passo 1 seguimos a demonstração em [53]. Primeiramente vamos precisar de um resultado que limita a distância de um ponto ao

conjunto viável pela norma das restrições h nesse ponto. Em [53] é usado um lema do artigo de Robinson [94] para assegurar esse resultado. As hipóteses deste lema são que Ω é uma caixa limitada, isto é $\Omega = \{x \in \mathbb{R}^n | l \leq x \leq u\}$, com $-\infty < l_i < u_i < \infty$, o que é assumido na definição do problema (4.1), e que vale MFCQ para todo

$$x \in V \equiv \Omega \cap \{x | h(x) = 0\}.$$

Lema 5.1.1. *Suponhamos que valha MFCQ para todo ponto $z \in V$ então existem $\delta(z) > 0$ e $\xi_0 > 0$, independente de z , tais que $\text{dist}(x, V) \leq \xi_0 \|h(x)\|$ para todo $x \in \Omega \cap B(z, \delta(z))$.*

Demonstração: Ver Corolário 1 de [94]. □

Acreditamos que um resultado mais geral, que utiliza a condição de qualificação mais fraca CRSC, possa ser mostrado. No Teorema 5.4 de [14] os autores provam um resultado muito semelhante ao Lema 5.1.1. Entretanto nós não conseguimos ver facilmente que é possível tirar a dependência de z em ξ no caso em que Ω é compacto.

O próximo lema assegura que o limitante do erro valha em todo Ω .

Lema 5.1.2. *Suponhamos que valha MFCQ para todo ponto $z \in V$, então existe $\xi > 0$, independente de z e x , tal que $\text{dist}(x, V) \leq \xi \|h(x)\|$ para todo $x \in \Omega$.*

Demonstração: Vamos considerar nessa demonstração que $\delta(z)$ e ξ_0 são dados pelo Lema 5.1.1. Pelo fato de Ω ser compacto e h ser contínua temos que o conjunto viável V é compacto. Como $\cup_{z \in V} B(z, \delta(z))$, é uma cobertura de V , a compacidade garante que existe uma subcobertura finita, ou seja

$$V \subset A \equiv \cup_{i=1, \dots, N_0} B(z_i, \delta(z_i))$$

para algum $N_0 \in \mathbb{N}$. Como a fronteira de A , que será denotada por $fr(A)$, é um conjunto fechado e V é compacto, existem pontos que realizam a menor distância entre $fr(A)$ e V , ou seja, existem $x_0 \in fr(A)$ e $z_0 \in V$ tais que $\|x_0 - z_0\| \leq \|x - z\|$ para todo $x \in fr(A)$ e $z \in V$. Como A é um conjunto aberto, temos que cada ponto de

V é interior de A , e portanto não está na fronteira de A . Portanto $\|x_0 - z_0\| > 0$. Tomando $\delta = \frac{\|x_0 - z_0\|}{2}$ temos que $V + \delta B \subset A$, onde B é a bola aberta unitária, e portanto $dist(x, V) \leq \xi_0 \|h(x)\|$ para todo $x \in \Omega \cap (V + \delta B)$.

Dessa forma temos que, se $\Omega \subset V + \delta B$, então $dist(x, V) \leq \xi_0 \|h(x)\|$ para todo $x \in \Omega$. Caso contrário, $\hat{\Omega} = \Omega \setminus (V + \delta B)$ é não vazio e compacto e portanto existem $c_1 \equiv \sup_{x \in \hat{\Omega}} dist(x, V) < \infty$ e $c_2 \equiv \inf_{x \in \hat{\Omega}} \|h(x)\| > 0$. Dessa forma temos que

$$dist(x, V) \leq c_1 = \frac{c_1}{c_2} c_2 \leq \frac{c_1}{c_2} \|h(x)\|, \quad \text{para todo } x \in \hat{\Omega}.$$

Portanto, se tomarmos $\xi = \xi_0 + \frac{c_1}{c_2}$ temos que

$$dist(x, V) \leq \xi \|h(x)\|, \quad \text{para todo } x \in \Omega.$$

□

Vamos demonstrar agora que o passo 1 está bem definido sob as hipóteses do Lema 5.1.2.

Teorema 5.1.1. *Suponhamos que V seja não vazio e que valha MFCQ para todos pontos de V , então o passo 1 do Algoritmo 5.1 está bem definido para qualquer que seja $x_k \in \Omega$.*

Demonstração: Se para cada $x \in \Omega$ definirmos $d_x : V \rightarrow \mathbb{R}$ como a função que mede a distância de cada ponto de V ao ponto x fixo, isto é

$$d_x(y) = \|x - y\|,$$

teríamos para cada $x \in \Omega$ uma função contínua definida num compacto. Dessa forma, o teorema de Weierstrass garante que existe $y(x) \in V$, não necessariamente único, tal que

$$\|x - y(x)\| = \min_{y \in V} d_x(y) = \min_{y \in V} \|x - y\| = dist(x, V) \leq \xi \|h(x)\|.$$

Considerando L_λ um limitante para $\|\lambda\|$ em Ω_λ , L_f a constante de Lipschitz de f em Ω , e $\beta = L_f \xi + (1 + r)L_\lambda$ vamos mostrar que tomando $y_k = y(x_k)$ então, para qualquer $\lambda_{k+1} \in \Omega_\lambda$, o par (y_k, λ_{k+1}) é aceitável pelo passo 1.

De fato, como y_k é viável, temos que

$$\|h(y_k)\| = 0 \leq r\|h(x_k)\|,$$

qualquer que seja $r > 0$. Deste modo temos que:

$$\begin{aligned} L(y_k, \lambda_{k+1}) - L(x_k, \lambda_k) &= (f(y_k) + \lambda_{k+1}^T h(y_k)) - (f(x_k) + \lambda_k^T h(x_k)) \\ &\leq |f(y_k) - f(x_k)| + \|\lambda_{k+1}\| \|h(y_{k+1})\| + \|\lambda_k\| \|h(x_k)\| \\ &\leq L_f \|y_k - x_k\| + L_\lambda r \|h(x_k)\| + L_\lambda \|h(x_k)\| \\ &\leq L_f \xi \|h(x_k)\| + (1 + r) L_\lambda \|h(x_k)\| \\ &\leq (L_f \xi + (1 + r) L_\lambda) \|h(x_k)\| \\ &\leq \beta \|h(x_k)\|. \end{aligned}$$

□

Algumas observações importantes serão ressaltadas a seguir. A primeira é que da segunda pra terceira desigualdade da demonstração da deterioração limitada do Lagrangiano poderíamos utilizar que $h(y_k) = 0$ em vez de $\|h(y_k)\| \leq r\|h(x_k)\|$. Optamos pela demonstração acima pois ela mostra que na verdade qualquer ponto y_k que satisfaça $\|h(y_k)\| \leq r\|h(x_k)\|$ e $\|y_k - x_k\| \leq \xi\|h(x_k)\|$ irá satisfazer a condição de deterioração limitada do Lagrangiano. Dessa forma é possível encontrar o ponto restaurado construindo uma sequência convergente a $y(x_k)$ pois ela irá satisfazer essas condições (com outra constante $\xi' = 2\xi$) a partir de certo instante. Além disso, se restaurarmos desse modo, a condição

$$L(y_k, \lambda_{k+1}) - L(x_k, \lambda_k) \leq \beta \|h(x_k)\|$$

não precisa ser testada no algoritmo, já que ela é automaticamente satisfeita para algum β suficientemente grande e y_k próximo de $y(x_k)$.

Em Martínez [78], outro ponto que satisfaz as condições do passo 1 é apresentado. As hipóteses para obtermos tal ponto são mais fortes do que as apresentadas

aqui. A condição de qualificação exigida é a independência linear, e não MFCQ, note que a presença da caixa faz com que estas condições não sejam equivalentes. Além disso, o teorema apresentado em [78] é local, no sentido de que só vale para pontos em que as restrições já estavam suficientemente próximas de se anular. Por outro lado, intuitivamente não devemos ser muito exigentes em manter a otimalidade quando estamos muito afastados da viabilidade, portanto este não é um fato que preocupa na restauração feita em [78] e é de se esperar que consigamos restaurar em qualquer ponto. Outro ponto positivo da restauração proposta em [78] é o apelo Newtoniano que ela apresenta. Isso tem um papel importante em demonstrações relativas à taxa de convergência do método híbrido quando combinamos o Algoritmo 5.1 com o método local de Birgin-Martínez. A seguir vamos enunciar o Lema 6.1 de [78], onde são apresentados tais fatos.

Lema 5.1.3. Lema 6.1 de [78]

Para cada $x \in \Omega$ defina:

$$T(x) = \{y \in \Omega; J_h(x)(y - x) + h(x) = 0\}.$$

Se $T(x) \neq \emptyset$ denotamos $y(x) = \operatorname{argmin}\{\|y - x\|; y \in T(x)\}$. Seja $r \in (0, 1)$, e suponhamos que todos os pontos de Ω sejam regulares. Então existem $\epsilon > 0, \xi > 0$ tais que, se $\|h(x)\| \leq \epsilon$ valem:

$$\begin{aligned} T(x) &\neq \emptyset \\ \|h(y(x))\| &\leq r\|h(x)\| \\ \|y(x) - x\| &\leq \xi\|h(x)\|. \end{aligned}$$

Demonstração: Ver Lema 6.1 de [78] □

Sendo assim, o ponto $y(x)$ cumpre as condições suficientes para satisfazer a condição de deterioração limitada do Lagrangiano, conforme a observação após a demonstração do Teorema 5.1.1. Mais uma vez temos que a condição de deterioração

limitada do Lagrangiano não precisa ser testada nas implementações práticas do algoritmo.

Para mostrar que o passo 2 está bem definido vamos mostrar que se $\rho \leq \frac{1-r}{2(\beta+1-r)}$ então ele será aceito.

Demonstração: De fato, neste caso teríamos que $\rho(\beta + 1 - r) - (1 - r) \leq -\frac{1-r}{2}$, e portanto, já utilizando que o passo 1 foi feito com sucesso, temos

$$\begin{aligned} \Phi(y_k, \lambda_{k+1}, \rho) - \Phi(x_k, \lambda_k, \rho) &= \rho (L(y_k, \lambda_{k+1}) - L(x_k, \lambda_k)) + \\ &\quad + (1 - \rho) (\|h(y_k)\| - \|h(x_k)\|) \\ &\leq \rho\beta \|h(x_k)\| - (1 - \rho)(1 - r)\|h(x_k)\| \\ &\leq (\rho(\beta + 1 - r) - (1 - r))\|h(x_k)\| \\ &\leq -\frac{1-r}{2}\|h(x_k)\| \\ &\leq \frac{1-r}{2}(\|h(y_k)\| - \|h(x_k)\|). \end{aligned}$$

Deste modo a sequência dos parâmetros de penalidade está afastada do zero pois $\rho_k \geq \min \left\{ \frac{1-r}{4(\beta+1-r)}, \rho_0 \right\}$, e conseqüentemente, como ρ_k é não crescente, concluímos que $\rho_k = \bar{\rho}$ é constante para k suficientemente grande. \square

O passo 3 está bem definido se $H(y_k, \lambda_{k+1})$ é positiva definida, pois neste caso o problema (5.1) consiste em minimizar um modelo quadrático estritamente convexo num hiperplano (observe que $d = 0$ é viável para (5.1)), o que se trata de um problema bem estabelecido na literatura.

Antes de demonstrar que o passo 4 está bem definido vamos provar algumas propriedades da direção de busca. Estas propriedades, bem como os lemas necessários para suas demonstrações, foram provadas por Haeser em [59] para o algoritmo original.

Lema 5.1.4. Lema 3.32 de [59]

Se $f : \mathbb{R} \rightarrow \mathbb{R}$ é continuamente diferenciável e $t \in \mathbb{R}$ então:

$$t \int_0^1 (f'(tx) - f'(0))dx = f(t) - f(0) - tf'(0)$$

Demonstração: Para $t = 0$ o resultado é evidente. Para $t \neq 0$ temos que

$$\begin{aligned} t \int_0^1 (f'(tx) - f'(0))dx &= t \int_0^1 f'(tx)dx - tf'(0) \int_0^1 1dx \\ &= t \left(\frac{f(t \cdot 1) - f(t \cdot 0)}{t} \right) - tf'(0) \\ &= f(t) - f(0) - tf'(0) \end{aligned}$$

□

Para $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ podemos aplicar o Lema em $f(t) = F_i(y + td)$ de forma a obtermos que:

$$F(y + td) = F(y) + tJ_F(y)d + t \int_0^1 (J_F(y + txd) - J_F(y)) d dx. \quad (5.4)$$

O próximo resultado que apresentaremos é o Lema 3.33 de [59], o qual demonstraremos novamente aqui apenas por conveniência e para acrescentar os detalhes de estarmos trabalhando com o Lagrangiano e não com a função objetivo e o fato de que a equação (5.5) vale para todo $t > 0$.

Lema 5.1.5. Lema 3.33 de [59]

Suponha que os gradientes de f e h são Lipschitz contínuos em Ω e que as matrizes $H(y_k, \lambda_{k+1})$ são escolhidas de forma que sejam uniformemente definidas positivas. Então existem constantes positivas γ , $\bar{\gamma}$ e τ tais que

$$\|h(y_k + td_k)\| \leq \|h(y_k)\| + \bar{\gamma}t^2\|d_k\|^2, \quad (5.5)$$

para todo $t \geq 0$, e

$$L(y_k + td_k, \lambda_{k+1}) \leq L(y_k, \lambda_{k+1}) - \gamma t\|d_k\|^2 \quad (5.6)$$

para todo $t \in [0, \tau]$.

Demonstração: Aplicando (5.4) com $F(y + td) = h(y_k + td_k)$ temos que:

$$h(y_k + td_k) = h(y_k) + tJ_h(y_k)d_k + t \int_0^1 (J_h(y_k + txd_k) - J_h(y_k)) d_k dx.$$

Pela viabilidade de d_k para (5.1) temos que $J_h(y_k)d_k = 0$ e portanto

$$\begin{aligned}
 h(y_k + td_k) &\leq h(y_k) + t \left| \int_0^1 (J_h(y_k + txd_k) - J_h(y_k)) d_k dx \right| \\
 &\leq h(y_k) + t \int_0^1 \|J_h(y_k + txd_k) - J_h(y_k)\| \|d_k\| dx \\
 &\leq h(y_k) + t \int_0^1 L_{J_h} \|y_k + txd_k - y_k\| \|d_k\| dx \\
 &\leq h(y_k) + L_{J_h} t^2 \|d_k\|^2 \int_0^1 |x| dx \\
 &\leq h(y_k) + \frac{L_{J_h}}{2} t^2 \|d_k\|^2.
 \end{aligned}$$

Denotando $\bar{\gamma} = \frac{L_{J_h}}{2}$ obtemos (5.5).

Para demonstrar (5.6) observe que do fato de d_k ser solução do subproblema (5.1) e $d = 0$ ser viável para (5.1) temos que:

$$L(y_k, \lambda_{k+1}) \geq d_k^T H(y_k, \lambda_{k+1}) d_k + \nabla_x L(y_k, \lambda_{k+1})^T d_k + L(y_k, \lambda_{k+1}),$$

portanto,

$$\nabla_x L(y_k, \lambda_{k+1})^T d_k \leq -d_k^T H(y_k, \lambda_{k+1}) d_k.$$

Pelo fato das matrizes $H(y_k, \lambda_{k+1})$ serem uniformemente definidas positivas temos que $d_k^T H(y_k, \lambda_{k+1}) d_k \geq \sigma_1 \|d_k\|^2$, onde $\sigma_1 > 0$ é um limitante inferior para o menor autovalor de $H(y_k, \lambda_{k+1})$ para todo k . Desse modo

$$\nabla_x L(y_k, \lambda_{k+1})^T d_k \leq -\sigma_1 \|d_k\|^2.$$

Observe que pela viabilidade do subproblema (5.1) o resultado acima é equivalente a

$$\nabla_x f(y_k)^T d_k \leq -\sigma_1 \|d_k\|^2.$$

Aplicando (5.4) com $F(y + td) = L(y_k + td_k, \lambda_{k+1})$ temos:

$$\begin{aligned}
 L(y_k + td_k, \lambda_{k+1}) &= L(y_k, \lambda_{k+1}) + t \nabla L(y_k, \lambda_{k+1})^T d_k + \\
 &\quad + t \int_0^1 (\nabla L(y_k + tx d_k, \lambda_{k+1}) - \nabla L(y_k, \lambda_{k+1}))^T d_k dx \\
 &\leq L(y_k, \lambda_{k+1}) - \sigma_1 t \|d_k\|^2 + \\
 &\quad + t \int_0^1 \left| (\nabla L(y_k + tx d_k, \lambda_{k+1}) - \nabla L(y_k, \lambda_{k+1}))^T d_k \right| dx \\
 &\leq L(y_k, \lambda_{k+1}) - \sigma_1 t \|d_k\|^2 + \\
 &\quad + t \int_0^1 \|(\nabla L(y_k + tx d_k, \lambda_{k+1}) - \nabla L(y_k, \lambda_{k+1}))\| \|d_k\| dx \\
 &\leq L(y_k, \lambda_{k+1}) - \sigma_1 t \|d_k\|^2 + t \int_0^1 L_{\nabla L} \|y_k + tx d_k - y_k\| \|d_k\| dx \\
 &\leq L(y_k, \lambda_{k+1}) - \sigma_1 t \|d_k\|^2 + L_{\nabla L} t^2 \|d_k\|^2 \int_0^1 x dx \\
 &\leq L(y_k, \lambda_{k+1}) - \left(\sigma_1 - \frac{L_{\nabla L} t}{2} \right) t \|d_k\|^2.
 \end{aligned}$$

Logo, para $\gamma = \frac{\sigma_1}{2}$ e $\tau = \frac{\sigma_1}{2L_{\nabla L}}$ temos que vale (5.6). □

De posse dessas propriedades, vamos mostrar agora que o passo 4 está bem definido sob as hipóteses das mesmas.

Demonstração: A condição (5.6) já garante que (5.2), a primeira condição do passo 4, está bem definida, pois qualquer $t \leq \tau$ satisfaz essa condição. Para (5.3), a segunda condição, vamos mostrar que se $t \leq \min \left\{ \frac{\bar{\rho}\gamma}{\bar{\gamma}(1-\bar{\rho})}, \tau \right\}$ então ele será aceito. De fato, nessas condições temos que $(1 - \bar{\rho})\bar{\gamma}t - \bar{\rho}\gamma \leq 0$ e que valem as condições (5.5) e (5.6). Sendo assim, como no passo 2 exigimos que $\Phi(y_k, \lambda_{k+1}, \rho_{k+1}) - \Phi(x_k, \lambda_k, \rho_{k+1}) \leq \frac{1}{2}(1 - r)(\|h(y_k)\| - \|h(x_k)\|)$, e provamos que $\rho_k \geq \bar{\rho}$, para todo k , temos que

$$\begin{aligned}
\Phi(y_k + td_k, \lambda_{k+1}, \rho_{k+1}) - \Phi(x_k, \lambda_k, \rho_{k+1}) &= \Phi(y_k + td_k, \lambda_{k+1}, \rho_{k+1}) - \Phi(y_k, \lambda_{k+1}, \rho_{k+1}) + \\
&\quad + \Phi(y_k, \lambda_{k+1}, \rho_{k+1}) - \Phi(x_k, \lambda_k, \rho_{k+1}) \\
&\leq \rho_{k+1} (L(y_k + td_k, \lambda_{k+1}) - L(y_k, \lambda_{k+1})) + \\
&\quad + (1 - \rho_{k+1}) (\|h(y_k + td_k)\| - \|h(y_k)\|) + \\
&\quad + \frac{1-r}{2} (\|h(y_k)\| - \|h(x_k)\|) \\
&\leq -\rho_{k+1}\gamma t \|d_k\|^2 + (1 - \rho_{k+1})\bar{\gamma}t^2 \|d_k\|^2 + \\
&\quad + \frac{1-r}{2} (\|h(y_k)\| - \|h(x_k)\|) \\
&\leq -\bar{\rho}\gamma t \|d_k\|^2 + (1 - \bar{\rho})\bar{\gamma}t^2 \|d_k\|^2 + \\
&\quad + \frac{1-r}{2} (\|h(y_k)\| - \|h(x_k)\|) \\
&\leq ((1 - \bar{\rho})\bar{\gamma}t - \bar{\rho}\gamma)t \|d_k\|^2 + \\
&\quad + \frac{1-r}{2} (\|h(y_k)\| - \|h(x_k)\|) \\
&\leq \frac{1-r}{2} (\|h(y_k)\| - \|h(x_k)\|).
\end{aligned}$$

Sendo assim, o passo 4 está bem definido, e mais ainda, a sequência $\{t_k\}$ está afastada do zero pois $t_k \geq \bar{t} \equiv \min \left\{ \frac{\bar{\rho}\gamma}{2\bar{\gamma}(1-\bar{\rho})}, \frac{\tau}{2}, 1 \right\}$. \square

Note que para testar esta segunda equação do passo 4 não há necessidade de conhecermos nem γ nem $\bar{\gamma}$. Basta a existência destas constantes para provarmos a possibilidade da condição desejada ser satisfeita para t suficientemente pequeno. Entretanto, no passo 4 devemos testar também se a equação (5.2) é satisfeita. Nesse caso temos um problema pois devemos conhecer a constante algorítmica γ , a qual pode depender do limitante para o menor autovalor das aproximações das Hessianas utilizadas. Caso tenhamos uma boa estimativa para σ_1 é possível encontrar γ de forma que a condição (5.6) valha pois basta tomar qualquer valor tal que $\gamma < \sigma_1$, e $\tau = \frac{\sigma_1 - \gamma}{L_{\nabla L}}$. Observe que também não é necessário conhecer τ , basta sua existência para garantir

que t suficientemente pequeno será aceito pelo passo 4.

Nada impede que a condição (5.6) valha para algum τ mesmo que $\gamma > \sigma_1$. Dessa forma poderíamos tentar ser mais exigentes com o decréscimo suficiente escolhendo $\gamma \geq \frac{\sigma_1}{2}$, mesmo se tivéssemos uma boa estimativa para σ_1 . Nesse caso, e no de não termos uma boa estimativa para $\sqrt{\sigma_1}$, uma alternativa seria usar uma estratégia que chamaremos aqui de busca dupla. Isso se faz necessário pois não teríamos certeza que existiria τ de forma que a condição (5.6) valha para o valor de γ escolhido. O algoritmo usando essa estratégia segue como o algoritmo global com pequenas modificações na inicialização e o passo 4 substituído pela estratégia de busca dupla. No passo 0 são acrescentados os seguintes procedimentos: “Escolha γ_0 inicial positivo qualquer, considere uma sequência decrescente tendendo a zero $\{s_j\}$ e faça $j = 0$ ”. Já o passo 4 passa a ser:

a) Considere as relações:

$$t > s_j, \quad (5.7)$$

$$L(y_k + td_k, \lambda_{k+1}) \leq L(y_k, \lambda_{k+1}) - \gamma_k t \|d_k\|^2 \quad \text{e} \quad (5.8)$$

$$\Phi(y_k + t_k d_k, \lambda_{k+1}, \rho_{k+1}) - \Phi(x_k, \lambda_k, \rho_{k+1}) \leq \frac{1}{2}(1-r)(\|h(y_k)\| - \|h(x_k)\|) \quad (5.9)$$

b) Faça $t = 1$

c) Se valem (5.7), (5.8) e (5.9):

Faça $t_k = t$, $\gamma_{k+1} = \gamma_k$ e vá ao passo 5.

d) Se vale (5.7) mas não valem (5.8) e/ou (5.9)

Faça $t = \frac{t}{2}$ e volte ao passo 4.c).

e) Se não vale (5.7) faça $j = j + 1$, $\gamma_k = \frac{\gamma_k}{2}$ e volte ao passo 4.b).

Provemos agora que esse novo passo 4 também está bem definido e que a partir de certo instante γ_k se mantém constante e portanto desse momento em diante

o passo 4 alternativo é igual ao passo 4 original.

Demonstração: Vamos mostrar que o passo 4.e) é realizado um número finito de vezes, portanto γ_k ficará constante a partir de certo momento. Nesse caso a busca linear será bem sucedida de algum momento em diante, pois caso contrário t_k seria pequeno suficiente e não valeria (5.7).

Primeiramente note que se t satisfaz (5.8) para $\gamma_k = \hat{\gamma}$ então t satisfaz (5.8) para $\gamma_k \leq \hat{\gamma}$, pois estamos exigindo um decréscimo menor neste caso. Suponhamos agora que k_0 seja tal que $\gamma_{k_0} < \sigma_1$.

Pela demonstração do Lema 5.1.5 sabemos que se $\hat{t} = \frac{\sigma_1}{2L_{\nabla L}}$ então todo $t \leq \hat{t}$ satisfaz a (5.8) para todo $k \geq k_0$. Mais ainda, denotando $\tau = \min\{\frac{\rho\sigma_1}{L_{J_h}(1-\bar{\rho})}, \hat{t}\}$ sabemos que $t \leq \tau$ satisfaz a (5.9).

Logo, para $s_{j_0} < \min\{\frac{\tau}{2}, 1\}$ existe $t \in \{2^{-i}, i \in \mathbb{N}\}$, $t > s_{j_0}$ que satisfaz (5.8) e (5.9) para $k \geq k_0$. Portanto será aceito pelo passo 4.c) e o passo 4 terminará em tempo finito. Desse modo vemos que se $k \geq k_0$ e $j \geq j_0$ o passo 4.e) não é mais feito e conseqüentemente γ se mantém constante a partir de certa iteração e t se mantém afastado do zero. \square

Uma vez mostrado que o algoritmo está bem definido vamos mostrar um teorema associado à viabilidade dos pontos limites do algoritmo.

Teorema 5.1.2. *Supondo que $\{x_k\}$ seja uma seqüência gerada pelo algoritmo, então*

$$\sum_{k=0}^{\infty} \|h(x_k)\| = \hat{\sigma} < \infty.$$

Além disso, todo ponto limite de $\{x_k\}$, ou de $\{y_k\}$, é viável.

Demonstração: Para $k \geq k_0$, onde $\rho_{k_0} = \bar{\rho}$ os passos 4 e 1 do algoritmo garantem que

$$\Phi(x_{k+1}, \lambda_{k+1}, \bar{\rho}) - \Phi(x_k, \lambda_k, \bar{\rho}) \leq \frac{1}{2}(1-r)(\|h(y_k)\| - \|h(x_k)\|) \leq -\frac{1}{2}(1-r)^2\|h(x_k)\|,$$

portanto:

$$\begin{aligned}\Phi(x_l, \lambda_l, \bar{\rho}) - \Phi(x_{k_0}, \lambda_{k_0}, \bar{\rho}) &= \sum_{k=k_0}^{l-1} \Phi(x_{k+1}, \lambda_{k+1}, \bar{\rho}) - \Phi(x_k, \lambda_k, \bar{\rho}) \\ &\leq \sum_{k=k_0}^{l-1} -\frac{1}{2}(1-r)^2 \|h(x_k)\|.\end{aligned}$$

Como $\Phi(x, \lambda, \bar{\rho})$ é contínua nas variáveis (x, λ) e $\Omega \times \Omega_\lambda$ é um compacto temos que Φ é limitada inferiormente e portanto $S = \sum_{k=k_0}^{\infty} \|h(x_k)\|$ é limitada superiormente. Como S é uma série crescente, temos que ela é convergente, e portanto $\sum_{k=0}^{\infty} \|h(x_k)\| = \hat{\sigma} < \infty$. Dessa forma temos que o termo geral $\|h(x_k)\|$ converge a zero e como $\|h(y_k)\| \leq \|h(x_k)\|$ também temos que $\|h(y_k)\|$ e, conseqüentemente, tanto $h(x_k)$ quanto $h(y_k)$, convergem a zero. Supondo que \bar{x} seja um ponto limite de $\{x_k\}$, ou seja existe $K_1 \subset \mathbb{N}$ tal que

$$\lim_{k \in K_1} x_k = \bar{x},$$

teremos pela continuidade de h que

$$h(\bar{x}) = \lim_{k \in K_1} h(x_k) = 0,$$

e portanto \bar{x} é viável. Com raciocínio completamente análogo para os pontos limites de $\{y_k\}$ concluímos a demonstração. \square

O próximo resultado que vamos demonstrar é que a direção d_k vai a zero. Este resultado segue da demonstração feita em [53] para o fato análogo quando considerada a função objetivo ao invés do Lagrangiano na função de mérito.

Demonstração: Pelos passos 1 e 4 temos que $L(x_{k+1}, \lambda_{k+1}) - L(y_k, \lambda_{k+1}) \leq -\gamma t_k \|d_k\|^2$ e $L(y_k, \lambda_{k+1}) - L(x_k, \lambda_k) \leq \beta \|h(x_k)\|$. Sendo assim

$$\begin{aligned}L(x_{k+1}, \lambda_{k+1}) - L(x_k, \lambda_k) &= L(x_{k+1}, \lambda_{k+1}) - L(y_k, \lambda_{k+1}) + L(y_k, \lambda_{k+1}) - L(x_k, \lambda_k) \\ &\leq -\gamma t_k \|d_k\|^2 + \beta \|h(x_k)\| \\ &\leq -\gamma \bar{t} \|d_k\|^2 + \beta \|h(x_k)\|.\end{aligned}$$

Deste modo, e pelo Teorema 5.1.2, temos que

$$\begin{aligned}
 L(x_{l+1}, \lambda_{l+1}) - L(x_0, \lambda_0) &= \sum_{k=0}^l (L(x_{k+1}, \lambda_{k+1}) - L(x_k, \lambda_k)) \\
 &\leq -\gamma \bar{t} \sum_{k=0}^l \|d_k\|^2 + \beta \sum_{k=0}^l \|h(x_k)\| \\
 &\leq -\gamma \bar{t} \sum_{k=0}^l \|d_k\|^2 + \beta \hat{\sigma}.
 \end{aligned}$$

Como L é contínua num compacto, devemos ter que $\sum_{k=0}^l \|d_k\|^2$ é limitada, e portanto $\|d_k\|$, e conseqüentemente d_k , convergem a zero. \square

Por fim, o último resultado que vamos mostrar se refere à qualidade dos pontos limites gerados pelo algoritmo.

Teorema 5.1.3. *Suponhamos que Ω é uma caixa limitada, que $d_k = d(y_k)$ é escolhida de acordo com (5.1), que existe $M > 0$ tal que $\|H(y_k, \lambda_{k+1})\| \leq M$ e que $\bar{y} = \lim_{k \in K_1} y_k$. Então \bar{y} é AKKT do problema original. Sendo assim, se \bar{y} satisfaz a CPLD então \bar{y} é KKT to problema original.*

Demonstração: Do Teorema 5.1.2 já sabemos que $h(y_k)$ converge a zero. Como as restrições de (5.1) são bem qualificadas, pois são lineares, valem as condições KKT para o subproblema, ou seja, existem $\mu_k, \nu_k^l > 0$ e $\nu_k^u > 0$ tais que

$$H(y_k, \lambda_{k+1})d_k + \nabla f(y_k) + J_h(y_k)^T \lambda_{k+1} + J_h(y_k)^T \mu_k - \nu_k^l + \nu_k^u = 0, \quad (5.10)$$

$$J_h(y_k)d_k = 0$$

$$l \leq y_k + d_k \leq u \quad (5.11)$$

$$[l - y_k - d_k]_i [\nu_k^l]_i = 0 \quad \text{e} \quad [y_k + d_k - u]_i [\nu_k^u]_i = 0. \quad (5.12)$$

De (5.11) temos que

$$-\|d_k\|_\infty + l \leq y_k \leq u + \|d_k\|_\infty,$$

De (5.12) temos que $[\nu_k^l]_i = 0$ se $[l - y_k - d_k]_i < 0$, ou seja $[\nu_k^l]_i = 0$ se $[l - y_k]_i < [d_k]_i$, logo $[\nu_k^l]_i = 0$ se $[l - y_k]_i < -\|d_k\|_\infty$ (pois $[l - y_k]_i < -\|d_k\|_\infty$ implica $[l - y_k]_i < [d_k]_i$). Analogamente $[\nu_k^u]_i = 0$ se $[y_k + d_k - u]_i < 0$, ou seja $[\nu_k^u]_i = 0$ se $[y_k - u]_i < -[d_k]_i$, portanto $[\nu_k^u]_i = 0$ se $[y_k - u]_i < -\|d_k\|_\infty$. Por fim, se definirmos $\gamma_k = \lambda_{k+1} + \mu_k$, da relação (5.10) temos que

$$\|\nabla f(y_k) + J_h(y_k)^T \gamma_k - \nu_k^l + \nu_k^u\| = \|H(y_k, \lambda_{k+1})d_k\|. \quad (5.13)$$

Sendo assim, se definirmos $\epsilon_k = \max\{\|H(y_k, \lambda_{k+1})d_k\|, \|d_k\|_\infty, \|h(y_k)\|\}$, teremos que valem

$$\|\nabla f(y_k) + J_h(y_k)^T \gamma_k - \nu_k^l + \nu_k^u\| \leq \epsilon_k,$$

$$\|h(y_k)\| \leq \epsilon_k,$$

$$l - \epsilon_k \leq y_k \leq u + \epsilon_k,$$

$$[l - y_k]_i < -\epsilon_k \Rightarrow [\nu_k^l]_i = 0,$$

e

$$[y_k - u]_i < -\epsilon_k \Rightarrow [\nu_k^u]_i = 0,$$

ou seja, temos que y_k é ϵ_k -KKT do problema original (4.1). Como $\|H(y_k, \lambda_{k+1})\| \leq M$ para todo k temos que $\|H(y_k, \lambda_{k+1})d_k\|$ converge a zero, uma vez que d_k converge a zero. Dessa forma temos que ϵ_k converge a zero e portanto \bar{y} é AKKT [15]. Consequentemente, por [19], teremos que, se as restrições em \bar{y} satisfazem a CPLD, então \bar{y} é KKT. \square

Uma observação importante é que as demonstrações das propriedades do método modificado seguem de perto as demonstrações feitas em [53] e os resultados são basicamente os mesmos. Isso já era de certa forma esperado pois na demonstração da boa definição do método mostramos que, com escolhas adequadas das constantes,

qualquer estimativa do multiplicador poderia ser aceita. Em particular, se o vetor nulo pertence a Ω_λ , poderíamos adotá-lo em todas as iterações. Nesse caso, nossa extensão coincide com o método original apresentado em [53]. Entretanto, se supusermos boas aproximações dos multiplicadores de Lagrange nosso método tem melhor expectativa de aceitar o passo 1 e assim obtermos melhores taxas de convergência.

Pelas considerações anteriores seria importante mostrar uma atualização natural dos multiplicadores de Lagrange e apresentar resultados de convergência também para as variáveis duais. Para fazer isso basta olhar com um pouco mais de atenção a demonstração do teorema anterior para ver que a relação (5.13) sugere a atualização $\lambda_{k+2} = \gamma_k = \lambda_{k+1} + \mu_k$. Para assegurar que λ_{k+1} permaneça em Ω_λ fazemos a projeção de $\lambda_k + \mu_k$ em Ω_λ . Como o subproblema de programação quadrática resolvido na fase de otimização pode fornecer os multiplicadores μ_k praticamente sem esforço algum, temos que esta atualização é muito barata computacionalmente. Além disso, com hipóteses adequadas, podemos provar a convergência da sequência $\{\lambda_k\}$, conforme apresentado no teorema a seguir.

Teorema 5.1.4. *Suponhamos que $y^* = \lim_{k \in K_1} y_k$ seja um ponto limite regular gerado pelo Algoritmo 5.1. Denotemos por λ^* o vetor de multiplicadores de Lagrange associado a y^* e às restrições $h(x) = 0$. Então a sequência $\{\lambda_{k+2}\}$ converge a λ^* .*

Demonstração: Pela regularidade de y^* temos que todo ponto suficientemente próximo de y^* também será regular. Logo, $k \in K_1$ suficientemente grande, a expressão

$$\nabla f(y_k) + J_h(y_k)^T \gamma_k - \nu_k^l + \nu_k^u = -H(y_k, \lambda_{k+1})d_k$$

fornece solução única para γ_k , ν_k^l e ν_k^u dada por:

$$\begin{pmatrix} \gamma_k \\ \nu_k^l \\ \nu_k^u \end{pmatrix} = -A_k \nabla f(y_k) - A_k H(y_k, \lambda_{k+1})d_k, \quad (5.14)$$

onde A_k é a pseudo-inversa de $\begin{pmatrix} J_h(y_k)^T & -I & I \end{pmatrix}$. Como a pseudo inversa é contínua e $H(y_k, \lambda_{k+1})d_k$ converge a zero se tomarmos o limite em (5.14) para $k \in K_1$ indo para

infinito temos que

$$\begin{pmatrix} \gamma^* \\ \nu_*^l \\ \nu_*^u \end{pmatrix} = -A_* \nabla f(y_*),$$

onde A_* é a pseudo-inversa de $\begin{pmatrix} J_h(y_*)^T & -I & I \end{pmatrix}$. Pela regularidade de y^* temos que $\gamma^* = \lambda^*$, já que a expressão dos multiplicadores em y^* é única. Portanto $\lim_{k \in K_1} \lambda_{k+2} = \lambda^*$. \square

Um exemplo clássico de que a função de mérito adotada pode sempre rejeitar o passo 1, independentemente de quão perto de uma solução quanto estivermos, foi apresentado por Maratos em [75]. A seguir vamos exibir este exemplo e mostrar que o método original de Fischer-Friedlander apresenta efeito Maratos mas nossa extensão não. Para isso vamos supor que o multiplicador λ_k utilizado é igual ao multiplicador na solução. Como já mostramos uma atualização que converge para o multiplicador exato, parece que com apenas um esforço algébrico um pouco maior o resultado valha mesmo que não soubéssemos o valor desse multiplicador.

Considere o problema

$$\begin{aligned} \text{Minimizar} \quad & 2(x_1^2 + x_2^2 - 1) - x_1 \\ \text{sujeito a:} \quad & (x_1^2 + x_2^2 - 1) = 0, \end{aligned} \tag{5.15}$$

A solução é evidentemente $x^* = (1; 0)$ associada ao multiplicador de Lagrange $\lambda^* = -\frac{3}{2}$. Entretanto, se estivermos em um ponto viável, não importa quão próximo da solução ele esteja, desde que não estejamos nela, o passo proveniente do modelo quadrático no espaço tangente não será aceito pela função de mérito usada por Fischer-Friedlander em [53]. Vejamos isso. O Lagrangiano e suas derivadas para o problema seriam:

$$\begin{aligned} L((x; y), \lambda) &= (2 + \lambda)(x_1^2 + x_2^2 - 1) - x_1 \\ \nabla_x L((x; y), \lambda) &= \begin{pmatrix} 2(2 + \lambda)x_1 - 1 \\ 2(2 + \lambda)x_2 \end{pmatrix} \end{aligned}$$

$$\nabla_{xx}^2 L((x; y), \lambda) = \begin{pmatrix} 2(2 + \lambda) & 0 \\ 0 & 2(2 + \lambda) \end{pmatrix}$$

$$\nabla_x L((x; y), \lambda^*) = \begin{pmatrix} x_1 - 1 \\ x_2 \end{pmatrix}$$

$$\nabla_{xx}^2 L((x; y), \lambda^*) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Vamos usar a Hessiana verdadeira e o multiplicador de Lagrange exato associado à solução do problema para construir o problema quadrático para encontrar a direção de busca. Considerando que estamos em um ponto viável podemos escrever $x = y = (\cos(\theta); \sin(\theta))$ e o subproblema seria

$$\begin{aligned} \text{Minimizar} \quad & \frac{1}{2} (d_1^2 + d_2^2) + (4 \cos(\theta) - 1)d_1 + 4 \sin(\theta)d_2 \\ \text{sujeito a:} \quad & 2 \cos(\theta)d_1 + 2 \sin(\theta)d_2 = 0. \end{aligned} \tag{5.16}$$

A solução de (5.16) é $d = (\sin^2(\theta); -\sin(\theta) \cos(\theta))$. A distância $\|y - x^*\| = 2|\sin(\frac{\theta}{2})|$ enquanto $\|y + d - x^*\| = 2 \sin^2(\frac{\theta}{2})$, logo

$$\frac{\|y + d - x^*\|}{\|y - x^*\|^2} = \frac{1}{2},$$

o que indica convergência quadrática. Entretanto

$$h(y + d) = \sin^2(\theta) > 0 = h(y)$$

e

$$f(y + d) = -\cos(\theta) + \sin^2(\theta) > -\cos(\theta) = f(y).$$

Sendo assim, se $\rho \in [0, 1]$ e $H : \mathbb{R} \rightarrow \mathbb{R}$ é uma função crescente qualquer, teríamos que nenhuma função de mérito do tipo $\rho f(x) + (1 - \rho)H(\|h(x)\|)$ aceitaria o ponto $y + d$. A Figura 5.1 ilustra essa situação.

Entretanto, se usássemos a função de mérito $\Phi(x, \lambda, \rho) = \rho L(x, \lambda) + (1 -$

$\rho)\|h(x)\|$ com $\lambda = \lambda^*$ e $\rho \geq \frac{2}{3}$ teríamos que $1 - \frac{3\rho}{2} \leq 0$ e conseqüentemente

$$\begin{aligned} \Phi\left(y + d, -\frac{3}{2}, \rho\right) - \Phi\left(y, -\frac{3}{2}, \rho\right) &= -\rho \cos(\theta) + \left(1 - \frac{3\rho}{2}\right) \text{sen}^2(\theta) - (-\rho \cos(\theta)) \\ &= \left(1 - \frac{3\rho}{2}\right) \text{sen}^2(\theta) \\ &\leq 0. \end{aligned}$$

Como $x = y$ temos que $h(y) - h(x) = 0$ e portanto

$$\Phi\left(y + d, -\frac{3}{2}, \rho\right) - \Phi\left(y, -\frac{3}{2}, \rho\right) \leq \frac{1}{2}(1 - r)(\|h(y)\| - \|h(x)\|).$$

Ou seja, com essas escolhas, o passo que faz com que a distância entre o iterando e a solução diminua quadraticamente será aceito. Infelizmente não temos como garantir que ao longo de todo o processo iterativo o Algoritmo mantenha $\rho_k \geq \frac{2}{3}$. Nesse caso o passo unitário também poderia ser rejeitado por nossa extensão mesmo para iterandos arbitrariamente próximos da solução. Essa observação é que nos motivou a desenvolver o método híbrido descrito na Seção 5.3.

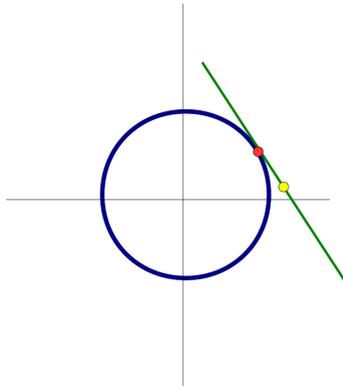


Figura 5.1: Representação gráfica do exemplo 5.15 que é um caso clássico que apresenta efeito Maratos. O passo $t = 1$ sempre é rejeitado pela função de mérito original de Fischer-Friedlander, mas é aceito com a nossa função de mérito.

5.2 Algoritmo Birgin-Martínez

Vamos agora apresentar o algoritmo local de Birgin e Martínez [29] e algumas ligeiras modificações dos resultados teóricos a ele associados além de refazer algumas das demonstrações dos mesmos.

Para nossa análise vamos definir a medida de otimalidade $G(x, \lambda) = P_\Omega(x - \nabla_x L(x, \lambda)) - x$, onde P_Ω é o operador de projeção em Ω com respeito à norma Euclidiana.

Com isto, dados $K_1, K_2, K_3 \geq 0$ e um par (x, λ) diremos que uma iteração da Restauração Inexata com tolerância $r \in [0, 1)$, $\eta \in [0, 1)$ está bem definida se for possível encontrar $y, z \in \Omega$ e $\mu \in \mathbb{R}^m$ tais que:

$$\|h(y)\| \leq r\|h(x)\|, \quad (5.17)$$

$$\|y - x\| \leq K_1\|h(x)\|, \quad (5.18)$$

$$\|J_h(y)(z - y)\| \leq K_2\|G(y, \lambda)\|^2, \quad (5.19)$$

$$\|P_\Omega(z - \nabla L(z, \lambda) - J_h(y)^T(\mu - \lambda)) - z\| \leq \eta\|G(y, \lambda)\|, \quad (5.20)$$

$$\|z - y\| + \|\mu - \lambda\| \leq K_3\|G(y, \lambda)\|. \quad (5.21)$$

No caso em que a iteração com tolerância r, η esteja bem definida vamos denotar um par (z, μ) resultante desta iteração por $N_{r,\eta}(x, \lambda)$. Se for possível gerar uma sequência $\{(z_k, \mu_k)\}$ tomando $(z_{k+1}, \mu_{k+1}) = N_{r_k, \eta_k}(z_k, \mu_k) \equiv N_k(z_k, \mu_k)$, vamos dizer que esta é uma sequência gerada pelo método local com a sequência de parâmetros $r_k, \eta_k \in [0, 1)$. Uma observação importante é que se $(r, \eta) \leq (\bar{r}, \bar{\eta})$ e a iteração com tolerância r, η estiver bem definida, então a iteração com tolerância $\bar{r}, \bar{\eta}$ também vai estar, pois os mesmos y, z e μ satisfazem as equações menos exigentes (5.17)-(5.21).

Assim como no método global, para análise de nossos resultados vamos supor que os gradientes das funções envolvidas sejam Lipschitz-contínuos e, por simplicidade, vamos adotar aqui a mesma constante de Lipschitz \bar{L} para todas as funções. Ou seja,

suponhamos que para todo $x, w \in \Omega$ e $i = 1, 2, \dots, m$ valem:

$$\begin{aligned}\|\nabla f(x) - \nabla f(w)\| &\leq \bar{L}\|x - w\| \\ \|\nabla h_i(x) - \nabla h_i(w)\| &\leq \bar{L}\|x - w\| \\ \|J_h^T(x) - J_h^T(w)\| &\leq \bar{L}\|x - w\| \\ \|h(x) - h(w) - J_h(x)(x - w)\| &\leq \bar{L}\|x - w\|^2.\end{aligned}$$

Vamos definir também as constantes:

$$\begin{aligned}c &= \max\{K_1, K_2, K_3\} \\ c_1 &= 2c + c\bar{L} \\ c_2 &= c\bar{L} \\ c_3 &= c + 2c^2 + c^2\bar{L} \\ c_4 &= c + c^2\bar{L}\end{aligned}$$

As principais características dos pontos resultantes de uma iteração do método local estão descritos no teorema a seguir.

