UNIVERSIDADE ESTADUAL DE CAMPINAS

Instituto de Matemática, Estatística e Computação Científica

LEONARDO DELARMELINA SECCHIN

Aplicação de métodos de Lagrangiano aumentado a problemas de otimização com restrições de complementariedade

Campinas

2018

Leonardo Delarmelina Secchin

Aplicação de métodos de Lagrangiano aumentado a problemas de otimização com restrições de complementariedade

Tese apresentada ao Instituto de Matemática, Estatística e Computação Científica da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Doutor em Matemática Aplicada.

Orientador: Roberto Andreani

Coorientador: Paulo José da Silva e Silva

Este exemplar corresponde à versão final da Tese defendida pelo aluno Leonardo Delarmelina Secchin e orientada pelo Prof. Dr. Roberto Andreani.

Campinas

2018

Agência(s) de fomento e nº(s) de processo(s): Não se aplica.
ORCID: https://orcid.org/0000-0002-9224-9051

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Matemática, Estatística e Computação Científica
Ana Regina Machado - CRB 8/5467

Se23a

Secchin, Leonardo Delarmelina, 1982-
Aplicação de métodos de Lagrangiano aumentado a problemas de otimização com restrições de complementariedade / Leonardo Delarmelina Secchin. – Campinas, SP : [s.n.], 2018.

Orientador: Roberto Andreani.
Coorientador: Paulo José da Silva e Silva.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

1. Programação não-linear. 2. Lagrangiano aumentado. 3. Problemas de complementaridade não-linear. I. Andreani, Roberto, 1961-. II. Silva, Paulo José da Silva e, 1973-. III. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. IV. Título.

Informações para Biblioteca Digital

Título em outro idioma: Application of augmented Lagrangian methods to mathematical problems with complementarity constraints
Palavras-chave em inglês:
Nonlinear programming
Augmented Lagrangian
Nonlinear complementarity problems
Área de concentração: Matemática Aplicada
Titulação: Doutor em Matemática Aplicada
Banca examinadora:
Roberto Andreani [Orientador]
Lucio Tunes dos Santos
Ernesto Julián Goldberg Birgin
Gabriel Haeser
Orizon Pereira Ferreira
Data de defesa: 05-02-2018
Programa de Pós-Graduação: Matemática Aplicada

Tese de Doutorado defendida em 05 de fevereiro de 2018 e aprovada pela banca examinadora composta pelos Profs. Drs.

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

Prof(a). Dr(a). LUCIO TUNES DOS SANTOS

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

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

Prof(a). Dr(a). ORIZON PEREIRA FERREIRA

As respectivas assinaturas dos membros encontram-se na Ata de defesa

Agradecimentos

Agradeço a Nino, meu orientador e amigo, pela confiança depositada em mim. Dono de uma inteligência ímpar e de uma vitalidade rara, é um exemplo que levarei para minha vida acadêmica.

Agradeço a todos os meus professores durante o doutorado, em especial, a Mario, Paulo e Sandra. Para além dos conteúdos curriculares, considero-os referências na pesquisa. A Paulo e Sandra, agradeço ainda pela "pesquisa extra-classe" que realizamos. Foi um aprendizado muito rico para mim!

Agradeço aos meus pais, Maria Celia e Norival, por todo o suporte e carinho dedicados a mim, sem os quais certamente eu não chegaria a lugar algum. Agradeço a Shirley, meu amor, sempre ao meu lado confortando-me.

Agradeço a todos os amigos que fiz em Campinas pelos momentos que passamos juntos.

Por fim, agradeço aos meus colegas do Departamento de Matemática Aplicada da UFES por possibilitarem meu afastamento para realização do doutorado.

Resumo

Problemas de programação matemática com restrições de complementariedade (MPCC) têm alto grau de degeneração: não satisfazem a maioria das condições de qualificação estabelecidas. Em particular, nenhum ponto viável cumpre a condição de Mangasarian-Fromovitz. Na ausência de complementariedade estrita, podemos esperar apenas que a condição de Guignard seja satisfeita com alguma generalidade. Como não há métodos computacionais conhecidos que convirjam a pontos KKT somente com a condição de Guignard, não há garantia alguma de que tais métodos convirjam a pontos KKT quando aplicados a MPCCs. Recentemente, Izmailov, Solodov e Uskov investigaram o comportamento de métodos de Lagrangiano aumentado em MPCCs. A despeito de sua boa performance global, estes métodos falham em casos simples. Dois anos antes, Andreani *et al.* observaram que seu método de Lagrangiano aumentado de segunda ordem não falhava em tais casos. Neste trabalho, retomamos o tema da convergência de métodos de Lagrangiano aumentado em MPCCs. Estendemos o resultado de Izmailov, Solodov e Uskov para o método de primeira ordem, e provamos que, de fato, o método de segunda ordem sempre evita situações indesejadas como a observada, convergindo a pontos M-estacionários sob hipóteses pouco exigentes. Testes computacionais foram realizados, corroborando a teoria.

Palavras-chave: Programação não-linear. Lagrangiano aumentado. Métodos de segunda ordem.

Abstract

Mathematical problems with complementarity constraints (MPCC) are highly degenerate problems: they do not satisfy the majority of the established Constraint Qualifications (CQ). In particular, no feasible point conforms to Mangasarian-Fromovitz CQ (MFCQ). In the absence of strict complementarity, only Guignard's condition can be expected with certain generality. Since there is not any known computational method that converges to KKT points using only Guignard's condition, there is not any guarantee that traditional optimization methods applied to MPCC converge to KKT points. Recently, Izmailov, Solodov and Uskov investigated the behavior of augmented Lagrangian methods applied to MPCC. Despite good global performance, these methods fail in simple cases. Two years before, Andreani *et al* observed that their second-order augmented Lagrangian method do not fail in such cases. In this research, we return to the theme of theoretical convergence of augmented Lagrangian methods. We extend the previous result of Izmailov, Solodov and Uskov for the first-order method, and we proved that, in fact, the second-order method always avoid the undesirable situations mentioned, converging to M-stationary points under mild assumptions. Computational tests were performed, corroborating the theory.

Keywords: Nonlinear programming. Augmented Lagrangian. Second-order methods.