Teorema 5.2.1. *Suponhamos que $(z, \mu) = N_{r,\eta}(x, \lambda)$, então valem as seguintes desigualdades:*

$$\begin{aligned}\|h(z)\| &\leq r\|h(x)\| + c_4 [\|G(x, \lambda)\| + (c_1 + c_2\|\lambda\|)\|h(x)\|]^2, \\ \|G(z, \mu)\| &\leq \eta [\|G(x, \lambda)\| + (c_1 + c_2\|\lambda\|)\|h(x)\|] + c_4 [\|G(x, \lambda)\| + (c_1 + c_2\|\lambda\|)\|h(x)\|]^2, \\ \|z - x\| &\leq c\|G(x, \lambda)\| + (c_3 + c_4\|\lambda\|)\|h(x)\|, \\ \|\mu - \lambda\| &\leq c\|G(x, \lambda)\| + (c_3 + c_4\|\lambda\|)\|h(x)\|.\end{aligned}$$

Demonstração: Esse resultado é demonstrado no Teorema 2.1. de [29].

□

Em [29] é definido o resíduo como uma medida do grau de “estacionaridade” de um ponto. Sendo assim o resíduo combina uma medida de viabilidade, no caso

$\|h(x)\|$, e uma medida de otimalidade, no caso $\|G(x, \lambda)\|$. Neste trabalho vamos medir o resíduo de um ponto (x, λ) como $R(x, \lambda) = \max\{\|h(x)\|, \|G(x, \lambda)\|\}$. Esta maneira de medir o resíduo é mais simples do que a maneira exposta em [29] onde o resíduo também é medido como uma norma de $(\|h(x)\|, \|G(x, \lambda)\|)$ mas esta é uma norma implícita que depende dos parâmetros r e η . Com o novo resíduo $R(x, \lambda)$ vamos simplificar as relações dadas pelo Teorema 5.2.1. Vejamos como faremos isto

$$\begin{aligned} \|G(x, \lambda) + (c_1 + c_2\|\lambda\|)\|h(x)\| &\leq R(x, \lambda) + (c_1 + c_2\|\lambda\|)R(x, \lambda) \\ &\leq (c_1 + c_2\|\lambda\| + 1)R(x, \lambda), \end{aligned}$$

consequentemente

$$c_4 [\|G(x, \lambda)\| + (c_1 + c_2\|\lambda\|)\|h(x)\|]^2 \leq c_4(c_1 + c_2\|\lambda\| + 1)^2 R(x, \lambda)^2.$$

Da mesma forma

$$\begin{aligned} c\|G(x, \lambda)\| + (c_3 + c_4\|\lambda\|)\|h(x)\| &\leq cR(x, \lambda) + (c_3 + c_4\|\lambda\|)R(x, \lambda) \\ &\leq (c + c_3 + c_4\|\lambda\|)R(x, \lambda). \end{aligned}$$

Deste modo, se $(z, \mu) = N_{r,\eta}(x, \lambda)$, as equações do teorema 5.2.1 se reduzem

a:

$$\|h(z)\| \leq rR(x, \lambda) + c_4(c_1 + c_2\|\lambda\| + 1)^2 R(x, \lambda)^2, \quad (5.22)$$

$$\|G(z, \mu)\| \leq \eta(c_1 + c_2\|\lambda\| + 1)R(x, \lambda) + c_4(c_1 + c_2\|\lambda\| + 1)^2 R(x, \lambda)^2, \quad (5.23)$$

$$\|z - x\| \leq (c + c_3 + c_4\|\lambda\|)R(x, \lambda), \quad (5.24)$$

$$\|\mu - \lambda\| \leq (c + c_3 + c_4\|\lambda\|)R(x, \lambda). \quad (5.25)$$

Vamos agora demonstrar um resultado análogo ao teorema 2.2 de [29]. Para isso, dado um par crítico $(\bar{x}, \bar{\lambda})$, ou seja,

$$h(\bar{x}) = 0,$$

$$G(\bar{x}, \bar{\lambda}) = 0,$$

vamos definir $M = 2\|\bar{\lambda}\| + 1$ e $c_5 = (c_1 + c_2M + 1)$.

Lema 5.2.1. *Suponhamos que a iteração da Restauração Inexata esteja bem definida com parâmetros r, η e $(z, \mu) = N_{r, \eta}(x, \lambda)$. Então, se $\max\{r, c_5\eta\} < \theta$ existem $\epsilon_{r, \eta, \theta}, \delta_{r, \eta, \theta} > 0$ tais que, se*

$$\|x - \bar{x}\| \leq \epsilon_{r, \eta, \theta}, \quad e \quad \|\lambda - \bar{\lambda}\| \leq \delta_{r, \eta, \theta}$$

então valem:

$$\|\lambda\| \leq M, \tag{5.26}$$

$$\|z - x\| \leq \beta R(x, \lambda), \tag{5.27}$$

$$\|\mu - \lambda\| \leq \beta R(x, \lambda), \tag{5.28}$$

$$R(z, \mu) \leq \theta R(x, \lambda). \tag{5.29}$$

Demonstração: Pela desigualdade triangular temos que

$$\|\lambda\| \leq \|\lambda - \bar{\lambda}\| + \|\bar{\lambda}\|.$$

Dessa forma tomando $\delta_0 = \|\bar{\lambda}\| + 1$ temos que se $\|\lambda - \bar{\lambda}\| \leq \delta_0$ então

$$\|\lambda\| \leq \delta_0 + \|\bar{\lambda}\| = M,$$

e consequentemente

$$(c + c_3 + c_4\|\lambda\|)R(x, \lambda) \leq (c + c_3 + c_4M)R(x, \lambda).$$

Definindo $\beta = (c + c_3 + c_4M)$ as relações (5.24) e (5.25) garantem que

$$\|z - x\| \leq \beta R(x, \lambda)$$

e

$$\|\mu - \lambda\| \leq \beta R(x, \lambda).$$

Ainda considerando $\|\lambda - \bar{\lambda}\| \leq \delta_0$ a relação (5.22) nos dá que

$$\begin{aligned} \|h(z)\| &\leq rR(x, \lambda) + c_4(c_1 + c_2\|\lambda\| + 1)^2 R(x, \lambda)^2 \\ &\leq [r + c_4c_5^2 R(x, \lambda)] R(x, \lambda). \end{aligned} \tag{5.30}$$

Sendo assim, pela continuidade de $R(x, \lambda)$, e o fato de que $R(\bar{x}, \bar{\lambda}) = 0$, temos que para $0 < \epsilon_{r,\eta,\theta}$ e $0 < \delta_{r,\eta,\theta} \leq \delta_0$ suficientemente pequenos temos que

$$\|h(z)\| \leq \theta R(x, \lambda),$$

sempre que $\|x - \bar{x}\| \leq \epsilon_{r,\eta,\theta}$ e $\|\lambda - \bar{\lambda}\| \leq \delta_{r,\eta,\theta}$.

Analogamente, se consideramos (5.23) temos que

$$\begin{aligned} \|G(z, \mu)\| &\leq \eta(c_1 + c_2\|\lambda\| + 1)R(x, \lambda) + c_4(c_1 + c_2\|\lambda\| + 1)^2 R(x, \lambda)^2 \\ &\leq [c_5\eta + c_4c_5^2 R(x, \lambda)] R(x, \lambda). \end{aligned} \quad (5.31)$$

Da mesma forma, temos que para $0 < \epsilon_{r,\eta,\theta}$ e $0 < \delta_{r,\eta,\theta} \leq \delta_0$ suficientemente pequenos vale que

$$\|G(z, \mu)\| \leq \theta R(x, \lambda),$$

sempre que $\|x - \bar{x}\| \leq \epsilon_{r,\eta,\theta}$ e $\|\lambda - \bar{\lambda}\| \leq \delta_{r,\eta,\theta}$.

Considerando então $0 < \epsilon_{r,\eta,\theta}$ e $0 < \delta_{r,\eta,\theta} \leq \delta_0$ suficientemente pequenos de modo que valham tanto (5.30) quanto (5.31) temos que

$$R(z, \mu) = \max\{\|h(z)\|, \|G(z, \mu)\|\} \leq \theta R(x, \lambda).$$

□

Uma observação importante é que se $r \leq \bar{r}$ então $r + c_4c_5^2R(x, \lambda) \leq \bar{r} + c_4c_5^2R(x, \lambda)$ e portanto se $\bar{r} + c_4c_5^2R(x, \lambda) \leq \theta$ teríamos também que $r + c_4c_5^2R(x, \lambda) \leq \theta$, analogamente, $\eta \leq \bar{\eta}$ e $c_5\bar{\eta} + c_4c_5^2R(x, \lambda) \leq \theta$ também implica que $c_5\eta + c_4c_5^2R(x, \lambda) \leq \theta$. Dessa forma $\epsilon_{r,\eta,\theta} \geq \epsilon_{\bar{r},\bar{\eta},\theta}$.

Com o auxílio dos resultados anteriores vamos agora demonstrar os principais teoremas relativos à convergência do método local.

Teorema 5.2.2. *Seja $(\bar{x}, \bar{\lambda})$ um par crítico. Consideremos as sequências de parâmetros $\{r_k\}$, $\{\eta_k\}$, limitadas superiormente por \bar{r} e $\bar{\eta}$ respectivamente, de forma que*

$$\max\{\bar{r}, c_5\bar{\eta}\} < \theta < 1.$$

Suponhamos também que a iteração da Restauração Inexata com esses parâmetros possa ser completada sempre que $\|x - \bar{x}\| \leq \epsilon_1 \leq \epsilon_{\bar{r}, \bar{\eta}, \theta}$ e $\|\lambda - \bar{\lambda}\| \leq \delta_1 \leq \delta_{\bar{r}, \bar{\eta}, \theta}$, onde $\epsilon_{\bar{r}, \bar{\eta}}$ e $\delta_{\bar{r}, \bar{\eta}, \theta}$ são referentes ao Lema 5.2.1. Então:

1. Existem $\epsilon_{2, \theta} \in (0, \epsilon_1]$ e $\delta_{2, \theta} \in (0, \delta_1]$ tais que se $\|x_0 - \bar{x}\| \leq \epsilon_{2, \theta}$ e $\|\lambda_0 - \bar{\lambda}\| \leq \delta_{2, \theta}$ então $\|x_k - \bar{x}\| \leq \epsilon_1$ e $\|\lambda_k - \bar{\lambda}\| \leq \delta_1$, para todo k , e portanto a sequência gerada pelo método local está bem definida;
2. $R(x_{k+1}, \lambda_{k+1}) \leq \theta R(x_k, \lambda_k)$ e $R(x_k, \lambda_k) \leq \theta^k R(x_0, \lambda_0)$;
3. A sequência $\{(x_k, \lambda_k)\}$ converge a um par crítico (x_*, λ_*) ;
4. $\|x_k - x_*\| \leq \frac{\beta \theta_k}{1 - \theta} R(x_0, \lambda_0)$ e $\|\lambda_k - \lambda_*\| \leq \frac{\beta \theta_k}{1 - \theta} R(x_0, \lambda_0)$, para todo k ;
5. Se tomássemos r e η dependendo de k , não crescentes e tendendo a zero então $R(x_k, \lambda_k)$ converge a zero Q -superlinearmente e tanto x_k quanto λ_k convergem R -superlinearmente;
6. Se tomássemos r e η nulos então $R(x_k, \lambda_k)$ converge Q -quadraticamente e tanto x_k quanto λ_k convergem R -quadraticamente.

Demonstração: Pelo fato de que $\epsilon_1 \leq \epsilon_{\bar{r}, \bar{\eta}, \theta}$ e $\delta_1 \leq \delta_{\bar{r}, \bar{\eta}, \theta}$, para todas as iterações na vizinhança $\|x - \bar{x}\| \leq \epsilon_1$ e $\|\lambda - \bar{\lambda}\| \leq \delta_1$ valem os resultados do Lema 5.2.1. Provemos então que se x_0 e λ_0 estão suficientemente próximos da solução então todas as demais iterações estarão na vizinhança em que vale o Lema 5.2.1. Para isso defina

$$\phi(\epsilon, \delta) = \max\{R(x, \lambda) \mid \|x - \bar{x}\| \leq \epsilon, \|\lambda - \bar{\lambda}\| \leq \delta\}.$$

Pela continuidade de $R(x, \lambda)$ e o fato de que $R(\bar{x}, \bar{\lambda}) = 0$ temos que

$$\lim_{\epsilon \rightarrow 0, \delta \rightarrow 0} \phi(\epsilon, \delta) = 0.$$

Portanto existem $\epsilon_{2, \theta} \leq \frac{\epsilon_1}{2}$ e $\delta_{2, \theta} \leq \frac{\delta_1}{2}$ tais que

$$\frac{\beta \phi(\epsilon_{2, \theta}, \delta_{2, \theta})}{1 - \theta} \leq \frac{\min\{\epsilon_1, \delta_1\}}{2}.$$

Dessa forma se $\|x_0 - \bar{x}\| \leq \epsilon_{2,\theta}$ e $\|\lambda_0 - \bar{\lambda}\| \leq \delta_{2,\theta}$ temos que

$$\epsilon_{2,\theta} + \frac{\beta R(x_0, \lambda_0)}{1 - \theta} \leq \epsilon_1,$$

$$\delta_{2,\theta} + \frac{\beta R(x_0, \lambda_0)}{1 - \bar{r}} \leq \delta_1.$$

Vamos mostrar agora por indução que se (x_k, λ_k) está bem definido então (x_{k+1}, λ_{k+1}) também estará. Para isso mostraremos que

$$\begin{aligned} R(x_k, \lambda_k) &\leq \theta^k R(x_0, \lambda_0), \\ \|x_k - \bar{x}\| &\leq \epsilon_{2,\theta} + \beta R(x_0, \lambda_0) \sum_{j=0}^{k-1} \theta^j \leq \epsilon_1, \\ \|\lambda_k - \bar{\lambda}\| &\leq \delta_{2,\theta} + \beta R(x_0, \lambda_0) \sum_{j=0}^{k-1} \theta^j \leq \delta_1. \end{aligned}$$

Para $k = 0$ o resultado é evidente. Assumindo agora que seja verdadeiro para algum $k = i$ provemos que também será verdadeiro para $k = i + 1$. Como a hipótese de indução nos diz que $\|x_i - \bar{x}\| \leq \epsilon_1$ e $\|\lambda_i - \bar{\lambda}\| \leq \delta_1$ então a iteração da Restauração Inexata, com parâmetros r_i e η_i , está bem definida, e conseqüentemente, com parâmetros \bar{r} e $\bar{\eta}$ também está, e portanto $(x_{i+1}, \lambda_{i+1}) = N_{\bar{r}, \bar{\eta}}(x_i, \lambda_i)$. Sendo assim, a relação (5.29) nos diz que

$$R(x_{i+1}, \lambda_{i+1}) \leq \theta R(x_i, \lambda_i).$$

Combinando então (5.29) e a hipótese de indução temos que

$$R(x_{i+1}, \lambda_{i+1}) \leq \theta R(x_i, \lambda_i) \leq \theta \theta^i R(x_0, \lambda_0) = \theta^{i+1} R(x_0, \lambda_0),$$

e portanto a primeira desigualdade já está demonstrada para $k = i + 1$.

Para as outras duas desigualdades notemos que, como $\sum_{j=0}^{k-1} \theta^j = \frac{1-\theta^k}{1-\theta}$, temos que $\sum_{j=0}^k \theta^j \leq \frac{1}{1-\theta}$ para todo k , e conseqüentemente

$$\beta R(x_0, \lambda_0) \sum_{j=0}^{k-1} \theta^j \leq \frac{\beta R(x_0, \lambda_0)}{1 - \theta} \leq \frac{\min\{\epsilon_1, \delta_1\}}{2}$$

para todo k . Usando isto, a desigualdade triangular, a hipótese de indução e a relação (5.27) temos que:

$$\begin{aligned}
 \|x_{i+1} - \bar{x}\| &\leq \|x_{i+1} - x_i\| + \|x_i - \bar{x}\| \\
 &\leq \beta R(x_i, \lambda_i) + \epsilon_{2,\theta} + \beta R(x_0, \lambda_0) \sum_{j=0}^{i-1} \theta^j \\
 &\leq \beta \theta^i R(x_0, \lambda_0) + \epsilon_{2,\theta} + \beta R(x_0, \lambda_0) \sum_{j=0}^{i-1} \theta^j \\
 &\leq \epsilon_{2,\theta} + \beta R(x_0, \lambda_0) \sum_{j=0}^i \theta^j \\
 &\leq \frac{\epsilon_1}{2} + \frac{\epsilon_1}{2} \leq \epsilon_1.
 \end{aligned}$$

Usando exatamente os mesmos argumentos temos que

$$\|\lambda_{i+1} - \bar{\lambda}\| \leq \delta_{2,\theta} + \beta R(x_0, \lambda_0) \sum_{j=0}^i \theta^j \leq \delta_1.$$

Dessa forma temos que as afirmações são verdadeiras pra $k = i + 1$, e portanto são válidas para todo k . Com isso provamos que se a iteração da Restauração Inexata estiver bem definida numa vizinhança de um ponto crítico então, se tomarmos um ponto inicial suficientemente próximo dessa solução, teremos que todas as iterações do método permanecem nessa vizinhança e portanto a sequência de iterandos está bem definida.

Vamos mostrar agora que a sequência $\{(x_k, \lambda_k)\}$ converge a um par crítico (x_*, λ_*) . De fato, como os iterandos estão na vizinhança onde vale o Lema 5.2.1 as relações (5.27)-(5.29) garantem que

$$\|x_{k+1} - x_k\| \leq \beta R(x_k, \lambda_k) \leq \beta \theta^k R(x_0, \lambda_0),$$

e

$$\|\lambda_{k+1} - \lambda_k\| \leq \beta R(x_k, \lambda_k) \leq \beta \theta^k R(x_0, \lambda_0).$$

Sendo assim, para todo k e $j = 0, 1, 2, \dots$, temos que

$$\begin{aligned} \|x_{k+j} - x_k\| &\leq \|x_{k+1} - x_k\| + \|x_{k+2} - x_{k+1}\| + \dots + \|x_{k+j} - x_{k+j-1}\| \\ &\leq \beta(\theta^k + \theta^{k+1} + \dots + \theta^{k+j-1})R(x_0, \lambda_0) \\ &\leq \frac{\theta^k}{1-\theta}\beta R(x_0, \lambda_0). \end{aligned} \quad (5.32)$$

Sendo assim, fazendo o limite quando k vai para infinito temos que $\|x_{k+j} - x_k\|$ vai a zero e portanto $\{x_k\}$ é uma sequência de Cauchy, e conseqüentemente, como Ω é fechado, $\{x_k\}$ é convergente para algum $x_* \in \Omega$. Seguindo exatamente o mesmo raciocínio temos que

$$\|\lambda_{k+j} - \lambda_k\| \leq \frac{\theta^k}{1-\theta}\beta R(x_0, \lambda_0), \quad (5.33)$$

e portanto $\{\lambda_k\}$ é uma sequência de Cauchy que converge a $\lambda_* \in \mathbb{R}^n$. Como $R(x, \lambda)$ é uma função continua e $R(x_k, \lambda_k) \leq \theta^k R(x_0, \lambda_0)$ se fizermos o limite de k tendendo a infinito obtemos que $R(x_*, \lambda_*) = 0$, ou seja (x_*, λ_*) é um par crítico.

Para a taxa de convergência basta fazer o limite de j tendendo a infinito em (5.32) e (5.33) para obtermos que:

$$\|x_* - x_k\| \leq \frac{\theta^k}{1-\theta}\beta R(x_0, \lambda_0)$$

e

$$\|\lambda_* - \lambda_k\| \leq \frac{\theta^k}{1-\theta}\beta R(x_0, \lambda_0).$$

Dessa forma, se denotarmos $\epsilon_k = \frac{\theta^k}{1-\theta}\beta R(x_0, \lambda_0)$ teremos que $\frac{\epsilon_{k+1}}{\epsilon_k} = \theta < 1$ e portanto tanto a sequência primal $\{x_k\}$ quanto a sequência dual $\{\lambda_k\}$ convergem R-linearmente.

Supondo agora que r_k e η_k sejam sequências não crescentes tendendo a zero podemos tomar $\{\theta_k\}$, também não crescente tendendo a zero, tal que $\max\{r_k, c_5\eta_k\} < \theta_k < 1$. Nessas condições vamos provar melhores taxas de convergência. Pela monotonicidade temos que $\theta = \theta_0$ é um limitante superior de θ_k e portanto valem os resultados do item anterior. Em particular já sabemos que as sequências $\{x_k\}$ e $\{\lambda_k\}$ geradas

pelo algoritmo convergem. Além disso, podemos redefinir $\{r_k\}$, $\{\eta_k\}$ e $\{\theta_k\}$, inserindo termos repetidos $r_k = r_{k-1}$, $\eta_k = \eta_{k-1}$ e $\theta_k = \theta_{k-1}$ sempre que necessário, de forma que sempre teremos que $\|x_k - x_*\| \leq \epsilon_{r_k, \theta_k}$ e $\|\lambda_k - \lambda_*\| \leq \delta_{r_k, \theta_k}$. Como x_k converge a x_* e λ_k converge a λ_* em um número finito de passos teremos que os iterandos estarão dentro da vizinhança de interesse. Portanto o número de inserções entre dois termos da sequência original é finito. Dessa forma valem os resultados do Lema 5.2.1 com $r = r_k$, $\eta = \eta_k$, $\theta = \theta_k$ em todas as iterações.

Pelo resultado (5.29) no Lema 5.2.1 temos que $R(x_{k+1}, \lambda_{k+1}) \leq \theta_k R(x_k, \lambda_k)$, logo a sequência dos resíduos $\{R(x_k, \lambda_k)\}$ converge Q-superlinearmente para zero. Vamos agora mostrar a convergência para os pontos primais e duais. Pelo fato da sequência $\{\theta_k\}$ ser não crescente temos que se $i \geq k$ então $\theta_i \leq \theta_k$ e portanto

$$R(x_{i+1}, \lambda_{i+1}) \leq \theta_i R(x_i, \lambda_i) \leq \theta_k R(x_i, \lambda_i).$$

Deste modo $R(x_{i+1}, \lambda_{i+1}) \leq \theta_k^{i-k} R(x_k, \lambda_k)$ e conseqüentemente

$$\begin{aligned} \|x_{k+j} - x_k\| &\leq \|x_{k+1} - x_k\| + \|x_{k+2} - x_{k+1}\| + \cdots + \|x_{k+j} - x_{k+j-1}\| \\ &\leq \beta R(x_k, \lambda_k) + \beta R(x_{k+1}, \lambda_{k+1}) + \cdots + \beta R(x_{k+j-1}, \lambda_{k+j-1}) \\ &\leq \beta R(x_k, \lambda_k)(1 + \theta_k + \theta_k^2 + \cdots + \theta_k^{j-1}) \\ &\leq \beta R(x_k, \lambda_k) \frac{1}{1 - \theta_k}. \end{aligned}$$

Analogamente

$$\|\lambda_{k+j} - \lambda_k\| \leq \beta R(x_k, \lambda_k) \frac{1}{1 - \theta_k}.$$

Fazendo o limite para j tendendo a infinito temos que

$$\|x_* - x_k\| \leq \frac{\beta}{1 - \theta_k} R(x_k, \lambda_k) \quad \text{e} \quad \|\lambda_* - \lambda_k\| \leq \frac{\beta}{1 - \theta_k} R(x_k, \lambda_k).$$

Definindo $\epsilon_k = \frac{\beta}{1 - \theta_k} R(x_k, \lambda_k)$ temos que

$$\frac{\epsilon_{k+1}}{\epsilon_k} = \frac{1 - \theta_k}{1 - \theta_{k+1}} \frac{R(x_{k+1}, \lambda_{k+1})}{R(x_k, \lambda_k)} \leq \theta_k.$$

Logo $\{\epsilon_k\}$ converge Q-superlinearmente para zero e conseqüentemente $\{x_k\}$ e $\{\lambda_k\}$ convergem R-superlinearmente para x_* .

Para uma taxa de convergência ainda superior suponhamos agora que os parâmetros r_k e η_k sejam nulos em todas as iterações. Dessa forma então as equações (5.30) e (5.31) se reduzem a

$$\begin{aligned}\|h(z)\| &\leq [c_4 c_5^2 R(x, \lambda)] R(x, \lambda), \\ \|G(z, \mu)\| &\leq [c_4 c_5^2 R(x, \lambda)] R(x, \lambda).\end{aligned}$$

Dessa forma podemos tomar $\theta_k = c_4 c_5^2 R(x_k, \lambda_k)$ e $\epsilon_{2, \theta_k} > 0$ e $0 < \delta_{2, \theta_k} \leq \delta_0$ constantes em todas as iterações. Tomando (x_0, λ_0) suficientemente próximo de $(\bar{x}, \bar{\lambda})$ temos $R(x_0, \lambda_0)$ tão pequeno quanto quisermos. Portanto se o ponto inicial está suficientemente próximo da solução temos que $\theta_0 < 1$. O Lema 5.2.1 garante que

$$R(x_{k+1}, \lambda_{k+1}) \leq \theta_k R(x_k, \lambda_k),$$

portanto, como $\theta_0 < 1$ temos que $R(x_k, \lambda_k)$ é decrescente e portanto θ_k também é. Deste modo o resultado anterior já garante que $R(x_k, \lambda_k)$ converge para zero, e consequentemente, θ_k também converge a zero e (x_k, λ_k) converge a (x_*, λ_*) . Do fato de que

$$\tilde{R}(x_{k+1}, \lambda_{k+1}) \leq \theta_k R(x_k, \lambda_k) \leq c_4 c_5^2 R(x_k, \lambda_k)^2$$

temos que $R(x_k, \lambda_k)$ converge quadraticamente a zero. Novamente, usando o fato de que

$$\|x_{k+1} - x_k\| \leq \beta R(x_k, \lambda_k), \quad \|\lambda_{k+1} - \lambda_k\| \leq \beta R(x_k, \lambda_k), \quad R(x_{k+1}, \lambda_{k+1}) \leq \theta_k R(x_k, \lambda_k)$$

e que θ_k é decrescente concluímos que

$$\|x_k - x_*\| \leq \frac{\beta}{1 - \theta_k} R(x_k, \lambda_k) \leq \frac{\beta}{1 - \theta_0} R(x_k, \lambda_k)$$

e

$$\|\lambda_k - \lambda_*\| \leq \frac{\beta}{1 - \theta_0} R(x_k, \lambda_k).$$

Sendo assim provamos que $\{x_k\}$ e $\{\lambda_k\}$ convergem R-quadraticamente para x_* . \square

Em [29] os autores indicam que iterações do método de Brent-Brown deve satisfazer as condições da restauração. Entretanto não são exibidas condições bem estabelecidas e uma demonstração formal sob essas condições para que a iteração da Restauração Inexata esteja bem definida. Embora os teoremas de convergência apresentem melhores propriedades quanto menores forem r e η é mais difícil satisfazer as condições do método local com tolerâncias menores. Os próximos resultados dessa seção tem como objetivo estabelecer condições particulares que nos permitem demonstrar rigorosamente a boa definição de iterações do método de Restauração Inexata Birgin-Martínez. Um ponto chave desses resultados é que nas demonstrações envolvidas exibimos uma maneira Newtoniana de conseguir satisfazer as equações referentes ao método local.

O primeiro teorema que demonstraremos resume as propriedades do método de Newton para sistemas não lineares que vamos precisar em nossa análise.

Teorema 5.2.3. *Consideremos a função suave $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ e o sistema não linear associado $F(x) = 0$. Suponhamos que x^* é uma solução do sistema, ou seja $F(x^*) = 0$. Suponhamos também que $J_F(x^*)$ é invertível, e que existam constantes positivas ϵ , M , L_F e L_J tais que*

$$\|J_F(x^*)^{-1}\| \leq M,$$

$$\|F(x) - F(y)\| \leq L_F\|x - y\|,$$

e

$$\|J_F(x) - J_F(y)\| \leq L_J\|x - y\|,$$

para todo x e y na bola fechada $\bar{B}(x^*, \epsilon)$. Então, se $\|y - x^*\| \leq \min \left\{ \frac{1}{8L_JM}, \epsilon \right\}$, valem:

a) $J_F(y)$ é invertível e portanto a iteração de Newton $z = y - J_F(y)^{-1}F(y)$ está bem definida.

b) $\|J_F(y)^{-1}\| \leq 2M$.

c) $\|z - x^*\| \leq 4L_JM\|y - x^*\|^2$.

$$d) \|F(z)\| \leq 16L_F L_J M^3 \|F(y)\|^2.$$

Demonstração: Primeiramente observemos que se $\|y - x^*\| \leq \min \left\{ \frac{1}{8L_J M}, \epsilon \right\}$ temos que

$$\|J_F(y) - J_F(x^*)\| \leq L_J \|y - x^*\| \leq \frac{1}{8M} \leq \frac{1}{8\|J_F(x^*)^{-1}\|}.$$

Mais ainda, se $y \neq x^*$, podemos usar (5.4) para obtermos que

$$\begin{aligned} \frac{\|F(y) - F(x^*) - J_F(x^*)(y - x^*)\|}{\|y - x^*\|} &\leq \frac{\|\int_0^1 (J_F(x^* + t(y - x^*)) - J_F(x^*)) (y - x^*) dt\|}{\|y - x^*\|} \\ &\leq \frac{\int_0^1 \|J_F(x^* + t(y - x^*)) - J_F(x^*)\| \|y - x^*\| dt}{\|y - x^*\|} \\ &\leq L_J \|y - x^*\| \int_0^1 |t| dt \\ &\leq \frac{1}{16M} \\ &\leq \frac{1}{8\|J_F(x^*)^{-1}\|}. \end{aligned}$$

Dessa forma os itens *a)*, *b)* e *c)* seguem como escólios e corolários do Teorema 5.4.4 de [81] e de seus resultados auxiliares usando a taxa linear envolvida nas passagens das demonstrações como sendo $r = 0.5$.

Para demonstrar o item *d)* vamos inicialmente mostrar que é válida a seguinte relação

$$\|y - x^*\| \leq 2\|J_F(x^*)^{-1}\| \|F(y)\|,$$

sempre que $\|y - x^*\| \leq \min \left\{ \frac{1}{\|J_F(x^*)^{-1}\| L_J}, \epsilon \right\}$. Para isso usamos novamente (5.4) para obtermos que

$$F(y) = J_F(x^*)(y - x^*) + \int_0^1 (J_F(x^* + t(y - x^*)) - J_F(x^*)) (y - x^*) dt.$$

e conseqüentemente

$$y - x^* = J_F(x^*)^{-1} F(y) - J_F(x^*)^{-1} \int_0^1 (J_F(x^* + t(y - x^*)) - J_F(x^*)) (y - x^*) dt.$$

Portanto,

$$\begin{aligned}
 \|y - x^*\| &\leq \|J_F(x^*)^{-1}\| \|F(y)\| + \|J_F(x^*)^{-1}\| \int_0^1 \|J_F(x^* + t(y - x^*)) - J_F(x^*)\| \|y - x^*\| dt, \\
 &\leq \|J_F(x^*)^{-1}\| \|F(y)\| + \|J_F(x^*)^{-1}\| \int_0^1 L_J \|(x^* + t(y - x^*)) - x^*\| \|y - x^*\| dt, \\
 &\leq \|J_F(x^*)^{-1}\| \|F(y)\| + L_J \|J_F(x^*)^{-1}\| \|y - x^*\|^2 \int_0^1 |t| dt.
 \end{aligned}$$

Dessa forma

$$\left(1 - \frac{L_J \|J_F(x^*)^{-1}\| \|y - x^*\|}{2}\right) \|y - x^*\| \leq \|J_F(x^*)^{-1}\| \|F(y)\|.$$

Para $\|y - x^*\| \leq \frac{1}{\|J_F(x^*)^{-1}\| L_J}$ temos que $\frac{1}{2} \leq \left(1 - \frac{L_J \|J_F(x^*)^{-1}\| \|y - x^*\|}{2}\right)$, o que implica na relação desejada.

De posse disso, e usando o item *c*), temos que

$$\begin{aligned}
 \|F(z)\| &= \|F(z) - F(x^*)\| \\
 &\leq L_F \|z - x^*\| \\
 &\leq 4L_F L_J \|J_F(x^*)^{-1}\| \|y - x^*\|^2 \\
 &\leq 16L_F L_J \|J_F(x^*)^{-1}\|^3 \|F(y)\|^2, \\
 &\leq 16L_F L_J M^3 \|F(y)\|^2.
 \end{aligned}$$

□

Para as próximas considerações vamos definir $\tilde{F} : \mathbb{R}^{n+m+n+m} \rightarrow \mathbb{R}^{n+m}$ por:

$$\tilde{F}(z, \mu, y, \lambda) = \begin{pmatrix} \nabla_x L(z, \lambda) + J_h(y)^T (\mu - \lambda) \\ J_h(y)(z - y) \end{pmatrix}.$$

Além disso vamos denotar $F_k(z, \mu) = \tilde{F}(z, \mu, y_k, \lambda_k)$ e supor que x^* é uma solução do problema original (4.1), que $J_h(x^*)$ tem posto completo e que x^* é interior a Ω . Nessas condições existe $\lambda^* \in \mathbb{R}^m$ tal que $\nabla_x L(x^*, \lambda^*) = 0$, portanto temos que $\tilde{F}(x^*, \lambda^*, x^*, \lambda^*) = 0$. Como x^* é interior existe $\delta_1 > 0$ tal que $B(x^*, 3\delta_1) \subset \Omega$.

Outra suposição que faremos até o fim desta seção é que valha também a condição suficiente de segunda ordem em x^* , ou seja, que $d^T \nabla_{xx}^2 L(x^*, \lambda^*) d > 0$ para todo $d \neq 0$ tal que $J_h(x^*) d = 0$. Neste caso temos que a matriz

$$\begin{pmatrix} \nabla_{xx}^2 L(x^*, \lambda^*) & J_h(x^*)^T \\ J_h(x^*) & 0 \end{pmatrix}$$

é não singular.

O próximo resultado mostra que perto de (x^*, λ^*) os sistemas $F_k(z, \mu) = 0$ tem solução.

Lema 5.2.2. *Existe $\epsilon_1 > 0$ tal que se $\|(y_k, \lambda_k) - (x^*, \lambda^*)\| \leq \epsilon_1$ então o sistema $F_k(z, \mu) = 0$ tem solução (z_k^*, μ_k^*) , onde $z_k^* = z(y_k, \lambda_k)$ e $\mu_k^* = \mu(y_k, \lambda_k)$ são funções contínuas.*

Demonstração: Primeiramente observemos que

$$J_{\tilde{F}_{z,\mu}}(z, \mu, y, \lambda) = \begin{pmatrix} \nabla_{xx}^2 L(z, \lambda) & J_h(y)^T \\ J_h(y) & 0 \end{pmatrix},$$

e que portanto $J_{\tilde{F}_{z,\mu}}(x^*, \lambda^*, x^*, \lambda^*)$ é não singular. Como $\tilde{F}(x^*, \lambda^*, x^*, \lambda^*) = 0$ o Teorema da Função Implícita garante que existe $\epsilon_1 > 0$ tal que para cada (y, λ) tal que $\|(y, \lambda) - (x^*, \lambda^*)\| \leq \epsilon_1$ existe um único par $(z(y, \lambda), \mu(y, \lambda))$ tal que $\|(z(y, \lambda), \mu(y, \lambda)) - (x^*, \lambda^*)\| \leq \epsilon_1$ e $\tilde{F}(z(y, \lambda), \mu(y, \lambda), y, \lambda) = 0$. Além disso as funções $z(y, \lambda)$ e $\mu(y, \lambda)$ são contínuas. Denotando $(y, \lambda) = (y_k, \lambda_k)$ concluímos a demonstração.

□

O próximo teorema mostra que perto de (x^*, λ^*) os sistemas $F_k(z, \mu) = 0$ se enquadram no teorema de convergência quadrática de Newton com constantes que independem de k . Nele é importante observar que $J_{F_k}(z, \mu) = J_{\tilde{F}_{z,\mu}}(z, \mu, y_k, \lambda_k)$.

Teorema 5.2.4. *Suponhamos que $\|(y_k, \lambda_k) - (x^*, \lambda^*)\| \leq \epsilon_1$, onde ϵ_1 é dado pelo Lema 5.2.2. Então existem constantes positivas ϵ_2 , M , L_F e L_J , independentes de k , tais que*

se $\|(z_k^*, \mu_k^*, y_k, \lambda_k) - (x^*, \lambda^*, x^*, \lambda^*)\| \leq \epsilon_2$ valem

$$J_{F_k}(z_k^*, \mu_k^*)^{-1} \leq M,$$

$$\|F_k(z, \mu) - F_k(s, \nu)\| \leq L_F \|(z, \mu) - (s, \nu)\|$$

e

$$\|J_{F_k}(z, \mu) - J_{F_k}(s, \nu)\| \leq L_J \|(z, \mu) - (s, \nu)\|,$$

sempre que (z, μ) e (s, ν) estejam na bola fechada $\bar{B}((z_k^*, \mu_k^*), \epsilon_2)$.

Demonstração: Pela continuidade de $J_{\tilde{F}_{z,\mu}}$, existe $\epsilon_2 > 0$ tal que $J_{\tilde{F}_{z,\mu}}(z, \mu, y, \lambda)$ é não singular para todo $\|(z, \mu, y, \lambda) - (x^*, \lambda^*, x^*, \lambda^*)\| \leq 2\epsilon_2$. Pela compacidade de $\bar{B}((x^*, \lambda^*, x^*, \lambda^*), 2\epsilon_2)$ existe $M > 0$ tal que $\|J_{\tilde{F}_{z,\mu}}(z, \mu, y, \lambda)^{-1}\| \leq M$ nessa mesma vizinhança. Em particular $\|J_{F_k}(z_k^*, \mu_k^*)^{-1}\| \leq M$ sempre que

$$\|(z_k^*, \mu_k^*, y_k, \lambda_k) - (x^*, \lambda^*, x^*, \lambda^*)\| \leq \epsilon_2.$$

Consideremos agora que $\|(z_k^*, \mu_k^*, y_k, \lambda_k) - (x^*, \lambda^*, x^*, \lambda^*)\| \leq \epsilon_2$ e que $(z, \mu) \in \bar{B}((z_k^*, \mu_k^*), \epsilon_2)$, dessa forma temos que

$$\begin{aligned} \|(z, \mu, y_k, \lambda_k) - (x^*, \lambda^*, x^*, \lambda^*)\| &\leq \|(z, \mu, y_k, \lambda_k) - (z_k^*, \mu_k^*, y_k, \lambda_k)\| + \\ &\quad + \|(z_k^*, \mu_k^*, y_k, \lambda_k) - (x^*, \lambda^*, x^*, \lambda^*)\| \\ &\leq \|(z, \mu) - (z_k^*, \mu_k^*)\| + \epsilon_2 \\ &\leq 2\epsilon_2. \end{aligned}$$

Como \tilde{F} e $J_{\tilde{F}_{z,\mu}}$ são suaves, a compacidade de $\bar{B}((x^*, \lambda^*, x^*, \lambda^*), 2\epsilon_2)$ também garante que essas funções são Lipschitz contínuas nessa bola. Denotando respectivamente por $L_F > 0$ e $L_J > 0$ as constantes de Lipschitz de \tilde{F} e $J_{\tilde{F}_{z,\mu}}$ temos que:

$$\begin{aligned} \|F_k(z, \mu) - F_k(s, \nu)\| &= \|\tilde{F}(z, \mu, y_k, \lambda_k) - \tilde{F}(s, \nu, y_k, \lambda_k)\| \\ &\leq L_F \|(z, \mu, y_k, \lambda_k) - (s, \nu, y_k, \lambda_k)\| \\ &\leq L_F \|(z, \mu) - (s, \nu)\| \end{aligned}$$

e analogamente

$$\|J_{F_k}(z, \mu) - J_{F_k}(s, \nu)\| \leq L_J \|(z, \mu) - (s, \nu)\|,$$

sempre que (z, μ, y_k, λ_k) e (s, ν, y_k, λ_k) estão em $\bar{B}((x^*, \lambda^*, x^*, \lambda^*), 2\epsilon_2)$. Sendo assim, como caso particular, os resultados são válidos sempre que (z, μ) e (s, ν) estejam na bola fechada $\bar{B}((z_k^*, \mu_k^*), \epsilon_2)$ e $\|(z_k^*, \mu_k^*, y_k, \lambda_k) - (x^*, \lambda^*, x^*, \lambda^*)\| \leq \epsilon_2$, o que conclui a demonstração. □

Combinando os Teoremas 5.2.3 e 5.2.4 concluímos que valem uniformemente os resultados de convergência quadrática do método de Newton para os sistemas $F_k(z, \mu) = 0$.

Corolário 5.2.1. *Sejam $\epsilon_1, \epsilon_2, M, L_F$ e L_J como no Teorema 5.2.4. Se*

$$\|(y_k, \lambda_k) - (x^*, \lambda^*)\| \leq \epsilon_1,$$

$$\|(z_k^*, \mu_k^*, y_k, \lambda_k) - (x^*, \lambda^*, x^*, \lambda^*)\| \leq \epsilon_2$$

e

$$\|(y_k, \lambda_k) - (z_k^*, \mu_k^*)\| \leq \min \left\{ \frac{1}{8L_J M}, \epsilon_2 \right\}$$

valem:

a) $J_{F_k}(y_k, \lambda_k)$ é invertível e portanto a iteração de Newton

$$\begin{pmatrix} z_k \\ \mu_k \end{pmatrix} = \begin{pmatrix} y_k \\ \lambda_k \end{pmatrix} - J_{F_k}(y_k, \lambda_k)^{-1} F(y_k, \lambda_k)$$

está bem definida.

b) $\|J_{F_k}(y_k, \lambda_k)^{-1}\| \leq 2M$.

c) $\|(z_k, \mu_k) - (z_k^*, \mu_k^*)\| \leq 4L_J M \|(y_k, \lambda_k) - (z_k^*, \mu_k^*)\|^2$.

d) $\|F_k(z_k, \mu_k)\| \leq 16L_F L_J M^3 \|F_k(y_k, \lambda_k)\|^2$.

Demonstração: Nas vizinhanças consideradas pelo corolário o Teorema 5.2.4 garante que as hipóteses do Teorema 5.2.3 são satisfeitas e portanto os resultados desejados seguem diretamente da tese do Teorema 5.2.3 para $F = F_k$, $y = (y_k, \lambda_k)$ e $x^* = (z_k^*, \mu_k^*)$. \square

A continuidade das funções $z_k^*(y, \lambda)$, $\mu_k^*(y, \lambda)$ e $\tilde{F}(z, \mu, y, \lambda)$ asseguram a validade das proposições a seguir.

Proposição 5.2.1. *Existe $0 < \delta_2 \leq \epsilon_1$ tal que $\|(y_k, \lambda_k) - (x^*, \lambda^*)\| \leq \delta_2$ implica que*

$$\|(z_k^*, \mu_k^*) - (x^*, \lambda^*)\| \leq \frac{1}{2} \min \left\{ \epsilon_2, \frac{1}{8L_J M} \right\}. \quad (5.34)$$

Proposição 5.2.2. *Dado $\eta \in (0, 1)$ existe $\delta_3 > 0$ tal que $\|(z, \mu, y, \lambda) - (x^*, \lambda^*, x^*, \lambda^*)\| \leq \delta_3$ implica que*

$$\tilde{F}(z, \mu, y, \lambda) \leq \min \left\{ \delta_1, \frac{\delta_1}{4M}, \frac{\eta}{16L_F L_J M^3} \right\}. \quad (5.35)$$

O próximo lema mostra que para (y_k, λ_k) perto de (x^*, λ^*) as distâncias entre os pontos de interesse no estudo de uma iteração de Newton para o sistema $F_k(z, \mu) = 0$ estão controladas. Com isso poderemos usar o Corolário 5.2.1 e a Proposição 5.2.2 nessas condições.

Lema 5.2.3. *Se $\|(y_k, \lambda_k) - (x^*, \lambda^*)\| \leq \epsilon \equiv \min \left\{ \frac{1}{16L_J M}, \epsilon_1, \frac{1}{2}\epsilon_2, \frac{\delta_1}{2}, \delta_2, \frac{\sqrt{2}}{2}\delta_3 \right\}$ então*

a) $\|(y_k, \lambda_k, y_k, \lambda_k) - (x^*, \lambda^*, x^*, \lambda^*)\| \leq \delta_3$

b) $\|(z_k^*, \mu_k^*, y_k, \lambda_k) - (x^*, \lambda^*, x^*, \lambda^*)\| \leq \epsilon_2$.

c) $\|(y_k, \lambda_k) - (z_k^*, \mu_k^*)\| \leq \min \left\{ \frac{1}{8L_J M}, \epsilon_2 \right\}$.

d) $\|z_k - x^*\| \leq \delta_1$.

e) $\|F_k(z_k, \mu_k)\| \leq \eta \|F_k(y_k, \lambda_k)\|$.