Lista de abreviaturas e siglas

min	"Mínimo entre" ou "Minimizar"
max	"Máximo entre" ou "Maximizar"
s.a	"sujeito a"
AKKT	Aproximadamente KKT
C-estacionário	Clarke-estacionário
CCP	Condição de qualificação de cone continuidade
CNO	Condição Necessária de Otimalidade
CQ	Condição de Qualificação
JPCG	*Jacobi Preconditioned Conjugate Gradients*
KKT	Karush-Kunh-Tucker
LA	Lagrangiano Aumentado
LICQ	Condição de qualificação de independência linear dos gradientes das restrições ativas
M-estacionário	Mordukhovich-estacionário
MPCC	Problema de programação matemática com restrições de complementariedade
RCPLD	Condição de qualificação de dependência linear positiva constante relaxada
S-estacionário	Fortemente estacionário
SPG	Gradiente espectral projetado
SSONC	Condição necessária forte de otimalidade de segunda ordem
TNLP	Problema não linear apertado

W-estacionário Fracamente estacionário

WSONC Condição necessária fraca de otimalidade de segunda ordem

Lista de símbolos

$(\beta)_+$	$(\beta)_+ = \max\{\beta, 0\} \quad (\beta \in \mathbb{R})$
$(z)_+$	$(z)_+ = ((z_1)_+, \ldots, (z_n)_+) \quad (z \in \mathbb{R}^n)$
$I_q(z)$	Conjunto dos índices i tais que $q_i(z) = 0$
$\mathbb{R}^n_+$	Conjunto dos pontos em $\mathbb{R}^n$ com todas as entradas não negativas
K, K_i	Conjuntos infinitos de índices em $\mathbb{N} = \{1, 2, \ldots\}$
$\operatorname{span} V$	Espaço gerado pelos vetores em V
$L(x, \mu)$	Função Lagrangiano usual
$\mathcal{L}(x, \lambda)$	Função Lagrangiano-MPCC
$\nabla_z q$	Gradiente de q em relação à variável z. Quando z estiver claro no contexto, podemos escrever simplesmente ∇q
∇Q	Matriz cujas colunas são os gradientes das componentes da função $Q : \mathbb{R}^n \to \mathbb{R}^m$, isto é, $\nabla Q = [\ \nabla Q_1 \ \cdots \ \nabla Q_m\ ]$
$\lambda_{\min}(A)$	Menor autovalor da matriz simétrica A
$\|\cdot\|$	Norma Euclidiana em $\mathbb{R}^n$
$\|\cdot\|_\infty$	Norma do supremo em $\mathbb{R}^n$, isto é, $\|z\|_\infty = \max_{i=1,\ldots,n} \|z_i\|$
C°	Polar do conjunto $C \subset \mathbb{R}^n$, isto é, $C^\circ = \{z \in \mathbb{R}^n \mid z^T d \leqslant 0,\ \forall d \in C\}$
$P_X(\cdot)$	Projeção ortogonal (Euclidiana) sobre o conjunto X
z_I	Vetor de $\mathbb{R}^{\|I\|}$ formado pelas entradas de $z \in \mathbb{R}^n$ com índices no conjunto ordenado I
μ	Vetor de multiplicadores para a função Lagrangiano usual
λ	Vetor de multiplicadores para a função Lagrangiano-MPCC

Sumário

Introdução **12**
1 Métodos de Lagrangiano aumentado para problemas gerais **17**
1.1 Condições de qualificação e otimalidade 18
1.2 Lagrangiano aumentado de primeira ordem 26
1.3 Lagrangiano aumentado de segunda ordem 32
1.4 Melhoramentos no cálculo de autopares 38
2 Problemas com restrições de complementariedade **41**
2.1 Condições de qualificação e estacionariedade para MPCC 41
2.2 Notas sobre condições sequenciais de otimalidade 54
3 Convergência de métodos de Lagrangiano aumentado em MPCCs . . **57**
3.1 Restrições de caixa 68
3.2 Restrições gerais 69
4 Resultados Numéricos **72**
4.1 Uma estratégia alternativa 79

5 Considerações Finais **84**
5.1 Trabalhos futuros 85

REFERÊNCIAS **88**

Anexos **103**
ANEXO A Resultados numéricos detalhados **104**

Introdução

Neste trabalho, estamos interessados no problema de *programação matemática com restrições de complementariedade* (MPCC[1]) dado por

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{s.a} \quad & G(x) \leqslant 0, \quad H(x) = 0 \\ & g(x) \geqslant 0, \quad h(x) \geqslant 0, \quad g(x)^T h(x) \leqslant 0, \end{aligned} \tag{MPCC}$$

onde $f : \mathbb{R}^n \to \mathbb{R}$, $G : \mathbb{R}^n \to \mathbb{R}^s$, $H : \mathbb{R}^n \to \mathbb{R}^q$ e g, $h : \mathbb{R}^n \to \mathbb{R}^m$ são funções duas vezes continuamente diferenciáveis. Se $g(x) \geqslant 0$ e $h(x) \geqslant 0$, a restrição

$$g(x)^T h(x) = \sum_{i=1}^{m} g_i(x) h_i(x) \leqslant 0 \tag{1}$$

garante que $g_i(x) = 0$ ou $h_i(x) = 0$, e é chamada *restrição de complementariedade*. Podemos escrevê-la de forma equivalente como

$$\begin{aligned} & g(x)^T h(x) = 0, \\ & g_i(x) h_i(x) \leqslant 0, \ i = 1, \ldots, m, \quad \text{ou} \qquad (2) \\ & g_i(x) h_i(x) = 0, \ i = 1, \ldots, m. \qquad (3) \end{aligned}$$

Ao estudarmos a aplicação de métodos que envolvem multiplicadores das restrições, como é o caso do método de Lagrangiano aumentado, é preferível trabalhar com a complementariedade na forma de desigualdades, pois o sinal do multiplicador associado pode ser definido *a priori*. Desta forma, adotaremos a escrita (1). Cabe ressaltar que não há, pelo menos em princípio, um motivo prático que justifique essa escolha, apenas a conveniência ao lidar com a teoria. Por uma questão de compatibilidade com a literatura, poderemos utilizar a expressão "restrições de complementariedade", no plural, para nos referirmos à restrição (1), ainda que seu uso seja mais adequado às expressões (2) e (3).

Problemas com restrições de complementariedade surgem em vários contextos, tais como: problemas de dois níveis [47, 48]; otimização de portfólio/análise de risco [7, 42]; máquinas de suporte vetorial [98]; otimização discreta [102, 136]; problemas com restrições

[1] Do inglês *Mathematical Program with Complementarity Constraints*.

de cardinalidade [43]; controle de tráfego urbano [141]; economia; problemas do setor elétrico [50, 73]; controle ótimo [2]. Veja ainda [53, 107, 119, 127] e suas referências.