Demonstração: Por hipótese $\|(y_k, \lambda_k) - (x^*, \lambda^*)\| \leq \frac{\sqrt{2}}{2}\delta_3$. Para demonstrar o item a), basta observar que

$$\|(y_k, \lambda_k, y_k, \lambda_k) - (x^*, \lambda^*, x^*, \lambda^*)\|^2 = \|(y_k, \lambda_k) - (x^*, \lambda^*)\|^2 + \|(y_k, \lambda_k) - (x^*, \lambda^*)\|^2 \leq 2 \frac{2\delta_3^2}{4}.$$

Para o item b) notamos primeiramente que $\|(y_k, \lambda_k) - (x^*, \lambda^*)\| \leq \epsilon_1$ e o Lema 5.2.2 garantem a existência de (z_k^*, μ_k^*) . Usando agora a hipótese de que $\|(y_k, \lambda_k) - (x^*, \lambda^*)\| \leq \min\{\delta_2, \frac{1}{2}\epsilon_2\}$ e a Proposição 5.2.1, a qual garante que $\|(z_k^*, \mu_k^*) - (x^*, \lambda^*)\| \leq \frac{\epsilon_2}{2}$, temos que

$$\begin{aligned} \|(z_k^*, \mu_k^*, y_k, \lambda_k) - (x^*, \lambda^*, x^*, \lambda^*)\|^2 &= \|(z_k^*, \mu_k^*) - (x^*, \lambda^*)\|^2 + \|(y_k, \lambda_k) - (x^*, \lambda^*)\|^2 \\ &\leq \frac{\epsilon_2^2}{4} + \frac{\epsilon_2^2}{4} \leq \epsilon_2^2. \end{aligned}$$

Para demonstrar o item c) notemos que pelo fato de que $\|(y_k, \lambda_k) - (x^*, \lambda^*)\| \leq \delta_2$ a Proposição 5.2.1 assegura que $\|(z_k^*, \mu_k^*) - (x^*, \lambda^*)\| \leq \frac{1}{2} \min\left\{\frac{1}{8L_J M}, \epsilon_2\right\}$. Como por hipótese também vale que $\|(y_k, \lambda_k) - (x^*, \lambda^*)\| \leq \frac{1}{2} \min\left\{\frac{1}{8L_J M}, \epsilon_2\right\}$ a desigualdade triangular garante o resultado desejado:

$$\|(y_k, \lambda_k) - (z_k^*, \mu_k^*)\| \leq \|(y_k, \lambda_k) - (x^*, \lambda^*)\| + \|(z_k^*, \mu_k^*) - (x^*, \lambda^*)\| \leq 2\frac{1}{2} \min\left\{\frac{1}{8L_J M}, \epsilon_2\right\}.$$

Pelo fato de que $\|(y_k, \lambda_k) - (x^*, \lambda^*)\| \leq \epsilon_1$ e pelos itens b) e c) mostramos que as hipóteses do Corolário 5.2.1 são satisfeitas, dessa forma temos a boa definição de (z_k, μ_k) . Ainda pelo Corolário 5.2.1 temos que $\|J_{F_k}(y_k, z_k)^{-1}\| \leq 2M$. Já o item a) assegura que $(y_k, \lambda_k, y_k, \lambda_k)$ cumpre as condições da Proposição 5.2.2, portanto $\|F_k(y_k, \lambda_k)\| \leq \frac{\delta_1}{4M}$. Dessa forma, como $\|y_k - x^*\| \leq \|(y_k, \lambda_k) - (x^*, \lambda^*)\| \leq \frac{\delta_1}{2}$, temos que:

$$\|z_k - x^*\| \leq \|z_k - y_k\| + \|y_k - x^*\| \leq \|J_{F_k}(y_k, \lambda_k)^{-1} F_k(y_k, \lambda_k)\| + \frac{\delta_1}{2} \leq 2M \frac{\delta_1}{4M} + \frac{\delta_1}{2} \leq \delta_1.$$

Portanto o item d) também está demonstrado

Mais uma vez pelo Corolário 5.2.1 temos que

$$\|F_k(z_k, \mu_k)\| \leq 16L_F L_J M^3 \|F_k(y_k, \lambda_k)\|^2.$$

Pelo item a) e a Proposição 5.2.2 temos que $\|F_k(y_k, \lambda_k)\| \leq \frac{\eta}{16L_F L_J M^3}$, e dessa forma a demonstração está concluída.

□

O próximo teorema conclui que, sob as hipóteses anteriores, a fase de otimização da iteração do Algoritmo Local de Birgin-Martínez está bem definida numa vizinhança de (x^*, λ^*) .

Teorema 5.2.5. *Se $\|(y_k, \lambda_k) - (x^*, \lambda^*)\| \leq \epsilon$ então*

a) $J_h(y_k)(z_k - y_k) = 0.$

b) $\|P_\Omega [z_k - \nabla_x L(z_k, \lambda_k) - J_h(y_k)^T(\mu_k - \lambda_k)] - z_k\| \leq \eta \|G(y_k, \lambda_k)\|.$

c) *Existe $K_3 > 0$ tal que $\|z_k - y_k\| + \|\mu_k - \lambda_k\| \leq K_3 \|G(y_k, \lambda_k)\|.$*

Demonstração: Como as equações $J_h(y_k)(z - y_k)$ são lineares em relação à (z, μ) o Método de Newton assegura que estas equações serão nulas para $z = z_k$. Portanto o item a) está demonstrado.

Como para $z = y_k$ também temos que as equações $J_h(y_k)(z - y_k)$ são nulas, temos que $\|F_k(z_k, \mu_k)\| = \|\nabla_x L(z_k, \lambda_k) + J_h(y_k)^T(\mu_k - \lambda_k)\|$ e $\|F_k(y_k, \lambda_k)\| = \|\nabla_x L(y_k, \lambda_k)\|.$

Pelos itens a) e e) do Lema 5.2.3, a Proposição 5.2.2, e o fato de que $\eta \in (0, 1)$ temos que

$$\|F_k(z_k, \mu_k)\| \leq \eta \|F_k(y_k, \lambda_k)\| \leq \|F_k(y_k, \lambda_k)\| \leq \delta_1.$$

Pelo item d) do Lema 5.2.3 também temos que $\|z_k - x^*\| \leq \delta_1$. Dessa forma

$$\|z_k - \nabla_x L(z_k, \lambda_k) - J_h(y_k)^T(\mu_k - \lambda_k) - x^*\| \leq \|z_k - x^*\| + \|F_k(z_k, \mu_k)\| \leq 2\delta_1$$

Por outro lado, a hipótese também garante que $\|y_k - x^*\| \leq \delta_1$, já que $\|y_k - x^*\| \leq \|(y_k, \lambda_k) - (x^*, \lambda^*)\| \leq \frac{\delta_1}{2}$. Portanto

$$\|y_k - \nabla_x L(y_k, \lambda_k) - x^*\| \leq \|y_k - x^*\| + \|F_k(y_k, \lambda_k)\| \leq 2\delta_1.$$

Sendo assim tanto $z_k - \nabla_x L(z_k, \lambda_k) - J_h(y_k)^T(\mu_k - \lambda_k)$ quanto $y_k - \nabla_x L(y_k, \lambda_k)$ são interiores a Ω .

Como a projeção coincide com o operador identidade quando aplicada em pontos interiores, temos que

$$\begin{aligned} \|P_{\Omega} [z_k - \nabla_x L(z_k, \lambda_k) - J_h(y_k)^T(\mu_k - \lambda_k)] - z_k\| &= \|\nabla_x L(z_k, \lambda_k) + J_h(y_k)^T(\mu_k - \lambda_k)\| \\ &= \|F_k(z_k, \mu_k)\| \\ &\leq \eta \|F_k(y_k, \lambda_k)\| \\ &\leq \eta \|G(y_k, \lambda_k)\|. \end{aligned}$$

Desse modo o item b) está demonstrado.

Para o item c) observemos que

$$\begin{pmatrix} z_k - y_k \\ \mu_k - \lambda_k \end{pmatrix} = -J_{F_k}(y_k, \lambda_k)^{-1} F_k(y_k, \lambda_k).$$

Logo, pelo item b) do Corolário 5.2.1,

$$\left\| \begin{pmatrix} z_k - y_k \\ \mu_k - \lambda_k \end{pmatrix} \right\| \leq \|J_{F_k}(y_k, \lambda_k)^{-1}\| \|F_k(y_k, \lambda_k)\| \leq 2M \|G(y_k, \lambda_k)\|. \quad (5.36)$$

Pela equivalência das normas em \mathbb{R}^{n+m} existe $c > 0$ tal que:

$$\|z_k - y_k\| + \|\mu_k - \lambda_k\| \leq c \left\| \begin{pmatrix} z_k - y_k \\ \mu_k - \lambda_k \end{pmatrix} \right\|. \quad (5.37)$$

Portanto, para $K_3 = 2cM$, as relações (5.36) e (5.37) implicam no resultado desejado. □

Sob as hipóteses correspondentes tanto a restauração sugerida por Martínez em [78] quanto a sugerida por Fischer e Friedlander em [53] satisfazem as condições exigidas pela restauração do método local. Mais do que isso, ambas as restaurações asseguram que o ponto restaurado está controladamente afastado da solução. Assim sendo, caso a fase de otimização esteja bem definida numa vizinhança, é possível garantir que o ponto restaurado permanece nessa vizinhança para iterandos suficientemente

próximos da solução. Vejamos estes fatos.

Primeiramente vejamos a restauração sugerida por Martínez em [78]. Caso todos os pontos de Ω sejam regulares, pelo Lema 5.1.3, temos que dado $r \in (0, 1)$ arbitrário, existem constantes positivas δ_4 e K_1 tais que se $\|x_k - x^*\| \leq \delta_4$ então

$$T(x_k) = \{y \in \Omega; J_h(x_k)(y - x_k) + h(x_k) = 0\} \neq \emptyset.$$

Nesse caso, definindo $y_k = \operatorname{argmin}\{\|y - x_k\|; y \in T(x_k)\}$, temos que

$$\begin{aligned} \|h(y_k)\| &\leq r\|h(x_k)\| \\ \|y_k - x_k\| &\leq K_1\|h(x_k)\|. \end{aligned}$$

Dessa forma as condições (5.17) e (5.18) estão satisfeitas. Pela continuidade de $h(x)$ temos que existe $0 < \delta_5 \leq \delta_4$ tal que $\|h(x_k)\| \leq \frac{\epsilon}{4K_1}$ sempre que $\|x_k - x^*\| \leq \delta_5$. Dessa forma, se $\|x_k - x^*\| \leq \delta \equiv \min\{\delta_5, \frac{\epsilon}{4}\}$ temos que

$$\|y_k - x^*\| \leq \|y_k - x_k\| + \|x_k - x^*\| \leq K_1\|h(x_k)\| + \frac{\epsilon}{4} \leq \frac{\epsilon}{2}. \quad (5.38)$$

Já a restauração sugerida por Fischer e Friedlander em [53] consiste em encontrar o ponto viável mais próximo de x_k . Uma condição suficiente para que esta restauração esteja bem definida é que a condição de Mangasarian-Fromovitz seja verificada em todos os pontos de Ω e que o conjunto viável seja não vazio. Para esta restauração condições (5.17) e (5.18) também estão satisfeitas, conforme demonstrado no Teorema 5.1.1.

Nesse caso a relação $\|y_k - x^*\| \leq \frac{\epsilon}{2}$ também pode ser obtida, se supormos que $\|x_k - x^*\| \leq \frac{\epsilon}{4}$. Isso é fácil de ver pois como o ponto solução é viável essa restauração garante que $\|y_k - x_k\| \leq \|x_k - x^*\|$ portanto:

$$\|y_k - x^*\| \leq \|y_k - x_k\| + \|x_k - x^*\| \leq 2\|x_k - x^*\| \leq \frac{\epsilon}{2}.$$

Teorema 5.2.6. *Se $\|x_k - x^*\| \leq \delta$ e $\|\lambda_k - \lambda^*\| \leq \frac{\sqrt{3}}{2}\epsilon$, então a iteração local está bem definida.*

Demonstração: Como vimos a restauração garante as condições (5.17) e (5.18). Para a fase de otimização basta observar que se $\|x_k - x^*\| \leq \delta$ então $\|y_k - x^*\| \leq \frac{\epsilon}{2}$, dessa forma

$$\|(y_k, \lambda_k) - (x^*, \lambda^*)\|^2 = \|y_k - x^*\|^2 + \|\lambda_k - \lambda^*\|^2 \leq \frac{1}{4}\epsilon^2 + \frac{3}{4}\epsilon^2 \leq \epsilon^2.$$

Portanto o Teorema 5.2.5 assegura as condições (5.19), (5.20) e (5.21). Sendo assim a boa definição está demonstrada. □

Resumindo então os últimos resultados vamos enunciar o seguinte teorema.

Teorema 5.2.7. *Suponha que exista um ponto crítico $(\bar{x}, \bar{\lambda})$ interior a Ω , que a Hessiana do Lagrangiano seja definida positiva neste ponto e que o Jacobiano das restrições neste ponto tenha posto completo. Então existe uma vizinhança de $(\bar{x}, \bar{\lambda})$ em que a iteração local está bem definida.*

Em [29] os autores sugerem resolver a parte da viabilidade encontrando y_k como uma solução aproximada de

$$\begin{aligned} & \underset{y}{\text{Minimizar}} \quad \|h(y)\|^2 \\ & \text{sujeito a:} \quad \|y - x_k\|_\infty \leq K_1 \|h(x_k)\|, \\ & \quad \quad \quad y \in \Omega. \end{aligned} \tag{5.39}$$

Para a parte da otimalidade z_k é escolhido como uma solução aproximada de

$$\begin{aligned} & \underset{z}{\text{Minimizar}} \quad L(z, \lambda_k) \\ & \text{sujeito a:} \quad J_h(y_k)(z - y_k) = 0, \\ & \quad \quad \quad \|z - y_k\|_\infty \leq \tilde{K}_3 \max\{1, \|y_k\|_\infty\}, \\ & \quad \quad \quad z \in \Omega, \end{aligned} \tag{5.40}$$

e $\mu_k = \lambda_k + \nu_k$, onde ν_k é o vetor dos multiplicadores de Lagrange associados a z_k para o problema (5.40).

Motivados pela demonstração dos resultados que garantem a boa definição do método local vamos propor uma nova alternativa de satisfazer as condições da iteração.

Para a fase de restauração encontramos y_k resolvendo inexatamente o sub-problema:

$$\begin{aligned} & \underset{y}{\text{Minimizar}} \quad \frac{1}{2} \|y - x_k\|^2 \\ & \text{sujeito a:} \quad h(y) = 0, \\ & \quad \quad \quad y \in \Omega. \end{aligned} \tag{5.41}$$

Como primeira iteração do subproblema (5.41) tentamos resolver

$$\begin{aligned} & \underset{y}{\text{Minimizar}} \quad \frac{1}{2} \|y - x_k\|^2 \\ & \text{sujeito a:} \quad J_h(x_k)(y - x_k) = -h(x_k), \end{aligned} \tag{5.42}$$

e depois projetamos a solução encontrada em Ω . A motivação para essa primeira tentativa é que ela está relacionada com a restauração proposta por Martínez em [78]. Caso $J_h(x_k)$ tenha posto completo a resolução de (5.42) pode ser feita de duas maneiras clássicas. A primeira delas é resolvendo o problema de ponto sela resultante do sistema KKT:

$$\begin{pmatrix} I & J_h(x_k)^T \\ J_h(x_k) & 0 \end{pmatrix} \begin{pmatrix} y - x_k \\ \bar{\lambda} \end{pmatrix} = \begin{pmatrix} 0 \\ -h(x_k) \end{pmatrix}.$$

A segunda maneira seria considerar uma fatoração

$$J_h(x_k)^T = Q \begin{pmatrix} R \\ 0 \end{pmatrix},$$

onde Q é uma matriz ortogonal $n \times n$ e R é uma matriz triangular superior $m \times m$ e tomar

$$y = x_k + Q \begin{pmatrix} -R^{-T}h(x_k) \\ 0 \end{pmatrix}.$$

Ambas as alternativas tem vantagens e desvantagens e devem ser escolhidas convenientemente de acordo com as dimensões do problema e a esparsidade de $J_h(x_k)$.

Quanto ao sentido de inexatidão na resolução do problema (5.41) consideramos o parâmetro $\bar{r} \in (0, 1)$, tal que $\bar{r} \approx 1$ e \bar{r} é independente de r , e colocamos como critério de parada que $h(y) \leq \bar{r}h(x_k)$. A ideia aqui em não usar o próprio r no critério de parada é que esperamos que perto da solução a restauração segundo Martínez seja realizada. Nesse caso a condição da restauração vai ser automaticamente satisfeita para qualquer r desde que estejamos suficientemente próximos da solução. Caso não seja possível encontrar solução de (5.42), ou a projeção da solução em Ω não satisfaça o critério de parada, usamos um método com convergência global, por exemplo Algen-can, para resolver (5.41) com esse critério de parada. Caso o método escolhido para a restauração também falhe o algoritmo declara o problema como infactível.

Para a parte da otimalidade z_k é escolhido como uma solução de

$$\begin{aligned} & \underset{z}{\text{Minimizar}} \quad \frac{1}{2}(z - y_k)^T \nabla_{xx}^2 L(y_k, \lambda_k)(z - y_k) + \nabla_x L(y_k, \lambda_k)(z - y_k) \\ & \text{sujeito a:} \quad J_h(y_k)(z - y_k) = 0, \\ & \quad \quad \quad z \in \Omega, \end{aligned} \tag{5.43}$$

e $\mu_k = \lambda_k + \nu_k$, onde ν_k é o vetor dos multiplicadores de Lagrange associados a z_k para o problema (5.43).

Uma observação importante é que, desprezando as restrições relativas ao conjunto Ω , a solução do problema (5.43) é a solução do sistema linear:

$$\begin{pmatrix} \nabla_{xx}^2 L(y_k, \lambda_k) & J_h(y_k)^T \\ J_h(y_k) & 0 \end{pmatrix} \begin{pmatrix} z - y_k \\ \mu - \lambda_k \end{pmatrix} = \begin{pmatrix} -\nabla_x L(y_k, \lambda_k) \\ 0 \end{pmatrix}. \tag{5.44}$$

Este sistema é exatamente o mesmo que usamos para obter (z_k, μ_k) a partir de uma iteração do Método de Newton para o sistema $F_k(z, \mu)$ no ponto (y_k, λ_k) . Isso faz com que nossa alternativa para satisfazer as condições da iteração do método local seja condizente com a teoria que apresentamos.

Antes de terminarmos essa seção vamos destacar também que nas condições sobre as quais demonstramos a boa definição da iteração do método local a convergência é superlinear. O ponto chave aqui é que a demonstração da boa definição da iteração local pode ser mostrada para qualquer que sejam os parâmetros $r > 0$ e $\eta > 0$ e, por outro lado, a construção do ponto que satisfaz a relação independe desses parâmetros. O próximo teorema formaliza essa observação.

Teorema 5.2.8. *Suponhamos que x^* seja uma solução do problema original (4.1) associada ao multiplicador λ^* , que x^* seja interior a Ω , que $J_h(x^*)$ tenha posto completo e que valham as condições de segunda ordem. Suponhamos também que a sequência (x_k, λ_k) é construída como solução do problemas (5.41) e (5.43) da maneira sugerida e que (x_k, λ_k) converge a (x^*, λ^*) . Então a convergência é superlinear.*

Demonstração: Como (x^*, λ^*) é um ponto interior a Ω a partir de certo momento todos os iterandos estarão no interior de Ω . Dessa forma podemos descartar as restrições associadas a Ω nos problemas (5.41) e (5.43) para k suficientemente grande. Consideremos agora duas sequências arbitrárias de termos positivos $\{r_k\}$ e $\{\eta_k\}$ não crescentes e convergentes a zero. Pelo Teorema 5.2.6 existe uma vizinhança de (x^*, λ^*) para a qual a iteração local com parâmetros r_0 e η_0 pode ser obtida como solução de (5.42) e (5.44). Como (x_k, λ_k) converge a (x^*, λ^*) temos que existe k_0 tal que para $k > k_0$ os iterandos estão nessa vizinhança. Ainda como consequência da convergência de (x_k, λ_k) podemos redefinir $\{r_k\}$ e $\{\eta_k\}$ inserindo um número finito de termos entre dois termos da sequência original de forma que valha que $(x_{k+1}, \lambda_{k+1}) = N_k(x_k, \lambda_k)$. Dessa forma ainda teríamos que $\{r_k\}$ e $\{\eta_k\}$ são não crescente e convergentes a zero, portanto o item 5 do Teorema 5.2.2 garante a convergência superlinear. \square

5.3 Estratégia híbrida combinando Fischer-Friedlander e Birgin-Martínez

Infelizmente, na Seção 5.1 não conseguimos provar que nossa modificação do Algoritmo global Fischer-Friedlander está livre do efeito Maratos. Pensando nisso resolvemos propor uma estratégia *whatchdog* combinando os Algoritmos Fischer-Friedlander e Birgin-Martínez de modo a conseguir um algoritmo global com convergência superlinear ou quadrática. Vamos chamar o algoritmo proveniente desta estratégia de Algoritmo global-local FF-BM. A ideia principal deste algoritmo é a seguinte: enquanto o passo local Birgin-Martínez estiver sendo bem sucedido fique com ele, caso ele fracasse, dê um passo conservador do algoritmo Fischer-Friedlander, para o qual a convergência global está garantida. O Algoritmo híbrido FF-BM está descrito a seguir.

Algoritmo FF-BM Sejam $r, \bar{\theta} \in [0, 1)$, $\beta, \gamma, K_1, K_2, K_3 > 0$, constantes algorítmicas, $r_s \in [0, 1)$, $\eta_s \in [0, 1)$ sequências de parâmetros positivos convergentes a zero e $\Omega_\lambda \subset \mathbb{R}^m$ um compacto. Defina também, como no método local,

$$G(x, \lambda) = P_\Omega(x - \nabla_x L(x, \lambda)) - x, \quad \text{e} \quad R(x, \lambda) = \max\{\|h(x)\|, \|G(x, \lambda)\|\}$$

e, como no método global,

$$\Phi(x, \lambda, \rho) = \rho L(x, \lambda) + (1 - \rho)\|h(x)\|.$$

Passo 0: Escolha $x_0 \in \Omega$, $\lambda_0 \in \Omega_\lambda$, $\rho_0 \in (0, 1)$. Faça $k = 0$ e $s = 0$.

Passo 1: Tente construir $(z_{s+1}, \mu_{s+1}) = N(z_s, \mu_s)$ por meio de uma iteração do Algoritmo local Birgin-Martínez com parâmetros $K_1, K_2, K_3, r_s, \eta_s$.

- a) Caso o método local falhe ou $R(z_{s+1}, \mu_{s+1}) > \bar{\theta}R(z_s, \mu_s)$ faça $s = s + 1$ e vá ao passo 2,
- b) Caso contrário, ou seja se $R(z_{s+1}, \mu_{s+1}) \leq \bar{\theta}R(z_s, \mu_s)$, faça $s = s + 1$ e recomece o passo 1.

Passo 2: Calcule y_k , λ_{k+1} e x_{k+1} usando uma iteração do Algoritmo global Fischer-Friedlander no ponto (x_k, λ_k) com parâmetros r, β, γ e compacto Ω_λ .
 Faça $k = k + 1$, $(z_s, \mu_s) = (x_k, \lambda_k)$ e vá ao passo 1.

Vamos dizer que o Algoritmo FF-BM está bem definido se, ou ele gera uma sequência infinita $\{(z_s, \mu_s)\}$ sendo executado exclusivamente o passo 1 a partir de certo momento, ou ele gera uma sequência infinita $\{(x_k, \lambda_k)\}$. Deste modo, para mostrar que ele está bem definido basta observar, se tivermos um mecanismo que acuse falha em uma iteração do método local em tempo finito, caso ele não seja bem sucedido antes, teremos a boa definição. Isso acontece pois, neste caso, o passo 1 seria executado em tempo finito e o passo 2 está associado ao método global, para o qual já mostramos a boa definição. No caso padrão de tentarmos conseguir uma iteração do método local resolvendo (5.39) e (5.40) existem vários algoritmos que exibem essa garantia de terminação em tempo finito.

Os teoremas de convergência do método híbrido global-local se dividem em duas partes, se a partir de um momento sempre há sucesso nas iterações locais ou se existem infinitas falhas no passo 1. No primeiro caso temos o resultado a seguir

Teorema 5.3.1. *Suponhamos que o algoritmo esteja bem definido e gere apenas um número finito de termos da sequência $\{x_k\}$. Então valem todas as propriedades do método local, em particular que as sequências $\{z_s\}$ e $\{\mu_s\}$ são R -linearmente convergentes para z_* e μ_* . Mais ainda, se os parâmetros r_k e η_k tendem a zero então $R(x_k, \lambda_k)$ converge Q -superlinearmente, $\{x_k\}$ e $\{\lambda_k\}$ convergem R -superlinearmente; e se r_k e η_k*

são nulos então $R(x_k, \lambda_k)$ converge Q -quadraticamente e $\{x_k\}$ e $\{\lambda_k\}$ convergem R -quadraticamente.

Demonstração: Como apenas um número finito de termos da sequência $\{y_k\}$ é gerado temos que o passo 1 é realizado indefinidamente a partir de certo momento. Nesse caso é como se tivéssemos aplicado o método local com ponto inicial sendo o último par (x_k, λ_k) gerado pelo algoritmo global. Como o passo 1 só é continuado se $R(z_{s+1}, \mu_{s+1}) \leq \bar{\theta}R(z_s, \mu_s)$ temos que qualquer ponto limite da sequência $\{(z_s, \mu_s)\}$ é crítico. Dessa forma, caso exista algum ponto limite de $\{(z_s, \mu_s)\}$, e existe no caso em que Ω é uma caixa limitada, haverá um iterando da sequência que estaria numa vizinhança onde vale o Teorema 5.2.2 e a partir desse momento a sequência pode ser pensada como uma sequência gerada pelo algoritmo local sob as hipóteses do Teorema 5.2.2, o que garante todos os resultados já demonstrados para este método. \square

No caso em que não podemos garantir que somente o método local é sempre aplicado a partir de certo instante temos resultados de convergência relativos à sequência gerada pelo método global.

Teorema 5.3.2. *Caso o algoritmo global-local gere infinitos termos de uma sequência $\{y_k\}$, então temos que $\{h(y_k)\}$ converge a zero e se y_* for um ponto limite de $\{y_k\}$, então y_* é AKKT. Consequentemente, se as restrições em y_* satisfazem a CPLD, então y_* é KKT.*

Demonstração: Não há o que provar pois neste caso a sequência foi gerada pelo Algoritmo global Fischer-Friedlander e vale o teorema de convergência deste algoritmo. \square

O próximo resultado mostra que, sob certas condições, podemos esperar que o método híbrido apresente a convergência local superlinear.

Teorema 5.3.3. *Suponhamos que existam $\epsilon_1 > 0$ e $\delta_1 > 0$ tais que a iteração local esteja bem definida sempre que $\|x - x_*\| \leq \epsilon_1$ e $\|\lambda - \lambda_*\| \leq \delta_1$ para algum ponto limite regular (x_*, λ_*) do algoritmo global-local. Então, se for possível construir uma sequência $\theta_k > \max\{r_k, c_5\eta_k\}$ tal que $\limsup \theta_k \leq \theta < \bar{\theta}$, teremos que a partir de certo momento apenas iterações do método local serão feitas e portanto herdamos a taxa de convergência garantida pelo Algoritmo local.*

Demonstração: Pelo fato de (x_*, λ_*) ser um ponto limite regular temos que ele é um par crítico. Pelo Teorema 5.2.2 existem $\epsilon_{2,\theta}$ e $\delta_{2,\theta}$ tais que se começarmos o método local em (z_s, μ_s) tais que $\|z_s - x_*\| \leq \epsilon_{2,\theta}$ e $\|\mu_s - \lambda_*\| \leq \delta_{2,\theta}$ então a sequência gerada pelo método local está bem definida. Por outro lado, como $\limsup \theta_s < \bar{\theta}$ temos que a partir de certo momento, que denotaremos por $s \geq s_0$, $\theta_s < \bar{\theta}$, a iteração local será sempre aceita pelo passo 1 caso esteja bem definida. Como (x_*, λ_*) é um ponto limite do algoritmo global-local, existe um iterando (z_s, μ_s) tal que $\|z_s - x_*\| \leq \epsilon_{2,\theta}$ e $s \geq s_0$, portanto a partir desse momento as iterações locais estão bem definidas e serão aceitas pelo passo 1, conseqüentemente apenas iterações do método local serão feitas. Pela hipótese de (x_*, λ_*) já ser um ponto limite de $\{(z_s, \mu_s)\}$ e o fato do Teorema 5.2.2 garantir que a sequência converge, temos que

$$\lim(z_s, \mu_s) = (x_*, \lambda_*).$$

Além disso, as taxas de convergência do método local, tanto para o resíduo, quanto para as sequências dual e primal, relacionadas aos parâmetros utilizados também são válidas □

Observações importantes: Embora nosso teorema de convergência do método local dependa do fato de que $\max\{r_k, c_5\eta_k\} < 1$ e a priori não conhecemos c_5 , as demonstrações feitas em [29] garantem a convergência da sequência apenas supondo $\max\{r_k, \eta_k\} < 1$. Lá também são demonstradas taxas de convergência lineares usando outra norma pra medir o resíduo. Além disso, quando fazemos os parâmetros r_k e η_k tenderem a zero temos que $\lim \theta_k = 0$, portanto $\limsup \theta_k < \bar{\theta}$ mesmo sem conhecermos

c_5 , portanto o algoritmo global-local tem convergência garantida nesse caso. Mais do que isso, nossa proposta de executar o passo local garante a possibilidade de tomar r_k e η_k desse modo. Por fim, como nos Capítulos 1 e 2, é possível substituir a CPLD por condições de qualificação mais fracas (R-CPLD, CRSC, CPG) em todos os resultados que afirmam que pontos AKKT são pontos KKT sob essas condições.

Capítulo 6

Restauração Inexata sem busca linear e otimização sem derivadas

Nessa seção nossa contribuição será propor um novo método de Restauração Inexata que não faça busca linear no passo de otimização. Os resultados deste capítulo fazem parte de um trabalho conjunto com A. Friedlander, J. M. Martínez e F. Sobral. A não ser pelo fato de não fazermos a busca linear, nosso método é baseado nas ideias centrais do Algoritmo de Fischer-Friedlander para resolver o problema alvo (4.1):

$$\begin{aligned} & \text{Minimizar } f(x) \\ & x \in \Omega \subset \mathbb{R}^n \\ & \text{sujeito a: } h(x) = 0, \end{aligned}$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ são funções duas vezes continuamente diferenciáveis e Ω é uma caixa. Nosso intuito em evitar a busca linear é fazer com que o método seja aplicável em problemas onde as derivadas de f não estejam disponíveis, embora existam, e portanto direções de descida para f não sejam conhecidas. Com isso não é possível garantir a boa definição da busca linear.

Nesse caso teremos que a função objetivo é difícil mas as restrições são simples. Como restrições simples estamos supondo funções para as quais existam métodos eficientes de restauração e que J_h esteja disponível. Embora o estudo das propriedades

do nosso método não dependa dessas características, acreditamos que essas são fundamentais para sua eficiência. Em problemas desse tipo esperamos encontrar a direção de otimização utilizando o método GSS (*generating set search*) para restrições lineares de Lewis, Torczon e seus colaboradores [67, 66, 64], que é um método bem estabelecido na literatura na minimização sem o uso de derivadas para problemas com restrições lineares. Por ser chave na solução de nosso subproblema vamos fazer uma breve introdução deste método, que de agora em diante denotaremos por GSS-KLT.

6.1 Método GSS-KLT

Nesta seção vamos fazer uma descrição superficial sobre o método GSS-KLT e exibiremos também algumas das suas propriedades teóricas demonstradas em [64]. Outras propriedades, não apresentadas em [64], que serão necessárias para nosso uso do GSS-KLT na resolução de nossos subproblemas também serão demonstradas aqui.

O problema alvo do GSS-KLT em [64] é:

$$\begin{aligned} & \text{Minimizar } f(x) \\ & x \in \mathbb{R}^n \\ & \text{sujeito a: } Ax \leq b, \end{aligned} \tag{6.1}$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é uma função duas vezes continuamente diferenciável, $A \in \mathbb{R}^{m \times n}$ e $b \in \mathbb{R}^m$ são dados. Não há hipótese do sistema $Ax \leq b$ ser não degenerado, portanto não há necessidade de tratarmos restrições de caixa separadamente, além do fato de que problemas com restrições de igualdade também estão incluídos nessa formulação.

As seguintes definições são feitas em [64] e serão de grande utilidade aqui:

Definições:

1. a_i^T é a i -ésima linha de A .

2. O conjunto das restrições quase ativas, ou seja aquelas cuja fronteira está a uma distância menor do que ϵ do ponto x , será denotado por:

$$I(x, \epsilon) = \{i | a_i^T x \geq b_i - \epsilon \|a_i\|\}.$$

3. O cone tangente às restrições quase ativas em x é dado por:

$$T(x, \epsilon) = \{v | a_i^T v \leq 0, \forall i \in I(x, \epsilon)\}.$$

De posse dessas definições passemos à descrição do Algoritmo GSS-KLT que, muito a grosso modo, consiste em:

1. Dados os parâmetros $0 < \theta < 1$, $\phi \geq 1$, κ^{min} , β^{min} e β^{max} positivos, uma função $\rho : \mathbb{R} \rightarrow \mathbb{R}_+$ contínua, tal que $\lim_{t \rightarrow 0} \frac{\rho(t)}{t} = 0$, um ponto viável x^0 e um tamanho de passo Δ^0 , faça $k = 0$.
2. Faça $\epsilon^k = \beta^{max} \Delta^k$.
3. Construa um conjunto finito G^k de direções de busca d_i^k tais que:

(a) $\beta^{min} \leq \|d_i^k\| \leq \beta^{max}$,

(b) G^k seja um conjunto gerador de $T(x^k, \epsilon^k)$,

(c) $\kappa^{min} \leq \inf_{\substack{v \in \mathbb{R}^n \\ v_k \neq 0}} \max_{d \in G^k} \frac{v^T d}{\|v_k\| \|d\|}$, sempre que $T(x, \epsilon) \neq \{0\}$, onde v_k é a projeção

de v em $T(x^k, \epsilon^k)$.

4. Se existir $x^k + \Delta^k d_i^k$ viável tal que $f(x^k + \Delta^k d_i^k) < f(x^k) + \rho(\Delta^k)$, fazer $x^{k+1} = x^k + \Delta^k d_i^k$ para algum i e $\Delta^{k+1} = \phi \Delta^k$.

Caso contrário, a iteração é chamada de mal sucedida e faça $x^{k+1} = x^k$ e $\Delta^{k+1} = \theta \Delta^k$

5. Fazer $k = k + 1$ e voltar ao passo 2

No método mais geral descrito em [64] é permitido que θ e ϕ dependam de k . É permitido também a avaliação de direções num conjunto de direções heurísticas H^k , o que está fortemente ligado ao melhor desempenho do método, mas não à sua convergência. Vários métodos diferentes, todos eles englobados pelo GSS-KLT, podem ser obtidos para escolhas particulares em vários itens do algoritmo [64]. Estas características especiais, podem ser impostas na geração dos conjuntos G^k e H^k , na atualização do tamanho de passo Δ^k e na atualização do ponto tentativo no caso em que $x^k + \Delta^k d_i^k$ não é viável, e também da função ρ . Com cada uma dessas particularidades, é possível demonstrar (veja [64]) que sob hipóteses adequadas e razoáveis $\liminf \Delta^k = 0$, e com algumas hipóteses a mais que $\lim \Delta^k = 0$.

Este resultado é importante pois garante que o critério de parada $\Delta^k \leq \Delta^{tol}$ será satisfeito em tempo finito, mas seu principal papel vem do Teorema 6.3 de [64] que relaciona Δ^k com a medida de otimalidade $\chi(x^k, \epsilon^k) = \|[-\nabla f(x^k)]_{T(x^k, \epsilon^k)}\|$, a projeção de $-\nabla f(x^k)$ em $T(x^k, \epsilon^k)$.

Teorema 6.1.1. - Teorema 6.3 de [64].

Suponha que o conjunto de nível $L = \{x | Ax \leq b \text{ e } f(x) \leq f(x^0)\}$ seja limitado, e que M seja a constante de Lipschitz de $\nabla f(x)$ num conjunto convexo que contenha L . Então, para toda iteração mal sucedida vale:

$$\|[-\nabla f(x^k)]_{T(x^k, \epsilon^k)}\| \leq \frac{1}{\kappa^{min}} \left(M \Delta^k \beta^{max} + \frac{\rho(\Delta^k)}{\Delta^k \beta^{min}} \right).$$

Demonstração: Veja Teorema 6.3 de [64]. □

Um fato conhecido é que um ponto viável x é KKT de (6.1) se e somente se $\chi(x, 0) = 0$, entretanto $\chi(x, \epsilon)$ não é uma função contínua nas variáveis (x, ϵ) . Pensando

em contornar isso também é usada a medida de otimalidade:

$$\bar{\chi}(x) = \max_{\substack{A(x+w) \leq b \\ \|w\| \leq 1}} -\nabla f(x)^T w.$$

Também é demonstrado em [64] que $\bar{\chi}(x^k) = O(\Delta^k)$ nas iterações mal sucedidas. Este resultado é mais forte que o demonstrado no Teorema 6.1.1, pois, além do fato de que $\bar{\chi}(x, 0) = 0$ se e somente se x é um ponto KKT de (6.1), temos que $\bar{\chi}(x)$ é contínua. Desta forma, temos que um ponto limite gerado pelo GSS-KLT é KKT no caso em que $\liminf \Delta^k = 0$, e que todo ponto limite é KKT no caso em que $\lim \Delta^k = 0$.

Como esperamos utilizar o método GSS-KLT na resolução de nossos subproblemas devemos interrompê-lo em um tempo finito, não esperando que gere uma sequência infinita. Para isso vamos utilizar como critério de parada que Δ^k seja menor que uma tolerância Δ^{tol} , e portanto as propriedades relativas ao iterando final são mais importante que as do ponto limite. Considerando isso, optamos por trabalhar com a medida $\chi(x)$ ao invés de $\bar{\chi}(x)$, uma vez que suas propriedades são suficientes para garantir as condições que necessitamos para nosso algoritmo de Restauração Inexata, além do fato de que ela se assemelha muito mais às condições de otimalidade sequenciais AGP e AKKT [59], com as quais estamos habituados a trabalhar. Os próximos resultados que demonstraremos não estão em [64] e são feitos aqui com o propósito único de mostrar ser adequado o uso do método GSS-KLT na resolução dos subproblemas para encontrar a direção de otimalidade no nosso algoritmo de Restauração Inexata.

Definindo $A_{x,\epsilon}$ como a matriz formada pelas linhas de A que estão quase ativas, ou seja pelas linhas a_i^T com $i \in I(x, \epsilon)$, temos que

$$T(x, \epsilon) = \{y | A_{x,\epsilon} y \leq 0\}.$$

Como $[-\nabla f(x)]_{T(x,\epsilon)}$ é o ponto de $T(x, \epsilon)$ mais próximo de $-\nabla f(x)$ temos que $[-\nabla f(x)]_{T(x,\epsilon)}$

é a solução de:

$$\begin{aligned} & \text{Minimizar } \frac{1}{2} \|y + \nabla f(x)\|^2 \\ & y \in \mathbb{R}^n \end{aligned} \tag{6.2}$$

$$\text{sujeito a: } A_{x,\epsilon} y \leq 0.$$

Como (6.2) é um problema com função objetivo estritamente convexa e conjunto viável convexo temos que y é a única solução de (6.2) se e somente se existe $\lambda \in \mathbb{R}^{\#I(x,\epsilon)}$ tal que valham as condições KKT de (6.2):

$$y = -(\nabla f(x) + A_{x,\epsilon}^T \lambda),$$

$$\lambda \geq 0,$$

$$A_{x,\epsilon} y \leq 0,$$

$$\lambda_i = 0 \text{ se } [a_{x,\epsilon}]_i^T y < 0.$$

Das duas primeiras equações temos que $[-\nabla f(x)]_{T(x,\epsilon)} = -(\nabla f(x) + A_{x,\epsilon}^T \lambda)$ e $\lambda \geq 0$.

Equivalentemente:

$$[-\nabla f(x)]_{T(x,\epsilon)} = -(\nabla f(x) + A^T \lambda) \quad \text{e} \quad \lambda \geq 0,$$

onde $\lambda \in \mathbb{R}^m$ é tal que $\lambda_i = 0$ se $i \notin I(x, \epsilon)$. Este fato deixa clara a afirmação de que um ponto viável x é KKT de (6.1) se e somente se $\chi(x, 0) = 0$, mas também nos diz mais a respeito da sequência de pontos em iterações mal sucedidas. Pelo Teorema 6.1.1 temos que os iterandos em iterações mal sucedidas satisfazem

$$\|\nabla f(x^k) + A_{x^k,\epsilon}^T \lambda\| \leq \frac{1}{\kappa^{min}} \left(M \Delta^k \beta^{max} + \frac{\rho(\Delta^k)}{\Delta^k \beta^{min}} \right),$$

$$\lambda \geq 0,$$

$$\lambda_i = 0 \text{ se } a_i^T x^k < b_i - \epsilon^k \|a_i\|.$$

Neste caso, definindo x^F como o ponto resultante de uma iteração mal sucedida em que $\Delta^k \leq \Delta^{tol}$ e $\bar{\epsilon} = \max\left\{\frac{1}{\kappa^{min}} \left(M \Delta^k \beta^{max} + \frac{\rho(\Delta^k)}{\Delta^k \beta^{min}} \right), \epsilon^k \|A\|_F\right\}$ temos que x^F é $\bar{\epsilon}$ -KKT do problema (6.1).

O último resultado dessa seção é o essencial na nossa análise do algoritmo de Restauração Inexata sem busca linear proposto na seção seguinte, caso pretendamos usar o GSS-KLT na resolução dos subproblemas.

Teorema 6.1.2. *Suponhamos que vamos resolver uma sequência de problemas (P_i) usando o GSS-KLT com tolerância Δ_i^{tol} , obtendo como solução x_i^F , o qual será $\bar{\epsilon}_i$ -KKT de P_i . Caso seja possível obter a mesma constante de Lipschitz M para os gradientes das funções objetivos e a sequência das matrizes das restrições tenha norma de Frobenius limitada por N , é possível fazer com que $\bar{\epsilon}_i$ tenda a zero controlando apenas o critério de parada Δ_i^{tol} .*

Demonstração: Para obter esse resultado basta que implementemos o GSS-KLT com alguns padrões na resolução de cada P_i . Um desses padrões seria utilizar os mesmos parâmetros β^{min} , β^{max} e a mesma função ρ , o que é natural no caso em que os problemas tenham escalas parecidas. Além disso devemos usar também o mesmo parâmetro κ^{min} , o que pode ser feito de maneira simples se as matrizes A são parecidas, ou incluindo um conjunto positivamente gerador de \mathbb{R}^n fixo em todos os conjuntos G^k na resolução de todos os problemas P_i . Com essas particularidades basta impor que o critério de parada Δ_i^{tol} tenda a zero para obtermos que cada um dos termos envolvidos no limitante de $\bar{\epsilon}_i$ vá a zero

$$\bar{\epsilon}_i \leq \max \left\{ \frac{1}{\kappa^{min}} \left(M \Delta^k \beta^{max} + \frac{\rho(\Delta^k)}{\Delta^k \beta^{min}} \right), \beta^{max} \Delta^k N \right\}.$$

□

Observe que para esse fato não precisamos conhecer explicitamente ∇f e $\bar{\epsilon}_i$, apenas é necessário que o critério de parada Δ_i^{tol} tenda a zero.

6.2 Restauração Inexata sem busca linear

Passaremos agora a exposição do nosso algoritmo de Restauração Inexata para resolver o problema original de programação não linear (4.1) sem o uso de busca linear no

cálculo da direção de otimalidade. Lembramos que o problema de interesse aqui acontece quando f é uma função difícil, mas h é uma função fácil, e que esperamos utilizar o método GSS-KLT no cálculo da direção de otimalidade, entretanto esses fatos não são requisitos na nossa análise do método. O objetivo em evitar a busca linear é que a garantia de que ela será bem sucedida em tempo finito está fortemente baseada no fato de termos uma direção de descida. Como no problema de interesse abordado neste capítulo não estão disponíveis as derivadas da função objetivo não é possível garantir que uma direção dada é de descida. Por isso introduzimos uma regularização que substitui a busca linear. Passemos então a descrição do algoritmo:

Algoritmo 6

Considere $\phi(x, \rho) = \rho f(x) + (1 - \rho)\|h(x)\|$. Sejam $r \in [0, 1)$, $\beta, \gamma > 0$ constantes algorítmicas e ϵ_k uma sequência de escalares não negativos.

Passo 0: Escolha $x_0 \in \Omega$, $\rho_0 \in (0, 1)$ e faça $k = 0$.

Passo 1: Escolha $y_k \in \Omega$ tal que

$$\|h(y_k)\| \leq r\|h(x_k)\|, \text{ ou seja, } \|h(y_k)\| - \|h(x_k)\| \leq -(1 - r)\|h(x_k)\|,$$

$$f(y_k) \leq f(x_k) + \beta\|h(x_k)\|$$

Passo 2: Escolha $\rho_{k+1} \in \{2^{-i}\rho_k : i \in \mathbb{N}\}$ o maior possível tal que

$$\Phi(y_k, \rho_{k+1}) - \Phi(x_k, \rho_{k+1}) \leq \frac{1}{2}(1 - r)(\|h(y_k)\| - \|h(x_k)\|).$$

Passo 3: Escolha $\theta_k \in \{2^i\gamma, i \in \mathbb{N}\}$, o menor possível, e uma direção d_k associada a θ_k de forma que d_k :

a) seja viável e ϵ_k -KKT para o problema

$$\begin{aligned} & \text{Minimizar} && f(y_k + d) + \theta_k \|d\|^2 \\ & && d \in \mathbb{R}^n \\ & \text{sujeito a:} && J_h(y_k)d = 0 \\ & && y_k + d \in \Omega, \end{aligned} \tag{6.3}$$

b) satisfaça as relações:

$$f(y_k + d_k) \leq f(y_k) - \gamma \|d_k\|^2 \tag{6.4}$$

$$\Phi(y_k + d_k, \lambda_{k+1}, \rho_{k+1}) - \Phi(x_k, \lambda_k, \rho_{k+1}) \leq \frac{1}{2}(1-r)(\|h(y_k)\| - \|h(x_k)\|). \tag{6.5}$$

Passo 4: Faça $x_{k+1} = y_k + d_k$, $k = k + 1$ e volte ao passo 1.

Teorema 6.2.1. *O Algoritmo 6 está bem definido.*

Demonstração: A boa definição dos passos 1 e 2 já está provada, uma vez que são os mesmos do Algoritmo 5.1 com a escolha $\lambda_k = 0 \forall k$. Conseqüentemente ainda obtemos que existe $\bar{\rho} > 0$ tal que $\rho_k \geq \bar{\rho}$ para todo k .

Para a boa definição do passo 3 basta observar que, uma vez fixado θ_k , se aplicarmos um método para resolver inexatamente o subproblema (6.3) cuja convergência a pontos ϵ_k -KKT seja garantida, por exemplo o método GSS-KLT, como mostramos anteriormente, obteremos a condição a).

Como $d = 0$ é viável para (6.3), se usarmos um método monótono, por exemplo GSS-KLT, e a origem como ponto inicial na resolução de (6.3), garantiremos que

$$f(y_k + d_k) + \theta_k \|d_k\|^2 \leq f(y_k).$$

Conseqüentemente, como $\gamma \leq \theta_k$ teremos que

$$f(y_k + d_k) + \gamma \|d_k\|^2 \leq f(y_k + d_k) + \theta_k \|d_k\|^2 \leq f(y_k),$$

e portanto a relação (6.4) também é satisfeita.