A aplicação de estratégias clássicas em otimização a MPCCs foi considerada em vários trabalhos [52, 59, 60, 72, 86, 99, 101], em especial, estratégias de pontos interiores [100, 104, 128], métodos de penalização [23–25, 84, 142], programação quadrática sequencial [61, 63, 89, 115, 124], e métodos de Lagrangiano aumentado [87, 105, 106, 143]. Também, estratégias específicas para (MPCC) foram desenvolvidas. Nas duas últimas décadas, particular atenção foi dada aos chamados *métodos de regularização* (também conhecidos como *métodos de relaxação*) [51, 80–82, 91, 92, 94, 95, 103, 104, 111, 115, 129, 137, 138], que se consistem em esquemas iterativos onde o subproblema é obtido "suavizando" a região viável (sobretudo as restrições de complementariedade). Seus melhores resultados de convergência foram estabelecidos supondo que os subproblemas são resolvidos exatamente. Até muito recentemente, todas as melhores variantes perdiam propriedades teóricas de convergência se resolvêssemos seus subproblemas inexatamente, o que é a realidade computacional [96]. Portanto não ofereciam, pelo menos no que diz respeito à garantia teórica, nada além do que os métodos gerais de otimização oferecem. No entanto, em um relatório técnico recente ainda não publicado [110], uma estratégia de regularização foi capaz de manter seus bons resultados teóricos de convergência sob inexatidão. Ainda que muito precoce, isso dá aos métodos de regularização um "status teórico" que lhes faltava. De qualquer forma, trabalhos recentes sobre a aplicação de métodos gerais a MPCCs mostram que o tema é pertinente. Em particular, o presente trabalho dedica-se à convergência de métodos de Lagrangiano aumentado a MPCCs.

Dada a grande literatura dedicada ao problema, o leitor poderia se perguntar: o que há de especial no modelo (MPCC) que justifique um estudo específico? De fato, (MPCC) pode ser encarado como um modelo de programação não linear padrão. No entanto, precisamente as restrições de complementariedade trazem dificuldades. No que segue, ilustraremos algumas delas.

Consideremos o problema de programação não linear padrão

$$\min_x \tilde{f}(x) \quad \text{s.a} \quad \tilde{g}(x) \leqslant 0, \quad \tilde{h}(x) = 0, \tag{4}$$

onde $\tilde{f} : \mathbb{R}^n \to \mathbb{R}$, $\tilde{g} : \mathbb{R}^n \to \mathbb{R}^s$ e $\tilde{h} : \mathbb{R}^n \to \mathbb{R}^q$ são funções continuamente diferenciáveis.

Sabemos que, em geral, encontrar minimizadores globais de (4) de forma eficiente é impraticável. O que podemos esperar dos métodos é que um ponto limite x^* satisfaça certas propriedades verificadas por minimizadores. Neste contexto, as condições de Karush-Kuhn-Tucker (KKT) [85] desempenham papel fundamental. Infelizmente, essas condições não são satisfeitas por todo minimizador: no problema

$$\min_x x \quad \text{s.a} \quad x^2 = 0,$$

por exemplo, a origem é o único minimizador, mas não é KKT. A fim de obtermos propriedades satisfeitas por todo minimizador, é comum agregarmos hipóteses às restrições, as conhecidas *Condições de Qualificação das restrições* (CQs). Uma dentre as várias CQs conhecidas na literatura é a *independência linear dos gradientes das restrições ativas* (LICQ[2]), que consiste em exigir que os gradientes das restrições que se verificam como igualdade em x^* (restrições ativas) sejam linearmente independentes. É possível mostrar que se o minimizador local x^* de (4) satisfaz LICQ então x^* é um ponto KKT [85], ou seja,

"KKT ou não-LICQ"

é uma condição verificada por todo minimizador. A implicação prática disso é que, supondo *a priori* que os minimizadores de (4) satisfazem LICQ, pontos KKT são candidatos a minimizadores locais de (4). Valendo-se disso, métodos computacionais, como Lagrangiano aumentado, buscam pontos limite que satisfaçam KKT. Aqui nos deparamos com a primeira dificuldade do modelo (MPCC): nenhum ponto viável satisfaz LICQ! De fato, os gradientes das restrições envolvidas na complementariedade, e ativas em um ponto viável x^*, são

$$\nabla g_i(x^*), \quad i \in I_g, \tag{5}$$

$$\nabla h_i(x^*), \quad i \in I_h, \tag{6}$$

$$\sum_{i=1}^{m} \nabla g_i(x^*)h_i(x^*) + \nabla h_i(x^*)g_i(x^*), \tag{7}$$

onde $I_g = \{i \mid g_i(x^*) = 0\}$ e $I_h = \{i \mid h_i(x^*) = 0\}$. Como x^* é viável, $I_g \cup I_h = \{1, \ldots, m\}$. Daí, multiplicando (6) por $g_i(x^*)$, $i \in I_h$, (5) por $h_i(x^*)$, $i \in I_g$, e somando sobre todos os índices i obtemos (7). Isso mostra que os vetores (5)–(7) não formam um conjunto linearmente independente.

Como vimos, pontos viáveis de (MPCC) não satisfazem LICQ. É possível mostrar que nenhuma outra condição de qualificação menos exigente comumente requerida pelos métodos em seus resultados de convergência é esperada em (MPCC) [56, 87]. Em particular, mostraremos no Exemplo 2.4 (página 48) que nenhum ponto do simples conjunto

$$X = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1, x_2 \geqslant 0,\ x_1 x_2 \leqslant 0\}$$

satisfaz uma CQ muito menos exigente que LICQ, conhecida como *propriedade de cone continuidade* (CCP[3]), recentemente introduzida em [15]. Até mesmo a condição de qualificação de Abadie, ainda menos exigente, falha em casos simples [56]. Esta condição consiste na igualdade entre o cone de direções tangentes ao conjunto viável e sua linearização, e será analisada no Exemplo 2.3 (página 47). Portanto não há garantia de que métodos

[2] Do inglês *Linear Independence Constraint Qualification*.
[3] Do inglês *Cone Continuity Property*.

tradicionais de otimização, como Lagrangiano aumentado, convirjam a pontos KKT de (MPCC). No Capítulo 3 exibiremos um exemplo onde o método de Lagrangiano aumentado definido em [5] pode convergir a um ponto que não é KKT, e nem satisfaz a CQ de Abadie.

A ausência de condições de qualificação está relacionada sobretudo às restrições de desigualdade $g_i(x) \geqslant 0$ e $h_i(x) \geqslant 0$ com índices $i \in I_g \cap I_h$, isto é, àquelas restrições complementares que se anulam simultaneamente. Parte importante da teoria dedica-se a cobrir justamente os casos onde $I_g \cap I_h$ é não vazio. Conceitos de estacionariedade menos exigentes que KKT foram definidos na literatura [64, 93, 104, 118, 123, 135, 145]. Neste trabalho consideramos os conceitos mais comuns, e que serão úteis em nossos resultados de convergência. Da mesma maneira, condições de qualificação específicas para (MPCC) foram definidas [45, 70, 71, 80, 82, 93, 130, 131, 135, 138, 144]. No Capítulo 2 trataremos essas questões.

Uma outra dificuldade que apontamos em MPCCs está relacionada à existência de boas condições sequenciais de otimalidade. Em programação não linear padrão, essas condições estão associadas ao critério de parada de algoritmos, prestam a unificar a teoria de convergência, e cada vez mais tem seu papel destacado na pesquisa (veja por exemplo [9–11, 15, 16, 18, 39, 74, 76, 77, 108])[4]. Considerando as noções de estacionariedade específicas para (MPCC), métodos de Lagrangiano aumentado têm razoável garantia de convergência teórica quando comparados a outros da literatura. Sabemos que eles geram pontos limite satisfazendo a condição sequencial de otimalidade conhecida como *Aproximadamente KKT* (AKKT), definida em [10] (veja ainda [39]). No caso de MPCCs, condições sequenciais específicas praticamente inexistem até o momento (de nosso conhecimento, há apenas uma proposta em [130]), e as existentes para programação não linear padrão parecem não se aplicar com sucesso. De fato, na Seção 2.2 exibiremos um exemplo onde um ponto viável é AKKT, mas apenas cumpre uma condição de estacionariedade mais fraca do que aquela garantida pelo método de Lagrangiano aumentado. Deste modo, o método gera sequências que possuem características adicionais úteis. Apesar da relevância do tema, condições sequenciais não serão alvo deste trabalho.

Vale ressaltar que, apesar dos obstáculos da teoria, métodos gerais de otimização não linear resolvem MPCCs satisfatoriamente [59, 60]. Portanto a contribuição deste trabalho é majoritariamente em resultados teóricos de convergência, ainda que testes numéricos sejam realizados.

Como já comentado, analisamos nesta tese a aplicação de métodos de Lagrangiano aumentado a MPCCs. No Capítulo 1 revisamos os métodos conhecidos como Algencan e Algencan-second. O primeiro foi proposto em [5], e serve como modelo

[4] A ideia de analisar otimalidade via sequências de pontos estacionários aproximados foi introduzida, pelo menos, desde os anos 1960. Veja, por exemplo, a introdução de [66]. A nosso ver, em contrapartida, o tratamento moderno pelas chamadas condições sequenciais de otimalidade coloca a convergência de algoritmos como questão central.

para outros do mesmo tipo. Já ALGENCAN-SECOND é um método com convergência a pontos de segunda ordem baseado em ALGENCAN, e proposto em [4]. No Capítulo 2 apresentamos os conceitos fundamentais relacionados ao problema (MPCC). A principal contribuição do nosso trabalho é dada no Capítulo 3. Nele estendemos os resultados de convergência obtidos por Izmailov, Solodov e Uskov [87] para ALGENCAN. Especificamente, mostramos que quando certa sequência de multiplicadores associados à restrição de complementariedade (1) é limitada, ALGENCAN converge a pontos KKT de (MPCC) mediante a condição de qualificação conhecida como *dependência linear positiva constante relaxada para MPCC*, enquanto que, anteriormente, resultado análogo foi estabelecido supondo uma CQ mais exigente (veja os Teoremas 3.1 e 3.2, página 58). Estabelecemos ainda resultado de convergência para ALGENCAN-SECOND, mostrando que, em geral, este método atinge pontos estacionários qualitativamente melhores que seu par de primeira ordem (Teorema 3.3, página 63). Essa era uma questão em aberto na literatura. Resultados numéricos são apresentados no Capítulo 4, e no Capítulo 5 tecemos nossas considerações finais. Por fim, reportamos os resultados numéricos em detalhes no anexo A.

Capítulo 1

Métodos de Lagrangiano aumentado para problemas gerais

Consideremos o problema geral de programação não linear

$$\min_x F(x) \quad \text{s.a} \quad G(x) \leqslant 0, \quad H(x) = 0, \tag{P}$$

onde $F : \mathbb{R}^n \to \mathbb{R}$, $G : \mathbb{R}^n \to \mathbb{R}^s$ e $H : \mathbb{R}^n \to \mathbb{R}^q$ são funções continuamente diferenciáveis. A função *Lagrangiano* associada ao problema (P) é definida pondo

$$L(x, \mu) = F(x) + (\mu^G)^T G(x) + (\mu^H)^T H(x)$$

para todos $x \in \mathbb{R}^n$ e $\mu = (\mu^G, \mu^H) \in \mathbb{R}^s_+ \times \mathbb{R}^q$. Definimos ainda a função *Lagrangiano aumentado* PHR (Powell-Hestenes-Rockafellar) [78, 125, 132] por

$$L_\rho(x, \mu) = F(x) + \frac{\rho}{2}\left[\left\|\left(\frac{\mu^G}{\rho} + G(x)\right)_+\right\|^2 + \left\|\frac{\mu^H}{\rho} + H(x)\right\|^2\right], \tag{1.1}$$

onde $x \in \mathbb{R}^n$, $\mu = (\mu^G, \mu^H) \in \mathbb{R}^s_+ \times \mathbb{R}^q$ e $\rho > 0$.

Métodos de Lagrangiano aumentado (LA) constituem uma importante classe de algoritmos destinados a resolver (P). São métodos iterativos baseados na minimização (aproximada) sucessiva da função Lagrangiano aumentado. Na função $L_\rho(x, \mu)$, as restrições são penalizadas com um "deslocamento", ou em outras palavras, são penalizadas as restrições modificadas

$$\frac{\mu^G}{\rho} + G(x) \leqslant 0 \quad \text{e} \quad \frac{\mu^H}{\rho} + H(x) = 0.$$

Se $\mu = 0$ então a penalização ocorre com as restrições originais. Nesse sentido, métodos de LA assemelham-se ao de penalização externa. De certa forma, herdam as boas propriedades de convergência deste último, ao mesmo tempo que minimizam seu mau comportamento numérico (consulte [39, 109] para uma discussão detalhada). Ademais, nos métodos de LA o vetor μ é atualizado a cada iteração na tentativa de aproximar os multiplicadores de

Lagrange da solução. No caso convexo, podem inclusive ser vistos como métodos proximais duais, possuidores de excelentes propriedades de convergência [29].

Cabe ressaltar que há na literatura outras definições de funções Lagrangianos aumentados (consulte por exemplo [134]). Métodos de LA que empregam (1.1) estão entre os mais utilizados. O problema com $L_\rho(x, \mu)$ é que suas derivadas segundas são descontínuas nos pontos onde restrições de desigualdade deslocadas são ativas. Essa é uma das justificativas para análises de métodos de LA com outras funções, como recentemente feito em [49]. Em [32], foram realizados testes numéricos comparando 65 métodos de LA com diferentes funções Lagrangianos aumentados. Os autores concluem que métodos baseados em (1.1) têm melhor desempenho, apesar de ponderarem que isso depende dos algoritmos utilizados para resolver os subproblemas. Neste trabalho, adotaremos exclusivamente a função Lagrangiano aumentado (1.1).