Sendo assim só nos resta demonstrar que a relação (6.5) pode ser satisfeita para θ_k suficientemente grande. Para isso lembremos que pelo fato de $J_h(y_k)d_k = 0$ e J_h ser Lipschitz a relação (5.5) nos garante que existe $\bar{\gamma} > 0$, independente de y_k , tal que

$$\|h(y_k + d_k)\| \leq \|h(y_k)\| + \bar{\gamma}\|d_k\|^2.$$

De posse disso vamos mostrar agora que se $\theta_k \geq \frac{(1-\bar{\rho})}{\bar{\rho}}\bar{\gamma}$ então a relação 6.5 também é satisfeita. Desse modo o passo 3 está bem definido e $\gamma \leq \theta_k \leq \bar{\theta}$, onde $\bar{\theta} = 2\frac{(1-\bar{\rho})}{\bar{\rho}}\bar{\gamma}$. De fato, se $\theta_k \geq \frac{(1-\bar{\rho})}{\bar{\rho}}\bar{\gamma}$ então $(1-\bar{\rho})\bar{\gamma} - \bar{\rho}\theta_k \leq 0$ e portanto

$$\begin{aligned} \Phi(y_k + d_k, \rho_{k+1}) - \Phi(x_k, \rho_{k+1}) &= \Phi(y_k + d_k, \rho_{k+1}) - \Phi(y_k, \rho_{k+1}) + \\ &\quad + \Phi(y_k, \rho_{k+1}) - \Phi(x_k, \rho_{k+1}) \\ &\leq \rho_{k+1} (f(y_k + d_k) - f(y_k)) + \\ &\quad + (1 - \rho_{k+1}) (\|h(y_k + d_k)\| - \|h(y_k)\|) + \\ &\quad + \frac{1-r}{2} (\|h(y_k)\| - \|h(x_k)\|) \\ &\leq -\rho_{k+1}\theta_k\|d_k\|^2 + (1 - \rho_{k+1})\bar{\gamma}\|d_k\|^2 + \\ &\quad + \frac{1-r}{2} (\|h(y_k)\| - \|h(x_k)\|) \\ &\leq -\bar{\rho}\theta_k\|d_k\|^2 + (1 - \bar{\rho})\bar{\gamma}\|d_k\|^2 + \\ &\quad + \frac{1-r}{2} (\|h(y_k)\| - \|h(x_k)\|) \\ &\leq ((1 - \bar{\rho})\bar{\gamma} - \bar{\rho}\theta_k)\|d_k\|^2 + \\ &\quad + \frac{1-r}{2} (\|h(y_k)\| - \|h(x_k)\|) \\ &\leq \frac{1-r}{2} (\|h(y_k)\| - \|h(x_k)\|). \end{aligned}$$

Dessa forma todos os passos estão bem definidos. □

A demonstração de que $\sum_{k=0}^{\infty} \|h(x_k)\| = \hat{\sigma} < \infty$, e portanto de que todos os pontos limite gerados pelo algoritmo são viáveis, também é a mesma da feita em Fischer-Friedlander. Isso acontece pois as relações usadas nesta demonstração também

são garantidas pelo método sem busca linear. De todo modo, por conveniência, vamos refazer os passos principais da demonstração aqui.

Demonstração: Para $k \geq k_0$, onde $\rho_{k_0} = \bar{\rho}$, os passos 3 e 1 do Algoritmo 6 garantem que

$$\Phi(x_{k+1}, \bar{\rho}) - \Phi(x_k, \bar{\rho}) \leq \frac{1}{2}(1-r)(\|h(y_k)\| - \|h(x_k)\|) \leq -\frac{1}{2}(1-r)^2\|h(x_k)\|,$$

portanto:

$$\begin{aligned} \Phi(x_l, \bar{\rho}) - \Phi(x_{k_0}, \bar{\rho}) &= \sum_{k=k_0}^{l-1} \Phi(x_{k+1}, \bar{\rho}) - \Phi(x_k, \bar{\rho}) \\ &\leq \sum_{k=k_0}^{l-1} -\frac{1}{2}(1-r)^2\|h(x_k)\|. \end{aligned}$$

Como $\Phi(x, \bar{\rho})$ é contínua em x e Ω é um compacto temos que Φ é limitada inferiormente e portanto $\sum_{k=k_0}^{\infty} \|h(x_k)\|$ é limitada superiormente. Como se trata de uma série crescente, temos que ela é convergente, e portanto $\sum_{k=0}^{\infty} \|h(x_k)\| = \hat{\sigma} < \infty$. Dessa forma temos que o termo geral $\|h(x_k)\|$ converge a zero e, como $\|h(y_k)\| \leq \|h(x_k)\|$, também temos que $\|h(y_k)\|$ converge a zero e conseqüentemente tanto $h(x_k)$ quanto $h(y_k)$ convergem a zero. Portanto a continuidade de h garante que todos os pontos limites gerados pelo algoritmo são viáveis

□

O fato da direção d_k tender a zero também é acarretado pelos mesmos motivos do que no algoritmo de Fischer-Friedlander. A seguir vamos fazer uma demonstração resumida desse fato.

Demonstração: Pelos passos 1 e 3 e o fato de $\sum_{k=0}^{\infty} \|h(x_k)\| = \hat{\sigma} < \infty$ temos que

$$\begin{aligned} f(x_{k+1}) - f(x_k) &= f(x_{k+1}) - f(y_k) + f(y_k) - f(x_k) \\ &\leq -\gamma\|d_k\|^2 + \beta\|h(x_k)\|. \end{aligned}$$

Sendo assim

$$\begin{aligned}
 f(x_{l+1}) - f(x_0) &= \sum_{k=0}^l (f(x_{k+1}) - f(x_k)) \\
 &\leq -\gamma \sum_{k=0}^l \|d_k\|^2 + \beta \sum_{k=0}^l \|h(x_k)\| \\
 &\leq -\gamma \sum_{k=0}^l \|d_k\|^2 + \beta \hat{\sigma}.
 \end{aligned}$$

Como f é contínua num compacto, devemos ter que $\sum_{k=0}^l \|d_k\|^2$ é limitada, e portanto $\|d_k\|$, e conseqüentemente d_k , convergem a zero. \square

Para os resultados referentes a otimalidade dos pontos limite vamos precisar mostrar algumas particularidades dos subproblemas (6.3) resolvidos no passo 3 do Algoritmo 6. A primeira dessas particularidades é que os gradientes das funções objetivo dos subproblemas são Lipschitz contínuos com a mesma constante $L_{\nabla f} + 2\bar{\theta}$ em todas as iterações. Vejamos isso:

$$\begin{aligned}
 \|\nabla [f(x+d) + \theta_k \|d\|^2] - \nabla [f(x+s) + \theta_k \|s\|^2]\| &= \\
 &= \|\nabla f(x+d) - \nabla f(x+s) + \theta_k (\nabla \|d\|^2 - \nabla \|s\|^2)\| \\
 &\leq \|\nabla f(x+d) - \nabla f(x+s)\| + \theta_k \|\nabla \|d\|^2 - \nabla \|s\|^2\| \\
 &\leq L_{\nabla f} \|(x+d) - (x+s)\| + \bar{\theta} \|2d - 2s\| \\
 &\leq (L_{\nabla f} + 2\bar{\theta}) \|d - s\|.
 \end{aligned}$$

A segunda particularidade é que $\|J_h(y_k)\|_F$ é limitada superiormente uma vez que $J_h(x)$ é contínua e Ω é um compacto. Deste modo já mostramos no Teorema 6.1.2 que se usarmos o método GSS-KLT, com algumas escolhas particulares bem razoáveis, na resolução dos subproblemas então é possível fazer com que ϵ_k tenda zero controlando apenas o critério de parada Δ_i^{tol} .

Por fim demonstraremos que todo ponto de acumulação \bar{y} de $\{y_k\}$ é AKKT, o que implica, com condições de qualificação bem fracas, que os pontos limites são KKT para o problema original (4.1). Para isso vamos supor que a sequência ϵ_k tenda a zero.

Demonstração: Pelo fato de d_k ser um ponto viável e ϵ_k -KKT de (6.3), existem $\lambda_k \in \mathbb{R}^m$, $\nu_k^l \geq 0$ e $\nu_k^u \geq 0$ tais que valem as relações:

$$\nabla f(y_k + d_k) + 2\theta_k d_k + J_h(y_k)^T \lambda_k - \nu_k^l + \nu_k^u = v_k, \quad (6.6)$$

$$J_h(y_k) d_k = 0,$$

$$l \leq y_k + d_k \leq u, \quad (6.7)$$

$$[l - y_k - d_k]_i [\nu_k^l]_i = 0 \quad \text{e} \quad [y_k + d_k - u]_i [\nu_k^u]_i = 0, \quad (6.8)$$

$$\nu_k^l, \nu_k^u \leq 0, \quad (6.9)$$

$$\|v_k\| \leq \epsilon_k. \quad (6.10)$$

Como f é duas vezes continuamente diferenciável temos que $\|\nabla^2 f(x)\| \leq W$, para todo $x \in \Omega$. Pelo teorema fundamental do cálculo sabemos que

$$\nabla f(y_k + d_k) = \nabla f(y_k) + \int_0^1 \nabla^2 f(y_k + td_k) d_k dt,$$

e portanto podemos reescrever a equação (6.6) como

$$\nabla f(y_k) + J_h(y_k)^T \lambda_k - \nu_k^l + \nu_k^u = v_k - \int_0^1 \nabla^2 f(y_k + td_k) d_k dt + 2\theta_k d_k.$$

Denotando $\bar{v}_k = v_k - \int_0^1 \nabla^2 f(y_k + td_k) d_k dt + 2\theta_k d_k$ temos que

$$\begin{aligned} \|\bar{v}_k\| &\leq \|v_k\| + \left\| \int_0^1 \nabla^2 f(y_k + td_k) d_k dt \right\| + 2\theta_k \|d_k\| \\ &\leq \epsilon_k + \int_0^1 \|\nabla^2 f(y_k + td_k) d_k\| dt + 2\bar{\theta} \|d_k\| \\ &\leq \epsilon_k + \int_0^1 \|\nabla^2 f(y_k + td_k)\| \|d_k\| dt + 2\bar{\theta} dt \|d_k\| \\ &\leq \epsilon_k + \int_0^1 W \|d_k\| dt + 2\bar{\theta} \|d_k\| \\ &\leq \epsilon_k + W \|d_k\| + 2\bar{\theta} \|d_k\|. \end{aligned}$$

De (6.7) temos que

$$-\|d_k\|e + l \leq y_k \leq u + \|d_k\|e,$$

onde e é um vetor de dimensão apropriada com todas as coordenadas iguais a 1. De (6.8) temos que $[\nu_k^l]_i = 0$ se $[l - y_k - d_k]_i < 0$, ou seja $[\nu_k^l]_i = 0$ se $[l - y_k]_i < [d_k]_i$, logo $[\nu_k^l]_i = 0$ se $[l - y_k]_i < -\|d_k\|$ (pois $[l - y_k]_i < -\|d_k\|$ implica $[l - y_k]_i < [d_k]_i$). Analogamente $[\nu_k^u]_i = 0$ se $[y_k + d_k - u]_i < 0$, portanto $[\nu_k^u]_i = 0$ se $[y_k - u]_i < -\|d_k\|$.

Sendo assim, se definirmos $\bar{\epsilon}_k = \max\{\epsilon_k + (W + 2\bar{\theta})\|d_k\|, \|d_k\|, \|h(y_k)\|\}$, teremos que valem

$$\|\nabla f(y_k) + J_h(y_k)^T \gamma_k - \nu_k^l + \nu_k^u\| \leq \bar{\epsilon}_k,$$

$$\|h(y_k)\| \leq \bar{\epsilon}_k,$$

$$l - \bar{\epsilon}_k e \leq y_k \leq u + \bar{\epsilon}_k e,$$

$$[l - y_k]_i < -\bar{\epsilon}_k \Rightarrow [\nu_k^l]_i = 0,$$

e

$$[y_k - u]_i < -\bar{\epsilon}_k \Rightarrow [\nu_k^u]_i = 0,$$

ou seja, temos que y_k é $\bar{\epsilon}_k$ -KKT do problema original (4.1). Do teorema anterior já sabemos que $\|h(y_k)\|$ converge a zero. Como já demonstramos que $\|d_k\|$ converge a zero, e ϵ_k é por hipótese uma sequência convergente a zero, temos que $\bar{\epsilon}_k$ também converge a zero, e portanto qualquer ponto limite \bar{y} de $\{y_k\}$ é AKKT. Consequentemente, por [19], teremos que, se as restrições em \bar{y} satisfazem a CPLD, então \bar{y} é KKT para o problema original (4.1). \square

Neste capítulo não fizemos uso dos multiplicadores de Lagrange no algoritmo. A primeira razão para isso é que para a construção do subproblema na fase de otimização não é necessária a utilização destes multiplicadores, já que não estamos trabalhando com modelos quadráticos. Entretanto, de todo modo seria interessante minimizar o Lagrangiano no espaço tangente e não somente a função objetivo, e nesse caso

necessitaríamos das variáveis duais. O grande empecilho para isso é que a alternativa natural de atualizar os multiplicadores utilizada no Capítulo 5 não está disponível no novo subproblema. Isso acontece porque não temos expressões para os multiplicadores dos subproblemas. Uma maneira de contornar isso seria inserir uma penalidade fixa e atualizar os multiplicadores como em métodos de Lagrangiano Aumentado. Esse ponto parece ser interessante de ser estudado no futuro.

Uma observação importante a ser feita aqui é que o preço que pagamos por evitar a busca linear é que devemos resolver possivelmente vários problemas de otimização para obter a direção de otimização. Como isso pode não ser uma tarefa fácil, é de se esperar que o método tratado aqui seja menos eficiente do que o método de Fischer-Friedlander quando este seja aplicável. Entretanto, ganhamos aqui uma amplitude maior de problemas possíveis de serem abordados, podendo ser uma opção para quando não tivermos as derivadas de f disponíveis.

Como dissemos no Capítulo 4, uma vantagem de métodos de Restauração Inexata em relação a métodos que tratam tanto a função objetivo quanto as restrições ao mesmo tempo, é que é possível explorar melhor a estrutura do problema no caso em que um dos objetivos, viabilidade ou otimalidade, seja mais fácil de ser tratado do que o outro. Isso acontece porque não perdemos tempo avaliando o objetivo difícil de ser tratado quando o foco é o objetivo fácil. No caso desta seção onde nos problemas alvo considerados não estão disponíveis as derivadas da função objetivo, o que geralmente implica que avaliações de função objetivo sejam muito custosas, não fazemos avaliações de função objetivo na fase de restauração.

Uma grande vantagem em nosso método aplicado a problemas de otimização sem o uso de derivadas é que vários dos métodos existentes para problemas deste tipo baseiam-se no uso de direções densas. Este argumento é bastante usado mesmo em problemas com restrições simples, no sentido em que as suas derivadas estão disponíveis e para as quais é fácil fazer uma restauração. Entretanto, existem várias críticas a essa abordagem pois, em teoria, métodos baseados nessas ideias podem não ser interessantes computacionalmente. Deste modo, embora a convergência desses métodos

envolva o argumento de direções densas, os experimentos computacionais comprovam boa eficiência apenas se tais direções são escolhidas de forma que a densidade de fato não seja um fator decisivo. Isso nos leva a crer que a densidade não é o argumento que de fato explica o comportamento desses métodos. Por exemplo, argumentos desse tipo demonstram a convergência de métodos para problemas de otimização global, para uma função contínua limitada inferiormente qualquer, embora a eficiência numérica desse método seja incontestavelmente ruim, [21]. Dessa forma consideramos que não usar o argumento de densidade é um diferencial de nosso método sobre os existentes na literatura.

Outra consideração que merece destaque é que em problemas em que métodos de interpolação, por exemplo BOBYQA [92], são mais adequados do que métodos de busca coordenada podemos modificar a fase de otimização do Algoritmo 6 para melhorar a eficiência de nosso método. Para isso basta trocar o Algoritmo GSS-KLT por algum método de interpolação com restrições lineares para resolver o subproblema (6.3). A razão de não termos explorado mais essa alternativa é que esse tipo de método não tem teoria de convergência bem estabelecida, portando não é possível garantir que eles forneçam pontos ϵ -KKT. Entretanto, em situações práticas em que esses métodos são bem sucedidos é observado que eles geram pontos aproximadamente KKT.

6.3 Experimentos numéricos

Testes computacionais referentes ao Algoritmo 6 são apresentados em [33] e serão reproduzidos a seguir.

Na fase de restauração foi usado Algencan para encontrar y_k resolvendo o seguinte problema:

$$\begin{aligned} & \underset{y \in \Omega}{\text{Minimizar}} && \frac{1}{2} \|y - x_k\|^2 \\ & \text{sujeito a:} && h(y) = 0. \end{aligned}$$

O critério de parada nessa fase foi que pelos menos $\|h(y)\| \leq 0.5\|h(x_k)\|$. Já o parâmetro r escolhido foi 0.99. Isso indica que estamos querendo dividir a infactibilidade por pelo

menos 2, mas caso Algencan não alcance essa melhora não vamos parar o Algoritmo 6.

Na fase de otimização o subproblema de otimização sem derivadas com restrições lineares foi resolvido com o Algoritmo GSS-KLT implementado no software HOPSPACK [89]. O valor do parâmetro de regularização inicial foi escolhido como $\gamma = 2^{-5}$ a não ser quando o algoritmo identifica que o parâmetro é inadequado e faz algum ajuste. O critério de parada foi:

$$\Delta_{k+1} = \min\{1.1^k, 0.1 \max\{\|h(y_k + d_k)\|, \|d_k\|\}\}.$$

O algoritmo declara convergência se

$$\|h(y_k + d_k)\| \leq 10^{-8}, \quad \|d_k\| \leq 10^{-3} \quad \text{e} \quad \Delta_k \leq 10^{-3}.$$

O algoritmo foi implementado em C++ num computador com 8GB de RAM, processador Dual Core Intel Core i7, 2.67GHz e sistema operacional Linux (Ubuntu) 64bits.

Confrontamos nosso algoritmo com HOPSPACK em 105 dos 116 problemas com restrições da coleção de testes de Hock e Schittkowski [61]. Os 11 problemas excluídos correspondem a problemas irrestritos ou a problemas que apresentaram falha no cálculo das derivadas das restrições na compilação. HOPSPACK é uma implementação de um método de Lagrangiano Aumentado para otimização sem uso de derivadas. Esse tipo de método possui teoria de convergência bem estabelecida [68], [69] e [42]. Entretanto esse tipo de método não explora o fato das restrições serem mais simples já que trabalha com função objetivo e restrições simultaneamente. Declaramos que o algoritmo resolveu um problema quando

$$\|h(\bar{x})\|_{\infty} \leq 10^{-8} \quad \text{e} \quad \frac{|f(\bar{x}) - f_L|}{\max\{1, |f(\bar{x})|, |f_L|\}} \leq 10^{-1}.$$

Os resultados obtidos estão apresentados na Tabela 6.1 e na Figura 6.3. Podemos ver que o método de restauração Inexata foi tanto mais eficiente quanto mais robusto que HOPSPACK.

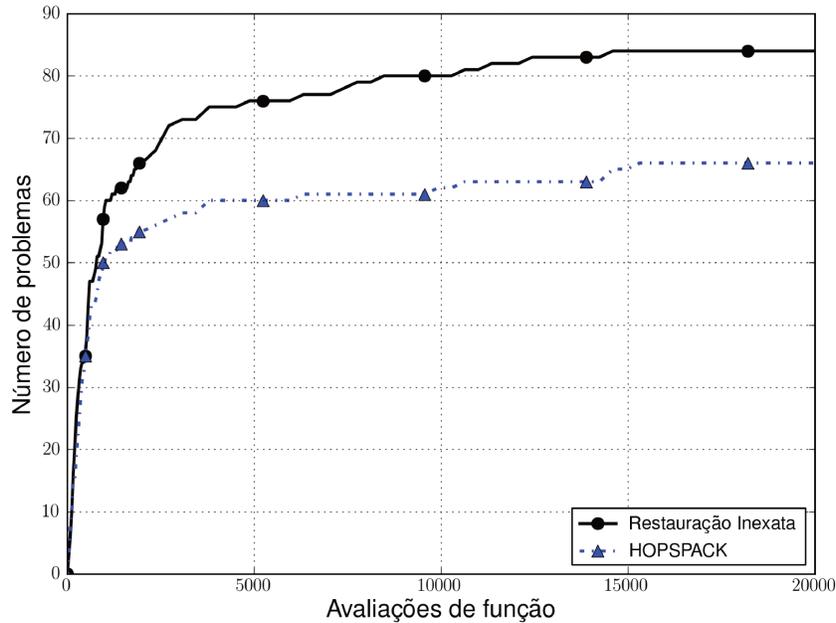


Figura 6.1: Evolução do número de problemas resolvidos pelo Algoritmo 6 e por HOPSPACK em função do número de avaliações de função objetivo.

Tabela 6.1: Comparação entre o valor de f obtido e o número de avaliações da função objetivo no Algoritmo 6 e em HOPSPACK. Asteriscos indicam pontos inviáveis encontrados pelo algoritmo correspondente.

Prob.	Algoritmo 6		HOPSPACK	
	f	#Aval. f	f	#Aval. f
6	3.44661E-12	206	4.8400E+00*	151
7	-1.73205E+00	346	6.93147E-01	325
8	-1.00000E+00	3	-1.0000E+00*	187
9	-5.00000E-01	198	-5.00000E-01	26
10	-1.00000E+00	958	-8.65109E-01	357
11	-8.49846E+00	260	-8.47917E+00	523
12	-3.00000E+01	589	-3.00000E+01	342

Tabela 6.1: Comparação entre o valor de f obtido e o número de avaliações da função objetivo no Algoritmo 6 e em HOPSPACK. Asteriscos indicam pontos inviáveis encontrados pelo algoritmo correspondente.

Prob.	Algoritmo 6		HOPSPACK	
	f	#Aval. f	f	#Aval. f
13	1.03474E+00	76398	9.9108E-01*	803
14	1.39346E+00	51	1.39411E+00	202
15	3.06500E+00	96	3.06500E+00	451
16	2.58234E-01	544	2.50000E-01	600
17	1.04614E+00	94	1.00004E+00	612
18	5.0224E+00*	4961	1.06635E+01	263
19	-6.96181E+03	138	-7.4719E+03*	1370
20	4.15105E+01	40	3.82187E+01	393
21	-9.99600E+01	144	-9.99600E+01	32
22	1.00061E+00	78	1.00000E+00	276
23	2.00000E+00	84	2.00000E+00	466
24	-1.00000E+00	125	-1.00000E+00	27
26	1.2332E-08*	9256	2.11600E+01	585
27	4.00000E+00	3695	4.00056E+00	1358
28	4.36831E-09	517	7.70336E-08	264
29	-2.26274E+01	885	-2.24711E+01	327
30	1.00004E+00	794	1.00000E+00	55
31	6.00000E+00	1603	6.00351E+00	921
32	1.00000E+00	155	1.00000E+00	51
33	-4.00000E+00	109	-4.58538E+00	381
34	-8.34032E-01	165	-2.28219E-01	582
35	1.11111E-01	287	1.11112E-01	340
36	-3.30000E+03	240	-3.30000E+03	60

Tabela 6.1: Comparação entre o valor de f obtido e o número de avaliações da função objetivo no Algoritmo 6 e em HOPSPACK. Asteriscos indicam pontos inviáveis encontrados pelo algoritmo correspondente.

Prob.	Algoritmo 6		HOPSPACK	
	f	#Aval. f	f	#Aval. f
37	-3.45600E+03	503	-3.45600E+03	102
39	-1.00000E+00	776	-1.00000E+00	830
40	-2.50000E-01	214	-2.5056E-01*	897
41	1.92593E+00	238	1.92593E+00	292
42	1.38579E+01	2055	1.40000E+01	779
43	-4.39930E+01	2657	-4.40000E+01	1134
44	-1.30000E+01	255	-1.30000E+01	57
46	1.25916E-05	1865	3.33763E+00	777
47	6.33827E-10	457	1.24954E+01	901
48	3.12666E-08	924	1.11743E-06	497
49	1.52106E-07	36006	1.42943E-04	1002
50	2.38728E-07	304	5.29367E-07	290
51	1.82282E-09	563	1.25372E-06	142
52	5.32665E+00	317	5.3267E+00*	311
53	4.09302E+00	252	4.09302E+00	216
54	-1.78438E-01	292998	-1.5590E-01*	-1
55	6.66667E+00	41	1.0000E+20*	-1
56	-3.75830E-01	45478	-1.00000E+00	2075
57	3.06462E-02	239	3.06380E-02	74
58	3.19440E+00	82	3.19226E+00	817
59	-7.80424E+00	988	-6.75457E+00	340
60	3.25682E-02	341	5.47089E-02	465
61	-1.43646E+02	155	-1.43000E+02	621

Tabela 6.1: Comparação entre o valor de f obtido e o número de avaliações da função objetivo no Algoritmo 6 e em HOPSPACK. Asteriscos indicam pontos inviáveis encontrados pelo algoritmo correspondente.

Prob.	Algoritmo 6		HOPSPACK	
	f	#Aval. f	f	#Aval. f
62	-2.62725E+04	849	-2.62716E+04	233
63	9.61715E+02	381	9.62606E+02	317
64	6.29984E+03	703	6.29985E+03	6253
65	9.53529E-01	944	1.00617E+00	379
66	5.18163E-01	502	5.33106E-01	566
68	-9.20425E-01	2012	-8.4354E-01*	1316
69	-9.56713E+02	20385	-9.5665E+02*	2471
70	2.69086E-01	669	7.74114E-03	3766
71	1.70141E+01	1767	1.7031E+01*	1939
72	7.27679E+02	2388	7.2059E+02*	9606
73	2.98944E+01	129	3.01595E+01	223
74	5.12650E+03	543	5.1447E+03*	46145
75	5.17441E+03	186	5.2331E+03*	22678
76	-4.68182E+00	560	-4.68182E+00	403
77	2.41505E-01	730	4.6807E+00*	1904
78	-2.91970E+00	1180	-2.8917E+00*	869
79	7.87768E-02	528	2.4186E-01*	1054
80	5.39498E-02	590	1.0000E+00*	557
81	5.39499E-02	2517	9.9999E-01*	557
83	-3.06655E+04	530	-3.06653E+04	1708
84	-5.28034E+01	1306	-5.27947E+01	2842
86	-3.23487E+01	562	-3.23152E+01	485
87	8.92760E+03	971	9.3254E+03*	16244

Tabela 6.1: Comparação entre o valor de f obtido e o número de avaliações da função objetivo no Algoritmo 6 e em HOPSPACK. Asteriscos indicam pontos inviáveis encontrados pelo algoritmo correspondente.

Prob.	Algoritmo 6		HOPSPACK	
	f	#Aval. f	f	#Aval. f
88	1.36268E+00	941	1.3674E+00*	1274
89	1.36269E+00	1715	1.3640E+00*	1945
90	1.36518E+00	3008	1.36531E+00	2655
91	1.36269E+00	3454	1.3619E+00*	4180
92	1.36386E+00	4655	1.3613E+00*	4004
93	1.3544E+02*	27232	1.37066E+02	129
95	1.56195E-02	12139	1.70816E-02	156
96	1.56195E-02	7401	1.70816E-02	156
97	4.07125E+00	11058	4.12011E+00	145
98	3.13581E+00	14366	4.12011E+00	145
99	-8.30871E+08	4	-7.4573E+08*	729
100	1.3619E+03*	24770	6.84327E+02	873
101	1.82638E+03	66903	1.81634E+03	14563
102	1.53792E+03	50936	9.20560E+02	15147
103	1.54704E+03	25457	5.43980E+02	14583
104	3.95117E+00	20730	3.95436E+00	9844
105	1.13841E+03	99757	1.13842E+03	10580
106	7.04925E+03	75091	1.13923E+04	30424
107	5.05545E+03	173	5.0628E+03*	7232
108	-5.00000E-01	7154	-5.00000E-01	99
109	1.1686E+04*	10	5.5010E+03*	57551
111	-4.0523E+01*	31196	-4.7780E+01*	-1
112	-4.77611E+01	8366	-4.77608E+01	730

Tabela 6.1: Comparação entre o valor de f obtido e o número de avaliações da função objetivo no Algoritmo 6 e em HOPSPACK. Asteriscos indicam pontos inviáveis encontrados pelo algoritmo correspondente.

Prob.	Algoritmo 6		HOPSPACK	
	f	#Aval. f	f	#Aval. f
113	2.43130E+01	2432	2.53962E+01	1887
114	-1.5648E+03*	21132	1.0000E+20*	-1
116	1.8779E+02*	2113	5.0000E+01*	9131
117	3.23487E+01	10557	5.40287E+01	7190
118	6.64820E+02	6183	6.64830E+02	3762
119	2.44900E+02	1018	2.44926E+02	944

Capítulo 7

Introdução ao problema de Equilíbrio de Nash

A Teoria de Jogos [83, 84] é uma ferramenta eficiente para modelar situações onde agentes (jogadores) têm que tomar decisões interagindo entre si, e por isso tem aplicações em diversos ramos da ciência. Um jogo consiste nos seguintes elementos básicos:

- Conjunto de jogadores $G = \{p_1, \dots, p_q\}$.
- Conjunto S_i de estratégias, ou decisões possíveis, do jogador p_i .
- Conjunto de resultados possíveis $S = S_1 \times S_2 \times \dots \times S_q$.
- Função de perda do jogador p_i

$$f_i : S \rightarrow \mathbb{R}.$$

Uma notação que será usada ao longo do texto é escrever S como $S_i \times S_{-i}$ para dividirmos as configurações possíveis como a jogada do jogador i , pertencente a S_i , e as jogadas dos demais, pertencentes a $S_{-i} \equiv S_1 \times \dots \times S_{i-1} \times S_{i+1} \times \dots \times S_q$. Além disso, escreveremos todo $x \in S$ fraccionado da forma $x = (x_i, x_{-i})$, onde $x_i \in S_i$ denota a estratégia do jogador i e $x_{-i} \in S_{-i}$ as estratégias dos demais agentes. No caso em que o conjunto de decisões de cada jogador seja finito denotaremos por $S_i =$

$\{x_{i1}, x_{i2}, \dots, x_{in_i}\}$ o conjunto das n_i jogadas possíveis do jogador i e o jogo será dito de estratégias puras. Ilustremos esta situação com um exemplo clássico, popularmente chamado de dilema do prisioneiro.

Considere que dois indivíduos, Bob e Will, estão sendo investigados sob a suspeita de serem os autores de um crime. O detetive do caso coloca cada um isolado do outro e explica que eles tem duas opções, confessar ou negar o crime. Ele explica também que a pena, em número de anos presos, que receberão será dada, em consequência de suas respostas, de acordo com a seguinte tabela:

		Will	
		Confessar	Negar
Bob	Confessar	(5,5)	(0,10)
	Negar	(10,0)	(1,1)

Ou seja, se ambos confessarem passarão cinco anos presos, se ambos negarem passarão um ano presos e se um confessar e o outro negar o que confessou será solto mas o que negou ficará dez anos preso. O detetive também fala que o outro prisioneiro também terá acesso as mesmas informações.

Neste exemplo os elementos do jogo seriam:

- Conjunto de jogadores $G = \{\text{Bob}, \text{Will}\}$.
- Conjunto $S_{\text{Bob}} = S_{\text{Will}} = \{\text{confessar}, \text{negar}\}$.
- Conjunto de resultados possíveis:

$$\{(\text{confessar}, \text{confessar}), (\text{confessar}, \text{negar}), (\text{negar}, \text{confessar}), (\text{negar}, \text{negar})\}.$$

- Funções de perda:

$$f_{\text{Bob}}(\text{confessar}, \text{confessar}) = f_{\text{Will}}(\text{confessar}, \text{confessar}) = 5,$$

$$f_{\text{Bob}}(\text{confessar}, \text{negar}) = f_{\text{Will}}(\text{negar}, \text{confessar}) = 0,$$

$$f_{\text{Bob}}(\text{negar}, \text{confessar}) = f_{\text{Will}}(\text{confessar}, \text{negar}) = 10,$$

$$f_{\text{Bob}}(\text{negar}, \text{negar}) = f_{\text{Will}}(\text{negar}, \text{negar}) = 1.$$

Como todos os jogadores sabem toda a estrutura do jogo, e sabem que o outro também sabe, dizemos que este é um jogo de informação completa. Além disso, como os jogadores estão isolados, e portanto não podem combinar suas ações, dizemos que este é um jogo de informação completa não cooperativo. Nós sempre trataremos deste tipo de jogo.

Uma pergunta pertinente agora seria, “Qual situação podemos esperar deste jogo?”. A resposta a esta pergunta não é simples, até porque existem mais de uma abordagem para tentar solucioná-la, e, dependendo da utilizada, podemos chegar a respostas diferentes. A chave da dificuldade encontra-se no fato de não termos um objetivo comum para os agentes, mas sim que cada um procura fazer o melhor para si mesmo. Ou seja, temos muitos objetivos num mesmo problema.

Uma alternativa para buscar soluções é nos colocarmos como um dos prisioneiros, por exemplo Bob, e fazer o seguinte raciocínio, se Will confessar qual seria a melhor solução que eu poderia tomar? Confessando ficaria preso 5 anos, negando ficaria 10, portanto preferiria confessar. E se Will negar? Confessando seria solto, negando ficaria 1 ano preso, portanto novamente preferiria confessar. Como, independente da decisão de Will, seria melhor para Bob confessar consideraremos que ele assim faça. Pela simetria deste jogo também podemos esperar que Will confesse. Portanto, a “solução” do jogo seria (confessar, confessar). Encontrar soluções descartando estratégias desta forma é o que é chamado de eliminação de estratégias estritamente dominadas.

Outra alternativa para tentar chegar a uma configuração esperada seria tratar o jogo como um problema multiobjetivo e adotar o conceito consagrado de ótimo a Pareto como solução do jogo.

Definição: Dizemos que $x^* \in S$ é um ponto ótimo a Pareto se não existe $x \in S$ tal que $f_i(x) \leq f_i(x^*)$, $\forall i$ e $f_j(x) < f_j(x^*)$ para algum j .

Informalmente estamos dizendo que não existe uma decisão em que todos jogadores, ao mesmo tempo, ganhem mais (ou percam menos) do que em x^* .

Claramente a situação (confessar, confessar), encontrada anteriormente, não é ótima a Pareto pois em (negar, negar) tanto Bob quanto Will ficariam presos menos tempo, um ano contra cinco na anterior. A opção (negar, negar) é um ótimo a Pareto pois em qualquer outra situação pelo menos um dos jogadores ficaria mais de um ano preso.

Uma terceira alternativa de analisar o jogo seria encontrar pontos de equilíbrio, ou seja, situações que acreditaríamos que seriam mantidas caso fossem atingidas. Chegamos, finalmente, ao conceito de equilíbrio de Nash.

Definição: Dizemos que $x^* \in S$ é um equilíbrio de Nash se: $f_i(x_i^*, x_{-i}^*) \leq f_i(x_i, x_{-i}^*)$, $\forall x_i \in S_i$ e $\forall i$.

Intuitivamente estamos falando que nenhum jogador tem motivação pra trocar de estratégia se os outros não trocaram. Ou seja, fixado x_{-i}^* , temos que x_i^* é um minimizador de f_i em S_i . No nosso exemplo podemos ver que (confessar, confessar) é um equilíbrio de Nash pois nenhum dos jogadores estaria disposto a trocar confessar por negar se soubesse que o outro confessou. Na verdade, toda situação que for obtida com eliminação de estratégias dominadas será um equilíbrio de Nash, embora nem todo equilíbrio de Nash de um jogo possa ser encontrado através deste tipo de eliminação [30]. Neste exemplo podemos ver também que não é verdadeiro que equilíbrio de Nash implica em ótimo a Pareto e que também nem a recíproca é válida.

Podemos observar que todos os conceitos introduzidos até agora independem do fato do número de estratégias ser finito. Na verdade, o caso em que S_i é contínuo é o de maior interesse em nosso trabalho. Vamos considerar que S_i seja um subconjunto

de \mathbb{R}^{n_i} descrito através de restrições de igualdade e desigualdade. No caso em que S_i dependa também das decisões dos outros jogadores, ou seja $S_i = S_i(x_{-i})$, o problema é chamado na literatura de problema de equilíbrio de Nash generalizado, e este é o caso com maior enfoque em nosso trabalho, portanto, vamos nos referir a ele simplesmente por problema de equilíbrio de Nash. Vamos denotar o problema de otimização a ser resolvido pelo jogador i , parametrizado pela decisão dos demais jogadores, para sua tomada de decisão por $PEN_i(x_{-i}^*)$:

$$\begin{aligned}
 & \text{Minimizar} && f_i(x_i, x_{-i}^*) \\
 & x_i \in \Omega_i \subset \mathbb{R}^{n_i} && \\
 & \text{sujeito a:} && h_i(x_i, x_{-i}^*) = 0 \\
 & && g_i(x_i, x_{-i}^*) \leq 0.
 \end{aligned} \tag{7.1}$$

Sendo assim, a partir de agora vamos considerar como definição de Equilíbrio de Nash a versão que envolve os problemas $PEN_i(x_{-i}^*)$ e é mais utilizada em otimização contínua.

Definição: Dizemos que $x^* \in \mathbb{R}^{n_1+n_2+\dots+n_q}$ é um ponto de equilíbrio de Nash (Generalizado) se para cada $i = 1, 2, \dots, q$ tivermos que x_i^* é solução, local ou global, de $PEN_i(x_{-i}^*)$.

Em [47] são discutidos vários tipos de métodos para encontrar equilíbrios de Nash. Os métodos que são mais utilizados na literatura são baseados em desigualdades variacionais explorando as condições KKT de cada um dos problemas (7.1). Outra alternativa que nos chama muito a atenção é o estudo de métodos do tipo Jacobi e Gauss Seidel. Acreditamos que este tipo de método tem um apelo muito forte sobre como equilíbrios são atingidos na vida real. Nossa motivação para adotar esse tipo de método é a natureza intrínseca deles de que o problema pode ser visto como um sistema dinâmico em função do tempo. Neste caso a decisão de cada agente seria a melhor resposta para seu problema considerando as decisões dos outros agentes no instante de tempo anterior.

7.1 Exemplos de problemas de equilíbrio com decisões contínuas

Nesta seção vamos introduzir dois exemplos de problemas com variáveis de decisão contínuas em que estamos interessados em encontrar pontos de equilíbrio. O primeiro de nossos objetivos aqui é poder ilustrar de forma simples a relação entre os problemas de otimização e os pontos de equilíbrio. O segundo objetivo é fornecer uma ilustração de uma aplicação de grande importância desse conceito, para isto escolhemos o modelo de equilíbrio geral de Arrow-Debreu. Estes exemplos servirão como base também no Capítulo 8, onde apresentaremos um método de Restauração Inexata para resolver problemas com restrições de equilíbrio.

Nosso primeiro exemplo é meramente ilustrativo e consiste em obter um modelo linear muito simples para estudarmos a situação de equilíbrio. Para contextualizar, vamos considerar que estamos modelando, de forma elementar, a situação em que duas empresas concorrentes devem tomar a decisão de quão agressivas serão em suas campanhas publicitárias. Vamos denotar por $x, y \in [0, 100]$ o nível de agressividade da primeira e da segunda empresa, respectivamente. Claramente a decisão ótima de cada empresa depende de uma medida natural de impacto no mercado e da decisão da empresa concorrente. Coeficientes de afeição e aversão à disputa característicos de cada empresa também são importantes nesse processo. Sendo assim uma maneira para descrever a decisão da primeira empresa é como sendo solução do problema:

$$\begin{array}{ll} \text{Minimizar} & (ax - by - c)^2 \\ & x \end{array} \tag{7.2}$$

$$\text{sujeito a: } x \in [0, 100],$$

e a decisão da segunda empresa como sendo solução de:

$$\begin{array}{ll} \text{Minimizar} & (a'x - b'y + c')^2 \\ & y \end{array} \tag{7.3}$$

$$\text{sujeito a: } y \in [0, 100],$$

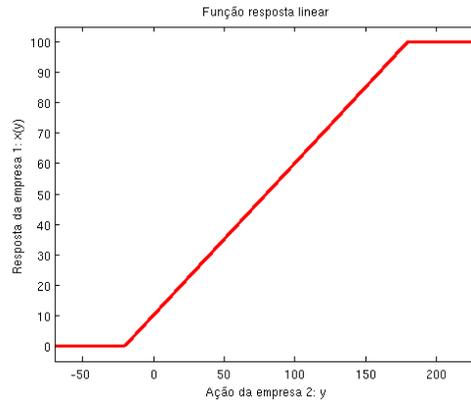


Figura 7.1: Resposta ótima $x(y)$ da primeira empresa à decisão y da segunda empresa.

onde os coeficientes a, a' correspondem ao fator de influência da decisão da primeira empresa respectivamente no seu próprio problema e no problema da segunda empresa, analogamente b, b' correspondem ao fator de influência da decisão da segunda empresa e c, c' são os parâmetros de agressividade inerente à cada empresa independentes da decisão da concorrente. Para que faça sentido que a decisão da empresa tenha algum impacto na sua função objetivo consideramos $a, b' \neq 0$. Portanto, fixado y , a melhor decisão para a empresa 1 seria $x(y) = \max\{0, \min\{\frac{c+by}{a}, 100\}\}$. Analogamente, a resposta ótima da empresa 2 à decisão x da primeira empresa, seria $y(x) = \max\{0, \min\{\frac{c'+a'x}{b'}, 100\}\}$. Dessa forma um par (x^*, y^*) seria um ponto de equilíbrio se e somente se $x(y^*) = x^*$ e $y(x^*) = y^*$, ou seja, cada uma das empresas está tendo a resposta ótima à decisão da outra.

Considerando $a = 2, b = 1$ e $c = 20$ temos que o gráfico da função resposta da primeira empresa é retratado na Figura 7.1. Escolhendo $a' = 1, b' = 3$, e $c' = 140$, a Figura 7.2 apresenta o gráfico da função resposta da segunda empresa.

Portanto o ponto de equilíbrio estaria na intersecção dos gráficos das duas funções respostas, já que neste caso teríamos que $x(y^*) = x^*$ e $y(x^*) = y^*$. Nesse caso o ponto de equilíbrio seria $x^* = 40$ e $y^* = 60$, conforme ilustrado na Figura 7.3.

O segundo exemplo que vamos abordar é o conceito de equilíbrio geral (ou

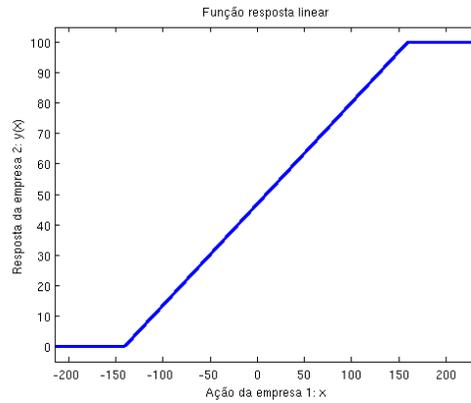


Figura 7.2: Resposta ótima $y(x)$ da segunda empresa à decisão x da primeira empresa.

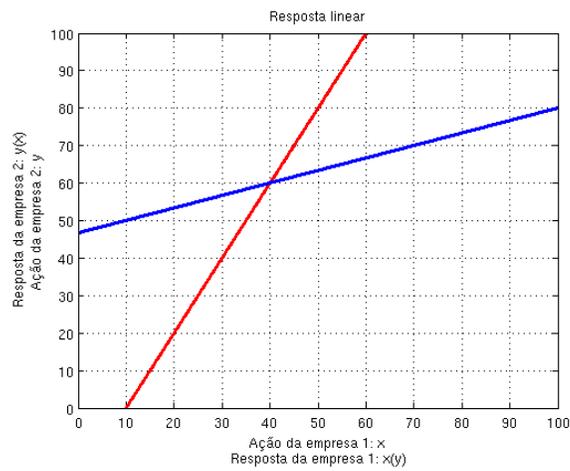


Figura 7.3: Gráfico mostrando simultaneamente as respostas $x(y)$ e $y(x)$ no intervalo $[0, 100]$, e o ponto de equilíbrio $(x^*, y^*) = (40, 60)$.

equilíbrio Walrasiano) de Arrow-Debreu [7]. Este conceito é uma das ferramentas mais importantes no estudo da microeconomia e foi um dos principais fatores a contribuir tanto para K. Arrow quanto para G. Debreu serem laureados com o Prêmio Nobel.

Para a formulação vamos supor que existam l tipos de bens econômicos, produtos e serviços, na economia. Os bens econômicos podem ser produzidos por unidades produtivas ou comprados e vendidos no mercado, tanto por firmas quanto por consumidores, por um preço $p_i \geq 0$ a ser determinado pela lei da oferta e demanda. Vamos denotar os mesmos produtos ou serviços com particularidades diferentes como bens econômicos diferentes. Desse modo o preço deles são determinados separadamente, mantendo apenas a coerência na relação entre eles. Por exemplo, uma tonelada de soja pode ser considerada como dois produtos diferentes se for comercializada hoje ou amanhã, ou se for plantada no Brasil mas comercializada no Brasil ou nos Estados Unidos. Evidentemente o preço delas está relacionado, uma vez que é possível substituir um produto pelo outro, mas isso a própria lei de mercado definirá.