Dentre os métodos de LA, estamos particularmente interessados no método denominado Algencan, desenvolvido em [5]. Algencan serve como modelo para métodos de LA baseados na função Lagrangiano aumentado (1.1). Também é a base para o método de segunda ordem conhecido como Algencan-second [4], objeto de estudo deste trabalho. Ademais, uma implementação robusta, eficiente e já bastante madura está disponível para uso sob licença livre pelo projeto TANGO [140]. Neste trabalho, portanto, vamos nos referir exclusivamente a esses dois métodos. Será comum diferenciarmos Algencan e Algencan-second como métodos de primeira e segunda ordens, respectivamente. Nas seções seguintes apresentamos ambos, revisando a teoria suficiente aos nossos propósitos.

1.1 Condições de qualificação e otimalidade

Como mencionado na introdução, não podemos esperar que, no caso geral não convexo, métodos de otimização sempre possam obter minimizadores globais do problema (P) de forma eficiente (consulte [109] para uma discussão detalhada). O razoável é esperar que métodos encontrem pontos que satisfaçam propriedades necessárias a um minimizador, ou seja, um candidato a minimizador. Em outras palavras, procuramos pontos que cumpram uma certa *Condição Necessária de Otimalidade* (CNO). Tais pontos são chamados *pontos estacionários*, e vale o seguinte:

"se x^* é um minimizador de (P) então x^* é um ponto estacionário com respeito à CNO considerada".

Uma boa CNO deve ser de fácil verificação, computável, e ao mesmo tempo caracterizar ao máximo minimizadores de (P). Neste sentido, "$x \in \mathbb{R}^n$" e "x é viável" são CNOs desinteressantes pois, apesar de facilmente verificáveis, não dizem muito a respeito dos

minimizadores; ao passo que "x é minimizador local" e "x é minimizador global" são CNOs que caracterizam bem os minimizadores, mas, evidentemente, são de difícil verificação.

Os conceitos de estacionariedade amplamente aceitos na literatura têm relação com as conhecidas condições de Karush-Kuhn-Tucker (KKT). Elas generalizam o fato de, em problemas irrestritos, o gradiente da função objetivo se anular em minimizadores.

Definição 1.1. *Dizemos que o ponto viável x^* de (P) é KKT se existe um vetor de multiplicadores $\mu^* = (\mu^{G,*}, \mu^{H,*}) \in \mathbb{R}^s_+ \times \mathbb{R}^q$ tal que*

$$\nabla_x L(x^*, \mu^*) = \nabla F(x^*) + \sum_{i=1}^{s} \mu_i^{G,*} \nabla G_i(x^*) + \sum_{i=1}^{q} \mu_i^{H,*} \nabla H_i(x^*) = 0, \tag{1.2}$$

$$\mu_i^{G,*} G_i(x^*) = 0,\ i = 1, \ldots, s. \tag{1.3}$$

Retomando o exemplo

$$\min_x x \quad \text{s.a} \quad x^2 = 0$$

exposto na introdução, vemos que as condições KKT não são por si mesmas uma CNO: no único minimizador $x^* = 0$, as equações (1.2) e (1.3) tornam-se

$$1 + 2 \cdot 0 \cdot \mu^* = 0,$$

o que não se verifica para qualquer escalar μ^*. Como já mencionado na introdução, a fim de obtermos uma CNO é comum agregarmos alguma hipótese adicional sobre as restrições, as chamadas *Condições de Qualificação (das restrições)* (CQs). Mais especificamente, se x^* é um minimizador local de (P) e vale uma CQ em x^* então x^* é KKT. Ou seja,

"KKT ou não-CQ"

é uma CNO.

São várias as CQs presentes na literatura. No que segue, definiremos três de nosso interesse. Para tanto, $I_G(x)$ denotará o conjunto dos índices das restrições de desigualdade ativas em x, isto é,

$$I_G(x) = \{i \mid G_i(x) = 0\}.$$

Diremos que

$$\{\nabla G_i(x)\}_{i \in I} \cup \{\nabla H_i(x)\}_{i \in J},$$

$I \subset I_G(x)$, $J \subset \{1, \ldots, q\}$, é *positivo linearmente dependente* se existir $(a, b) \in \mathbb{R}^{|I|}_+ \times \mathbb{R}^{|J|}$ não nulo tal que

$$\sum_{i \in I} a_i \nabla G_i(x) + \sum_{i \in J} b_i \nabla H_i(x) = 0.$$

Dada uma multifunção $\Gamma : \mathbb{R}^n \rightrightarrows \mathbb{R}^q$, denotaremos o *limite exterior de Painlevé-Kuratowski* quando $x \to x^*$ por

$$\limsup_{x \to x^*} \Gamma(x) := \left\{ z^* \in \mathbb{R}^q \mid \exists (x^k, z^k) \to (x^*, z^*) \text{ com } z^k \in \Gamma(x^k) \right\}.$$

Diremos que Γ é *semicontínua exterior* em x^* se $\limsup_{x \to x^*} \Gamma(x) \subset \Gamma(x^*)$ [133].

Definição 1.2. *1. [85] Dizemos que um ponto viável x^* satisfaz a condição de qualificação de* independência linear dos gradientes das restrições ativas *(LICQ[1]) se os gradientes das restrições ativas em x^**

$$\nabla G_i(x^*),\ i \in I_G(x^*), \quad \nabla H_i(x^*),\ i = 1, \ldots, q,$$

são linearmente independentes.

2. [12] Seja $J \subset \{1, \ldots, q\}$ um subconjunto de índices tal que $\{\nabla H_i(x^*)\}_{i \in J}$ é uma base para o espaço gerado por $\nabla H_i(x^*)$, $i = 1, \ldots, q$. Dizemos que um ponto viável x^* satisfaz a condição de qualificação de dependência linear positiva constante relaxada *(RCPLD[2]) se existe uma vizinhança $\mathcal{U}(x^*)$ de x^* tal que*

- *$\{\nabla H_i(x)\}_{i=1,\ldots,q}$ tem o mesmo posto para todo $x \in \mathcal{U}(x^*)$;*
- *Para todo subconjunto $I \subset I_G(x^*)$, se*

$$\{\nabla G_i(x^*)\}_{i \in I} \cup \{\nabla H_i(x^*)\}_{i \in J}$$

é positivo linearmente dependente, então

$$\{\nabla G_i(x)\}_{i \in I} \cup \{\nabla H_i(x)\}_{i \in J}$$

é linearmente dependente para todo $x \in \mathcal{U}(x^*)$.

3. [15] Dizemos que um ponto viável x^* satisfaz a condição de qualificação de cone continuidade *(CCP) se a multifunção $\mathcal{K} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ definida por*

$$\mathcal{K}(x) = \left\{ \sum_{i \in I_G(x^*)} \mu_i^G \nabla G_i(x) + \sum_{i=1}^{q} \mu_i^H \nabla H_i(x) \mid \mu^G_{I_G(x^*)} \geqslant 0,\ \mu^H \in \mathbb{R}^q \right\}$$

é semicontínua exterior em x^*, isto é, se

$$\limsup_{x \to x^*} \mathcal{K}(x) \subset \mathcal{K}(x^*).$$

[1] Do inglês *Linear Independence Constraint Qualification.*
[2] Do inglês *Relaxed Constant Positive Linear Dependence.*

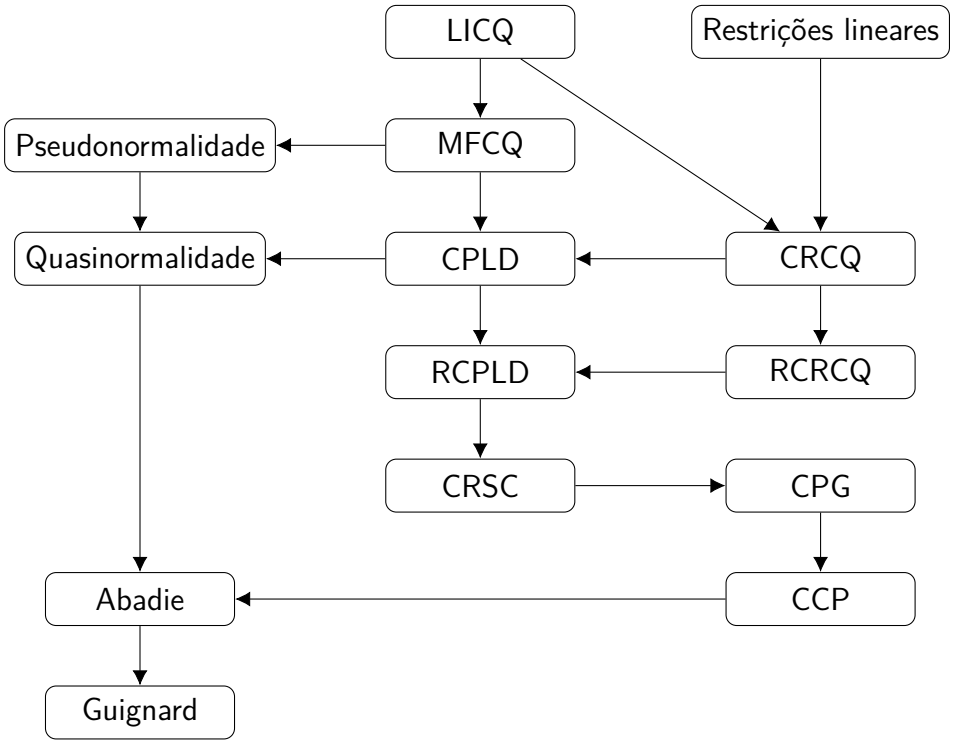

Figura 1 – Relações entre CQs. As setas indicam implicações lógicas.

Sem a pretensão de fazermos uma revisão de todas as CQs, listamos algumas a seguir indicando referências para o leitor interessado[3]: Mangasarian-Fromovitz (MFCQ) e linearidade das restrições [85]; dependência linear positiva constante (CPLD) [14, 126]; posto constante (CRCQ) [88]; posto constante relaxado (RCRCQ) [112]; *constant rank of the subspace component* (CRSC) e *constant positive generator* (CPG) [13]; quasinormalidade [79]; pseudonormalidade [30, 122]; Abadie [28]; e Guignard [67, 69]. A Figura 1 traz a hierarquia entre as CQs mencionadas.

É interessante pontuarmos que as condições KKT podem ser entendidas do ponto de vista geométrico. De fato, as expressões (1.2) e (1.3) podem ser escritas como

$$-\nabla F(x^*) \in \mathcal{T}^{\text{lin}}(x^*)^\circ,$$

onde $\mathcal{T}^{\text{lin}}(x^*)^\circ = \mathcal{K}(x^*)$ é o cone da definição de CCP (item 3 da Definição 1.2). Esse conjunto é obtido, via Lema de Farkas, tomando o polar do cone

$$\mathcal{T}^{\text{lin}}(x^*) = \left\{ d \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} d^T \nabla G_i(x^*) \leqslant 0, & \forall i \in I_G(x^*) \\ d^T \nabla H_i(x^*) = 0, & i = 1, \ldots, q \end{array} \right\}$$

que, por sua vez, é a linearização do conjunto de direções tangentes ao conjunto viável no ponto x^*. Assim, a grosso modo, KKT garante que não é possível caminharmos ao longo de uma direção de descida para F a partir de x^* ($-\nabla F(x^*)$ é uma dessas direções) ao

[3] As siglas são referentes aos correspondentes em inglês. Preferimos não traduzir algumas expressões por julgarmos não terem boa correspondência no português.

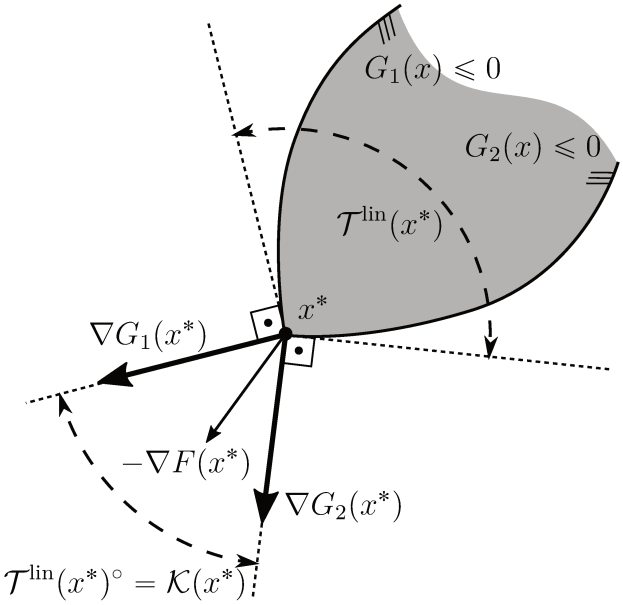

Figura 2 – Interpretação geométrica das condições KKT no caso de restrições de desigualdade. O cone $\mathcal{T}^{\text{lin}}(x^*)$ contém todas as direções tangentes ao conjunto viável no minimizador x^*. Caminhando ao longo de qualquer direção de descida para F a partir de x^*, nos distanciamos do conjunto viável.

mesmo tempo que mantemos viabilidade (veja a Figura 2). Mais geralmente, essa ideia geométrica simples vale diretamente com o cone de direções tangentes $\mathcal{T}(x^*)$. No entanto, esse cone não é computável, e logo justifica-se o uso de sua linearização. Nesse ponto, as condições de qualificação são a ponte entre as direções tangentes reais e aquelas no cone linearizado. Em última análise, todas as CQs garantem que $\mathcal{T}^{\text{lin}}(x^*)^\circ = \mathcal{T}(x^*)^\circ$ (essa é precisamente a CQ menos exigente Guignard). Para uma descrição geométrica abrangente, consulte os livros [28, 85], ou ainda [20].