Cada uma das s unidades produtivas existentes deve decidir qual plano de produção adotar. Vamos denotar essa decisão por $y_j \in Y_j \subset \mathbb{R}^l$, cuja entrada positiva em uma coordenada indica a produção do bem correspondente e uma coordenada negativa indica o consumo ou uso de algum bem ou serviço. Por exemplo, se $y_3 = (4, 2, -5, -7)$ temos que a terceira firma está produzindo 4 unidades do bem 1 e 2 do bem dois. Para isso, a unidade produtiva está utilizando 5 unidades do bem 3 como matéria prima e está contratando 7 unidades de mão-de-obra associada ao bem econômico 4. O conjunto Y_j indica o conjunto de possíveis decisões para a firma j . Por exemplo, se são necessárias pelo menos 3 unidades de mão-de-obra associada ao bem econômico 4 para cada unidade do bem 2 produzido, então a equação $3y_{32} \leq y_{34}$ está contemplada por Y_j . Tomando como dados os preços dos bens econômicos, os quais indicaremos pelo vetor $p \in \mathbb{R}^l$, o problema da unidade produtiva j seria de maximizar a receita dentro dos possíveis planos de produção, ou seja:

$$\begin{aligned} & \text{Maximizar } p^T y_j \\ & y_j \\ & \text{sujeito a: } y_j \in Y_j. \end{aligned} \tag{7.4}$$

Para cada um dos t consumidores, os quais não podem produzir bens econômicos a não ser mão de obra, vamos denotar a variável de decisão por $x_i \in X_i \subset \mathbb{R}^l$. Cada coordenada de x_i indica a quantidade do respectivo bem econômico adquirida pelo consumidor, e portanto as únicas coordenadas que podem estar abaixo de uma dotação inicial em x_i serão aquelas associadas à prestação de serviço do indivíduo i , as quais serão não negativas. Para tratar o caso de consumidores que negociam seus pertences podemos criar uma firma cujo dono seja o indivíduo i e que não crie nenhum bem, apenas negocie os existentes. Vamos considerar como patrimônio inicial do indivíduo i o vetor $\xi_i \in \mathbb{R}^l$, cujas entradas negativas indicam um dívida inicial do indivíduo i . A porção da empresa j pertencente ao consumidor i será denotada por $\alpha_{ij} \geq 0$, de modo que $\sum_{i=1}^t \alpha_{ij} = 1$. O conjunto X_i representa todos os possíveis vetores de decisão do indivíduo i , sem a restrição orçamentária. Por exemplo, X_i limita a compra de um determinado bem à quantidade disponível no mercado e a quantidade de serviço prestado a um limite de horas diárias dado pela legislação e/ou pela capacidade física do indivíduo. Já a restrição orçamentária é que, tomando os preços como dados, os gastos sejam menores que a receita, ou seja:

$$p^T x \leq \sum_{j=1}^s \alpha_{ij} p^T y_j.$$

Como a firma artificial j' considerada no caso do consumidor que negocia seus pertences tem como conjunto viável $\{y \in \mathbb{R}^l; y \leq \xi_i\}$ e o vetor de preços é não negativo temos que a decisão ótima é $y' = \xi_i$. Como o indivíduo i é o único proprietário da firma j' temos que $\alpha_{i,j'} = 1$, logo, sua receita com possíveis negociações do patrimônio inicial seria $p^T \xi_i$. Dessa forma podemos desconsiderar o problema da empresa j' e incluir explicitamente a receita inicial do consumidor i em seu patrimônio. Portanto, se considerarmos como sendo $u_i(x_i)$ a função de utilidade do consumidor i , seu problema, parametrizado pelo

vetor de preços e decisões das unidades produtivas, pode ser escrito como:

$$\begin{aligned} & \text{Maximizar } u_i(x_i) \\ & \quad x_i \\ & \text{sujeito a: } x_i \in X_i, \\ & \quad p^T x \leq p^T \xi_i + \sum_{j=1}^s \alpha_{ij} p^T y_j. \end{aligned} \tag{7.5}$$

Tendo em mente esses problemas diremos que (x^*, y^*, p^*) é dito um equilíbrio Walrasiano se satisfaz as seguintes condições:

EW-1. Para todo $j \in \{1, \dots, s\}$, y_j^* é solução do problema (7.4) da j -ésima unidade produtiva para $p = p^*$.

EW-2. Para todo $i \in \{1, \dots, t\}$, x_i^* é solução do problema (7.5) do i -ésimo consumidor para $y = y^*$ e $p = p^*$.

EW-3.

$$p^* \geq 0, \quad z^* \leq 0 \quad \text{e} \quad p^{*T} z^* = 0,$$

onde

$$z \equiv z(x, y) = \sum_{i=1}^t (x_i - \xi_i) - \sum_{j=1}^s y_j$$

é a variável de excesso de demanda e $z^* = z(x^*, y^*)$.

Uma observação importante é que como $p^* \geq 0$ e $z^* \leq 0$, o produto escalar $p^{*T} z^* = \sum_{h=1}^l p_h^* z_h^*$ é uma soma de termos não positivos. Portanto, $p^{*T} z^* = 0$ se e somente se cada termo $p_h^* z_h^*$ for nulo para todo h . Como consequência temos que numa situação de equilíbrio $z_h^* < 0$ somente se $p_h^* = 0$. Ou seja, no equilíbrio é possível atender a demanda de todos os consumidores ($z^* \leq 0$) e só há excesso de oferta ($z_h^* < 0$) se o preço do bem econômico for nulo. Isso reflete a lei da oferta e da procura que diz que se há um excesso de demanda então o preço do bem sobe de forma a eliminar tal procura abundante. Por outro lado, se há uma disponibilidade maior de um bem do que a quantidade demandada por ele, o preço do bem cai, estimulando assim um aumento na demanda. Como o preço de um produto é não negativo, mesmo que haja

excesso de oferta de um bem econômico o preço desse produto não poderá cair mais se já for nulo, ou seja se já estiver no seu limitante inferior. Portanto, numa situação de equilíbrio, só pode haver uma de duas possibilidades. Ou deve ocorrer a igualdade entre a demanda e a oferta para cada bem econômico, ou o bem econômico não tem valor, circunstância em que poderá ocorrer excesso de oferta. Obviamente a primeira situação é mais comum, já que não esperamos preços nulos em sistemas competitivos.

A seguir vamos apresentar duas formulações imediatas para tratar o problema de encontrar pontos de equilíbrio. A primeira consiste em reformular a condição (EW-3) considerando um problema de otimização auxiliar onde um agente artificial, que representa o mercado, resolve o seguinte problema:

$$\begin{aligned} & \text{Maximizar } p^T z^* \\ & p \\ & \text{sujeito a: } p \geq 0. \end{aligned} \tag{7.6}$$

Vamos dizer que (x^*, y^*, p^*) satisfaz EW-3' se p^* é solução de (7.6). A seguir vamos demonstrar a equivalência entre EW-3 e EW-3'.

Teorema 7.1.1. *Uma tripla $(x^*, y^*, p^*) \in \mathbb{R}^{tl+sl+l}$ satisfaz EW-3 se e somente se satisfaz EW-3'.*

Demonstração: Se (x^*, y^*, p^*) satisfaz tanto EW-3 quanto EW-3' temos que $z^* \leq 0$ pois esta condição está explícita em EW-3 e é necessária para que o problema (7.6) seja limitado, ou seja, que apresente solução $p^* \in \mathbb{R}^l$. Se p^* é solução de (7.6) então $p^{*T} z^* \geq p^T z^*$ para todo $p \geq 0$, em particular

$$p^{*T} z^* \geq 0^T z^* = 0. \tag{7.7}$$

Por outro lado, como $z^* \leq 0$ e $p^* \geq 0$ temos que

$$p^{*T} z^* \leq 0 \tag{7.8}$$

e portanto a função objetivo de (7.6) é limitada superiormente por zero. Sendo assim, por (7.7) e (7.8), temos que p^* é solução de (7.6) se e somente se $p^{*T} z^* = 0$. Deste

modo concluímos que (EW-3) e (EW-3)' são equivalentes. \square

Um detalhe a ser ressaltado aqui é que $p = 0$ é uma solução de (7.6), entretanto o vetor de preços no equilíbrio nunca será nulo se estivermos trabalhando sob a hipótese de que a função utilidade de algum dos consumidores seja não saciada. Tal hipótese diz que se não tivermos a restrição orçamentária, então para qualquer vizinhança aberta de x existe x' nessa vizinhança tal que $u(x') > u(x)$. Por exemplo, um consumidor sempre gostaria de adquirir mais de um certo produto se isso não compromettesse seu orçamento. Desse modo se $p = 0$ os consumidores não teriam restrições orçamentárias e portanto o problema (7.5) não teria solução e conseqüentemente nenhuma tripla $(x, y, 0)$ seria um equilíbrio Walrasiano.

Considerando a equivalência entre EW-3 e EW-3' temos que uma maneira de obter um equilíbrio Walrasiano é resolver um problema de equilíbrio de Nash cujos agentes do jogo são as s unidades produtivas, os t consumidores e o mercado, e seus respectivos problemas de máxima satisfação são dados por (7.4), (7.5) e (7.6).

A segunda alternativa para obter pontos de equilíbrio Walrasiano baseia-se no fato de não esperarmos preços nulos, e portanto restringe-se ao fato de procurarmos pontos onde a oferta coincide com a demanda e que cumpram as propriedades (EW-1) e (EW-2). Ou seja, estamos procurando um trio (x^*, y^*, p^*) que seja solução do problema:

$$\begin{aligned} & \text{Minimizar} \quad \left\| \sum_{i=1}^t (x_i - \xi_i) - \sum_{j=1}^s y_j \right\|^2 \\ & \quad x, y, p \\ & \text{sujeito a:} \quad y_j \text{ é solução de (7.4) para todo } j \in \{1, \dots, s\}, \\ & \quad x_i \text{ é solução de (7.5) para todo } i \in \{1, \dots, t\}. \end{aligned} \tag{7.9}$$

Essa abordagem é interessante pois retrata um exemplo da situação que muito nos interessa, e que será investigada no Capítulo 8, que é o problema com restrições de equilíbrio. Nesse problema, fixado p , as restrições exigem que (x, y) seja um equilíbrio de Nash. Essa interpretação nos remete também à questão da estabilidade do Equilíbrio Walrasiano, que também discutiremos brevemente na Seção 7.3. Essa situação estaria associada a um processo iterativo no qual os preços dos bens econômicos

iriam se ajustando de forma que ao longo do processo as decisões dos agentes, firmas e consumidores, convergiriam para um ponto de equilíbrio.

A seguir vamos apresentar alguns exemplos simples dessa formulação. No primeiro exemplo que abordaremos, apenas uma firma e um consumidor estão se relacionando. Suponhamos que o único bem produzido pela firma seja desejado pelo consumidor de acordo com uma preferência não saciada, por exemplo $u(x) = x$. Consideremos também que o consumidor tenha uma quantidade de dinheiro em espécie igual a 100 reais. Para a firma, as hipóteses são que a quantidade produzida não dependa de nenhum bem material ou humano e portanto que a capacidade de produção seja ilimitada. Dessa forma os problemas das partes seriam:

- Consumidor:

$$\begin{aligned} &\text{Maximizar } x \\ &x \\ &\text{sujeito a: } px \leq 100 \\ &x \geq 0 \end{aligned}$$

- Firma:

$$\begin{aligned} &\text{Maximizar } py \\ &y \\ &\text{sujeito a: } y \geq 0 \end{aligned}$$

Para analisar as possíveis configurações para esse sistema vamos dividir em dois casos possíveis, $p = 0$ e $p > 0$. No primeiro caso notemos que se o preço é nulo o problema do consumidor é ilimitado, e para a firma é indiferente o quanto produzir. Nesse caso não temos pontos de equilíbrio. Uma situação de convergência assintótica seria quando a decisão do consumidor seja denotada por x_k e a da firma por y_k com $x_k < y_k$ e $x_k \rightarrow \infty$. Nesse caso teríamos excesso de oferta em cada período e portanto o preço se manteria nulo. Já o consumo aumentaria a cada período ($x_{k+1} > x_k$), e como a firma é indiferente à quantidade produzida, pode ocorrer que haja excesso de oferta em todos os períodos o que, dito com falta de rigor, manteria o ciclo convergente para o equilíbrio ilimitado.

Para o caso em que $p > 0$ a decisão ótima do consumidor seria $x^* = \frac{100}{p}$ mas o problema da firma seria ilimitado. Portanto não existiriam situações de equilíbrio. Para o caso assintótico podemos ver que esse caso convergirá para o anterior pois dado $p_k > 0$ é natural supor que $x_k = \frac{100}{p_k}$ e $y_k \gg x_k$. Como teremos excesso de oferta o preço p_k convergirá a zero, o que pode ser visto se igualarmos x_k e y_k ($\frac{100}{p_k} \approx \infty$). Desse modo, x_k também tomaria valores cada vez mais elevados, resultando na situação anterior.

O exemplo anterior não apresenta um equilíbrio pois desrespeita uma hipótese muito natural que consiste em limitar a capacidade de produção de cada firma. Vejamos o que acontece se acrescentarmos a restrição para a firma que sua produção deva obedecer $0 \leq y \leq 10$. Nesse caso, ainda não poderíamos ter o preço de equilíbrio nulo, pois o problema do consumidor continuaria ilimitado. Entretanto, tampouco a situação ilimitada poderia ocorrer agora, já que se supormos que x_k converge para infinito não podemos ter que $y_k > x_k$, já que y é limitado superiormente. Nesse caso a partir de certo momento haveria excesso de procura e portanto o preço p subiria, deixando de ser nulo. Para o caso em que $p > 0$ teríamos que $x^* = \frac{100}{p}$ e $y^* = 10$. Igualando x^* e y^* obteríamos o preço de equilíbrio $p^* = 10$ e portanto teríamos um equilíbrio finito $(x^*, y^*, p^*) = (10, 10, 10)$.

No próximo exemplo consideramos um modelo um pouco mais elaborado onde incluiremos mais um agente, o dono da firma. Mais ainda, vamos supor que a produção de cada unidade do produto fabricado pela firma necessite de mão de obra para ser construído. As variáveis que usaremos nesse exemplo estão descritas a seguir.

- x - Quantidade de horas trabalhadas pelo consumidor.
- y - Quantidade de horas trabalhadas pelo proprietário da firma.
- z - Quantidade de unidades do produto adquirida pelo consumidor.
- w - Quantidade de unidades do produto adquirida pelo proprietário.
- s - Quantidade de unidades do produto produzida pela firma.

- t - Quantidade de horas de trabalho do consumidor contratadas pela firma.
- r - Quantidade de horas de trabalho do proprietário contratadas pela firma.
- p_1 - Preço da hora de trabalho do consumidor.
- p_2 - Preço da hora de trabalho do proprietário.
- p_3 - Preço de uma unidade do produto.

Considerando que a satisfação do consumidor em adquirir o produto seja dez vezes maior do que sua aversão ao trabalho e que a legislação limite o tempo de trabalho em 100 horas no período considerado, temos que o problema do consumidor seria:

$$\begin{aligned} &\text{Maximizar} && z - \frac{x}{10} \\ &&& x, z \\ &\text{sujeito a:} && p_3 z \leq p_1 x \\ &&& 0 \leq x \leq 100 \\ &&& z \geq 0. \end{aligned}$$

Para o proprietário vamos considerar que sua satisfação em adquirir o produto seja cinco vezes maior do que sua aversão ao trabalho, que também está limitado à 100 horas. Dessa forma o seu problema seria:

$$\begin{aligned} &\text{Maximizar} && w - \frac{y}{5} \\ &&& w, y \\ &\text{sujeito a:} && p_3 w \leq p_2 y + p_3 s - p_1 t - p_2 r \\ &&& 0 \leq y \leq 100 \\ &&& w \geq 0. \end{aligned}$$

Já para a firma vamos supor que a quantidade de unidades produzidas seja limitada pela quantidade de horas contratadas do consumidor mais duas vezes as do proprietário e que essas horas também estejam limitadas pela legislação como nos pro-

blemas do consumidor e do proprietário. Sendo assim o problema da firma é:

$$\begin{aligned} &\text{Maximizar} && p_3s - p_1t - p_2r \\ &&& s, t, r \\ &\text{sujeito a:} && 0 \leq s \leq t + 2r \\ &&& 0 \leq t \leq 100 \\ &&& 0 \leq r \leq 100. \end{aligned}$$

Se $p_3 = 0$ então o problema do consumidor seria ilimitado, portanto não teríamos situações de equilíbrio. Desse modo vamos assumir que $p_3 > 0$ de agora em diante. A resposta do consumidor está associada à relação entre p_1 e p_3 . Caso $\frac{p_1}{p_3} < \frac{1}{10}$ então o consumidor chega à conclusão que não vale a pena trabalhar. Isso acontece pois o dinheiro ganho com o trabalho não é suficiente para compensar sua aversão ao trabalho em relação à satisfação do consumo. Ou seja ele teria que trabalhar mais do que 10 horas para conseguir comprar uma unidade do produto, portanto sua função utilidade nesse caso seria negativa. Dessa forma, a decisão ótima do consumidor, no caso em que $\frac{p_1}{p_3} < \frac{1}{10}$, seria $x^* = z^* = 0$. Analogamente, se $\frac{p_1}{p_3} = \frac{1}{10}$ o consumidor seria indiferente sobre quanto trabalhar e o conjunto das soluções ótimas seria $\{(x, \frac{x}{10}); x \in [0, 100]\}$. Quando a relação dos preços supera a relação de satisfação e aversão, ou seja $\frac{p_1}{p_3} > \frac{1}{10}$, então a decisão ótima do consumidor é trabalhar o máximo possível para poder consumir também o máximo possível. Nesse caso teríamos que $x^* = 100$ e $z^* = \frac{p_1}{p_3}100$.

Para o proprietário da firma a análise pode ser feita da mesma forma, apenas substituindo a relação de aversão por $\frac{1}{5}$ e acrescentando a receita da firma em seu orçamento. Desse modo concluímos que a decisão ótima do proprietário seria $y^* = 0$ e $w^* = \frac{p_3s - p_1t - p_2r}{p_3}$ se $\frac{p_2}{p_3} < \frac{1}{5}$; $y^* \in [0, 100]$ arbitrário e $w^* = \frac{y^*}{5} + \frac{p_3s - p_1t - p_2r}{p_3}$ se $\frac{p_2}{p_3} = \frac{1}{5}$; e $y^* = 100$ e $w^* = \frac{p_2}{p_3}100 + \frac{p_3s - p_1t - p_2r}{p_3}$ caso $\frac{p_2}{p_3} > \frac{1}{5}$.

Agora vamos analisar o problema da firma já identificando os casos de equilíbrio. Para $p_3 > p_1$ temos que $t^* = 100$, para $p_3 = p_1$ temos que qualquer $t^* \in [0, 100]$ é solução e para $p_3 < p_1$ temos que $t^* = 0$. Como numa situação de equilíbrio o número de horas trabalhadas pelo consumidor deve ser igual ao número contratado pela firma, ou seja $x^* = t^*$, chegamos a conclusão que a única situação

possível é que $p_1 \leq p_3 \leq 10p_1$ e $x^* = t^* = 100$. Analisando agora as relações envolvendo as horas trabalhadas pelo proprietário concluímos que no equilíbrio devemos ter $\frac{p_2}{2} \leq p_3 \leq 5p_2$ e $y^* = r^* = 100$. Dessa forma verificamos, a partir dos problemas da firma, do consumidor e do proprietário, respectivamente, que $s^* = 300$, $z^* = \frac{p_1}{p_3}100$, $w^* = 300 - \frac{p_1}{p_3}100$. Como a última relação de equilíbrio, $z^* + w^* = s^*$, também é verificada obtemos os pontos de equilíbrio.

No último exemplo que apresentaremos modificamos a função de utilidade do consumidor e do proprietário para que essas se assemelhem mais à função de Cobb-Douglas [37] que é uma das mais aceitas em modelos econômicos. Nessa funções os fatores de interesse estão relacionados de acordo com um produto e não com uma soma. Essa formulação faz com que a distribuição entre grupos de bens econômicos seja geométrica, mostrando que nem todo grupo pode ser substituído por outro. Usualmente a função de Cobb-Douglas é descrita na forma

$$f(x) = x_1^{a_1} \times x_2^{a_2} \times \cdots \times x_n^{a_n},$$

com $\sum_{i=1}^n a_i = 1$. No nosso exemplo vamos relaxar a condição da soma e considerar como função utilidade do consumidor e do proprietário como sendo, respectivamente, $z^{10}(100 - x)$ e $w^5(100 - y)$. Portanto o problema do consumidor passa a ser

$$\begin{aligned} &\text{Maximizar} && z^{10}(100 - x) \\ &&& x, z \\ &\text{sujeito a:} && p_3 z \leq p_1 x \\ &&& 0 \leq x \leq 100 \\ &&& z \geq 0. \end{aligned}$$

Considerando os preços p_1 , p_2 e p_3 não nulos e analisando as condições KKT temos que

$x^* = \frac{10}{11}100$ e $z^* = \frac{p_1}{p_3} \frac{10}{11}100$. Do mesmo modo temos que o problema do proprietário é

$$\begin{aligned} &\text{Maximizar } w^5(100 - y) \\ &\quad w, y \\ &\text{sujeito a: } p_3 w \leq p_2 y + p_3 s - p_1 t - p_2 r \\ &\quad 0 \leq y \leq 100 \\ &\quad w \geq 0. \end{aligned}$$

e a solução é $y^* = \frac{5}{6}100 - \frac{1}{6} \left(\frac{p_3}{p_2} s - \frac{p_1}{p_2} t - r \right)$ e $w^* = \frac{p_2}{p_3} 5(100 - y^*)$. Para que o problema da firma tenha solução com t e r afastados dos extremos devemos ter $p_3 = p_1 = \frac{p_2}{2}$. Nesse caso podemos escolher t e r de forma a satisfazer a igualdade entre horas trabalhadas e contratadas e a expressão de s é dada por $s = t + 2r$ quaisquer que sejam t e r escolhidos. Portanto $x^* = z^* = t^* = \frac{10}{11}100$, $y^* = r^* = \frac{5}{6}100$, $w^* = \frac{10}{6}100$ e $s^* = \frac{85}{33}100$. Como mais uma vez temos que $z^* + w^* = s^*$ temos que esse é um ponto de equilíbrio. O único grau de liberdade no preço dos bens econômicos pode ser visto como apenas um fator de escala da moeda utilizada.

7.2 Introdução às desigualdades variacionais

Nesta seção vamos fazer um rápido resumo sobre os aspectos relacionados às desigualdades variacionais que consideramos mais importantes encontrados na bibliografia que estudamos (sobretudo [49, 50]). Esse tema é o mais tradicional para abordar problemas de Equilíbrio, mas não será usado como ferramenta ao longo dos capítulos posteriores. Entretanto acreditamos que seja possível combinar ideias relativas aos métodos de desigualdades variacionais com as técnicas que desenvolvemos. Levando em consideração a relevância desse tema e as possíveis adaptações futuras das ferramentas introduzidas neste texto, decidimos exibir esse breve resumo sobre o tema.

Em [49] encontramos o seguinte conceito:

Definição: Dados $K \subset \mathbb{R}^n$ e $F : K \rightarrow \mathbb{R}^n$ o problema $VI(K, F)$ consiste em encontrar

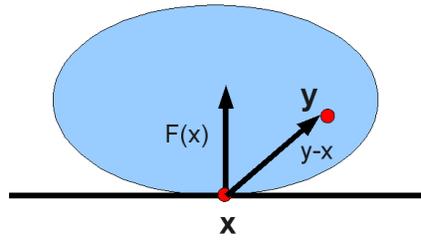


Figura 7.4: Ângulo agudo entre $F(x)$ e todas as direções viáveis.

$x \in K$ tal que

$$(y - x)^T F(x) \geq 0, \quad \forall y \in K. \quad (7.10)$$

O conjunto solução do problema é denotado por $SOL(K, F)$.

Uma interpretação geométrica da definição anterior no caso de K convexo é que queremos encontrar $x \in K$ tal que $F(x)$ faça um ângulo agudo com toda direção viável em x , ou seja, vetores que somados a x ainda permaneceriam em K (Ver Figura 7.4).

Esta interpretação geométrica mostra claramente uma relação entre desigualdades variacionais e otimização quando K é convexo. Se considerarmos $F(x) = \nabla f(x)$ a condição acima é evidentemente a condição necessária de primeira ordem para que um ponto x seja um minimizador local de f em K . No caso em que f também é convexa a condição é também suficiente.

Posteriormente vamos fazer colocações sobre o caso em que K não é convexo e fazer uma nova definição do conceito de desigualdades variacionais que generalizará a definição dada por (7.10). Antes disso vejamos diversas aplicações que podem ser modeladas como um problema de desigualdades variacionais.

O primeiro exemplo que vemos é o de um sistema de equações não lineares. Encontrar $x \in \mathbb{R}^n$ tal que $F(x) = 0$ é equivalente a encontrar $x \in SOL(\mathbb{R}^n, F)$. Vamos demonstrar essa afirmação. O fato de que se $F(x) = 0$ então $x \in SOL(\mathbb{R}^n, F)$ é evidente.

A recíproca é verdadeira pois se $y = -F(x) + x$ então $(y - x)^T F(x) = -\|F(x)\|^2$, logo, para que este produto interno seja não negativo devemos ter que $F(x) = 0$.

Como segundo exemplo, um sistema de desigualdades não lineares $F(x) \leq 0$ também pode ser escrito como um problema de desigualdades variacionais. Para isso introduzimos as variáveis artificiais $z \in \mathbb{R}^n$ e procuramos soluções para a desigualdade variacional associada à função $\bar{F}(x) = \begin{pmatrix} F(x) + z^2 \\ 0 \end{pmatrix}$ em $K = \mathbb{R}^{2n}$. A equivalência é imediata pois $\bar{F}(x) = 0 \Leftrightarrow F(x) \leq 0$.

O problema de complementaridade, definido em [49] como o caso em que K é um cone (ou seja, K é tal que $\alpha x \in K$ para todo $\alpha \geq 0$ e $x \in K$), fornece outros exemplos. Neste caso o conjunto solução do problema será o conjunto:

$$\{x \in K; x^T F(x) = 0 \text{ e } y^T F(x) \geq 0, \quad \forall y \in K\}.$$

Podemos reescrever facilmente a afirmação acima denotando por

$$K^* \equiv \{d \in \mathbb{R}^n; d^T v \geq 0 \quad \forall v \in K\}$$

o cone dual de K . Sendo assim teríamos que caso K seja um cone então

$$x \in SOL(K, F) \Leftrightarrow x \in K, \quad x^T F(x) = 0 \text{ e } F(x) \in K^*. \quad (7.11)$$

A seguir vamos demonstrar esse fato.

Demonstração: Primeiramente suponhamos que $x \in SOL(K, F)$, o que implica que $x \in K$. Pelo fato de K ser um cone, temos que 0 e $2x$ também estão em K . Portanto, considerando $y \in \{0, 2x\}$ em (7.10) obtemos que $-x^T F(x) \geq 0$ e $x^T F(x) \geq 0$, logo $x^T F(x) = 0$. Além disso, para qualquer que seja $y \in K$, temos que $y^T F(x) = (y - x)^T F(x) \geq 0$ e portanto vale que $F(x) \in K^*$. Desse modo obtemos a primeira implicação:

$$x \in SOL(K, F) \Rightarrow x \in K, \quad x^T F(x) = 0 \text{ e } F(x) \in K^*.$$

Para a implicação contrária basta observar que $-x^T F(x) = 0$ e que se $y \in K$ então $y^T F(x) \geq 0$. Logo, somando essas relações obtemos que $(y - x)^T F(x) \geq 0$ e, como

$x \in K$ por hipótese, concluímos que $x \in SOL(K, F)$. \square

No caso especial em que $K = \mathbb{R}_+^n$ temos que o cone dual $K^* = \mathbb{R}_+^n$, e portanto o conjunto solução seria formado pelos vetores x tais que $0 \leq x \perp F(x) \geq 0$, o que implica que $x_i, F_i(x) \geq 0$ e $x_i F_i(x) = 0$ para todo i . Essa situação é chamada de problema de complementaridade não linear.

O caso em que $K = \mathbb{R}^{n_1} \times \mathbb{R}_+^{n_2}$ e $F : \mathbb{R}^{n_1} \times \mathbb{R}_+^{n_2} \rightarrow \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$ é decomposta da forma $F(u, v) = (G(u, v), H(u, v))$, onde $G : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \rightarrow \mathbb{R}^{n_1}$ e $H : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \rightarrow \mathbb{R}^{n_2}$, é dito um problema de complementaridade mista. Neste caso $SOL(K, F) = \{(u, v); G(u, v) = 0 \text{ e } 0 \leq v \perp H(u, v) \geq 0\}$. Demonstremos esse fato.

Demonstração: Se $x = (u, v) \in SOL(K, F)$ então

$$0 \leq (y - x)^T F(x) = (y_1 - u)^T G(u, v) + (y_2 - v)^T H(u, v).$$

Tomando y da forma $(-G(u, v) + u, v)$ temos que $y \in K$ e $G(u, v) = 0$. Se y é da forma $(u, 2v)$ também temos que $y \in K$ e que $v^T H(u, v) \geq 0$. Já para y da forma $(u, 0)$ temos mais uma vez que $y \in K$ e que $v^T H(u, v) \leq 0$, portanto $v^T H(u, v) = 0$. Denotando e_i um elemento da base canônica de \mathbb{R}^{n_2} temos que $(u, v + e_i)$ pertence a K e conseqüentemente $H(u, v) \geq 0$. Dessa forma concluímos que $0 \leq v \perp H(u, v) \geq 0$. Reciprocamente, como $(u, v) \in \mathbb{R}^{n_1} \times \mathbb{R}_+^{n_2}$, temos que se y pertence a K então

$$(y_1 - u)^T G(u, v) + (y_2 - v)^T H(u, v) = y_2^T H(u, v) \geq 0.$$

Sendo assim concluímos que $(u, v) \in SOL(K, F)$ e que a equivalência é verdadeira. \square

Uma das aplicações mais claras do problema de complementaridade mista é a resolução de um sistema KKT proveniente de um problema de otimização. Se considerarmos o problema:

$$\begin{aligned} & \text{Minimizar } f(x) \\ & \quad x \in \mathbb{R}^n \\ & \text{sujeito a: } h(x) = 0, \\ & \quad g(x) \leq 0, \end{aligned}$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$, o sistema KKT associado seria equivalente a resolver o problema de complementaridade mista com $G((x, \lambda), \mu) = (\nabla f(x) + J_h(x)^T \lambda + J_g(x)^T \mu; h(x))$, $H((x, \lambda), \mu) = -g(x)$ e $K = \mathbb{R}^{n+m} \times \mathbb{R}_+^p$.

Por fim vamos formular a aplicação de desigualdades variacionais que talvez seja a mais importante em trabalhos relacionados com o exposto nesse texto. Vamos mostrar que as condições de otimalidade para os problemas de otimização envolvidos numa situação de equilíbrio de Nash podem se reduzir a um problema de desigualdades variacionais. Se considerarmos todas as funções objetivo envolvidas (f_i s) sendo convexas, $K = K_1 \times K_2 \times \dots \times K_q \subset \mathbb{R}^{n_1+n_2+\dots+n_q}$ como sendo o conjunto de situações possíveis do jogo, $F : K \rightarrow \mathbb{R}^{n_1+n_2+\dots+n_q}$, como $F(x) = (\nabla f_1(x); \nabla f_2(x); \dots; \nabla f_n(x))$ e K convexo (o que acontece se e somente se cada K_i for convexo) então x será um equilíbrio de Nash se e somente se $x \in SOL(K, F)$.

Demonstração: Se x é um equilíbrio de Nash então x_i é solução de (7.1) para todo $i \in \{1, 2, \dots, n\}$, portanto, pela convexidade de K_i e f_i devemos ter $(y_i - x_i)^T \nabla f_i(x) \geq 0$ para todo $y_i \in K_i$ e todo $i \in \{1, 2, \dots, n\}$. Somando essas desigualdades obtemos $(y - x)^T F(x) \geq 0$. Para provar a recíproca basta considerar os vetores y da forma $y = (y_i, x_{-i})$, onde $y_i \in K_i$ é arbitrário, e teremos que

$$\begin{aligned} 0 &\leq (x_1 - x_1)^T \nabla f_1(x) + \dots + (y_i - x_i)^T \nabla f_i(x) + \dots + (x_n - x_n)^T \nabla f_n(x) \\ &\leq (y_i - x_i)^T \nabla f_i(x). \end{aligned}$$

Fazendo isso para todo $i \in \{1, 2, \dots, n\}$, a convexidade dos problemas de cada agente garante que x_i seja solução (7.1) para todo i e, conseqüentemente, que x seja um equilíbrio de Nash. \square

Nos parágrafos anteriores mostramos uma imensa gama de problemas que podem ser vistos como casos particulares de problemas de desigualdades variacionais. Por esse motivo parece ser muito importante estudar formulações e extensões desse conceito assim como métodos para solucioná-las.

Como conseqüência do fato de que $(y - x)^T F(x) \geq 0$, e que a igualdade

é atingida quando $y = x$ podemos fazer uma reformulação muito útil da definição de desigualdades variacionais. Explicitamente obtemos que $x \in SOL(K, F)$ se e somente se $x \in K$ e é solução de:

$$\begin{aligned} & \text{Minimizar } (y - x)^T F(x) \\ & y \end{aligned} \tag{7.12}$$

sujeito a: $y \in K$.

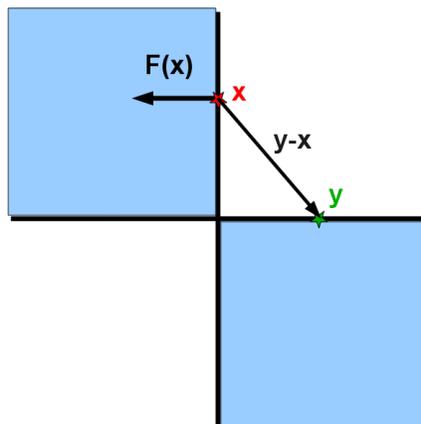
Se redefinirmos $SOL(K, F)$ exigindo apenas que seus elementos sejam minimizadores locais de (7.12) não estaríamos modificando a definição anterior no caso em que K é convexo. Essa equivalência é consequência direta do fato que o problema (7.12) tem função objetivo linear e conjunto viável convexo, portanto todo minimizador local é minimizador global.

Para ilustrar essa extensão para problemas em que K não é convexo vamos analisar o que acontece em um exemplo particular. Consideraremos neste exemplo que

$$K = \{(x_1, x_2) \in \mathbb{R}^2; x_1 x_2 \leq 0\}.$$

Para verificar se um ponto de K pertence a $SOL(K, F)$ temos que considerar três casos. O primeiro deles é se x é um ponto interior a K , portanto existe $\epsilon > 0$ tal que $\|y - x\| \leq \epsilon$ implica que y pertence a K . Nesse caso devemos ter que $v^T F(x) \geq 0$ para todo v tal que $\|v\| \leq \epsilon$ e conseqüentemente concluímos que $F(x) = 0$. O segundo caso é se $x = 0$. Nessa situação temos que se $y \in K$ então $-y \in K$, e, mais ainda, devemos ter que $y^T F(x) \geq 0$ e $-y^T F(x) \geq 0$ para todo $y \in K$. Sendo assim também temos que deve valer $F(x) = 0$. O terceiro caso é se x está em um dos eixos mas não na origem. Esse é o caso em que a generalização inclui pontos que não satisfaziam a definição anterior. Para esta nova definição o vetor $F(x)$ deve estar entrando perpendicularmente no quadrante em que x está localizado, entretanto a definição com minimização global exige que $F(x)$ seja nulo. Por exemplo, se $x = (0, 1)^T$ e $F(x) = (-1, 0)^T$ então x é um minimizador local de (7.12), mas para $y = (1, 0)^T$ temos $(y - x)^T F(x) = -1 \leq 0$. (Ver Figura 7.5)

Essa nova definição contempla o caso de ser uma generalização de uma condição necessária de otimalidade. Ou seja, se considerarmos $F(x) = \nabla f(x)$ estamos

Figura 7.5: Região viável $xy \leq 0$

exigindo que não exista uma sequência $y_k \in K$ convergindo a x tal que $(y_k - x)^T \nabla f(x) < 0$.

Voltando ao exemplo anterior, se olharmos para as condições de otimalidade do problema:

$$\begin{aligned} &\text{Minimizar } f(x, y) \\ &\text{sujeito a: } xy \leq 0, \end{aligned}$$

claramente deveríamos considerar como condição de otimalidade que nenhuma direção viável seja de descida. Entretanto direções que saem de K não são viáveis e portanto não deve ser consideradas nessa análise.

Outra maneira de ver a definição (7.10) aplicada a problemas não convexos envolveria considerar K como sendo um conjunto $K(x)$ que depende do ponto x em questão. Neste caso o problema é chamado de desigualdades quasi-variacionais. Na situação de interesse $K(x)$ é a translação $x + T(x, K)$, onde

$$T(x, K) \equiv \left\{ d \in \mathbb{R}^n; \exists \{y_k\} \subset \mathbb{R}^n \text{ e } t_k > 0 \text{ tais que } y_k \rightarrow x, t_k \rightarrow 0 \text{ e } \frac{y_k - x}{t_k} \rightarrow d \right\}$$

é o cone tangente de K em x . Uma ilustração dessa situação está apresentada na Figura 7.6. Para esse conjunto $K(x)$ teríamos que, o ponto x ser solução da desigualdade

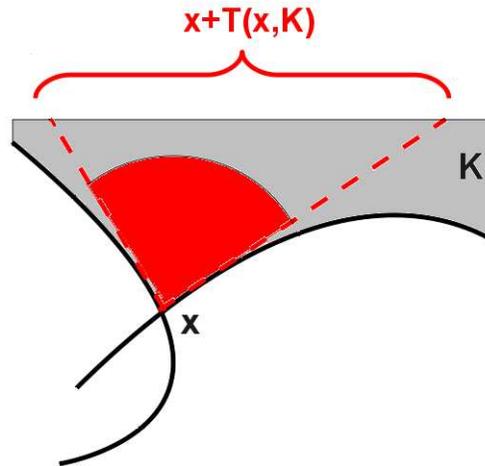


Figura 7.6: Cone tangente de K em x para K não convexo.

quasi-variacional com $F(x) = \nabla f(x)$ é uma condição necessária para este ponto ser um minimizador local do problema de otimização.

Uma situação interessante também modelada por desigualdades quasi-variacionais seria o conceito de equilíbrio de Nash generalizado no caso em que $F(x) = (\nabla f_1(x); \nabla f_2(x); \dots; \nabla f_n(x))$ e $K(x) = \prod_{i=1}^N K_i(x_{-i})$.

Uma observação relevante aqui é que o problema de otimização proposto é teórico e resolvê-lo só nos permite averiguar se x é de fato uma solução do problema de desigualdades variacionais mas não ajuda a encontrar uma solução, pois, para montarmos (7.12) precisamos já conhecer x .

Olhar para o problema de desigualdades variacionais sob o ponto de vista de resolver o problema (7.12) nos dá uma vantagem especial de podermos aplicar resultados conhecidos para problemas de otimização nesse contexto. Talvez o resultado mais importante ocorra quando K é descrito por igualdades e desigualdades, ou seja $K \equiv \{x \in \mathbb{R}^n; h(x) = 0 \text{ e } g(x) \leq 0\}$. Nesse caso, se vale alguma condição de qualificação, sabemos que uma condição necessária para um ponto ser um minimizador local

é que valham as condições KKT, ou seja:

$$\begin{aligned} F(x) + J_h(x)^T \lambda + J_g(x)^T \mu &= 0; \\ h(x) &= 0 \\ 0 \leq \mu \perp -g(x) &\geq 0. \end{aligned} \tag{7.13}$$

Essa formulação nos garante que métodos para resolver problemas de complementaridade mista podem ser usados para resolver as condições KKT de um problema de desigualdades variacionais.

Os métodos mais tradicionais para resolver este tipo de problema consistem em considerar uma função ψ tal que $\psi(a, b) = 0 \Leftrightarrow (a, b) \geq 0$ e $ab = 0$ (essa classe de funções são chamadas de funções C) e converter o problema em resolver um sistema não linear. As funções C mais utilizadas são a função mínimo $\psi_{min}(a, b) = \min\{a, b\}$ e a função de Fischer-Burmeister $\psi_{FB}(a, b) = \sqrt{a^2 + b^2} - (a + b)$. Embora existam funções C continuamente diferenciáveis em todo \mathbb{R}^2 , por exemplo $\psi_{FB}^2(a, b)$, razões teóricas e experimentos computacionais atestam que é preferível trabalhar com funções não diferenciáveis. Por essa razão, métodos de resolução de sistemas não suaves tornam-se importantes na resoluções de desigualdades variacionais.

Vários métodos do tipo Newton e Quasi-Newton são estudados utilizando o conceito de derivadas segundo Clarke, por exemplo em [46]. A razão disso é que várias das funções C conhecidas não só permitem a utilização desse conceito como também explicitam maneiras de como conseguir aproximações de Newton para este tipo de estrutura. Para esses protótipos de métodos são provados teoremas de convergência tanto local como global seguindo o mesmo espírito dos resultados tradicionais para problemas diferenciáveis, assim como resultados relativos à taxa de convergência.

Uma última observação é que o Algoritmo de Restauração Inexata local Birgin-Martínez pode ser aplicado a problemas de desigualdades variacionais.

7.3 Métodos do tipo Jacobi

Nesta seção vamos discutir superficialmente o uso de métodos do tipo Jacobi ([58], [45], [102]) na resolução de sistemas lineares e para encontrar problemas de Equilíbrio de Nash. Primeiramente vamos ilustrar o funcionamento do método graficamente e vamos ressaltar algumas características interessantes para que o método seja bem sucedido. Depois daremos interpretações de que o uso deste tipo de método está implícito na formulação de alguns problemas. Por fim vamos apresentar alguns resultados de convergência para esse tipo de métodos em situações bem particulares.

O método de Jacobi para resolução de sistemas lineares, os quais denotaremos por $Ax = b$, onde $A \in \mathbb{R}^{n \times n}$ e $b \in \mathbb{R}^n$ são dados e $x \in \mathbb{R}^n$ é um vetor de incógnitas, consiste num processo iterativo para aproximarmos a solução (x^*) do sistema. Dado um ponto inicial x^0 a sequência x^k gerada pelo método é dada por:

$$x^{k+1} = D^{-1}(b + Nx^k),$$

onde D é a matriz diagonal cujas entradas são as mesmas de A , as quais vamos supor durante toda esta seção ser não nulas, e $N = D - A$. Ou de uma maneira mais explícita:

$$x_i^{k+1} = \frac{1}{a_{ii}} \left(b_i - \sum_{j \neq i} a_{ij} x_j^k \right).$$

Uma ideia intuitiva do método de Jacobi é que estamos encontrando a i -ésima coordenada de x resolvendo a i -ésima equação do sistema linear considerando como sendo boas aproximações para as demais coordenadas da solução as variáveis correspondentes encontradas no instante anterior do processo. Uma ilustração do funcionamento do método é exposta na Figura 7.7.

Para análise de convergência do método vamos denotar $M = D^{-1}N$ e provar alguns resultados clássicos.

Lema 7.3.1. *Se A for estritamente diagonalmente dominante, isto é, $|a_{ii}| > \sum_{j \neq i} |a_{ij}|$ $\forall 1 \leq i \leq n$, então $\|M\|_\infty < 1$.*

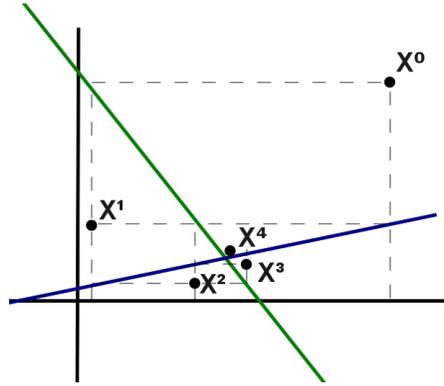


Figura 7.7: Processo iterativo de Jacobi para sistemas lineares.

Demonstração: Seja $v \in \mathbb{R}^n$ tal que $\|v\|_\infty = 1$. Então :

$$\|Mv\|_\infty \leq \max_{1 \leq i \leq n} \left| \sum_{j=1}^n m_{ij} v_j \right| \leq \max_{1 \leq i \leq n} \sum_{j=1}^n |m_{ij}| |v_j| \leq \max_{1 \leq i \leq n} \sum_{j=1}^n |m_{ij}|.$$

Lembrando que $m_{ii} = 0$ e $m_{ij} = -\frac{a_{ij}}{a_{ii}}$ para $j \neq i$, temos que para todo i tal que $1 \leq i \leq n$,

$$\sum_{j=1}^n |m_{ij}| = \sum_{j \neq i} \frac{|a_{ij}|}{|a_{ii}|} = \frac{1}{|a_{ii}|} \sum_{j \neq i} |a_{ij}| < 1.$$

Consequentemente,

$$\|M\|_\infty \leq \max_{\|v\|=1} \{\|Mv\|_\infty\} \leq \max_{1 \leq i \leq n} \sum_{j=1}^n |m_{ij}| < 1.$$

□

Teorema 7.3.1. *Se A é estritamente diagonalmente dominante então A é não singular e os iterandos do método de Jacobi convergem para a solução, qualquer que seja o ponto inicial x^0 .*

Demonstração: Seja $v \in \mathbb{R}^n$ tal que $Av = 0$ e $k \in \{1, 2, \dots, n\}$ tal que $|v_k| = \max_{1 \leq i \leq n} |v_i|$. Provemos então que $v = 0$. De fato, caso contrário teríamos que:

$$|a_{kk}| = \left| - \sum_{j \neq k} a_{kj} \frac{v_j}{v_k} \right| \leq \sum_{j \neq k} |a_{kj}| \frac{|v_j|}{|v_k|} \leq \sum_{j \neq k} |a_{kj}|,$$

o que contraria a hipótese de A ser estritamente diagonalmente dominante. Dessa forma o núcleo de A contém somente o vetor nulo e conseqüentemente A é não singular. Portanto, dado $b \in \mathbb{R}^n$ arbitrário, existe $x^* \in \mathbb{R}^n$ tal que $Ax^* = b$.

Como $Dx^{k+1} = Nx^k + b$ e $Dx^* = Nx^* + b$ temos que $D(x^{k+1} - x^*) = N(x^k - x^*)$, portanto

$$x^{k+1} - x^* = M(x^k - x^*) = M^{k+1}(x^0 - x^*). \quad (7.14)$$

Dessa forma temos que

$$\|x^{k+1} - x^*\|_\infty = \|M^{k+1}(x^0 - x^*)\|_\infty \leq \|M\|_\infty^{k+1} \|x^0 - x^*\|_\infty.$$

Como, pelo Lema 7.3.1, $\|M\|_\infty^k$ converge a zero concluímos que x^k converge à solução do sistema. \square

Da demonstração do Teorema 7.3.1 podemos ver que a convergência é linear com taxa $\|M\|_\infty$. As próximas considerações que faremos são sobre a convergência do resíduo no processo iterativo de Jacobi.

Teorema 7.3.2. *Suponhamos que existe x^* solução do sistema $Ax = b$. Se $x^k - x^*$ está no núcleo de $M \equiv D^{-1}(D - A)$ para algum k , ou se A e M comutam entre si e $\|M\| < 1$, ou se A e $N \equiv D - A$ são não singulares e $\|J\|\|J^{-1}\|\|M\| < 1$, onde $J = AM$, então o resíduo $r^k = Ax^k - b$ converge a zero pelo menos Q -linearmente.*

Demonstração: Como visto na relação (7.14) da demonstração do Teorema 7.3.1 sabemos que $x^{k+1} - x^* = M(x^k - x^*)$ (isto não depende da propriedade de A ser diagonalmente dominante, bastava a existência de uma solução x^*). Logo, se $x^k - x^*$ está no núcleo de M para algum k , teríamos que $x^{k+1} = x^*$ e daí por diante teríamos que $x^s = x^*$, para todo $s > k$. Conseqüentemente $Ax^s - b = 0$ para todo $s > k$ e a tese estaria demonstrada.