Enunciamos no teorema seguinte o fato das CQs da Definição 1.2 garantirem que minimizadores sejam pontos KKT.

Teorema 1.1. *Seja x^* um minimizador local de (P). Se x^* satisfaz alguma das condições de qualificação LICQ, RCPLD ou CCP, então x^* é ponto KKT de (P).*

Demonstração. O resultado é consequência direta do Corolário 3.3 de [15] e das implicações da Figura 1. □

As condições KKT (1.2) e (1.3) usam informações apenas das derivadas primeiras das funções do problema. Dizemos assim que são condições de otimalidade de *primeira ordem*. Consideremos agora o problema

$$\min_x -x^2 \quad \text{s.a} \quad -1 \leqslant x \leqslant 1.$$

A origem é maximizador estrito e ponto KKT com multiplicador $\mu^* = (0, 0)$. Aqui, as condições KKT não detectam que a origem não é minimizador pois não levam em

consideração a *curvatura* do gráfico da função $-x^2$. Para tanto, podemos recorrer às derivadas segundas. Neste pequeno exemplo, a derivada segunda da função Lagrangiano na origem é $L''(0,\mu) = -2$, indicando uma curvatura "para baixo" no gráfico. Isso nos motiva a considerar condições de otimalidade de *segunda ordem*, para as quais dedicaremos o resto desta seção.

Definição 1.3. *Seja x^* um ponto KKT do problema (P). Dizemos que x^* satisfaz a* condição necessária forte de segunda ordem *(SSONC[4]) se*

$$d^T \nabla^2_{xx} L(x^*, \mu^*) d \geqslant 0 \quad \textit{para todo} \quad d \in C^S(x^*),$$

onde μ^* é vetor de multiplicadores KKT associados a x^* e

$$\mathcal{C}^S(x^*) := \{d \in \mathcal{T}^{\text{lin}}(x^*) \mid d^T \nabla F(x^*) \leqslant 0\} = \left\{ d \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} d^T \nabla F(x^*) \leqslant 0, & \\ d^T \nabla G_i(x^*) \leqslant 0, & \forall i \in I_G(x^*) \\ d^T \nabla H_i(x^*) = 0, & i = 1, \ldots, q \end{array} \right\}$$

é o cone crítico *(ou* cone de direções críticas*) em x^*.*

Geometricamente, todas as direções tangentes ao conjunto viável no ponto x^* e que são de descida para F a partir de x^* pertencem ao cone $\mathcal{C}^S(x^*)$. Para uma descrição geométrica consistente, consulte [28, 85].

A fim de obtermos condições necessárias de otimalidade envolvendo SSONC, novamente agregamos condições de qualificação às restrições. Infelizmente, nem toda CQ da Figura 1 (página 21) cumpre esse propósito. Particularmente, MFCQ não é uma CQ para SSONC [22] (e logo, nenhuma CQ implicada por MFCQ será, veja a Figura 1). Por outro lado, vale o próximo resultado.

Teorema 1.2 ([113, Teorema 6])**.** *Seja x^* um minimizador local de (P). Se x^* satisfaz RCRCQ então x^* cumpre SSONC.*

O Teorema 1.2 estende resultado análogo de [8], com CRCQ. Naturalmente, o mesmo vale para LICQ e Linearidade (veja a Figura 1). LICQ, inclusive, é a única condição de qualificação para SSONC utilizada em livros clássicos de otimização, como [28]. Extensões e uma discussão detalhada sobre CQs para SSONC são apresentadas em [3]. Mais recentemente, uma outra CQ para SSONC, chamada de *cone continuidade forte de segunda ordem* (SCCP2) [11], foi definida. Um fato interessante é que SCCP2 e LICQ são independentes entre si.

Infelizmente, há um inconveniente com a condição SSONC: verificar que um ponto x^* a cumpre é uma tarefa de alto custo computacional [117]. Ademais, Gould e Toint [68] mostraram que não podemos esperar convergência a pontos SSONC em

[4] Do inglês *Strong Second-Order Necessary Condition*.

métodos de barreira, mesmo sob hipóteses muito fortes. O mesmo ocorre com métodos de Lagrangiano aumentado [19]. Portanto, é comum lidarmos com uma condição de segunda ordem menos exigente, que enunciamos a seguir.

Definição 1.4. *Seja x^* um ponto KKT do problema (P). Dizemos que x^* satisfaz a* condição necessária fraca de segunda ordem *(WSONC[5]) se*

$$d^T \nabla^2_{xx} L(x^*, \mu^*) d \geqslant 0 \quad \textit{para todo} \quad d \in C^W(x^*),$$

onde μ^* é vetor de multiplicadores KKT associados a x^* e

$$\mathcal{C}^W(x^*) := \left\{ d \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} d^T \nabla G_i(x^*) = 0, & \forall i \in I_G(x^*) \\ d^T \nabla H_i(x^*) = 0, & i = 1, \ldots, q \end{array} \right\}$$

é o cone crítico fraco *em x^*.*

Verificar WSONC consiste na resolução de um problema quadrático de otimização somente com restrições de igualdade, uma tarefa computacionalmente viável [117]. Além disso, algoritmos de segunda ordem costumam garantir convergência a pontos que satisfazem WSONC. Portanto, até o momento, parece consenso na literatura que WSONC é a condição adequada para análise de convergência de métodos de segunda ordem.

É possível mostrar que $C^W(x^*) \subset C^S(x^*)$ para todo ponto KKT x^* de (P) (esta inclusão pode ser estrita, por exemplo, no problema $\min\limits_{x \leqslant 0} x^3$ em $x^* = 0$). Assim, SSONC implica WSONC, e logo qualquer CQ para a primeira condição é também uma CQ para segunda. Em particular, vale o Teorema 1.2 para WSONC.

Infelizmente, como ocorre com SSONC, MFCQ não é uma CQ para WSONC [22]. A fim de garantir WSONC, ou mesmo SSONC, uma das estratégias adotadas na literatura é agregar à MFCQ hipóteses adicionais [26, 27]. Nesse sentido, destacamos a seguinte propriedade sobre o posto dos gradientes das restrições:

Definição 1.5 ([17])**.** *Dizemos que x^* satisfaz a propriedade de* posto constante fraco *(WCR[6]) se*

$$\{\nabla G_i(x)\}_{i \in I_G(x^*)} \cup \{\nabla H_i(x)\}_{i=1,\ldots,q}$$

tem o mesmo posto para todo x em uma vizinhança de x^*.