Para o segundo caso vamos multiplicar a igualdade (7.14) por A pela esquerda em ambos os lados de forma a obtermos

$$r^{k+1} = Ax^{k+1} - Ax^* = AM(x^k - x^*). \quad (7.15)$$

Portanto, se A e M comutam, teríamos que $r^{k+1} = Mr^k$ o que implica que $\|r^{k+1}\| \leq \|M\|\|r^k\|$. Dessa forma, supondo $\|M\| < 1$, o que acontece se usarmos a norma infinito e A for estritamente diagonalmente dominante, temos a convergência linear dos resíduos.

Para o último caso vamos observar que as matrizes $J = AD^{-1}N$ e $M = D^{-1}N$ são não singulares, pois A e N são invertíveis por hipótese. Dessa forma podemos definir $x^{-1} = M^{-1}x^0 - N^{-1}b$, $e^k = x^k - x^*$ e da relação (7.15) teríamos que $e^{k-1} = J^{-1}r^k$, para todo $k = 0, 1, 2, \dots$. Considerando isso e continuando a desenvolver a relação (7.15) usando (7.14) obteríamos que, para todo $k \geq 0$,

$$\|r^{k+1}\| \leq \|J\|\|e^k\| \leq \|J\|\|M\|\|e^{k-1}\| \leq \|J\|\|M\|\|J^{-1}r^k\| \leq \|J\|\|J^{-1}\|\|M\|\|r^k\|.$$

Como por hipótese $\|J\|\|J^{-1}\|\|M\| < 1$ temos que o teorema é verdadeiro também para este terceiro caso, o que conclui a demonstração. \square

Um comentário pertinente aqui é que os resultados anteriores tratam de condições suficientes e que as suas teses podem ser válidas mesmo quando as hipóteses não sejam satisfeitas. Por exemplo, para a matriz $A = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$ o método de Jacobi atingirá a solução do sistema $Ax = b$ em no máximo duas iterações, independentemente de quais sejam o vetor b e o ponto inicial x^0 . Isso acontece pois a solução do sistema seria $x^* = \begin{pmatrix} b_1 \\ b_2 - 2b_1 \end{pmatrix}$ e a expressão do iterando x^{k+1} seria $x^{k+1} = \begin{pmatrix} b_1 \\ b_2 - 2x_1^k \end{pmatrix}$. Entretanto a matriz A não é diagonalmente dominante, portanto não satisfaz a hipótese do Teorema 7.3.1.

No segundo exemplo vamos considerar a matriz $A = \begin{pmatrix} 3 & 2 \\ 1 & 2 \end{pmatrix}$ e o vetor $b = \begin{pmatrix} 4 \\ 4 \end{pmatrix}$. Dessa forma as expressões dos iterandos do método de Jacobi seriam:

$$x^{2k} = \begin{pmatrix} \frac{x_1^0}{3^k} \\ 2 - \frac{2-x_2^0}{3^k} \end{pmatrix} \text{ e } x^{2k+1} = \begin{pmatrix} \frac{\frac{2}{3}(2-x_2^0)}{3^k} \\ 2 - \frac{\frac{1}{2}x_1^0}{3^k} \end{pmatrix}.$$

Para o cálculo do resíduo obtemos que

$$A(x^{2k} - x^*) = \frac{1}{3^k} A \begin{pmatrix} x_1^0 \\ -(2 - x_2^0) \end{pmatrix} \quad \text{e} \quad A(x^{2k+1} - x^*) = \frac{1}{3^k} A \begin{pmatrix} \frac{2}{3}(2 - x_2^0) \\ -\frac{1}{2}x_1^0 \end{pmatrix}.$$

Logo, para $x^0 = \begin{pmatrix} -3 \\ 1 \end{pmatrix}$, temos que os iterandos são todos diferentes da solução e

$$\frac{\|A(x^{2k+1} - x^*)\|}{\|A(x^{2k} - x^*)\|} = \frac{5}{11} \quad \text{e} \quad \frac{\|A(x^{2k+2} - x^*)\|}{\|A(x^{2k+1} - x^*)\|} = \frac{11}{15}.$$

Portanto o resíduo converge linearmente a zero, e a taxa de convergência é realizada em todas as iterações. Entretanto podemos ver que as hipóteses do Teorema 7.3.2 não são satisfeitas, já que $M = \begin{pmatrix} 0 & -\frac{2}{3} \\ -\frac{1}{2} & 0 \end{pmatrix}$ é não singular e $x^k \neq x^*$, $\forall k$; as matrizes A e M não comutam pois

$$AM = \begin{pmatrix} -1 & -2 \\ -1 & -\frac{2}{3} \end{pmatrix} \neq \begin{pmatrix} -\frac{2}{3} & -\frac{4}{3} \\ -\frac{3}{2} & -1 \end{pmatrix} = MA$$

e o último caso também não é satisfeito pois $\|J\|\|J^{-1}\|\|M\| = 4 \geq 1$.

É importante notar que quanto menor for $\|M\|$ tanto as hipóteses requeridas no Teorema 7.3.1 quanto as no Teorema 7.3.2 têm maior chance de serem satisfeitas. Portanto, como pode ser visto no Lema 7.3.1, quão mais diagonalmente dominante for a matriz A maior a chance de o método de Jacobi ser eficiente tanto para a convergência na variável x quanto no resíduo $Ax - b$. Geometricamente, em \mathbb{R}^2 , dizer que A é diagonalmente dominante significa que a reta descrita pela primeira equação do sistema está próxima de ser vertical e que a descrita pela segunda equação está próxima de ser horizontal.

Uma visão possível é interpretar o método de Jacobi como uma situação onde a equação 1 define x_1 como função de x_2 e, analogamente, a segunda equação define x_2 em função de x_1 . Nesse caso, quanto mais diagonalmente dominante seja A teremos que mais estável será o iterando de Jacobi, no sentido que variações em

x^k causam pouco efeito em x^{k+1} . É muito comum quando se estuda métodos do tipo Jacobi mostrar que em algumas situações é possível reordenar o sistema de equações de modo que a matriz A passa a ser diagonalmente dominante e que portanto o método de Jacobi é convergente. Nós acreditamos que em muitas aplicações reais essa informação de qual equação, ou conjunto de equações, determinam algumas variáveis em função das outras está disponível. Além disso, acreditamos também que em diversas situações do cotidiano seja mais natural que as variáveis escritas em função das demais tenha um comportamento suave do que apresente mudanças bruscas com variação dos parâmetros. Dessa forma, intuímos que muitas vezes o modelo matemático do problema deve poder introduzir a estrutura que assegure um sistema estável. A Figura 7.8 ilustra a diferença na reordenação de um sistema.

Para sistemas não lineares $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ também existem generalizações de métodos do tipo Jacobi e Gauss-Siedel. Muitas vezes a análise de convergência desses métodos são baseadas em teoremas de ponto fixo. Generalizações do conceito de dominância diagonal exigem que a variação na k -ésima coordenada de F seja comandada por variações em x_k em relação às outras coordenadas. Em [82] é encontrado um estudo detalhado sobre uma generalização desse tipo. Um método do tipo SOR que utiliza estratégias quasi-Newton aplicadas em blocos do sistema não linear é introduzido em [77].

Para problemas de equilíbrio de Nash o método de Jacobi consiste em, dada uma aproximação da solução $x^k \in \mathbb{R}^{n_1+n_2+\dots+n_q}$, obter $x^{k+1} \in \mathbb{R}^{n_1+n_2+\dots+n_q}$ tal que x_i^{k+1} é solução de $PEN_i(x_{-i}^k)$ descrito em (7.1) (ver [47]). Os resultados de convergência para esse tipo de método geralmente apresentam hipóteses muito restritivas. O caso mais simples de convergência desse método é apresentado no próximo teorema.

Teorema 7.3.3. *Consideremos um problema de Equilíbrio de Nash em que, para todo i , a função f_i associada ao jogador i seja contínua tanto em relação a x_i quanto a x_{-i} e que o domínio de cada $PEN_i(x_{-i})$ não dependa de x_{-i} . Suponhamos também que $\{x^k\}$ seja uma sequência gerada pelo método de Jacobi com a característica de que x_i^{k+1} é*

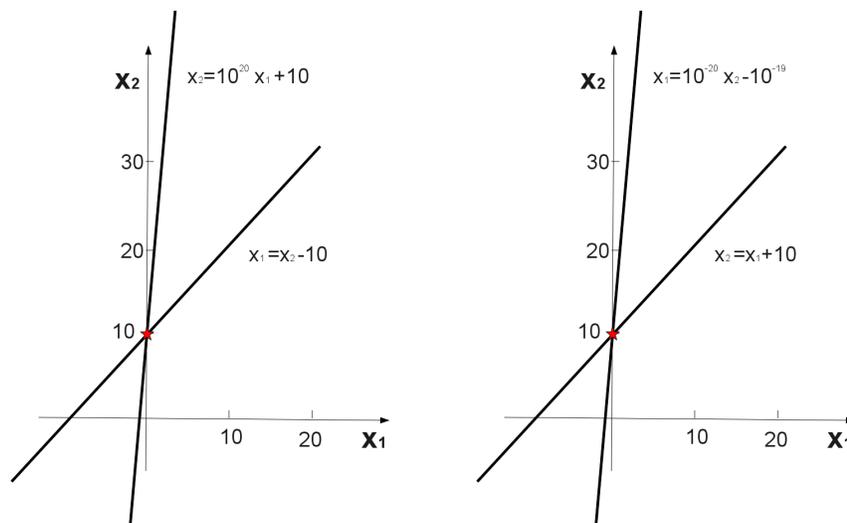


Figura 7.8: Ilustrações para sistema linear envolvendo as equações $10^{20}x_1 - x_2 = -10$ e $x_1 + x_2 = 10$. Na primeira formulação obtemos um sistema em que a variável x_2 é obtida da primeira equação, o que implica que x_2 sofre variações muito bruscas. Na segunda formulação, x_1 é extraído da primeira equação, o que torna a função resultante muito mais suave. Na primeira formulação o método de Jacobi não converge, mas para a segunda situação o método é convergente.

um minimizador global de $PEN_i(x_{-i}^k)$ para todo i e todo k . Então, se $\{x^k\}$ converge a x^* temos que x^* é um ponto de Equilíbrio de Nash para o sistema.

Demonstração: Denotando o conjunto viável do problema do agente i por D_i temos que

$$f_i(x_i^{k+1}, x_{-i}^k) \leq f(y_i, x_{-i}^k); \quad \forall y_i \in D_i,$$

para todo i e k . Passando o limite para k indo para infinito obtemos que

$$f_i(x_i^*, x_{-i}^*) \leq f(y_i, x_{-i}^*); \quad \forall y_i \in D_i.$$

Dessa forma, para todo i , temos que x_i^* é minimizador global de $PEN_i(x_{-i}^*)$. Sendo assim, concluímos que x^* é um ponto de Equilíbrio de Nash para o sistema. \square

Para o caso do Equilíbrio de Nash Generalizado, em que o conjunto viável de $PEN_i(x_{-i})$ depende de x_{-i} , os argumentos de convergência são muito mais elaborados. Sob fortes hipóteses, sobretudo sobre a convexidade, são apresentados resultados de convergência para métodos desse tipo em [48]. Algumas aplicações onde a convergência é garantida são expostas em [51]. Algumas propriedades mais interessantes sobre a convergência podem ser encontradas em métodos que utilizam regularizações, por exemplo os encontrados em [45], [52] e [48]. Mais recentemente, um método do tipo Jacobi com região de confiança foi introduzido em [102].

Para problemas de Arrow-Debreu é muito comum estudar a evolução do excesso de demanda em função do preço dos bens econômicos ($Z(p)$). Um sistema se diz estável quando existe um processo dinâmico tal que os preços dos bens econômicos se ajustam ao longo do tempo até que o excesso de demanda é zerado. Esse é um aspecto muito importante da economia teórica tanto para o caso descritivo quanto para o caso normativo. Um equilíbrio não estável não deve ser atingido em uma situação real e portanto não é tão relevante. Muitas vezes a função de excesso é calculada com um método do tipo Jacobi e o preço é atualizado com um método do tipo Newton para anular $Z(p)$. A estabilidade do sistema está relacionada com a existência e unicidade de

equilíbrios no sistema, assim como de hipóteses relacionadas à dominância da diagonal da Jacobiana de $Z(p)$. Um texto muito completo que aborda esses e outros aspectos do Equilíbrio Geral é o livro de W. Bryant [32].

Capítulo 8

Restauração Inexata para problemas com restrições de equilíbrio

Problemas de programação em dois níveis são aqueles que possuem a particularidade de que uma restrição pode ser descrita pelo fato de uma das variáveis ser solução de outro problema de otimização paramétrico. Em [10] foi proposto um algoritmo de Restauração Inexata no estilo de [80] e [78] para problemas deste tipo. A restauração usa a estrutura de otimização do problema envolvido na restrição e a fase de otimização no espaço tangente usa uma linearização das condições KKT do problema do segundo nível. A linearização das condições KKT serve apenas para manter a “estacionaridade” encontrada na fase de viabilidade. Como a restauração foi feita pensando no problema de otimização, este método provavelmente é mais adequado do que os métodos que desde o início substituem o problema do segundo nível por suas condições KKT. Na Seção 8.1 combinamos as principais ideias expostas em [10] e em [53] de forma a introduzir um novo algoritmo de Restauração Inexata para problemas de programação em dois níveis.

Na Seção 8.2 estendemos essa ideia para problemas com restrições do tipo Equilíbrio de Nash. Essa situação é mais geral pois várias das restrições envolvem

problemas de otimização e, além disso, todos eles são interligados entre si. No algoritmo proposto, a restauração é feita através de métodos do tipo Jacobi e o espaço tangente da fase de otimização é resultante da linearização dos sistemas KKT de todos os problemas envolvidos nas restrições.

O problema que denotamos por Equilíbrio Inverso de Nash é um caso importante dos problemas com restrições de equilíbrio. Nossa motivação para estudar esses problemas vem de duas situações relevantes que tem como restrições problemas de otimização. No primeiro caso de interesse gostaríamos de descobrir os parâmetros usados pelos agentes do sistema em seus respectivos problemas de otimização de forma que uma observação feita a priori seja de fato um Equilíbrio de Nash. Olhado de forma geral, este problema é um problema de quadrados mínimos com restrições de equilíbrio. O segundo caso é a situação normativa onde queremos determinar alguns parâmetros para que o equilíbrio atingido pelo sistema seja o melhor possível para uma função objetivo específica. Este caso pode ser usado para modelar uma situação onde o governo tem que determinar parâmetros para que o mercado se comporte de maneira eficiente para a sociedade. Vamos ilustrar essas situações para problemas de economias de Arrow-Debreu.

8.1 Um novo método de RI para programação em dois níveis

Uma classe especial de problema de programação não linear é chamada de programação em dois níveis. Problemas com esta estrutura aparecem naturalmente provenientes de modelagens nas mais diversas áreas, tais como engenharia mecânica [55], economia e ecologia [23], química [40], etc [41]. Tais problemas são geralmente escritos da seguinte maneira:

$$\begin{aligned}
 & \text{Minimizar } F(x, y) \\
 & \quad x, y \\
 & \text{sujeito a: } H(x) = 0, \\
 & \quad x \in \Omega_x, \\
 & \quad y \text{ é solução do problema:} \tag{8.1} \\
 & \left. \begin{array}{l} \text{Minimizar } f(x, y) \\ \quad y \\ \text{sujeito a: } h(x, y) = 0, \\ \quad y \in \Omega_y, \end{array} \right\} (PSN)
 \end{aligned}$$

onde $F : \mathbb{R}^{n+m} \rightarrow \mathbb{R}$, $H : \mathbb{R}^n \rightarrow \mathbb{R}^q$, $f : \mathbb{R}^{n+m} \rightarrow \mathbb{R}$, $h : \mathbb{R}^{n+m} \rightarrow \mathbb{R}^p$ são funções suaves e $\Omega_x \subset \mathbb{R}^n$ e $\Omega_y \subset \mathbb{R}^m$ são conjuntos simples.

O que chama a atenção nesse tipo de formulação é a restrição que impõe que uma das variáveis (y) seja solução de um outro problema de otimização, parametrizado por x , chamado de problema do segundo nível (PSN). Há duas maneiras clássicas de abordar este problema. A primeira procura escrever a variável y como uma função do parâmetro x . Isso claramente não é sempre possível, uma vez que o conjunto dos minimizadores do PSN pode não ser unitário e, portanto tal função não estaria bem definida. Mesmo assumindo a existência de tal função, podemos ter certos inconvenientes como apresentaremos no próximo exemplo, o qual é descrito em [98] e reproduzido em [35].

$$\begin{array}{l}
 \text{Minimizar } x - 2y \\
 \quad x, y \\
 \text{sujeito a: } -x + 3y - 4 \leq 0, \\
 \quad y \text{ é solução do problema:} \\
 \left. \begin{array}{l}
 \text{Minimizar } x + y \\
 \quad y \\
 \text{sujeito a: } x - y \leq 0 \\
 \quad -x - y \leq 0 \\
 \quad y \geq 0
 \end{array} \right\} (PSN) \tag{8.2}
 \end{array}$$

Podemos ver que o a solução do PSN para pontos factíveis do problema do primeiro nível é $y(x) = |x|$; $x \in [-1, 2]$. Portanto, mesmo em um caso simples em que todas as funções envolvidas são lineares, pode acontecer que a função $\mathcal{F}(x) = F(x, y(x))$ seja não suave. Desse modo pode acontecer que a reformulação do (8.1) não tenha a estrutura padrão de um problema de programação não linear suave. Embora o exemplo não esteja na forma padrão (8.1) é fácil ver que se acrescentarmos variáveis de folga continuamos com a mesma situação.

Outra alternativa muito utilizada para abordar este problema é substituir o PSN por suas condições KKT. No caso que Ω_y é o octante positivo de \mathbb{R}^m temos que o

problema reformulado é:

$$\begin{aligned}
 & \text{Minimizar } F(x, y) \\
 & \quad x, y, \mu, \gamma \\
 & \text{sujeito a: } x \in \Omega_x \\
 & \quad H(x) = 0, \\
 & \quad \nabla_y f(x, y) + J_{h_y}(x, y)^T \mu - \gamma = 0, \\
 & \quad h(x, y) = 0, \\
 & \quad y_1 \gamma_1 = 0, \\
 & \quad \vdots \\
 & \quad y_m \gamma_m = 0, \\
 & \quad y \geq 0, \\
 & \quad \gamma \geq 0.
 \end{aligned}$$

Definindo $C : \mathbb{R}^{n+m+p+m} \rightarrow \mathbb{R}^{m+p+m}$ por

$$C(x, y, \mu, \gamma) = \begin{pmatrix} \nabla_y f(x, y) + J_{h_y}(x, y)^T \mu - \gamma \\ h(x, y) \\ y_1 \gamma_1 \\ \vdots \\ y_m \gamma_m \end{pmatrix}$$

o problema reformulado apresenta a estrutura tradicional dos problemas de otimização

$$\begin{aligned}
 & \text{Minimizar } F(x, y) \\
 & \quad x, y, \mu, \gamma \\
 & \text{sujeito a: } H(x) = 0, \\
 & \quad C(x, y, \mu, \gamma) = 0, \\
 & \quad x \in \Omega_x, \\
 & \quad y \geq 0, \\
 & \quad \gamma \geq 0.
 \end{aligned} \tag{8.3}$$

Infelizmente esta reformulação só é equivalente ao problema original no caso em que o PSN é convexo e suas soluções são KKT. Além disso, no problema reformulado não estamos considerando a estrutura da restrição. Outro inconveniente do problema reformulado é abordar restrições de complementaridade, as quais reconhecidamente na literatura apresentam dificuldades teóricas e computacionais [73]. Essas dificuldades surgem do fato que o posto da matriz Jacobiana das restrições não se mantém constante em qualquer vizinhança da origem. Se $C(y, \gamma) = y\gamma$ então $\nabla C(y, \gamma) = \begin{pmatrix} \gamma \\ y \end{pmatrix}$ e conseqüentemente $\nabla C(0, 0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. Se considerarmos a restrição $y \geq 0$, temos que seu gradiente é $\begin{pmatrix} 0 \\ -1 \end{pmatrix}$ e portanto um subconjunto formado pelas restrições ativas em $(0, 0)^T$ é $\nabla C(0, 0)$ e $(0, -1)^T$, o qual é PLD. Por outro lado, o ponto (ϵ, ϵ) , com $\epsilon > 0$, está tão próximo da origem quanto desejarmos, mas os gradientes das restrições associadas a C e a positividade de y neste ponto são $(\epsilon, \epsilon)^T$ e $(0, -1)^T$, formando assim um conjunto PLI. Este fato faz com que nenhuma das condições de qualificação associadas a um algoritmo prático, por exemplo a CPLD, se cumpra na origem.

Teoremas de convergência típicos em otimização apresentam uma afirmação do tipo: “Todo (Algum) ponto limite é estacionário, ou as restrições nesse ponto não satisfazem uma certa condição de qualificação associada ao algoritmo proposto”. Algoritmos com essa propriedade podem sempre convergir para a origem, pois ela nunca satisfaz a condição de qualificação requerida. Portanto temos que estes resultados não nos dizem muito quando estamos trabalhando com um problema com restrições de complementaridade. Mais do que isso, experiências práticas mostram que estes pontos costumam ser pontos atratores para os algoritmos, o que é mais um forte inconveniente deste tipo de restrição.

Outra característica que seria desejável para os métodos que tratam de restrições deste tipo é que a origem só seja um ponto limite se for KKT em cada um dos eixos. Isso não pode ser alcançado por nenhum algoritmo que trate igualmente as

variáveis, conforme podemos ver no exemplo descrito em [9], reproduzido a seguir:

$$\begin{aligned} & \text{Maximizar } x + y \\ & \quad x, y \\ & \text{sujeito a: } xy = 0 \\ & \quad x, y \geq 0. \end{aligned}$$

Neste exemplo, caso o ponto inicial esteja localizado na bissetriz $x = y$, a simetria do problema faz com que o único modo de reduzir a inviabilidade, tratando x e y da mesma maneira, é caminhar em direção à origem. Uma vez na origem, novamente pela simetria, o algoritmo não teria nenhum estímulo para preferir caminhar por direções que tratem x e y de maneira diferente. Dessa forma é de se esperar que qualquer algoritmo indiferente pela posição das variáveis tenha a origem como único ponto limite, embora não valham as condições KKT neste ponto para maximizar $x + y$, nem com as restrições $x = 0$ e $y \geq 0$ nem com $x \geq 0$ e $y = 0$.

Em [10] foi proposto um algoritmo de Restauração Inexata para abordar problemas de dois níveis que não usa nem técnicas de otimização não diferenciável nem a substituição direta do problema do segundo nível por suas condições KKT. Como em problemas de dois níveis uma restrição envolve resolver um problema de otimização, é de se esperar que muitas vezes seja mais fácil tratar da função objetivo do que das restrições. Dessa forma, métodos de Restauração Inexata podem ser interessantes para resolver este tipo de problema.

Outro ponto chave do método proposto em [10] é a utilização da estrutura do problema na restauração e a maneira de manter limitada a deterioração da viabilidade durante a otimização. A restauração é feita fixando o iterando corrente x^k e calculando y^k como solução do problema do segundo nível parametrizado por x^k . Já para preservar a viabilidade fazemos a fase de otimização no espaço tangente das condições KKT do problema do segundo nível.

Como no passo de otimização os pontos viáveis do subproblema são aqueles pertencentes a uma linearização das condições KKT, o ponto resultante da otimização tem uma deterioração limitada da “estacionaridade”. Entretanto, a restauração foi feita

pensando no problema de otimização do segundo nível, portanto manter a “estacionariedade” está muito relacionado a manter a otimalidade deste problema. Acreditamos que estas ideias podem ser formalizadas usando que, sob condições razoáveis, a norma do sistema KKT é um “error bound” para o problema original. Provavelmente esse fato de aproveitarmos a estrutura de otimização seja uma das grandes vantagens deste método sobre os que desde o início substituem o problema do segundo nível pelas condições KKT.

O algoritmo descrito em [10] e os teoremas relativos à sua convergência seguem os conceitos e resultados apresentados em [78] e [80] para otimização tradicional. Nesta seção vamos propor um método de Restauração Inexata para problemas de dois níveis aproveitando as ideias principais de [10] que acabamos de listar, mas usando a estrutura, e conseqüentemente as vantagens, de nossa modificação do método de Fischer-Friedlander que apresentamos na Seção 5.1 do Capítulo 5.

Para a descrição do Algoritmo 8.1 vamos definir

$$s = (x, y, \mu, \gamma) \in \mathbb{R}^{n+m+p+m},$$

$$L(s, \lambda) = F(x, y) + \lambda^T \begin{pmatrix} H(x) \\ C(s) \end{pmatrix}$$

e

$$\Phi(s, \lambda, \rho) = \rho L(s, \lambda) + (1 - \rho) \left\| \begin{pmatrix} H(x) \\ C(s) \end{pmatrix} \right\|.$$

Algoritmo 8.1

Sejam $r \in [0, 1)$, $\beta, K_1 > 0$ constantes algorítmicas, $\Omega \subset \Omega_x \times \mathbb{R}^{m+p+m} \subset \mathbb{R}^{n+m+p+m}$ e $\Omega_\lambda \subset \mathbb{R}^{q+m+p+m}$ um conjunto compacto.

Passo 0: Escolher $s_0 \in \Omega$, $\lambda_0 \in \Omega_\lambda$, $\rho_0 \in (0, 1)$ e fazer $k = 0$.

Passo 1 - Restauração:

- Primeira parte: Obter x_k^R restaurando em relação à restrição $H(x) = 0$.
- Fixar x_k^R e obter $s_k^R \in \Omega$ e $\lambda_{k+1} \in \Omega_\lambda$ resolvendo o problema do segundo nível.

O ponto s_k^R é aceito na fase de restauração se cumpre as relações:

$$\left\| \begin{pmatrix} H(x_k^R) \\ C(s_k^R) \end{pmatrix} \right\| \leq r \left\| \begin{pmatrix} H(x_k) \\ C(s_k) \end{pmatrix} \right\|,$$

$$L(s_k^R, \lambda_{k+1}) \leq L(s_k, \lambda_k) + \beta \left\| \begin{pmatrix} H(x_k) \\ C(s_k) \end{pmatrix} \right\|.$$

Passo 2: Escolher $\rho_{k+1} \in \{2^{-i}\rho_k : i \in \mathbb{N}\}$ o maior possível tal que

$$\Phi(s_k^R, \lambda_{k+1}, \rho_{k+1}) - \Phi(s_k, \lambda_k, \rho_{k+1}) \leq \frac{1}{2}(1-r) \left(\left\| \begin{pmatrix} H(x_k^R) \\ C(s_k^R) \end{pmatrix} \right\| - \left\| \begin{pmatrix} H(x_k) \\ C(s_k) \end{pmatrix} \right\| \right).$$

Passo 3: Computar a direção de busca para otimalidade $d_k \in \mathbb{R}^{n+m+p+m}$ como solução do modelo quadrático:

$$\begin{aligned} & \text{Minimizar} && d^T B_k d + \nabla_s L(s_k^R, \lambda_{k+1})^T d + L(s_k^R, \lambda_{k+1}) \\ & && d \in \mathbb{R}^{n+2m+p} \\ & \text{sujeito a:} && \begin{pmatrix} J_H(x_k^R) \\ J_C(s_k^R) \end{pmatrix} d = 0, \\ & && s_k^R + d \in \Omega. \end{aligned} \tag{8.4}$$

Passo 4: Determinar $t_k \in \{2^{-i} : i \in \mathbb{N}\}$ o maior possível tal que

$$L(s_k^R + t_k d_k, \lambda_{k+1}) \leq L(s_k^R, \lambda_{k+1}) - K_1 t_k \|d_k\|^2 \tag{8.5}$$

$$\Phi(s_k^R + t_k d_k, \lambda_{k+1}, \rho_{k+1}) - \Phi(s_k, \lambda_k, \rho_{k+1}) \leq \frac{1}{2}(1-r) \left(\left\| \begin{pmatrix} H(x_k^R) \\ C(s_k^R) \end{pmatrix} \right\| - \left\| \begin{pmatrix} H(x_k) \\ C(s_k) \end{pmatrix} \right\| \right). \tag{8.6}$$

Passo 5: Fazer $s_{k+1} = s_k^R + t_k d_k$, $k = k + 1$ e voltar ao passo 1.

Vamos agora mostrar condições suficientes para que o algoritmo esteja bem definido. O ponto chave para isto é demonstrar que existem situações em que podemos

garantir que a restauração proposta pode ser feita. Para isto serão usadas hipóteses muito fortes, serão pedidos complementaridade estrita, regularidade e condições suficientes de segunda ordem. Entretanto estas condições são suficientes e nada impede que mesmo quando elas não sejam satisfeitas o algoritmo ainda esteja bem definido. A fase da restauração será a mesma da apresentada em [34], embora a nossa demonstração seja diferente. Vamos reproduzi-la aqui para que o texto fique mais completo e para expormos alguns detalhes não apresentados em [34]. Para isto consideremos as seguintes hipóteses:

H1. Para cada $x \in \Omega_x$ é possível encontrar $x^R \in \Omega_x$ tal que $H(x^R) \leq H(x)$ e $\|x^R - x\| \leq \beta_0 \|H(x)\|$, para algum $\beta_0 > 0$ independente de x .

- Essa primeira hipótese afirma que é possível restaurar em relação à restrição H . No Capítulo 5 já demos condições para que possamos garantir esse passo.

H2. Para cada x fixado, o problema do segundo nível tem solução única $y(x)$ e que este é o único ponto estacionário do PSN.

H3. A função que a cada x associa $y(x)$ é contínua e os gradientes das restrições ativas do PSN(x) são LIs em $y(x)$.

- Sob a hipótese H3 existem $\mu(x)$ e $\gamma(x)$, unicamente determinados, e funções contínuas de x , tais que

$$C(x, y(x), \mu(x), \gamma(x)) = 0.$$

H4. Para todo $x \in \Omega_x$ e $z \in \mathbb{R}^m$, com $z \neq 0$,

$$z^T \nabla_y^2 l(x, y(x), \mu(x), \gamma(x)) z > 0,$$

onde $l(x, y, \mu, \gamma) = f(x, y) + \mu^T h(x, y)$.

H5. Vale a complementaridade estrita para o problema do segundo nível, ou seja, para todo $y(x)$ solução do PSN(x) temos:

$$y_i(x) = 0 \Rightarrow \gamma_i(x) > 0$$

e

$$\gamma_i(x) = 0 \Rightarrow y_i(x) > 0,$$

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

H6. Todos os iterandos de nosso algoritmo permanecem num compacto Ω . Além disso, $V \subset \Omega$, onde

$$V = \{(x, y(x), \mu(x), \gamma(x)); x \in \Omega_x\},$$

é o conjunto dos pontos que anulam as condições KKT do PSN.

Pelo fato de Ω_x ser um compacto temos que, se as funções $y(x)$, $\mu(x)$ e $\gamma(x)$ forem contínuas então V é compacto. Neste caso é fácil implementar o algoritmo, colocando salvaguardas suficientemente grandes nas variáveis y, μ e γ , de modo que todos os iterandos permaneçam num compacto Ω . Dessa forma, assumiremos a validade da hipótese H6 sem nos preocupar de que esta é uma hipótese sobre a sequência gerada e não sobre o problema, uma vez que temos meios razoáveis para garantir este fato.

Uma observação importante aqui é que a hipótese H4 é uma condição suficiente de segunda ordem mais exigente que a usual, já que exige que a Hessiana do Lagrangiano avaliada na solução seja definida positiva em todo o espaço.

Teorema 8.1.1. *Supondo que valham as hipóteses H2-H5, o Jacobiano de $C(x, y, \mu, \gamma)$ em relação a (y, μ, γ) é não singular em $(x, y(x), \mu(x), \gamma(x))$, para todo $x \in \Omega_x$.*

Demonstração: Calculando o Jacobiano temos:

$$J_{C_{y,\mu,\gamma}}(x, y, \mu, \gamma) = \begin{pmatrix} \nabla_y^2 l(x, y, \mu, \gamma) & J_{h_y}(x, y)^T & -I \\ J_{h_y}(x, y) & 0 & 0 \\ \text{diag}(\gamma) & 0 & \text{diag}(y) \end{pmatrix}.$$

Multiplicando $J_{C_{y,\mu,\gamma}}$ por um vetor arbitrário $\begin{pmatrix} u \\ v \\ w \end{pmatrix} \in \mathbb{R}^{m+p+m}$ pertencente

ao núcleo obtemos as relações:

$$\nabla_y^2 l(x, y, \mu, \gamma)u + J_{h_y}(x, y)^T v - w = 0 \quad (8.7)$$

$$J_{h_y}(x, y)u = 0 \quad (8.8)$$

$$\text{diag}(\gamma)u + \text{diag}(y)w = 0 \quad (8.9)$$

De (8.9) temos que

$$\gamma_i u_i + y_i w_i = 0,$$

para todo $i = 1, \dots, m$. Considerando de agora em diante que $x \in \Omega_x$ e $(y, \mu, \gamma) = (y(x), \mu(x), \gamma(x))$, podem ocorrer dois casos. No primeiro $y_i(x) > 0$ e conseqüentemente, pela complementaridade, $\gamma_i(x) = 0$. Nesta situação obtemos que $w_i = 0$, o que implica que o produto $u_i w_i = 0$. No segundo caso $y_i(x) = 0$ e, em virtude da complementaridade estrita H5, $\gamma_i(x) > 0$ o que mostra que $u_i = 0$ e conseqüentemente que o produto $u_i w_i$ é novamente nulo. Dessa forma podemos afirmar que $u_i w_i = 0$ sempre, qualquer que seja i . Isso implica que os vetores u e w são ortogonais.

Multiplicando a relação (8.7) à esquerda por u^T obtemos que

$$u^T \nabla_y^2 l(x, y(x), \mu(x), \gamma(x))u = 0,$$

pois provamos que $u^T w = 0$ e a relação (8.8) nos diz que $u^T J_{h_y}(x, y(x))^T = 0$. Pela condição suficiente de segunda ordem exigida H4 obtemos que $u = 0$.

Usando o fato de que $u = 0$ a relação (8.9) implica que $y_i(x)w_i = 0$, ou seja, $w_i = 0$ sempre que a restrição $y_i(x) \geq 0$ não esteja ativa. Por fim, usando o fato de que $u = 0$ e olhando para (8.7) descartando as colunas referentes às variáveis w_i tais que $y_i(x) > 0$, obtemos uma combinação linear nula dos gradientes das restrições ativas em

$y(x)$ para o problema do segundo nível:

$$(J_{h_y}(x, y(x))^T - I^*) \begin{pmatrix} v \\ w^* \end{pmatrix} = 0,$$

onde w^* corresponde às coordenadas de w não descartadas e I^* às colunas correspondentes da matriz identidade $m \times m$. Pela hipótese da regularidade H3 obtemos que $(v, w^*) = 0$, e conseqüentemente $(u, v, w) = 0$.

Como o núcleo da matriz $J_{C_{y,\mu,\gamma}}(x, y(x), \mu(x), \gamma(x))$ é composto apenas pelo vetor nulo, temos que ela é não singular e obtemos o resultado desejado. \square

Como dito anteriormente, as hipóteses H2-H5 são suficientes para que $J_{C_{y,\mu,\gamma}}(x, y(x), \mu(x), \gamma(x))$ seja não singular, mas não são necessárias. Um caso importante onde isso acontece é quando o PSN é linear, ou seja:

$$\begin{aligned} &\text{Minimizar } f(x, y) \equiv c(x)^T y \\ &\quad y \\ &\text{sujeito a: } A(x)y - b(x) = 0, \\ &\quad y \geq 0. \end{aligned}$$

Nessa situação é evidente que a hipótese da Hessiana do Lagrangiano ser definida positiva H4 é falsa, uma vez que as segundas derivadas do problema em relação a y são nulas. Entretanto, continua a valer a tese do Teorema 8.1.1 conforme demonstramos a seguir.

Teorema 8.1.2. *Suponhamos que o problema do segundo nível seja linear e que valham as hipóteses H2, H3 e H5, então o Jacobiano de $C(x, y, \mu, \gamma)$ em relação a (y, μ, γ) é não singular em $(x, y(x), \mu(x), \gamma(x))$, para todo $x \in \Omega_x$.*

Demonstração: Denotando $A(x) = A$, já sabemos que

$$J_{C_{y,\mu,\gamma}}(x, y, \mu, \gamma) = \begin{pmatrix} 0 & A^T & -I \\ A & 0 & 0 \\ \text{diag}(\gamma) & 0 & \text{diag}(y) \end{pmatrix}.$$

Dessa forma, para um vetor arbitrário $\begin{pmatrix} u \\ v \\ w \end{pmatrix} \in \mathbb{R}^{m+p+m}$ pertencente ao núcleo de $J_{C_{y,\mu,\gamma}}$ obtemos as relações:

$$A^T v - w = 0, \quad (8.10)$$

$$Au = 0, \quad (8.11)$$

$$\text{diag}(\gamma)u + \text{diag}(y)w = 0. \quad (8.12)$$

Pela unicidade da solução do $PSN(x)$ (H2), o Teorema Fundamental da Programação Linear [72] garante que $y(x)$ é um vértice do conjunto viável do PSN. Já a complementaridade estrita H5 implica que $y(x)$ seja uma solução não degenerada. Dessa forma podemos afirmar que existem exatamente p coordenadas não nulas em $y(x)$, as quais, sem perda de generalidade, vamos supor serem as p primeiras. Dessa forma, a complementaridade estrita garante que $\gamma_i(x) = 0$ para todo $i = 1, \dots, p$ e $\gamma_i(x) > 0$ para todo $i = p + 1, \dots, m$.

De (8.12) temos que

$$\gamma_i(x)u_i + y_i(x)w_i = 0,$$

para todo $i = 1, \dots, m$, portanto, pela afirmação anterior, temos que $w_i = 0$ para todo $i = 1, \dots, p$ e $u_i = 0$ para todo $i = p + 1, \dots, m$.

Como as p primeiras variáveis de $y(x)$ estão associadas a uma solução básica temos que as p primeiras colunas de A são LI. Pelo fato das $m-p$ últimas coordenadas de u serem nulas, a equação (8.11) se reduz a uma combinação linear nula das p primeiras colunas de A , portanto $u_i = 0$ para todo $i = 1, \dots, p$ e conseqüentemente $u = 0$.

Da mesma maneira, podemos descartar as p primeiras coordenadas de w de modo que a equação (8.10) se resuma a

$$(A^T \quad -I^*) \begin{pmatrix} v \\ w^* \end{pmatrix} = 0,$$

onde w^* corresponde às $m - p$ últimas coordenadas de w e I^* às colunas correspondentes da matriz identidade $m \times m$. Sendo assim, temos uma combinação linear nula dos gradientes das restrições ativas em $y(x)$. Pela hipótese da regularidade H3 obtemos que $(v, w^*) = 0$, e conseqüentemente $(u, v, w) = 0$ o que implica que $J_{C_{y,\mu,\gamma}}(x, y(x), \mu(x), \gamma(x))$ é não singular. \square

Nosso próximo resultado fornece um limitante local para a distância entre a estimativa corrente e a solução do problema de dois níveis baseado na norma do sistema KKT no ponto corrente. Este fato está demonstrado no Lema 3.1.3 de [34] mas acreditamos que a demonstração usada aqui é mais simples.

Lema 8.1.1. *Sejam $\Omega_x \subset \mathbb{R}^n$ e $\Omega \subset \mathbb{R}^{n+m+p+m}$ conjuntos compactos e $x \in \Omega_x$. Suponhamos que valham as hipóteses H2 e H3 e que $J_{C_{y,\mu,\gamma}}(x, y(x), \mu(x), \gamma(x))$ seja não singular. Então, existem $\xi_0 > 0$ e $\delta > 0$, independentes de x , tais que*

$$\|(y, \mu, \gamma) - (y(x), \mu(x), \gamma(x))\| \leq \xi_0 \|C(x, y, \mu, \gamma)\|,$$

para todo $(x, y, \mu, \gamma) \in \Omega$ tal que $\|(y, \mu, \gamma) - (y(x), \mu(x), \gamma(x))\| \leq \delta$.

Demonstração: Ao longo dessa demonstração vamos denotar o jacobiano $J_{C_{y,\mu,\gamma}}$ no ponto $(x, y(x), \mu(x), \gamma(x))$ simplesmente por $J(x)$. Como $J(x)$ é não singular temos que $\|J(x)u\| > 0$ para todo $u \in \mathbb{R}^{m+p+m}$ tal que $\|u\| = 1$. Já que $J(x)$ é uma função contínua temos que a função que associa um par (x, u) a $\|J(x)u\|$ é contínua. Dessa forma, como a esfera unitária e Ω_x são compactos, temos que existe $K > 0$ tal que:

$$\|J(x)u\| \geq 2K$$

para todo $u \in \mathbb{R}^{m+p+m}$ tal que $\|u\| = 1$ e $x \in \Omega_x$. Conseqüentemente para qualquer que seja $u \in \mathbb{R}^{m+p+m}$ e $x \in \Omega_x$ vale que $\|J(x)u\| \geq 2K\|u\|$.

Como as funções $C(x, y, \mu, \gamma), y(x), \mu(x), \gamma(x)$ e $J(x)$ são contínuas temos

que a função

$$\begin{aligned}
 r(x, y, \mu, \gamma) &\equiv C(x, y, \mu, \gamma) - C(x, y(x), \mu(x), \gamma(x)) + \\
 &\quad - J_{C_{x,y,\mu,\gamma}}(x, y(x), \mu(x), \gamma(x)) \begin{pmatrix} x - x \\ y - y(x) \\ \mu - \mu(x) \\ \gamma - \gamma(x) \end{pmatrix} \\
 &= C(x, y, \mu, \gamma) - J(x) \begin{pmatrix} y - y(x) \\ \mu - \mu(x) \\ \gamma - \gamma(x) \end{pmatrix}
 \end{aligned}$$

também é contínua.

A função $g : \mathbb{R}^{n+m+p+m} \rightarrow \mathbb{R}^{m+p+m}$ definida por

$$g(x, y, \mu, \gamma) = \begin{cases} \frac{r(x,y,\mu,\gamma)}{\|(y,\mu,\gamma)-(y(x),\mu(x),\gamma(x))\|}; & \text{se } (y, \mu, \gamma) \neq (y(x), \mu(x), \gamma(x)) \\ 0; & \text{se } (y, \mu, \gamma) = (y(x), \mu(x), \gamma(x)) \end{cases}$$

é contínua pois, se $(y, \mu, \gamma) \neq (y(x), \mu(x), \gamma(x))$ então g é uma razão de funções contínuas e para o ponto $(y(x), \mu(x), \gamma(x))$ a diferenciabilidade de C nos diz que vale o limite $g(x, y, \mu, \gamma) \rightarrow 0 = g(x, y(x), \mu(x), \gamma(x))$ quando $(x, y, \mu, \gamma) \rightarrow (x, y(x), \mu(x), \gamma(x))$. Como Ω é compacto e g é contínua temos que g é uniformemente contínua. Dessa forma, existe $\delta > 0$, independente de x , tal que $\|(x, y, \mu, \gamma) - (x, y(x), \mu(x), \gamma(x))\| \leq \delta$ implica que $\|g(x, y, \mu, \gamma) - g(x, y(x), \mu(x), \gamma(x))\| \leq K$, e conseqüentemente que

$$\|r(x, y, \mu, \gamma)\| \leq K \|(y, \mu, \gamma) - (y(x), \mu(x), \gamma(x))\|.$$

Portanto, lembrando que $C(x, y(x), \mu(x), \gamma(x)) = 0$, temos que na vizi-

nhança de raio δ de $(y(x), \mu(x), \gamma(x))$ vale que:

$$\begin{aligned} \|C(x, y, \mu, \gamma)\| &= \left\| J(x) \begin{pmatrix} y - y(x) \\ \mu - \mu(x) \\ \gamma - \gamma(x) \end{pmatrix} + r(x, y, \mu, \gamma) \right\| \\ &\geq \left\| J(x) \begin{pmatrix} y - y(x) \\ \mu - \mu(x) \\ \gamma - \gamma(x) \end{pmatrix} \right\| - \|r(x, y, \mu, \gamma)\| \\ &\geq 2K\|(y, \mu, \gamma) - (y(x), \mu(x), \gamma(x))\| - K\|(y, \mu, \gamma) - (y(x), \mu(x), \gamma(x))\|. \end{aligned}$$

Sendo assim, se $\xi_0 = \frac{1}{K}$ temos o resultado desejado. \square

O Lema 8.1.1 traz um resultado análogo ao lema proposto por Robinson em [94] conjuntamente com o resultado que nos permite considerar a vizinhança envolvida independente do ponto corrente.

Lema 8.1.2. *Seja $x \in \Omega_x$ e suponhamos que valham as hipóteses H2 e H3 e que $J_{C_{y,\mu,\gamma}}(x, y(x), \mu(x), \gamma(x))$ seja não singular. Então, existe $\xi > 0$, independente de (x, y, μ, γ) , tal que*

$$\|(y, \mu, \gamma) - (y(x), \mu(x), \gamma(x))\| \leq \xi \left\| \begin{pmatrix} H(x) \\ C(s) \end{pmatrix} \right\|,$$

para todo $(x, y, \mu, \gamma) \in \Omega$.

Demonstração:

Vamos considerar nessa demonstração δ como no Lema 8.1.1,

$$d(x, y, \mu, \gamma) = \|(x, y, \mu, \gamma) - (x, y(x), \mu(x), \gamma(x))\|,$$

e

$$\bar{\Omega} = \{s \in \Omega; d(s) \geq \delta\}.$$

Como $y(x)$, $\mu(x)$ e $\gamma(x)$ são contínuas temos claramente que d é uma função contínua. Dessa forma temos que $\bar{\Omega}$ é compacto.

A demonstração pode ser dividida em duas possibilidades.

Na primeira $\bar{\Omega} = \emptyset$, e portanto

$$\|(y, \mu, \gamma) - (y(x), \mu(x), \gamma(x))\| \leq \xi_0 \|C(x, y, \mu, \gamma)\| \leq \xi_0 \left\| \begin{pmatrix} H(x) \\ C(s) \end{pmatrix} \right\|$$

para todo $(x, y, \mu, \gamma) \in \Omega$ e não nos resta nada a demonstrar.

No segundo caso, $\bar{\Omega}$ é não vazio e compacto. Sendo assim, como $C(s)$ também é contínua e $C(s) > 0$ se $s \in \bar{\Omega}$, existem $c_1 \equiv \max_{s \in \bar{\Omega}} d(s) < \infty$ e $c_2 \equiv \min_{s \in \bar{\Omega}} \|C(s)\| > 0$. Dessa forma temos que

$$\|(x, y, \mu, \gamma) - (x, y(x), \mu(x), \gamma(x))\| \leq c_1 = \frac{c_1}{c_2} c_2 \leq \frac{c_1}{c_2} \|C(x, y, \mu, \gamma)\|,$$

para todo $(x, y, \mu, \gamma) \in \bar{\Omega}$. Portanto, se tomarmos $\xi = \xi_0 + \frac{c_1}{c_2}$ temos que

$$\|(x, y, \mu, \gamma) - (x, y(x), \mu(x), \gamma(x))\| \leq \xi \|C(x, y, \mu, \gamma)\| \leq \xi \left\| \begin{pmatrix} H(x) \\ C(s) \end{pmatrix} \right\|,$$

para todo $(x, y, \mu, \gamma) \in \Omega$. □

A demonstração de que o passo 1 está bem definido segue de perto a prova que a restauração do Algoritmo 5.1 também está bem definida. O Lema 8.1.2 é análogo ao Lema 5.1.2 e o restante da demonstração está exposta a seguir por conveniência.

Teorema 8.1.3. *Seja $s_k \in \Omega$ e suponhamos que valham as hipóteses H1, H2 e H3 e que $J_{C_{y,\mu,\gamma}}(s(x_k^R))$ seja não singular. Então, existe $\beta > 0$, independente de s_k , tal que*

$$\begin{aligned} \left\| \begin{pmatrix} H(x_k^R) \\ C(s_k^R) \end{pmatrix} \right\| &\leq r \left\| \begin{pmatrix} H(x_k) \\ C(s_k) \end{pmatrix} \right\| \\ L(s_k^R, \lambda_{k+1}) &\leq L(s_k, \lambda_k) + \beta \left\| \begin{pmatrix} H(x_k) \\ C(s_k) \end{pmatrix} \right\|. \end{aligned}$$

Demonstração: Considerando L_λ um limitante para $\|\lambda\|$ em Ω_λ , L_F a constante de Lipschitz de F em Ω , e $\beta = L_F \xi + (1 + r)L_\lambda$ vamos mostrar que tomando $s_k^R =$

$(x_k^R, y(x_k^R), \mu(x_k^R), \gamma(x_k^R))$ então, para qualquer $\lambda_{k+1} \in \Omega_\lambda$, o par (s_k^R, λ_{k+1}) é aceitável pelo passo 1.

Como $C(s_k^R) = 0$, temos que

$$\begin{aligned} \left\| \begin{pmatrix} H(x_k^R) \\ C(s_k^R) \end{pmatrix} \right\| &= \|H(x_k^R)\| \\ &= r \|H(x_k)\| \\ &\leq r \left\| \begin{pmatrix} H(x_k) \\ C(s_k) \end{pmatrix} \right\|. \end{aligned}$$

Vamos agora demonstrar a deterioração limitada do Lagrangiano.

$$\begin{aligned} L(s_k^R, \lambda_{k+1}) - L(s_k, \lambda_k) &= \left(F(s_k^R) + \lambda_{k+1}^T \begin{pmatrix} H(x_k^R) \\ C(s_k^R) \end{pmatrix} \right) - \left(F(s_k) + \lambda_k^T \begin{pmatrix} H(x_k) \\ C(s_k) \end{pmatrix} \right) \\ &\leq |F(s_k^R) - F(s_k)| + \|\lambda_{k+1}\| \left\| \begin{pmatrix} H(x_k^R) \\ C(s_k^R) \end{pmatrix} \right\| + \|\lambda_k\| \left\| \begin{pmatrix} H(x_k) \\ C(s_k) \end{pmatrix} \right\| \\ &\leq L_F \|s_k^R - s_k\| + L_\lambda \left\| \begin{pmatrix} H(x_k^R) \\ C(s_k^R) \end{pmatrix} \right\| + L_\lambda \left\| \begin{pmatrix} H(x_k) \\ C(s_k) \end{pmatrix} \right\| \\ &\leq (L_F \xi + (1+r)L_\lambda) \left\| \begin{pmatrix} H(x_k) \\ C(s_k) \end{pmatrix} \right\| \\ &\leq \beta \left\| \begin{pmatrix} H(x_k) \\ C(s_k) \end{pmatrix} \right\|. \end{aligned}$$

□

Portanto, se redefinirmos $\beta = 2\beta$, temos que uma sequência convergente para $(x_k, y(x_k), \mu(x_k), \gamma(x_k))$ será tal que a partir de um certo índice todos os iterandos vão satisfazer o passo 1.

Para mostrar que o passo 2 está bem definido vamos mostrar que se $\rho \leq \frac{1-r}{2(\beta+1-r)}$ então ele será aceito.

Demonstração: Neste caso teríamos que $\rho(\beta + 1 - r) - (1 - r) \leq -\frac{1-r}{2}$, e portanto, já utilizando que o passo 1 foi feito com sucesso, temos

$$\begin{aligned}
 \Phi(s_k^R, \lambda_{k+1}, \rho) - \Phi(s_k, \lambda_k, \rho) &= \rho (L(s_k^R, \lambda_{k+1}) - L(s_k, \lambda_k)) + \\
 &\quad + (1 - \rho) \left(\left\| \begin{pmatrix} H(x_k^R) \\ C(s_k^R) \end{pmatrix} \right\| - \left\| \begin{pmatrix} H(x_k) \\ C(s_k) \end{pmatrix} \right\| \right) \\
 &\leq \rho\beta \left\| \begin{pmatrix} H(x_k) \\ C(s_k) \end{pmatrix} \right\| - (1 - \rho)(1 - r) \left\| \begin{pmatrix} H(x_k) \\ C(s_k) \end{pmatrix} \right\| \\
 &\leq (\rho(\beta + 1 - r) - (1 - r)) \left\| \begin{pmatrix} H(x_k) \\ C(s_k) \end{pmatrix} \right\| \\
 &\leq -\frac{1-r}{2} \left\| \begin{pmatrix} H(x_k) \\ C(s_k) \end{pmatrix} \right\| \\
 &\leq \frac{1-r}{2} \left(\left\| \begin{pmatrix} H(x_k^R) \\ C(s_k^R) \end{pmatrix} \right\| - \left\| \begin{pmatrix} H(x_k) \\ C(s_k) \end{pmatrix} \right\| \right).
 \end{aligned}$$

□

Pela demonstração anterior a sequência dos parâmetros de penalidade está afastada do zero pois $\rho_k \geq \min \left\{ \frac{1-r}{4(\beta+1-r)}, \rho_0 \right\}$, e conseqüentemente, como ρ_k é não crescente, concluímos que $\rho_k = \bar{\rho}$ é constante para k suficientemente grande.

O passo 3 está bem definido se $B_k = B(s_k^R, \lambda_{k+1})$ é positiva definida pois neste caso o problema (5.1) consiste em minimizar modelo quadrático estritamente convexo num hiperplano (observe que $d = 0$ é viável para (5.1)), o que se trata de um problema bem estabelecido na literatura.

Considerando agora que os passos 1, 2 e 3 foram realizados com sucesso vamos demonstrar a boa definição do passo 4.

Lema 8.1.3. *Suponha que os gradientes de F e de $\begin{pmatrix} H \\ C \end{pmatrix}$ são Lipschitz contínuos em*

Ω , então existem constantes positivas K_1 , K_2 e τ tais que

$$\left\| \begin{pmatrix} H(x_k^R + td_{x_k}) \\ C(s_k^R + td_k) \end{pmatrix} \right\| \leq \left\| \begin{pmatrix} H(x_k^R) \\ C(s_k^R) \end{pmatrix} \right\| + K_2 t^2 \|d_k\|^2, \quad (8.13)$$

para todo $t \geq 0$, e

$$L(s_k^R + td_k, \lambda_{k+1}) \leq L(s_k, \lambda_{k+1}) - K_1 t \|d_k\|^2 \quad (8.14)$$

para todo $t \in [0, \tau]$.

Demonstração: Basta aplicar o Lema 5.1.5 para as funções correspondentes. \square

A relação (8.14) já garante que a primeira condição do passo 4 é satisfeita para $t_k \leq \tau$. Para demonstrar que a segunda condição também é satisfeita para t suficientemente pequeno vamos considerar que $\sigma_1 > 0$ é um limitante inferior para os autovalores de B_k , que $L_{\nabla L}$ é a constante de Lipschitz para ∇L em Ω e que o passo 2 foi realizado com sucesso. Analogamente ao que demonstramos para o Algoritmo 5.1, nessas condições temos que $\nabla L(s_k^R, \lambda_{k+1})^T d_k \leq -\sigma_1 \|d_k\|^2$ e que

$$\begin{aligned} L(s_k^R + td_k, \lambda_{k+1}) &= L(s_k^R, \lambda_{k+1}) + t \nabla L(s_k^R, \lambda_{k+1})^T d_k + \\ &\quad + t \int_0^1 (\nabla L(s_k^R + tx d_k, \lambda_{k+1}) - \nabla L(s_k^R, \lambda_{k+1}))^T d_k dx \\ &\leq L(s_k^R, \lambda_{k+1}) - \sigma_1 t \|d_k\|^2 + \frac{L_{\nabla L}}{2} t^2 \|d_k\|^2. \end{aligned}$$

Portanto, todo t tal que $t \leq \min \left\{ \frac{\bar{\rho} k_1}{k_2(1-\bar{\rho})}, \tau \right\}$ será aceito pelo passo 4 e consequentemente também obtemos que a sequência $\{t_k\}$ está afastada do zero pois $t_k \geq \bar{t} \equiv \min \left\{ \frac{\bar{\rho} k_1}{2k_2(1-\bar{\rho})}, \frac{\tau}{2}, 1 \right\}$.

Uma vez mostrado que o algoritmo está bem definido vamos mostrar um teorema associado à viabilidade dos pontos limite do algoritmo.

Teorema 8.1.4. *Suponhamos que valham as hipóteses H1 a H6 e que $\{s_k\}$ seja uma sequência gerada pelo Algoritmo 8.1, então*

$$\sum_{k=0}^{\infty} \left\| \begin{pmatrix} H(x_k) \\ C(s_k) \end{pmatrix} \right\| = \hat{\sigma} < \infty.$$

Demonstração: De forma análoga ao Teorema 5.1.2 temos que para $l \geq k_0$, onde $\rho_{k_0} = \bar{\rho}$, vale que

$$\Phi(s_l, \lambda_l, \bar{\rho}) - \Phi(s_{k_0}, \lambda_{k_0}, \bar{\rho}) \leq \sum_{k=k_0}^{l-1} -\frac{1}{2}(1-r)^2 \left\| \begin{pmatrix} H(x_k) \\ C(s_k) \end{pmatrix} \right\|.$$

Como $\Phi(x, \lambda, \bar{\rho})$ é contínua nas variáveis (s, λ) , as hipóteses de compacidade garantem que $\sum_{k=0}^{\infty} \left\| \begin{pmatrix} H(x_k) \\ C(s_k) \end{pmatrix} \right\| = \hat{\sigma} < \infty$. □

Como consequência do Teorema 8.1.4 concluímos que $\left\| \begin{pmatrix} H(x_k) \\ C(s_k) \end{pmatrix} \right\|$, $H(s_k)$, $C(s_k)$, $H(s_k^R)$ e $C(s_k^R)$ convergem a zero e que qualquer ponto limite de $\{s_k\}$, ou de $\{s_k^R\}$, é viável para o problema reformulado (8.3).

Para demonstrar que a direção d_k vai a zero também prosseguimos da mesma maneira que no Algoritmo 5.1.

Teorema 8.1.5. *Suponhamos que valham as hipóteses H1 a H6 e que $\{s_k\}$ seja uma sequência gerada pelo Algoritmo 8.1, então d_k converge a zero.*

Demonstração: Somando $L(s_{l+1}, \lambda_{l+1}) - L(s_k, \lambda_k)$, usando as relações dos passos 1 e 4 e o Teorema 8.1.4 temos que

$$L(s_{l+1}, \lambda_{l+1}) - L(s_0, \lambda_0) \leq -\gamma \bar{t} \sum_{k=0}^l \|d_k\|^2 + \beta \hat{\sigma}.$$

Como L é contínua e estamos assumindo que os iterandos estão num compacto temos que d_k converge a zero. □

Por fim, podemos concluir, das condições de otimalidade do subproblema (8.4) e do Teorema 8.1.5, que todo ponto limite da sequência $\{s_k^R\}$ é AKKT para o problema reformulado (8.3). Entretanto a afirmação mais forte de que sob a CPLD, ou qualquer outra condição de qualificação prática, os pontos limites são KKT não tem grande relevância. Isso é uma consequência do fato já discutido de que o problema com restrições de complementaridade não satisfaz estas condições na origem, a qual muitas vezes é um atrator para os algoritmos.

Uma implementação simples foi feita em Matlab para resolver o problema:

$$\begin{aligned} & \text{Minimizar } 0 \\ & \quad x, y \\ & \text{sujeito a: } y \text{ é solução de:} \\ & \quad \text{Minimizar } \sin\left(\pi y + \frac{3\pi}{2}\right) \\ & \quad \quad y \\ & \quad \text{sujeito a: } 0 \leq y \leq 3. \end{aligned}$$

Tomando como ponto inicial $(x^0, y^0) = (0, 1.2)$ o Algoritmo 8.1 convergiu para uma solução do problema $(x^*, y^*) = (0, 2)$. Já uma implementação de PQS para o problema reformulado, substituindo o PSN por suas condições KKT, convergiu para o ponto $(0, 1)$, o qual não é nem viável para o problema original. A questão chave aqui é que o ponto $y^0 = 1.2$ está próximo do ponto estacionário $y = 1$, mas esse ponto não é atrativo para um algoritmo baseado em direções de descida. Resolvendo o PSN o ponto $y^* = 2$ é facilmente encontrado. Isso ilustra o porquê de acreditarmos que o Algoritmo 8.1 terá melhor desempenho que métodos que substituem diretamente o problema do segundo nível por suas condições KKT. Em nosso algoritmo usamos as condições KKT para o PSN como condição de parada na restauração mas a alternativa em si de encontrar o ponto restaurado usa a estrutura de minimização do problema do segundo nível. A Figura 8.1 ilustra a função objetivo do PSN.

De maneira independente a esse trabalho, A. Friedlander e F. Gomes propuseram, para uma aplicação particular em otimização de treliças, um algoritmo muito semelhante ao exposto nesta seção. Os resultados deles são apresentados em [55] mas

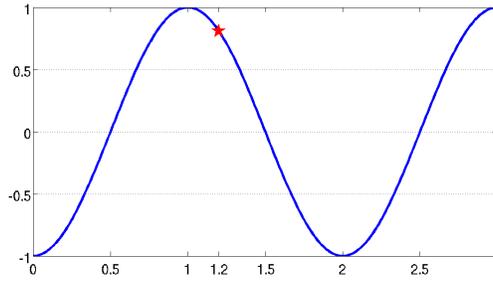


Figura 8.1: Gráfico da função objetivo do problema do segundo nível. O ponto $y^0 = 1.2$ está próximo do ponto estacionário $y = 1$, mas esse ponto não é atrativo para um algoritmo baseado em direções de descida. Resolvendo o PSN o ponto $y^* = 2$ é facilmente encontrado.

não é feita uma análise de convergência formal para o algoritmo proposto. Os testes computacionais de [55] são muito promissores e nos dão motivos suficientes para acreditar que implementações de nosso algoritmo para problemas específicos possam ser bem sucedidas.

8.2 Algoritmo para restrições de equilíbrio

Nessa seção nossos problemas alvo são aqueles em que queremos descobrir ou selecionar alguns parâmetros de forma que um sistema específico esteja em equilíbrio. Um caso ilustrativo é considerar no primeiro exemplo da Seção 7.1 que queremos descobrir quais são algumas das constantes a, b, c, a', b', c' usadas pelos agentes da disputa publicitária. As decisões de quão agressiva é cada uma das empresas são observáveis, e portanto são consideradas como dados do problema. Denotando por x^* e y^* as observações conhecidas nosso problema de interesse seria:

$$\begin{aligned}
 &\text{Minimizar} && (x - x^*)^2 + (y - y^*)^2 \\
 &a, b, c, a', b', c', x, y \\
 &\text{sujeito a:} && (x, y) \text{ é um ponto de equilíbrio de Nash para o sistema (7.2) - (7.3).}
 \end{aligned}
 \tag{8.15}$$

Outro caso ilustrativo seria o caso normativo, onde alguns dos parâmetros do problema sejam definidos por um agente regulador de modo a obter a melhor situação para um certo objetivo. Por exemplo, ainda considerando o caso do sistema (7.2) - (7.3), se os parâmetros a, b, c, a', b' forem conhecidos e intrínsecos de cada empresa e o parâmetro c' estipulado pelo governo, é possível que o governo escolha o valor de c' de forma que na situação de equilíbrio entre as duas empresas o nível total de agressividade seja o menor possível. Ou seja, assumindo $a = 2, b = 1$ e $c = 20$ os parâmetros de decisão da primeira empresa, $a' = 1$ e $b' = 3$ os parâmetro intrínsecos da segunda empresa, o governo toma a decisão c'^* como solução de

$$\begin{aligned} &\text{Minimizar } x + y \\ &\quad c', x, y \\ &\text{sujeito a: } c' \geq 0 \end{aligned}$$

(x, y) é um ponto de equilíbrio de Nash para o sistema (7.2) - (7.3).

No caso geral nosso problema de interesse terá a seguinte estrutura:

$$\begin{aligned} &\text{Minimizar } F(x, \theta) \\ &\quad x, \theta \\ &\text{sujeito a: } H(\theta) = 0, \\ &\quad \theta \in \Theta, \\ &\quad x_i \text{ seja solução de } P_i(x_i, x_{-i}, \theta) \text{ para todo } i = 1, 2, \dots, q. \end{aligned} \tag{8.16}$$

onde $x \equiv (x_1, \dots, x_q) \equiv (x_i, x_{-i}) \in \mathbb{R}^{n_1 + \dots + n_q}$, $\theta \in \Theta \subset \mathbb{R}^{n_\theta}$, as funções $F : \mathbb{R}^{n_1 + \dots + n_q + n_\theta} \rightarrow \mathbb{R}$ e $H : \mathbb{R}^{n_\theta} \rightarrow \mathbb{R}^{n_H}$ são suaves e $P_i(x_i, x_{-i}, \theta)$ é o problema do jogador i parametrizado tanto pelas decisões dos demais jogadores (x_{-i}) quanto pelo vetor de parâmetros θ . Cada problema $P_i(x_i, x_{-i}, \theta)$ tem formulação semelhante aos problemas descritos em (7.1) mas envolvem também os parâmetros θ conforme apresentamos a seguir:

$$\begin{aligned} &\text{Minimizar } f_i(x_i, x_{-i}, \theta) \\ &\quad x_i \in \mathbb{R}^{n_i} \\ &\text{sujeito a: } h_i(x_i, x_{-i}, \theta) = 0 \\ &\quad x_i \geq 0, \end{aligned} \tag{8.17}$$

onde, para todo i , as funções $f_i : \mathbb{R}^{n_1+\dots+n_q+n_\theta} \rightarrow \mathbb{R}$ e $h_i : \mathbb{R}^{n_1+\dots+n_q+n_\theta} \rightarrow \mathbb{R}^{n_{h_i}}$ são suaves.

Para o caso particular em que só temos um agente envolvido no equilíbrio, ou seja $q = 1$, o problema (8.16) se reduz a um problema de programação em dois níveis (8.1). Analogamente ao que foi feito para os problemas de programação em dois níveis vamos definir, para cada $i = 1, 2, \dots, q$, a medida das condições KKT para o problema do agente i será denotada por $C_i : \mathbb{R}^{(n_1+\dots+n_q)+n_\theta+n_{h_i}+n_i}$ e expressa por

$$C_i(x, \theta, \mu_i, \gamma_i) = \begin{pmatrix} \nabla_i f_i(x, \theta) + J_{h_i}(x, \theta)^T \mu_i - \gamma_i \\ h_i(x, \theta) \\ x_i \gamma_i \end{pmatrix},$$

onde $\nabla_i f_i$ e J_{h_i} denotam respectivamente o gradiente de f_i e a Jacobiana de h_i em relação as variáveis x_i . Além disso vamos definir $C : \mathbb{R}^{(n_1+\dots+n_q)+n_\theta+(n_{h_1}+\dots+n_{h_q})+(n_1+\dots+n_q)}$ como uma medida das condições KKT para todos os problemas envolvidos nas restrições de Equilíbrio de Nash

$$C(x, \theta, \mu, \gamma) = \begin{pmatrix} C_1(x, \theta, \mu_1, \gamma_1) \\ \vdots \\ C_q(x, \theta, \mu_q, \gamma_q) \end{pmatrix}.$$

Dessa forma uma reformulação do problema com restrições de equilíbrio seria:

$$\begin{aligned} & \text{Minimizar} && F(x, \theta) \\ & && x, \theta \\ & \text{sujeito a:} && H(\theta), \\ & && C(x, \theta, \mu, \gamma) = 0, \\ & && x, \gamma \geq 0, \\ & && \theta \in \Theta. \end{aligned} \tag{8.18}$$

O algoritmo que proporemos para abordar o problema (8.16) será semelhante ao Algoritmo 8.1. Para defini-lo vamos denotar as variáveis

$$s = (x, \theta, \mu, \gamma) \in \Omega \subset \mathbb{R}^{(n_1+\dots+n_q)+n_\theta+(n_{h_1}+\dots+n_{h_q})+(n_1+\dots+n_q)},$$

o Lagrangiano

$$L(s, \lambda) = F(x, \theta) + \lambda^T \begin{pmatrix} H(\theta) \\ C(s) \end{pmatrix}$$

e a função de mérito

$$\Phi(x, \lambda, \rho) = \rho L(s, \lambda) + (1 - \rho) \left\| \begin{pmatrix} H(\theta) \\ C(s) \end{pmatrix} \right\|.$$

Algoritmo 8.2

Sejam $r \in [0, 1)$, $\beta, K_1 > 0$ constantes algorítmicas, $n = n_1 + \dots + n_q$, $m = 2n_1 + n_{h_1} + \dots + 2n_q + n_{h_q} + n_1 + \dots + n_q$, $\Omega \subset \mathbb{R}^n \times \Theta \times \mathbb{R}^m$ e $\Omega_\lambda \subset \mathbb{R}^{n_H+m}$ um compacto.

Passo 0: Escolher $s_0 \in \Omega$, $\lambda_0 \in \Omega_\lambda$, $\rho_0 \in (0, 1)$ e fazer $k = 0$.

Passo 1 - Restauração:

- Primeira parte: Obter θ_k^R restaurando em relação à restrição $H(\theta) = 0$.
- Obter $s_k^R \in \Omega$ e $\lambda_{k+1} \in \Omega_\lambda$ fixando θ_k^R e fazendo uma iteração do tipo Jacobi para o problema de equilíbrio. Ou seja, fixar θ_k^R e as variáveis x_{-i_k} não referentes ao agente i e resolver o problema de cada agente $P_i(x_i, x_{-i_k}, \theta_k^R)$.

O ponto s_k^R é aceito na fase de restauração se cumpre as relações:

$$\left\| \begin{pmatrix} H(\theta_k^R) \\ C(s_k^R) \end{pmatrix} \right\| \leq r \left\| \begin{pmatrix} H(\theta_k) \\ C(s_k) \end{pmatrix} \right\| \quad (8.19)$$

$$L(s_k^R, \lambda_{k+1}) \leq L(s_k, \lambda_k) + \beta \left\| \begin{pmatrix} H(\theta_k) \\ C(s_k) \end{pmatrix} \right\|. \quad (8.20)$$

Passo 2: Escolher $\rho_{k+1} \in \{2^{-i}\rho_k : i \in \mathbb{N}\}$ o maior possível tal que

$$\Phi(s_k^R, \lambda_{k+1}, \rho_{k+1}) - \Phi(s_k, \lambda_k, \rho_{k+1}) \leq \frac{1}{2}(1 - r) \left(\left\| \begin{pmatrix} H(\theta_k^R) \\ C(s_k^R) \end{pmatrix} \right\| - \left\| \begin{pmatrix} H(\theta_k) \\ C(s_k) \end{pmatrix} \right\| \right).$$

Passo 3: Computar a direção de busca para otimalidade d_k como solução do modelo quadrático:

$$\begin{aligned} & \text{Minimizar} && d^T B_k d + \nabla_s L(s_k^R, \lambda_{k+1})^T d + L(s_k^R, \lambda_{k+1}) \\ & && d \in \mathbb{R}^{n+n_\theta+m} \\ & \text{sujeito a:} && \begin{pmatrix} J_H(\theta_k^R) \\ J_C(s_k^R) \end{pmatrix} d = 0, \\ & && s_k^R + d \in \Omega. \end{aligned} \tag{8.21}$$

Passo 4: Determinar $t_k \in \{2^{-i} : i \in \mathbb{N}\}$ o maior possível tal que

$$L(s_k^R + t_k d_k, \lambda_{k+1}) \leq L(s_k^R, \lambda_{k+1}) - \gamma t_k \|d_k\|^2, \tag{8.22}$$

$$\Phi(s_k^R + t_k d_k, \lambda_{k+1}, \rho_{k+1}) - \Phi(s_k, \lambda_k, \rho_{k+1}) \leq \frac{1}{2}(1-r) \left(\left\| \begin{pmatrix} H(\theta_k^R) \\ C(s_k^R) \end{pmatrix} \right\| - \left\| \begin{pmatrix} H(\theta_k) \\ C(s_k) \end{pmatrix} \right\| \right). \tag{8.23}$$

Passo 5: Fazer $s_{k+1} = s_k^R + t_k d_k$, $k = k + 1$ e voltar ao passo 1.

Na definição do Algoritmo 8.2 poderíamos ser menos específicos na descrição da fase de restauração e apenas exigir que as condições (8.19) e (8.20) fossem satisfeitas. Desse modo qualquer alternativa apresentada para encontrar s_k^R e λ_{k+1} que garanta que essas condições sejam satisfeitas implica na boa definição do passo 1 de nosso algoritmo. A nossa proposta de restauração é um método do tipo Jacobi pois ele apresenta uma motivação natural, que discutimos na Seção 7.3. Como a convergência dos métodos desse tipo ainda não é muito bem estabelecida na literatura e depende muito da aplicação em que está sendo usada, vamos simplesmente supor que após uma iteração de Jacobi obtemos que:

$$\|C(s_k^R)\| \leq r \|C(s_k)\|.$$

Condições suficientes para que essa relação seja satisfeita para sistemas lineares foram apresentadas na Seção 7.3.

Vamos supor também também que $\|H(\theta_k^R)\| \leq r\|H(\theta_k)\|$, já que apresentamos no Capítulo 5 (ver Lemas 5.1.2 e 5.1.3 e Teorema 5.1.1) condições suficientes para que essa relação seja satisfeita. Desse modo temos que:

$$\begin{aligned} \left\| \begin{pmatrix} H(\theta_k^R) \\ C(s_k^R) \end{pmatrix} \right\|^2 &= \|H(\theta_k^R)\|^2 + \|C(s_k^R)\|^2 \\ &\leq r^2\|H(\theta_k)\|^2 + r^2\|C(s_k)\|^2 \\ &\leq r^2 \left\| \begin{pmatrix} H(\theta_k) \\ C(s_k) \end{pmatrix} \right\|^2. \end{aligned}$$

Portanto, sob essas hipóteses, temos que a condição (8.19) está satisfeita. Dessa forma basta mostrar que a condição (8.20) é satisfeita pelo método do tipo Jacobi para mostrar que o passo 1 está bem definido. Conseqüentemente, só nos resta mostrar a condição de deterioração limitada para que o algoritmo como um todo também esteja bem definido, uma vez que todos os outros passos são exatamente iguais aos correspondentes no Algoritmo 8.1.

Analogamente ao estudo feito para problemas em dois níveis vamos listar nossas hipóteses para análise do método.

H8.21. Para qualquer s_k em Ω o processo de restauração garantirá que $\|H(\theta_k^R)\| \leq r\|H(\theta_k)\|$ e $\|C(s_k^R)\| \leq r\|C(s_k)\|$.

H8.22. Para qualquer que seja $i = 1, 2, \dots, q$ o problema do jogador i , descrito em (8.17) e denotado por $P_i(z_i, x_{-i}, \theta)$, tem solução única e esta é o único ponto estacionário do problema. Vamos denotar essa solução e os multiplicadores associados por $\bar{x}_i(x_{-i}, \theta)$, $\bar{\mu}_i(x_{-i}, \theta)$, $\bar{\gamma}_i(x_{-i}, \theta)$, ou simplesmente por \bar{x}_i , $\bar{\mu}_i$ e $\bar{\gamma}_i$. Além disso, vamos assumir que \bar{x}_i , $\bar{\mu}_i$ e $\bar{\gamma}_i$ sejam funções contínuas.

H8.23. Para qualquer que seja $i = 1, 2, \dots, q$ o Jacobiano de $C_i(x, \theta, \mu_i, \gamma_i)$ em relação às variáveis x_i, μ_i, γ_i , calculado no ponto $(\bar{x}_i, x_{-i}, \theta, \bar{\mu}_i, \bar{\gamma}_i)$, o qual é denotado por $J_{C_i}(\bar{x}_i, x_{-i}, \theta, \bar{\mu}_i, \bar{\gamma}_i)$, seja não singular.

- Na seção anterior já exibimos condições suficientes para que essa hipótese seja cumprida.

H8.24. Todos os iterandos do algoritmo se mantêm num compacto Ω .

De forma completamente análoga ao Lema 8.1.1 temos o próximo resultado.

Lema 8.2.1. *Suponhamos que valham as Hipóteses H8.22 a H8.24. Então, existem $\xi_{0i} > 0$ e $\delta_i > 0$, independentes de x_i e θ , tais que*

$$\|(x_i, \mu_i, \gamma_i) - (\bar{x}_i, \bar{\mu}_i, \bar{\gamma}_i)\| \leq \xi_{0i} \|C_i(x, \theta, \mu_i, \gamma_i)\|,$$

para todo $(x, \theta, \mu_i, \gamma_i)$ tal que $\|(x_i, \mu_i, \gamma_i) - (\bar{x}_i, \bar{\mu}_i, \bar{\gamma}_i)\| \leq \delta_i$.

Demonstração: Ao longo dessa demonstração vamos denotar o Jacobiano $J_{C_i}(\bar{x}_i, x_{-i}, \theta, \bar{\mu}_i, \bar{\gamma}_i)$ simplesmente por $J(x_{-i}, \theta)$. Como $J(x_{-i}, \theta)$ é não singular temos que existe $K > 0$ tal que:

$$\|J(x_{-i}, \theta)u\| \geq 2K\|u\|$$

para todo $u \in \mathbb{R}^{n+m}$ e $x \in \Omega$ e $\theta \in \Theta$.

Definindo

$$r(x, \theta, \mu_i, \gamma_i) = C_i(x_i, x_{-i}, \theta, \mu_i, \gamma_i) - C_i(\bar{x}_i, x_{-i}, \theta, \bar{\mu}_i, \bar{\gamma}_i) - J(x_{-i}, \theta) \begin{pmatrix} x_i - \bar{x}_i \\ \mu_i - \bar{\mu}_i \\ \gamma_i - \bar{\gamma}_i \end{pmatrix},$$

temos que $r(x, \theta, \mu_i, \gamma_i)$ também é contínua.

Utilizando esse fato, a diferenciabilidade de C_i e a compacidade de Ω temos que a função g_i definida por

$$g_i(x, \theta, \mu_i, \gamma_i) = \begin{cases} \frac{r(x, \theta, \mu_i, \gamma_i)}{\|(x_i, \mu_i, \gamma_i) - (\bar{x}_i, \bar{\mu}_i, \bar{\gamma}_i)\|}; & \text{se } (x_i, \mu_i, \gamma_i) \neq (\bar{x}_i, \bar{\mu}_i, \bar{\gamma}_i) \\ 0; & \text{se } (x_i, \mu_i, \gamma_i) = (\bar{x}_i, \bar{\mu}_i, \bar{\gamma}_i) \end{cases}$$

é uniformemente contínua para $x \in \Omega$ e $\theta \in \Theta$. Dessa forma, existem $\delta_i > 0$, independentes de x e θ , tais que

$$\|r(x, \theta, \mu_i, \gamma_i)\| \leq K_i \|(x_i, \mu_i, \gamma_i) - (\bar{x}_i, \bar{\mu}_i, \bar{\gamma}_i)\|.$$

Consequentemente, na vizinhança de raio δ_i de $(\bar{x}_i, \bar{\mu}_i, \bar{\gamma}_i)$ vale que:

$$\|C_i(x, \theta, \mu_i, \gamma_i)\| \geq K_i \|(x_i, \mu_i, \gamma_i) - (\bar{x}_i, \bar{\mu}_i, \bar{\gamma}_i)\|.$$

Sendo assim, se $\xi_{0i} = \frac{1}{K_i}$ temos o resultado desejado. \square

Considerando que vale o Lema 8.2.1 para os problemas de todos os jogadores, vamos denotar $\xi_0 = \max\{\xi_{01}, \xi_{02}, \dots, \xi_{0q}\}$ e $\delta = \min\{\delta_1, \delta_2, \dots, \delta_q\}$. Como $\sum_{i=1}^q \|(x, \mu, \gamma) - (\bar{x}, \bar{\mu}, \bar{\gamma})\|^2 = \sum_{i=1}^q \|(x_i, \mu_i, \gamma_i) - (\bar{x}_i, \bar{\mu}_i, \bar{\gamma}_i)\|^2$ temos que

$$\|(x, \mu, \gamma) - (\bar{x}, \bar{\mu}, \bar{\gamma})\| \leq \delta \Rightarrow \|(x_i, \mu_i, \gamma_i) - (\bar{x}_i, \bar{\mu}_i, \bar{\gamma}_i)\| \leq \delta \text{ para todo } i \in \{1, \dots, q\}.$$

Desse modo, usando o Lema 8.2.1, concluímos que

$$\begin{aligned} \|(x, \mu, \gamma) - (\bar{x}, \bar{\mu}, \bar{\gamma})\|^2 &= \sum_{i=1}^q \|(x_i, \mu_i, \gamma_i) - (\bar{x}_i, \bar{\mu}_i, \bar{\gamma}_i)\|^2 \\ &\leq \sum_{i=1}^q \xi_{0i}^2 \|C_i(x, \theta, \mu_i, \gamma_i)\|^2 \\ &\leq \xi_0^2 \sum_{i=1}^q \|C_i(x, \theta, \mu_i, \gamma_i)\|^2 \\ &\leq \xi_0^2 \|C(x, \theta, \mu, \gamma)\|^2 \end{aligned}$$

para todo $(x, \theta, \mu, \gamma) \in \Omega$ tal que $\|(x, \mu, \gamma) - (\bar{x}, \bar{\mu}, \bar{\gamma})\| \leq \delta$.

O próximo resultado é equivalente ao Lema 8.1.2 para a situação de equilíbrio.

Lema 8.2.2. *Suponhamos que valham as Hipóteses H8.22 a H8.24. Então, existe $\xi > 0$ tal que*

$$\|(x, \mu, \gamma) - (\bar{x}, \bar{\mu}, \bar{\gamma})\| \leq \xi \|C(x, \theta, \mu, \gamma)\|$$

para todo $(x, \theta, \mu, \gamma) \in \Omega$.

Demonstração: Primeiramente observamos que o limitante local para o erro, $\|(x, \mu, \gamma) - (\bar{x}, \bar{\mu}, \bar{\gamma})\| \leq \xi_0 \|C(x, \theta, \mu, \gamma)\|$ sempre que $\|(x, \mu, \gamma) - (\bar{x}, \bar{\mu}, \bar{\gamma})\| \leq \delta$, cumpre o papel do Lema 8.1.1 da seção anterior. Como a função distância $d(x, \theta, \mu, \gamma) = \|(x, \mu, \gamma) -$

$(\bar{x}, \bar{\mu}, \bar{\gamma})\|$ é contínua e Ω é compacto, a demonstração segue exatamente igual à referente ao Lema 8.1.2. \square

Com isso podemos demonstrar que a relação (8.20) é satisfeita.

Teorema 8.2.1. *Suponhamos que valham as Hipóteses H8.21 a H8.24. Então o ponto $s_k^R = (\bar{x}_k, \theta_k, \bar{\mu}_k, \bar{\gamma}_k)$ satisfaz as condições da restauração para qualquer que seja $\lambda_{k+1} \in \Omega_\lambda$.*

Demonstração: A condição (8.19) já está demonstrada e a condição (8.20) é consequência imediata do Lema 8.2.2 e das propriedades de Lipschitz das funções envolvidas, assim como foi feito no Teorema 8.1.3. \square

Como já dissemos, todos os demais passos do Algoritmo 8.2 são iguais ao Algoritmo 8.1 e portanto estão bem definidos. Além disso, podemos observar que os passos realizados pelos Algoritmos 8.2, 8.1, e 5.1 garantem que as mesmas relações são satisfeitas. A única diferença entre eles é a maneira de medir a inviabilidade. Da mesma forma, podemos enunciar o teorema a seguir, já que sua demonstração segue exatamente às feitas para os resultados análogos referentes aos Algoritmos 5.1 e 8.1, enunciados nos Teoremas 5.1.2, 8.1.4, 5.1.3 e 8.1.5.

Teorema 8.2.2. *Supondo que valham as hipóteses H8.21 a H8.24 e que $\{s_k\}$ seja uma sequência gerada pelo Algoritmo 8.2, então:*

a) *Existem $\bar{\rho}, \bar{t} > 0$ tais que $\rho_k \geq \bar{\rho}$ e $t_k \geq \bar{t}$ para todo k .*

b)

$$\sum_{k=0}^{\infty} \left\| \begin{pmatrix} H(\theta_k) \\ C(s_k) \end{pmatrix} \right\| = \hat{\sigma} < \infty.$$

Consequentemente todo ponto limite é viável para o problema reformulado (8.18).

c) *Todo ponto limite da sequência $\{s_k^R\}$ é AKKT para o problema reformulado (8.18).*

Experimentos computacionais ilustrativos

Para dar um indício de como nosso algoritmo funciona vamos testá-lo em alguns exemplos simples que relataremos brevemente a seguir. Nossos testes numéricos não têm o intuito de fornecer uma análise científica rigorosa sobre o Algoritmo 8.2. O objetivo aqui é simplesmente verificar a possibilidade de fazer uma implementação do Algoritmo.

Nosso primeiro exemplo será o problema (8.15) utilizado na introdução desta seção. Para diminuirmos os graus de liberdade do problema, vamos supor conhecidos primeiramente os parâmetros $a = 2$, $b = 1$, $c = 20$, $a' = 1$ e $b' = 3$, e apenas o parâmetro c' desconhecido. Portanto o problema a ser resolvido é:

$$\begin{array}{l}
 \text{Minimizar} \quad (x - 40)^2 + (y - 60)^2 \\
 \quad \quad \quad c', x, y \\
 \\
 \text{sujeito a:} \quad x \text{ é solução do problema:} \\
 \quad \quad \quad \left. \begin{array}{l}
 \text{Minimizar} \quad (2x - y - 20)^2 \\
 \quad \quad \quad x \\
 \text{sujeito a:} \quad 0 \leq x \leq 100
 \end{array} \right\} \text{(Problema da empresa 1)} \\
 \\
 \quad \quad \quad y \text{ é solução do problema:} \\
 \quad \quad \quad \left. \begin{array}{l}
 \text{Minimizar} \quad (x - 3y + c')^2 \\
 \quad \quad \quad y \\
 \text{sujeito a:} \quad 0 \leq y \leq 100
 \end{array} \right\} \text{(Problema da empresa 2)}
 \end{array}$$

Uma implementação simples para resolver esse problema foi feita em *Matlab* e alguns detalhes estão descritos a seguir. Na restauração usamos a expressão exata dos problemas de cada empresa, $x(y) = \max\{0, \min\{\frac{20+y}{2}, 100\}\}$ e $y(x, c') = \max\{0, \min\{\frac{c'+x}{3}, 100\}\}$. Duas alternativas para parar a restauração foram usadas. Uma delas foi aplicar apenas uma iteração do método de Jacobi e a outra foi aplicar Jacobi até obtermos que a condição (8.19) fosse satisfeita com $r = 0.001$. No passo da otimização utilizamos a função *fmincon* para resolver os subproblemas. Os multiplicadores de Lagrange foram sempre tomados como nulos e o modelo quadrático

foi construído com a Hessiana verdadeira. As restrições lineares referentes ao espaço tangente são constantes nesse caso e a matriz que as determina é

$$J_C = \begin{pmatrix} 2 & -1 & 0 \\ 1 & -3 & 1 \end{pmatrix}.$$

Numa segunda versão consideraremos apenas o parâmetro b' desconhecido e adotamos $c' = 140$ conforme descrito a seguir.

$$\begin{array}{l} \text{Minimizar} \\ b', x, y \end{array} (x - 40)^2 + (y - 60)^2$$

sujeito a: x é solução do problema:

$$\left. \begin{array}{l} \text{Minimizar} \\ x \end{array} (2x - y - 20)^2 \right\} \text{(Problema da empresa 1)}$$

$$\text{sujeito a: } 0 \leq x \leq 100$$

y é solução do problema:

$$\left. \begin{array}{l} \text{Minimizar} \\ y \end{array} (x - b'y + 140)^2 \right\} \text{(Problema da empresa 2)}$$

$$\text{sujeito a: } 0 \leq y \leq 100$$

Nesse segundo caso temos uma não linearidade nas restrições do problema reformulado nas variáveis x, y, b' . A implementação feita para esse caso é basicamente a mesma feita para o problema anterior. Na fase de restauração apenas substituímos a expressão de $y(x, c')$ por $y(x, b') = \max\{0, \min\{\frac{140+x}{b'}, 100\}\}$. Já na fase de otimização as restrições lineares são

$$J_C = \begin{pmatrix} 2 & -1 & 0 \\ 1 & -b'_k & y_r^k \end{pmatrix},$$

onde y_r^k é a decisão da segunda empresa obtida depois da fase de restauração.

Como ponto inicial adotamos $(x, y) = (35, 70)$ e para os parâmetros desconhecidos escolhemos na primeira versão $c' = 150$, e na segunda, $b' = 3.5$. Em todos os testes foi usado como critério de parada que a norma da diferença entre dois iterandos

consecutivos fosse menor que 10^{-5} . O algoritmo convergiu para a solução em todos os testes. A convergência foi obtida em 17 iterações restaurando com apenas uma iteração do método de Jacobi. Quando utilizamos o segundo critério de parada na restauração obtivemos convergência com 6 iterações.

Nossos próximos exemplos serão relativos a encontrar um equilíbrio em problemas de Arrow-Debreu, descritos na seção 7.1. No primeiro exemplo vamos considerar o caso de equilíbrio mais simples que descrevemos. Nesse caso temos que os problemas envolvidos, já acrescentando variáveis de folga, são:

- Problema do consumidor:

$$\begin{aligned} &\text{Maximizar } x \\ &\quad x, z \\ &\text{sujeito a: } px + z - 100 = 0 \\ &\quad x, z \geq 0. \end{aligned} \tag{8.24}$$

- Problema da firma:

$$\begin{aligned} &\text{Maximizar } py \\ &\quad y, w \\ &\text{sujeito a: } y + w = 10 \\ &\quad y, w \geq 0. \end{aligned} \tag{8.25}$$

As condições KKT desses problemas são, respectivamente:

$$\begin{aligned} \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} p \\ 1 \end{pmatrix} \lambda_1 + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \lambda_2 + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \lambda_3 &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ px + z - 100 &= 0 \\ x\lambda_2 &= 0 \\ z\lambda_3 &= 0 \\ x, z, \lambda_2, \lambda_3 &\geq 0; \end{aligned}$$

e

$$\begin{aligned} \begin{pmatrix} p \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} \lambda_4 + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \lambda_5 + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \lambda_6 &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ y + w - 10 &= 0 \\ y\lambda_5 &= 0 \\ w\lambda_6 &= 0 \\ y, w, \lambda_5, \lambda_6 &\geq 0. \end{aligned}$$

A linearização dessas condições em $\bar{p}, \bar{x}, \bar{z}, \bar{y}, \bar{w}, \bar{\lambda}$ é:

$$\bar{\lambda}_1(p - \bar{p}) + \bar{p}(\lambda_1 - \bar{\lambda}_1) + (\lambda_2 - \bar{\lambda}_2) = 0 \quad (8.26)$$

$$(\lambda_1 - \bar{\lambda}_1) + (\lambda_3 - \bar{\lambda}_3) = 0 \quad (8.27)$$

$$\bar{p}(x - \bar{x}) + \bar{x}(p - \bar{p}) + (z - \bar{z}) = 0 \quad (8.28)$$

$$\bar{\lambda}_2(x - \bar{x}) + \bar{x}(\lambda_2 - \bar{\lambda}_2) = 0 \quad (8.29)$$

$$\bar{\lambda}_3(z - \bar{z}) + \bar{z}(\lambda_3 - \bar{\lambda}_3) = 0 \quad (8.30)$$

$$(p - \bar{p}) + (\lambda_4 - \bar{\lambda}_4) + (\lambda_5 - \bar{\lambda}_5) = 0 \quad (8.31)$$

$$(\lambda_4 - \bar{\lambda}_4) + (\lambda_6 - \bar{\lambda}_6) = 0 \quad (8.32)$$

$$(y - \bar{y}) + (w - \bar{w}) = 0 \quad (8.33)$$

$$\bar{\lambda}_5(y - \bar{y}) + \bar{y}(\lambda_5 - \bar{\lambda}_5) = 0 \quad (8.34)$$

$$\bar{\lambda}_6(w - \bar{w}) + \bar{w}(\lambda_6 - \bar{\lambda}_6) = 0 \quad (8.35)$$

$$p, x, y, z, w, \lambda_2, \lambda_3, \lambda_5, \lambda_6 \geq 0. \quad (8.36)$$

Dado um preço \bar{p} e encontrando as outras variáveis, resolvendo os problemas do consumidor e da firma com \bar{p} fixado, o que corresponde à fase de restauração, temos que $\bar{x} = \frac{100}{\bar{p}}$, $\bar{z} = 0$, $\bar{\lambda}_1 = -\frac{1}{\bar{p}}$, $\bar{\lambda}_2 = 0$, $\bar{\lambda}_3 = -\bar{\lambda}_1$, $\bar{y} = 10$, $\bar{w} = 0$, $\bar{\lambda}_4 = -\bar{p}$, $\bar{\lambda}_5 = 0$ e $\bar{\lambda}_6 = \bar{p}$. Para forçar que a lei da oferta e da demanda esteja cumprida vamos minimizar a folga entre a quantidade consumida e a produzida. Para isso vamos usar o enfoque

apresentado em (7.9). No nosso caso teremos:

$$\begin{aligned} & \text{Minimizar } (x - y)^2 \\ & x, y, z, w, p \\ & \text{sujeito a: } (x, z) \text{ é solução de (8.24),} \\ & (y, w) \text{ é solução de (8.25),} \\ & p \geq 0. \end{aligned} \tag{8.37}$$

Portanto, na fase de otimização resolveríamos um modelo quadrático da função objetivo de (8.37) sujeito à linearização das condição KKT dos problemas (8.24) e (8.25). Nesse caso o modelo quadrático coincide com a função objetivo. Já o sistema das restrições pode ser reduzido com algumas eliminações algébricas. Para isso usaremos os valores encontrados na fase de restauração para construir a linearização.

Por (8.29) e (8.34) temos que $\lambda_2 = \lambda_5 = 0$. De (8.30) e (8.36) temos que $z = w = 0$ e conseqüentemente, por (8.31), (8.32) e (8.33) respectivamente, temos que $\lambda_4 = -p$, $\lambda_6 = p$ e $y = 10$. Dessa forma a solução do problema consiste em encontrar o ponto viável cuja coordenada x mais se aproxime de 10. Explorando isso temos que uma maneira de tentar encontrar uma solução do problema é assumir $x = 10$ e verificar se as demais equações tem solução. Da equação (8.28) temos que

$$p = \bar{p} \left(2 - \frac{10}{\bar{x}} \right) = \bar{p} \left(2 - \frac{\bar{p}}{10} \right). \tag{8.38}$$

Dessa forma teríamos de (8.26) e (8.27), respectivamente, que $\lambda_1 = -\frac{\bar{\lambda}_1 \bar{p}}{10} = -\frac{1}{10}$ e $\lambda_3 = -\lambda_1$. Portanto as equações (8.26) a (8.35) tem solução. Mais do que isso, se tomarmos $p_0 \in (0, 20)$ e definirmos a sequência

$$p_{k+1} = p_k \left(2 - \frac{p_k}{10} \right)$$

temos que $p_{k+1} - 10 = \frac{p_k - 10}{10} (p_k - 10)$ e portanto $p_k \in (0, 20)$ para todo k e p_k converge a 10. Dessa forma a positividade exigida em (8.36) é sempre satisfeita.

Pelas considerações acima temos que em uma versão sem busca linear de nosso método a convergência para o equilíbrio $(x^*, y^*, p^*) = (10, 10, 10)$ está garantida.

Em uma implementação simples feita em Matlab obtivemos os mesmos resultados. Considerando como ponto inicial para o preço todos os pontos inteiros do intervalo $(0, 20)$ o algoritmo aceitou o passo $t_k = 1$ em todas as iterações de todos os testes.

O problema abordado no exemplo anterior é estável, no sentido de que as soluções tanto do problema da firma como do problema do consumidor variam suavemente se variarmos os parâmetros dos problemas numa vizinhança do equilíbrio. Isso não acontece no exemplo mais completo de equilíbrio de Arrow-Debreu que exibimos na Seção 7.1. A seguir vamos lembrar os problemas envolvidos na situação de equilíbrio e analisar o comportamento numérico de nosso algoritmo nesse caso.

- Problema do consumidor:

$$\begin{aligned} & \text{Maximizar} && z^{10}(100 - x) \\ & && x, z \\ & \text{sujeito a:} && p_3 z \leq p_1 x \\ & && 0 \leq x \leq 100 \\ & && z \geq 0. \end{aligned} \tag{8.39}$$

- Problema do proprietário:

$$\begin{aligned} & \text{Maximizar} && w^5(100 - y) \\ & && w, y \\ & \text{sujeito a:} && p_3 w \leq p_2 y + p_3 s - p_1 t - p_2 r \\ & && 0 \leq y \leq 100 \\ & && w \geq 0. \end{aligned} \tag{8.40}$$

- Problema da firma:

$$\begin{aligned} & \text{Maximizar} && p_3 s - p_1 t - p_2 r \\ & && s, t, r \\ & \text{sujeito a:} && 0 \leq s \leq t + 2r \\ & && 0 \leq t \leq 100 \\ & && 0 \leq r \leq 100. \end{aligned} \tag{8.41}$$

A alternativa para encontrarmos os pontos de equilíbrio será novamente resolver o problema (7.9), o que nesse caso particular passa a ser:

$$\begin{aligned}
 &\text{Minimizar} && (x - t)^2 + (y - r)^2 + (z + w - s)^2 \\
 &x, y, z, w, s, r, t, p \\
 &\text{sujeito a:} && (x, z) \text{ é solução de (8.39),} \\
 &&& (y, w) \text{ é solução de (8.40),} \\
 &&& (s, r, t) \text{ é solução de (8.41),} \\
 &&& p \geq 0.
 \end{aligned} \tag{8.42}$$

Analisando o comportamento do algoritmo percebemos que muitas vezes a solução do problema era encontrada mas não era reconhecida. Isso acontecia pois a solução era encontrada na fase de otimização mas o critério de parada está associado à fase de restauração. Como o problema da firma tem infinitas soluções possíveis para $p = p^*$, e apenas uma delas é a real solução do problema de equilíbrio, era comum ver que a solução era perdida nessa fase. Um fator que complica ainda mais obter a solução é o fato dela ser instável, no sentido que pequenas variações em p fazem com que a solução do problema da firma seja bem diferente. Esta é uma situação que contraria a continuidade exigida nas hipóteses dos teoremas de convergência associados ao nosso algoritmo. Entretanto, mesmo assim tentamos fazer modificações que deixassem mais estável uma implementação do algoritmo para esse problema particular. A modificação mais simples foi simplesmente colocar como ponto inicial ao resolver o problema da firma os valores encontrados nos problemas do consumidor e do proprietário. Como a função *Linprog* do *Matlab* usa um algoritmo de pontos interiores para resolver o PL muitas vezes a solução passava a ser mantida pela fase de restauração. A segunda modificação que fizemos foi substituir o problema da firma por um problema regularizado que uma vez na face ótima desse preferência ao ponto que mais force que a lei da oferta e da demanda fosse cumprida. Portanto, nessa situação, o problema da firma de fato

resolvido foi

$$\begin{aligned}
 & \text{Maximizar} && p_3 s - p_1 t - p_2 r - \theta ((s - z - w)^2 + (t - x)^2 + (r - y)^2) \\
 & && s, t, r \\
 & \text{sujeito a:} && 0 \leq s \leq t + 2r \\
 & && 0 \leq t \leq 100 \\
 & && 0 \leq r \leq 100,
 \end{aligned} \tag{8.43}$$

onde θ é o fator de regularização que incentiva que a diferença entre oferta e demanda seja a menor possível.

O algoritmo foi implementado colocando salvaguardas nos preços de forma que $p_1 = 1$ (fator de escala), $p_2, p_3 \in [0.01, 10]$. Foram realizados cem testes com preços iniciais da forma $p = (1, a, b)^T$, onde a e b são números aleatórios uniformemente distribuídos em $[0.01, 10]$. Por razões de escala a função objetivo do consumidor foi substituída pela raiz décima da descrita em (8.39) e do proprietário pela raiz quinta da descrita em (8.40). Considerando uma tolerância de 5% tanto nas decisões quanto nos preços encontrados a solução verdadeira foi obtida pelo algoritmo em 68 vezes simplesmente optando pelo novo ponto inicial no PL e em 56 vezes pela formulação utilizando a regularização.

Nossos últimos exemplos também são baseados nos modelos de equilíbrio geral de Arrow-Debreu. O exemplo a seguir trata-se de um problema de Equilíbrio Inverso de Nash. Vamos voltar ao exemplo (8.37) mas vamos considerar desconhecida a quantidade de dinheiro M do consumidor. Entretanto, vamos considerar que sejam observados os valores de x^* , y^* e p^* . Portanto temos que descobrir M para que esses valores correspondam a um ponto de equilíbrio. Os problemas do sistema estão descritos a seguir.

- Problema do consumidor:

$$\begin{aligned}
 & \text{Maximizar} && x \\
 & && x \\
 & \text{sujeito a:} && px \leq M \\
 & && x \geq 0.
 \end{aligned} \tag{8.44}$$

- Problema da firma:

$$\begin{aligned} &\text{Maximizar } py \\ &\quad y \end{aligned} \tag{8.45}$$

$$\text{sujeito a: } 0 \leq y \leq 10.$$

- Problema do mercado:

$$\begin{aligned} &\text{Maximizar } (y - x)p \\ &\quad p \end{aligned} \tag{8.46}$$

$$\text{sujeito a: } p \geq 0.$$

- Logo o problema de equilíbrio passa a ser:

$$\begin{aligned} &\text{Minimizar } (x - 10)^2 + (y - 10)^2 + (p - 10)^2 \\ &\quad x, y, p, M \end{aligned} \tag{8.47}$$

$$\text{sujeito a: } x \text{ é solução de (8.44),} \tag{8.47}$$

$$y \text{ é solução de (8.45),}$$

$$p \text{ é solução de (8.46).}$$

Em uma implementação simples de Matlab recuperamos a quantidade $M = 100$ em diversos testes para vários pontos iniciais diferentes. Aqui não linearizamos as condições KKT do problema do mercado por ele ser instável e ilimitado para pequenos desvios do valor de equilíbrio. Além disso, na fase de restauração, atualizamos o preço pela seguinte expressão:

$$p = p \max \left\{ 0.9, \min \left\{ \frac{x}{y}, 1.1 \right\} \right\}.$$

Dessa forma, quando temos excesso de demanda, o preço sobe, e quando temos excesso de oferta, o preço cai.

Vamos considerar agora que existam dois consumidores, duas firmas e dois produtos a serem comercializados no mercado. Os consumidores não trabalham e possuem uma quantidade de dinheiro inicial, cem unidades monetárias para o primeiro e cento e cinquenta para o segundo. Além disso, o primeiro consumidor só pode comprar

o produto 1, o qual é produzido apenas pela firma 1. Já o consumidor 2 pode adquirir qualquer um dos dois bens e a firma 2 só produz o bem 2. Além disso, vamos assumir a função de utilidade do segundo consumidor como sendo da forma Cobb-Douglas. As variáveis envolvidas na modelagem, bem como o problema de cada agente, estão descritos a seguir.

- x - Quantidade do bem 1 adquirido pelo consumidor 1.
- y - Quantidade do bem 1 adquirido pelo consumidor 2.
- z - Quantidade do bem 2 adquirido pelo consumidor 2.
- w - Quantidade do bem 1 produzido pela firma 1.
- r - Quantidade do bem 2 produzido pela firma 2.
- a - Coeficiente de satisfação do consumidor 2 em relação ao produto 1.
- Problema do consumidor 1:

$$\begin{aligned} & \text{Maximizar } x \\ & \quad x \\ & \text{sujeito a: } p_1x \leq 100 \\ & \quad x \geq 0. \end{aligned} \tag{8.48}$$

- Problema do consumidor 2:

$$\begin{aligned} & \text{Maximizar } y^a z \\ & \quad y, z \\ & \text{sujeito a: } p_1y + p_2z \leq 150 \\ & \quad y, z \geq 0. \end{aligned} \tag{8.49}$$

- Problema da firma 1:

$$\begin{aligned} & \text{Maximizar } p_1w \\ & \quad w \\ & \text{sujeito a: } 0 \leq w \leq 10. \end{aligned} \tag{8.50}$$

- Problema da firma 2:

$$\begin{aligned} & \text{Maximizar } p_2 r \\ & \quad r \\ & \text{sujeito a: } 0 \leq r \leq 10. \end{aligned} \tag{8.51}$$

Considerando p_1 e p_2 positivos e analisando as condições KKT dos problemas chegamos que $x = \frac{100}{p_1}$, $y = \frac{a}{a+1} \frac{150}{p_1}$, $z = \frac{150}{(a+1)p_2}$, $w = r = 10$. Igualando r e z obtemos que $p_2 = \frac{15}{a+1}$. Igualando demanda e oferta para o bem 1 temos que $p_1 = 10 + \frac{15}{a+1}$. Para $a = 1$ obtemos que $p_1 = 17,50$, $p_2 = 7,50$, $x = \frac{40}{7}$, $y = \frac{30}{7}$, $z = 10$.

Entretanto, o problema que queremos resolver nesse caso é o problema de Equilíbrio Inverso de Nash em que observamos as decisões x, y, z, w, r e os preços p_1 e p_2 mas estamos interessados em descobrir qual o parâmetro a utilizado pelo consumidor 2. Para resolver esse problema vamos considerar o problema de equilíbrio geral como um problema de Equilíbrio de Nash reformulando a condição (EW-3) pelo problema (7.6), conforme descrito no Capítulo 7. Nesse caso o problema do mercado seria

$$\begin{aligned} & \text{Maximizar } (x + y - w)p_1 + (z - r)p_2 \\ & \quad p \\ & \text{sujeito a: } p \geq 0. \end{aligned} \tag{8.52}$$

Dessa forma, definido

$$F(x, y, z, w, r, p, a) = \left\| \begin{pmatrix} x - \frac{40}{7} \\ y - \frac{30}{7} \\ z - 10 \\ w - 10 \\ r - 10 \\ p_1 - 7,50 \\ p_2 - 17,50 \end{pmatrix} \right\|^2$$

o problema de Equilíbrio Inverso de Nash seria:

$$\begin{aligned}
 & \text{Minimizar} && F(x, y, z, w, r, p, a) \\
 & x, y, z, w, r, p, a \\
 & \text{sujeito a:} && x \text{ é solução de (8.48),} \\
 & && (y, z) \text{ é solução de (8.49),} \\
 & && w \text{ é solução de (8.50),} \\
 & && r \text{ é solução de (8.51),} \\
 & && p \text{ é solução de (8.52).}
 \end{aligned} \tag{8.53}$$

Diversos testes foram feitos para diferentes pontos iniciais no intervalo $[0.7, 1.3]$. Muitas vezes a fase de otimização não foi bem sucedida por falha da função *quadprog* do Matlab. Para superar esse inconveniente aceitamos o passo fornecido pela fase de otimização modificando a variável a de acordo com o sentido prático do problema. Atualizamos a conforme a expressão a seguir

$$a = a \min \left\{ 0.9, \max \left\{ \frac{y_{obs}}{y}, 1.1 \right\} \right\},$$

onde y_{obs} é a observação referente à variável y . Com essa atualização estamos procurando atender à ideia intrínseca do problema de que se o valor y consumido foi menor do que o observado então o valor do parâmetro de interesse utilizado no modelo deve aumentar, incentivando um maior consumo. Analogamente, para y maior que y_{obs} diminuimos a de modo a inibir o consumo de y no modelo. Para a fase de restauração foi usada uma atualização semelhante para os preços, já que geralmente o problema do mercado é ilimitado. Com essas escolhas a solução foi encontrada em todos os testes.

Uma das características mais importantes dos métodos de Restauração Inexata é essa liberdade para usar em cada fase um algoritmo que se adeque ao problema. Embora tenhamos especificado na teoria o uso de métodos de Jacobi para a restauração e do modelo quadrático para a otimização basta que melhoremos sensivelmente o objetivo correspondente na fase em questão sem deteriorar o outro para que o método seja aplicável. Isso permite que ajustemos os métodos utilizados na resolução dos subproble-

mas para melhor explorar seu significado prático, conforme fizemos nos dois exemplos anteriores.

No próximo exemplo vamos ainda usar a base do exemplo anterior para ilustrar a situação normativa. Nesse caso o governo tem que decidir quanto de subsídio dar ao consumidor 1 para que a satisfação entre os consumidores não seja tão desigual. Para $a = 1$ a satisfação do consumidor 2 no exemplo anterior é 7.5 vezes maior do que a do consumidor 1. Tentando diminuir essa diferença vamos supor que a função objetivo do governo seja deixar a satisfação do consumidor 2 como sendo duas vezes a do consumidor 1. Sendo assim, vamos considerar

$$F(x, y, z, w, r, p, s) = (2x - yz)^2,$$

onde s é um parâmetro de subsídio que o governo usa para regular o consumo, de forma que o problema do consumidor 1 passa a ser:

$$\begin{aligned} &\text{Maximizar } x \\ &\quad x \\ &\text{sujeito a: } \quad sp_1x \leq 100 \\ &\quad x \geq 0. \end{aligned} \tag{8.54}$$

Dessa forma temos que $x = \frac{100}{sp_1}$, $y = \frac{75}{p_1}$, $z = \frac{75}{p_2}$, $w = r = 10$. Igualando r e z obtemos que $p_2 = 7.50$. Igualando demanda e oferta para o bem 1 temos que $x + y = 10$. Considerando verdadeira a igualdade desejada $2x = yz$ temos que $x = 5y$. Portanto $x = \frac{50}{6}$ e $y = \frac{10}{6}$. Da expressão de y concluímos que $p_1 = 45$ e da expressão de x obtemos que $s = \frac{4}{15}$.

Numericamente procuramos determinar o valor do parâmetro s resolvendo

o problema de equilíbrio equivalente

$$\begin{aligned}
 & \text{Minimizar} && (2x - yz)^2 \\
 & x, y, z, w, r, p, s \\
 & \text{sujeito a:} && x \text{ é solução de (8.54),} \\
 & && (y, z) \text{ é solução de (8.49),} \\
 & && w \text{ é solução de (8.50),} \\
 & && r \text{ é solução de (8.51),} \\
 & && p \text{ é solução de (8.52),} \\
 & && s \in [0, 1].
 \end{aligned} \tag{8.55}$$

Novamente fizemos ajustes na fase de otimização, nesse caso consideramos:

$$s = \min \left\{ 0.9, \max \left\{ \frac{2x}{yz}, 1.1 \right\} \right\}.$$

Ou seja, se a satisfação do consumidor 1 está abaixo do desejado o valor de s diminui o que representa um aumento do subsídio. Caso contrário, se a satisfação estiver acima do desejado, o subsídio diminui. Com essas modificações a solução analítica foi encontrada para todos os pontos iniciais testados.

8.3 LOVO aplicada ao Equilíbrio Inverso de Nash

Nesta última seção vamos mostrar um exemplo simples em que juntamos as principais ideias expostas nessa tese.

O problema de interesse consiste em modelar uma situação de equilíbrio em que o preço de alguns dos bens desejáveis não sejam definidos pela lei de mercado. Desse modo, cada agente deve maximizar sua função de utilidade sujeito a restrições de risco referentes aos bens cujo preço é uma variável aleatória independente das decisões dos agentes envolvidos na situação de equilíbrio.

A alternativa que propomos para encontrar uma solução para esse problema é usar a técnica de Restauração Inexata utilizada nessa seção. A particularidade aqui

é que na fase de restauração o problema de risco de cada agente será resolvido com a estratégia LOVO apresentada nos capítulos 2 e 3.

O exemplo que usaremos para ilustrar essa ideia está descrito a seguir. Vamos considerar um mercado com 1 consumidor e 1 firma. Assumiremos que o consumidor possa adquirir 2 ativos, cujos preços são dados por fatores externos e 1 bem produzido pela firma, que terá o preço dado pela situação de mercado. A lista das variáveis envolvidas nessa formulação é:

- x_i , $i = 1, 2$ - Quantidade do ativo externo i adquirida pelo consumidor.
- y - Quantidade adquirida pelo consumidor do bem produzido pela firma.
- z - Quantidade produzida pela firma.
- p - Preço do bem produzido pela firma.
- q_i^0 , $i = 1, 2$ - Preço inicial do ativo externo i .
- q_i^j , $i = 1, 2$, $j = 1, 2, 3$ - Preço do ativo externo i sob o cenário futuro j .
- T - Tolerância para a restrição do tipo VaR.
- M - Dinheiro disponível para o consumidor.
- W - Produção máxima para a firma.

Vamos assumir a medida de satisfação do consumidor como sendo o produto entre o dinheiro que ele possui e a quantidade adquirida do bem negociado no sistema. Como a quantidade de dinheiro decorrente do investimento nos ativos externos é uma variável aleatória, vamos assumir que a função de utilidade do consumidor seja a satisfação futura média. Como medida de risco vamos exigir que a satisfação em pelo menos dois dos três cenários futuros seja maior do que T . Além disso vamos denotar a satisfação no cenário j por $f_j(x, y) = (q_1^j x_1 + q_2^j x_2)y$ e, como no Capítulo 1, vamos adotar que

$$f_{j_1(x,y)}(x, y) \leq f_{j_2(x,y)}(x, y) \leq f_{j_3(x,y)}(x, y).$$

Dessa forma o problema do consumidor passa a ser:

$$\begin{aligned}
 & \underset{x, y}{\text{Maximizar}} \quad \left(x_1 \frac{\sum_{j=1}^3 q_1^j}{3} + x_2 \frac{\sum_{j=1}^3 q_2^j}{3} \right) y \\
 & \text{sujeito a:} \quad q_1^0 x_1 + q_2^0 x_2 + py \leq M \\
 & \quad \quad \quad f_{j_2(x,y)}(x, y) \geq T \\
 & \quad \quad \quad f_{j_3(x,y)}(x, y) \geq T \\
 & \quad \quad \quad x, y \geq 0.
 \end{aligned} \tag{8.56}$$

Já o problema da firma é:

$$\begin{aligned}
 & \underset{z}{\text{Maximizar}} \quad pz \\
 & \text{sujeito a:} \quad 0 \leq z \leq W.
 \end{aligned} \tag{8.57}$$

Sendo assim o problema de equilíbrio pode ser formulado como:

$$\begin{aligned}
 & \underset{x, y, z, p}{\text{Minimizar}} \quad (y - z)^2 \\
 & \text{sujeito a:} \quad (x, y) \text{ é solução de (8.56),} \\
 & \quad \quad \quad z \text{ é solução de (8.57),} \\
 & \quad \quad \quad p \geq 0.
 \end{aligned} \tag{8.58}$$

Uma implementação em Matlab foi feita para resolver esse problema com os parâmetros $q_1^0 = q_2^0 = 0.5$, $q^1 = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$, $q^2 = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$, $q^3 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $T = 1000$, $M = 100$, $W = 10$. Foi usado o Algoritmo 2.3.1 para resolver o problema LOVO do consumidor na fase de restauração. Na fase de otimização fixamos os índices $i_2(x^R, y^R)$ e $i_3(x^R, y^R)$ para construir a linearização das condições *KKT* do problema (8.56). Foram feitos vários testes com o valor inicial de p no intervalo $[2, 8]$. Em todos eles foram encontradas soluções do tipo $(x^*, y^*, z^*, p^*) = (x, 10, 10, 5)$ com $x_1 + x_2 = 100$. Todos esses pontos são minimizadores locais do problema (8.58).

Capítulo 9

Conclusões e perspectivas futuras

Nesse trabalho estudamos as técnicas de otimização do tipo LOVO e de Restauração Inexata e demos contribuições importantes sobre esses temas. Vários métodos foram propostos e suas propriedades teóricas analisadas. Problemas práticos foram resolvidos com essas propostas e exemplos ilustrativos foram fornecidos. Dentre as aplicações dos métodos desenvolvidos está o Equilíbrio Inverso de Nash que foi aqui descrito e exemplificado em problemas de equilíbrio geral de Arrow-Debreu.

Apresentamos algoritmos do tipo Lagrangiano Aumentado para tratar o problema de otimização (de grande porte) de carteiras com custos de transação. A estrutura deste problema de otimização é diferente da estrutura padrão de programação não-linear suave. A falta de suavidade de nosso problema aparece de uma forma muito particular, que nos permite explorar plenamente as características principais do paradigma de otimização do tipo LOVO. Experimentos numéricos de grande porte, que foram realizados em [26] e reproduzidos aqui no Capítulo 3, atestam pela aplicabilidade dessa teoria. Embora a teoria referente a essa aplicação seja coerente com processos de otimização para minimizadores globais, os testes computacionais foram feitos utilizando *softwares* cujos resultados de convergência são referentes a pontos estacionários. Um trabalho futuro interessante é implementar um método para otimização global nos subproblemas dos Algoritmos 2.2.1, 2.3.1, e 2.4.1.

No Capítulo 5 apresentamos um novo método de Restauração Inexata que combina uma modificação no Algoritmo global de Fischer-Friedlander com o método local de Birgin-Martínez. Nosso método preserva as boas propriedades teóricas dos dois métodos. Um exemplo ilustrativo de que nossa modificação no método global pode evitar o Efeito Maratos foi exibido. As relações entre o passo 1 do método global e as condições do método local foram estudadas. Sob hipóteses adequadas mostramos que o método híbrido proposto cumpre os requisitos necessários para garantir a convergência global sem perder a eficiência do método local. Experimentos computacionais mais contundentes para validar a praticidade do algoritmo proposto estão dentro das possibilidades de trabalhos futuros relacionados com este texto.

Um novo método de Restauração Inexata sem busca linear fortemente relacionado com problemas de otimização em que as derivadas da função objetivo não estão disponíveis foi apresentado. Nesse trabalho mostramos que a filosofia da Restauração Inexata pode se adequar bem a esses problemas já que podemos usar todas as ferramentas da otimização suave na fase de restauração. O desempenho do método proposto foi comparado ao método de Lagrangiano Aumentado, *HOPSPACK*, em uma grande quantidade de problemas teste. Os resultados, exibidos em [33] e reproduzidos no Capítulo 6, se mostraram muito animadores. O uso de multiplicadores de Lagrange no método é um ponto a ser estudado no futuro. Um estudo sobre como tratar a restauração com um método do tipo Quase-Newton também está entre nossos objetivos futuros. Com isso pretendemos estender o uso do método para problemas em que as derivadas das restrições também não estejam disponíveis e analisar os aspectos tanto da convergência global quanto da local. Outro trabalho futuro interessante é o uso de métodos de interpolação na fase de otimização do Algoritmo 6.

No Capítulo 8 estendemos o método de Restauração Inexata de Fischer-Friedlander para problemas em dois níveis seguindo a mesma filosofia apresentada em [10] para outro método de Restauração Inexata. A análise de convergência teórica do método foi feita e um exemplo ilustrativo foi exibido. Uma implementação de um método muito semelhante foi feita independentemente a este trabalho por A. Friedlander

e F. Gomes para problemas de estruturas. O sucesso dos resultados numéricos obtidos para esse problema particular são muito encorajadores, indicando que implementações de nosso algoritmo possam ser eficientes. Trabalhos futuros na direção de resolver problemas práticos com essa técnica parecem ser naturais.

Por fim, introduzimos uma nova maneira para trabalhar com problemas com restrições de equilíbrio utilizando métodos de Restauração Inexata. Nossa proposta é uma extensão do método para problemas em dois níveis. Exemplos ilustrativos foram apresentados, muitos deles relacionados com a aplicação em economia do modelo de Arrow-Debreu para equilíbrio geral. Nossas ferramentas podem auxiliar na descoberta de parâmetros para obtenção de uma situação de equilíbrio. Intitulamos esse tema de Equilíbrio Inverso de Nash e acreditamos que esse seja o assunto de muitas aplicações relevantes. No último exemplo apresentado combinamos as principais ideias do texto, aplicando a otimização do tipo LOVO para resolver o problema de cada agente envolvido na situação de equilíbrio. Vários dos aspectos usados nessa formulação tem um horizonte repleto de investigações e inovações a serem trabalhados no futuro.

Referências Bibliográficas

- [1] J. Abadie e J. Carpentier. Generalization of the Wolfe Reduced Gradient Method to the Case of Nonlinear Constraints *Optimization*, R. Fletcher (editor), Academic Press, London, pp. 37-47, 1969.
- [2] C. S. Adjiman, I. P. Androulakis, C. D. Maranas e C. A. Floudas. A global optimization method α -BB for process design. *Computers & Chemical Engineering* 20, pp. 419-424, 1996.
- [3] C. S. Adjiman, S. Dallwig, C. A. Floudas e A. Neumaier. A global optimization method, α -BB, for general twice-differentiable constrained NLPs – I. Theoretical Advances, *Computers & Chemical Engineering* 22, pp. 1137-1158, 1998.
- [4] C. S. Adjiman, I. P. Androulakis e C. A. Floudas. A global optimization method, α -BB, for general twice-differentiable constrained NLPs – II. Implementation and computational results. *Computers & Chemical Engineering* 22, pp. 1159-1179, 1998.
- [5] R. F. Almgren. Optimal execution with nonlinear impact functions and trading-enhanced risk. *Applied Mathematical Finance* 10, pp. 1-18, 2003
- [6] R. F. Almgren e N. Chriss. Optimal execution of portfolio transaction. *Journal of Risk* 3, pp. 5-39, 2001.
- [7] K. J. Arrow e G. Debreu. Existence of an equilibrium for a competitive economy. *Econometrica* 22, pp. 265-290, 1954.

- [8] R. Andreani, E. G. Birgin, J. M. Martínez e M. L. Schuverdt. On Augmented Lagrangian methods with general lower-level constraints. *SIAM Journal on Optimization* 18, pp. 1286-1309, 2007.
- [9] R. Andreani, E. G. Birgin, J. M. Martínez e M. L. Schuverdt. Second-order negative-curvature methods for box-constrained and general constrained optimization. *Computational Optimization and Applications* 45, pp. 209-236, 2010.
- [10] R. Andreani, S. Castro, J. Chela, A. Friedlander e S. Santos. An inexact-restoration method for nonlinear bilevel programming problems. *Computational Optimization & Applications* 43, Issue: 3, pp. 307-328, 2009.
- [11] R. Andreani, C. Dunder e J. M. Martínez. Order-value optimization: formulation and solution by means of a primal Cauchy method. *Mathematical Methods of Operations Research* 58, pp. 387-399, 2003.
- [12] R. Andreani, C. Dunder e J. M. Martínez. Nonlinear-programming reformulation of the Order-value optimization problem. *Mathematical Methods of Operations Research* 61, pp. 365-384, 2005.
- [13] R. Andreani, G. Haeser, M.L. Schuverdt e P.J.S. Silva. A relaxed constant positive linear dependence constraint qualification applied to an augmented Lagrangian method. *Mathematical Programming*, 2011 (DOI:10.1007/s10107-011-0456-0).
- [14] R. Andreani, G. Haeser, M.L. Schuverdt e P.J.S. Silva. Two new weak constraint qualifications and applications. *Submetido, disponível em Optimization Online*, 07/2011.
- [15] R. Andreani, G. Haeser e J. M. Martínez. Sequential optimality conditions for constrained optimization. *Optimization*, 60-5, pp. 627-641, 2011.
- [16] R. Andreani, J. M. Martínez, L. Martínez e F. Yano. Continuous optimization methods for structure alignments. *Mathematical Programming* 112, pp. 93-124, 2008.

- [17] R. Andreani, J. M. Martínez, L. Martínez e F. Yano. Low Order-value optimization and applications. *Journal of Global Optimization* 43, pp. 1-10, 2009.
- [18] R. Andreani, J. M. Martínez, M. Salvatierra e F. Yano. Quasi-Newton methods for Order-value optimization and value-at-risk calculations. *Pacific Journal of Optimization* 2, pp. 11-33, 2006.
- [19] R. Andreani, J. M. Martínez e M. L. Schuverdt. On the relation between the Constant Positive Linear Dependence condition and quasinormality constraint qualification. *Journal of Optimization Theory and Applications* 125, pp. 473-485, 2005.
- [20] I. P. Androulakis, C. D. Maranas e C. A. Floudas. α -BB: A global optimization method for general constrained nonconvex problems. *Journal of Global Optimization* 7, pp. 337-363, 1995.
- [21] Anonymous, A new algorithm for optimization. *Mathematical Programming* 1, pp. 124-128, 1972.
- [22] L. Armijo. Minimization of functions having Lipschitz continuous first partial derivatives. *Pacific Journal of Mathematics*, 16 (1): 1-3, 1966.
- [23] J.F. Bard, J. Plummer, e J.C. Sourie. A bilevel programming approach to determining tax credits for biofuel production. *European Journal of Operational Research*, pp.30-46, 2000.
- [24] A. Ben Tal e M. Teboulle. An old-new concept of convex risk measures: the optimized certainty equivalent. *Mathematical Finance* 17, pp. 449-476, 2007.
- [25] M. J. Best e J. Hlouskova. Portfolio selection and transactions costs. *Computational Optimization & Applications* 24, pp. 95-116, 2003.
- [26] E. G. Birgin, L. F. Bueno, N. Krejić e J. M. Martínez. Low Order-Value approach for solving VaR-constrained optimization problems. *Journal of Global Optimization*, aceito para publicação, 2010.

- [27] E. G. Birgin, R. Castillo e J. M. Martínez. Numerical comparison of Augmented Lagrangian algorithms for nonconvex problems. *Computational Optimization & Applications*, 31, pp. 31-56, 2005.
- [28] E. G. Birgin, C. A. Floudas, J. M. Martínez. Global minimization using an Augmented Lagrangian method with variable lower-level constraints. *Mathematical Programming*, aceito para publicação (DOI: 10.1007/s10107-009-0264-y).
- [29] E. Birgin, J. M. Martínez. Local Convergence of an Inexact-Restoration Method and Numerical Experiments. *Journal of Optimization Theory and Applications*, 127, número 2, novembro 2005.
- [30] H. Bortolossi, G. Garbaggio, B. Sartini. *Uma introdução à teoria econômica dos jogos*. Publicações matemáticas IMPA, 2007.
- [31] J. P. Bouchaud, Y. Gefen, M. Potters e M. Wyart. Fluctuations and response in financial markets: The subtle nature de 'random' price changes. *Quantitative Finance* 4, pp. 57-62, 2004.
- [32] W. D. A. Bryant. *General equilibrium: theory and evidence*. World Scientific, 2010.
- [33] L. F. Bueno, A. Friedlander, J. M. Martínez e F. Sobral. Inexact-Restoration method for Derivative-Free Optimization with smooth constraints. Em preparação, 2011.
- [34] S. Castro. Técnica de Restauração Inexata Aplicada à Resolução de Problemas de Programação Matemática em Dois Níveis. *Tese de doutorado, Universidade Estadual de Campinas*, 2004.
- [35] S. Castro, E. Fancello e A. Friedlander. *Programação não linear em dois níveis: aplicação em engenharia mecânica*. SBMAC, 2007.

- [36] M. Celis, J. E. Dennis, e R. A. Tapia. A trust region strategy for nonlinear equality constrained optimization. *Numerical Optimization* 1984 (P. Boggs, R. Byrd and R. Schnabel, eds), Philadelphia: SIAM, pp. 71-82, 1985.
- [37] C. W. Cobb e P. H. Douglas. A Theory of Production. *American Economic Review*, 18 (Supplement): 139-165, 1928.
- [38] T. H. Cormen, C. E. Leiserson, R. L. Rivest e C. Stein. *Introduction to Algorithms*. Second Edition, The MIT Press, 2001.
- [39] J. Daniëlsson, B. N. Jorgensen, C. G. de Vries e X. Yang. Optimal portfolio allocation under the probabilistic VaR constraint and incentives for financial innovation. *Annals of Finance* 4, pp. 345-367, 2008.
- [40] S. Dempe. *Foundations of Bilevel Programming*. Kluwer Academic Publishers, Dordrecht, 2002.
- [41] S. Dempe. Annotated Bibliography on Bilevel Programming and Mathematical Programs with Equilibrium Constraints. *Optimization*, 52, 333-359, 2003.
- [42] M. A. Diniz-Ehrhardt, J. M. Martínez e L. G. Pedroso. Derivative-free methods for nonlinear programming with general lower-level constraints. *Computational & Applied Mathematics*, 30, pp. 19-52, 2011.
- [43] M. M. El-Alem. A Global Convergence Theory for a Class of Trust Region Algorithms for Constrained Optimization, *Ph.D. thesis, Department of Mathematical Sciences, Rice University, Houston, TX, 1988*.
- [44] L. Eldén. Algorithms for the regularization of ill-conditioned least squares problems. *BIT* 17, 134-145, 1977.
- [45] F. Facchinei, Y. Eldar, D. P. Palomar, J. S. Pang. Convex Optimization, Game Theory, and Variational Inequality Theory in Multiuser Communication Systems. *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 35-49, 2010.

- [46] F. Facchinei, A. Fischer e V. Piccialli. Generalized Nash equilibrium problems and Newton methods. *Mathematical Programming Series B* 117, n. 1-2, pp.163–194, 2009.
- [47] F. Facchinei, C. Kanzow. Generalized Nash equilibrium problems. *4OR* 5, 173-210, 2007.
- [48] F. Facchinei e J.S. Pang. Nash equilibria: The variational approach. *Convex optimization in signal processing and communications*. (Palomar D.P. e Eldar Y. eds.) Cambridge University Press, 2009.
- [49] F. Facchinei e J. S. Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems, volume I*. Springer, 2003.
- [50] F. Facchinei e J. S. Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems, volume II*. Springer, 2003.
- [51] F. Facchinei, J.S. Pang, G. Scutari e C. Wang. Distributed power allocation with rate constraints in gaussian parallel interference channels. *IEEE T. Inform. Theory* 54, 3471-3489, 2008.
- [52] F. Facchinei, V. Piccialli, M. Sciandrone. On a class of generalized Nash equilibrium problems. *DIS Technical Report*, “Sapienza“ Universit‘ di Roma, Rome, Italy.
- [53] A. Fischer e A. Friedlander. A new line search inexact restoration approach for nonlinear programming. *Computational Optimization & Applications* 46, 333-346, 2010.
- [54] R. Fletcher, S. Leyffer e P. Toint. A brief history of filter methods. *Technical report*, *Mathematics and Computer Science Division, Argonne National Laboratory*, 2006.
- [55] A. Friedlander e F. Gomes. Solution of a truss topology bilevel programming problem by means of an inexact restoration method. *Computational & Applied Mathematics*, 30, N. 1, pp. 109-125, 2011.

- [56] A. A. Gaivoronski e G. Pflug. Finding optimal portfolios with constraints on value at risk. *Technical Report, Department of Industrial Economics and Technology Management, NTNU - Norwegian University of Science and Technology, Norway*, 1999.
- [57] A. A. Gaivoronski e G. Pflug. Value at risk in portfolio Optimization: Properties & computational approach. *Journal of Risk* 7, pp. 1-31, 2005.
- [58] G. H. Golub e C. F. Van Loan. *Matrix Computations*. Baltimore: The Johns Hopkins University Press, 1996.
- [59] G. Haeser. Condições sequenciais de otimalidade. *Tese de doutorado, Universidade Estadual de Campinas*, 2009.
- [60] M. R. Hestenes. Multiplier and gradient methods. *Journal of Optimization Theory and Applications* 4, pp. 303-320, 1969.
- [61] W. Hock e K. Schittkowski. Test examples for nonlinear programming codes. *Lecture Notes in Economics and Mathematical Systems* 187, Springer-Verlag, Berlin, Heidelberg, New York, 1981.
- [62] R. Horst, P. M. Pardalos e M. V. Thoai. *Introduction to Global Optimization* Kluwer Book Series: Nonconvex Optimization and its Applications, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2000.
- [63] P. Jorion. *Value at Risk: The New Benchmark for Managing Financial Risk*. McGraw-Hill, 2001.
- [64] T. Kolda, R. Lewis, e V. Torczon. Stationarity results for generating set search for linearly constrained optimization. *SIAM Journal on Optimization*, 17, Issue 4, 943-968, 2006.
- [65] N. Krejić, M. Kumaresan e A. Rožnjik. VaR optimal portfolio with transaction costs. *University of Novi Sad, Department of Mathematics and Informatics*, 2008.

- [66] R. Lewis, A. Shepherd e V. Torczon. Implementing generating set search methods for linearly constrained minimization. *SIAM Journal on Scientific Computing*, 29, Issue 6, pp. 2507-2530, 2007.
- [67] R. Lewis e V. Torczon. Active set identification for linearly constrained minimization without explicit derivatives. *SIAM Journal on Scientific Computing*, 29, Issue 6, pp. 2507-2530, 2007.
- [68] R. M. Lewis e V. Torczon. A globally convergent Augmented Lagrangian pattern search algorithm for optimization with general constraints and simple bounds. *SIAM Journal on Optimization* 12, pp. 1075-1089, 2002.
- [69] R. M. Lewis e V. Torczon. A direct search approach to nonlinear programming problems using an Augmented Lagrangian method with explicit treatment of linear constraints. *Technical Report WMCS-2010-01, College of William & Mary, Department of Computer Sciences*, 2010.
- [70] F. Lillo, J. D. Farmer e R. N. Mantegna. Master curve for price impact function. *Nature* 421, pp. 129-130, 2003.
- [71] M. S. Lobo, M. Fazel e S. Boyd. Portfolio optimization with linear and fixed transaction costs. *Annals of Operation Research* 152, pp. 341-365, 2007.
- [72] D. G. Luenberger. *Linear and Nonlinear Programming, 2nd Edition*. Addison-Wesley Inc., Reading, Massachusetts, 1984.
- [73] Z.-Q. Luo, J. S. Pang, e D. Ralph. *Mathematical Programs with Equilibrium Constraints*. Cambridge University Press, 1996.
- [74] R. Mansini e G. Speranza. An exact approach for portfolio selection with transaction costs and rounds. *IIE Transactions* 37, pp. 919-929, 2005.

- [75] N. Maratos. Exact Penalty Function Algorithms for Finite-Dimensional and Control Optimization Problems, *PhD thesis, Department of Computing and Control, University of London, 1978.*
- [76] H. Markowitz. Portfolio Selection. *The Journal of Finance* 7, pp. 77-91, 1952.
- [77] J. M. Martínez. SOR - Secant methods. *SIAM Journal on Numerical Analysis*. 31, pp. 217-226, 1994.
- [78] J. M. Martínez. Inexact restoration method with Lagrangian tangent decrease and new merit function for nonlinear programming. *Journal of Optimization Theory and Applications* 111, pp. 39-58, 2001.
- [79] J. M. Martínez. Order-value optimization and new applications. *Sixth International Congress on Industrial and Applied Mathematics: Zürich, Switzerland, July 16-20, 2007 - Invited Lectures*, R. Jeltsch e G. Wanner eds., European Mathematical Society, Zürich, pp. 279-296, 2009.
- [80] J. M. Martínez e E. A. Pilotta. Inexact restoration algorithms for constrained optimization. *Journal of Optimization Theory and Applications* 104, pp. 135-163, 2000.
- [81] J. M. Martínez e S. A. Santos. *Métodos Computacionais de Otimização*. 1 ed. Rio de Janeiro: Sociedade Brasileira de Matemática, 1995.
- [82] J. J. Moré. Nonlinear Generalizations of Matrix Diagonal Dominance with Application to Gauss-Seidel Iterations. *SIAM Journal on Numerical Analysis*, 9, No. 2, pp. 357-378, 1972.
- [83] O. Morgenstern e J. von Neumann. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.
- [84] J. Nash Jr.. Equilibrium Points in n -person Games. *Proceedings of the National Academy of Sciences of the United States of America* pp. 48-49, 1950.

- [85] J. Nash Jr.. The Bargaining Problem. *Econometrica* 18, no. 2, pp. 155-162, 1950.
- [86] J. Nash Jr.. Non-cooperative Games. *Annals of Mathematics* 54, pp. 286-295, 1951.
- [87] J. Nash Jr., Two-person Cooperative Games, *Econometrica* pp. 128-140, 1953.
- [88] J. Outrata, M. Kočvara e J. Zowe. *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints*. Kluwer Academic Publishers, 1998.
- [89] T. D. Plantenga. *HOPSPACK 2.0 User Manual*. Relatório técnico SAND2009-6265. Sandia National Laboratories, Albuquerque, NM and Livermore, CA, 2009.
- [90] M. Potapchik, L. Tuncel e H. Wolkowicz. Large scale portfolio Optimization with piecewise linear transaction costs. *Optimization Methods & Software* 23, pp. 929-952, 2008.
- [91] M. J. D. Powell. A method for nonlinear constraints in minimization problems. *Optimization*, R. Fletcher (ed.), Academic Press, New York, NY, pp. 283-298, 1969.
- [92] M. J. D. Powell. *The BOBYQA algorithm for bound constrained optimization without derivatives*. Cambridge NA Report NA2009/06, University of Cambridge, Cambridge, 2009.
- [93] L. Qi e Z. Wei. On the constant positive linear dependence condition and its application to SQP methods. *SIAM Journal on Optimization* 10, pp. 963-981, 2000.
- [94] S. Robinson. Stability theory for systems of inequalities, Part II: Differentiable Nonlinear Systems. *SIAM Journal of numerical analysis* 13, No. 4, pp. 497-513, 1976.
- [95] R. T. Rockafellar. Augmented Lagrange multiplier functions and duality in non-convex programming. *SIAM Journal on Control and Optimization* 12, pp. 268-285, 1974.

- [96] R. T. Rockafellar e S. Uryasev. Conditional value-at-risk for general loss distributions. *Journal de Banking & Finance* 26, pp. 1443-1471, 2002.
- [97] A. M. So, J. Zhang e Y. Ye. Stochastic Combinatorial Optimization with Controllable Risk Aversion Level. *Mathematics of Operations Research* 34, pp. 522-537, 2009.
- [98] L. N. Vicente. Bilevel programming: Introduction, History and Overview. *Relatório em FTP: dial.uwaterloo.ca*, 1994.
- [99] R. B. Wilson. A Simplicial Method for Convex Programming. *PhD thesis, Harvard University*, 1963.
- [100] P. Wolfe. Methods of nonlinear programming, em Recent Advances. *Mathematical Programming* (R.L.Graves e P. Wolfe eds.), McGraw Hill, New York, 67-86, 1963.
- [101] S. J. Wright. An algorithm for degenerate nonlinear programming with rapid local convergence, 2005. *SIAM Journal of Optimization*, 15, No. 3, pp. 673-696, 2005.
- [102] Y. Yuan. A Trust Region Algorithm for Nash Equilibrium Problems. *Pacific Journal of Optimization*, 7, Number 1, pp. 125-138, 2011.
- [103] W. I. Zangwill. Nonlinear Programming via Penalty Functions. *Management Science* 13, 5, pp. 344-358, 1967.
- [104] <http://www.ime.usp.br/~egbirgin/tango/>
- [105] <http://www.ime.usp.br/~egbirgin/>