Sozinha, WCR não é uma condição de qualificação [17], mas em conjunto com MFCQ torna-se uma CQ para WSONC:

Teorema 1.3 ([17, Teorema 3.1])**.** *Seja x^* um minimizador local de (P). Se x^* satisfaz simultaneamente MFCQ e WCR então x^* cumpre WSONC. Ou seja,*

[5] Do inglês *Weak Second-Order Necessary Condition.*
[6] Do inglês *Weak Constant Rank.*

"WSONC ou não-(MFCQ e WCR)"

é uma CNO de segunda ordem.

Uma condição de qualificação associada a WSONC, e ainda menos exigente que ambas "MFCQ e WCR" e RCRCQ, foi definida em [11]. Trata-se de uma extensão da noção de cone continuidade (Definição 1.2, página 20) à segunda ordem.

Definição 1.6. *Para cada $x \in \mathbb{R}^n$, consideremos o cone*

$$\mathcal{C}^W(x, x^*) = \left\{ d \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} d^T \nabla G_i(x) = 0, & \forall i \in I_G(x^*) \\ d^T \nabla H_i(x) = 0, & i = 1, \ldots, q \end{array} \right\}.$$

Dizemos que x^* satisfaz a condição de qualificação de cone continuidade de segunda ordem *(CCP2) se a multifunção $\mathcal{K}_2 : \mathbb{R}^n \rightrightarrows \mathbb{R}^n \times \mathbb{R}^{n \times n}$ definida por*

$$\mathcal{K}_2(x) = \bigcup_{\substack{\mu_i^G \geqslant 0,\, \forall i \in I_G(x^*), \\ \mu^H \in \mathbb{R}^q}} \left\{ \begin{array}{l} \left(\displaystyle\sum_{i \in I_G(x^*)} \mu_i^G \nabla G_i(x) + \sum_{i=1}^{q} \mu_i^H \nabla H_i(x),\, \mathcal{M} \right) \text{ tal que} \\ -\mathcal{M} + \displaystyle\sum_{i \in I_G(x^*)} \mu_i^G \nabla^2 G_i(x) + \sum_{i=1}^{q} \mu_i^H \nabla^2 H_i(x) \\ \text{é semidefinida positiva em } \mathcal{C}^W(x, x^*) \end{array} \right\}$$

é semicontínua exterior em x^*, isto é, se

$$\limsup_{x \to x^*} \mathcal{K}_2(x) \subset \mathcal{K}_2(x^*).$$

O conjunto $\mathcal{C}^W(x, x^*)$ pode ser visto como uma perturbação do cone crítico fraco $\mathcal{C}^W(x^*)$ ao redor do ponto viável x^*. Em particular, $\mathcal{C}^W(x^*, x^*) = \mathcal{C}^W(x^*)$. A condição CCP2 está para WSONC assim como CCP está para KKT, no sentido que WSONC admite a forma geométrica

$$\big(-\nabla F(x^*), -\nabla^2 F(x^*)\big) \in \mathcal{K}_2(x^*),$$

e é uma condição de qualificação para WSONC, como enunciado a seguir.

Teorema 1.4 ([11, Corolário 4.3])**.** *Seja x^* um minimizador local de (P). Se x^* satisfaz CCP2 então x^* cumpre WSONC. Ou seja,*

"WSONC ou não-CCP2"

é uma CNO de segunda ordem.

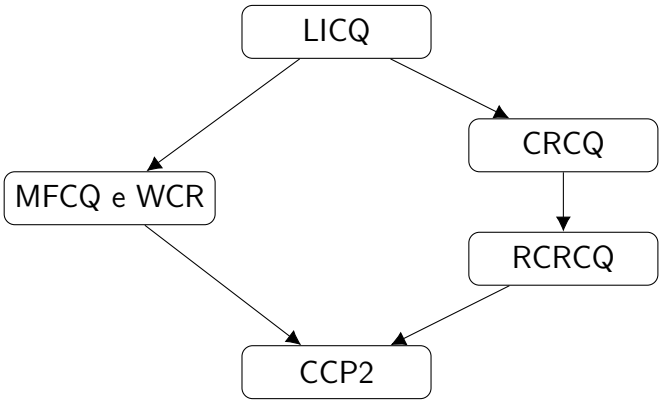

Figura 3 – Hierarquia entre CQs para WSONC. As setas indicam implicações lógicas.

A Figura 3 resume as condições de qualificação para WSONC mencionadas neste trabalho.

Antes de finalizarmos esta seção, cabe enfatizarmos que, assim como KKT, WSONC ou mesmo SSONC não são suficientes para otimalidade: no problema

$$\min_x \ x^3 \quad \text{s.a} \quad x \geqslant -1,$$

a origem satisfaz SSONC mas não é minimizador local, apesar de valer LICQ.

1.2 Lagrangiano aumentado de primeira ordem

Em Algencan, separamos as restrições do problema (P) (página 17) em restrições "difíceis" e "fáceis" [5, 35]. Via de regra, as restrições ditas difíceis são aquelas que não podem ser tratadas diretamente, e de forma eficiente, por algoritmos práticos. Em nosso caso, tais restrições são tratadas via penalização. Por um abuso de notação, vamos considerar que as restrições de (P) a serem penalizadas são exatamente $G(x) \leqslant 0$ e $H(x) = 0$. Por outro lado, as restrições consideradas fáceis são aquelas com estrutura simples, e as tratamos diretamente. Vamos denotar por $\Omega \subset \mathbb{R}^n$ o conjunto formado por essas restrições. Assim, a cada iteração de Algencan o subproblema a ser resolvido é

$$\min_x \ L_\rho(x, \mu) \quad \text{s.a} \quad x \in \Omega, \tag{1.4}$$

com ρ e μ fixos.

Um caso importante de restrições consideradas simples é quando todas elas são lineares (veja [6, 33]). Um caso ainda mais simples, e que ocorre com frequência, é quando Ω possui apenas restrições de caixa, isto é,

$$\Omega = \{x \in \mathbb{R}^n \mid \ell \leqslant x \leqslant u\}, \tag{1.5}$$

onde $\ell < u$. Atualmente, a implementação de Algencan fornecida pelo projeto TANGO [140] trabalha apenas com restrições de caixa. Também, o método de segunda ordem

Algencan-second só é definido para este caso. Por estes motivos, vamos supor durante todo este trabalho que Ω seja formado apenas por restrições de caixa. Isso facilitará o entendimento. Porém, é importante salientar que há trabalhos visando uma ampliação, tanto para o método de primeira ordem [33], quanto para o de segunda [35]. Mais ainda, uma nova estratégia de Lagrangiano aumentado que mantém restrições de igualdade nos subproblemas e penaliza restrições de desigualdades e de caixa foi recentemente proposta [31].

Passamos então ao estudo do subproblema (1.4) onde Ω é como em (1.5). Definimos o *gradiente projetado contínuo* como

$$\mathcal{G}_P(x) = P_\Omega\left(x - \nabla_x L_\rho(x,\mu)\right) - x,$$

onde $P_X(\cdot)$ é a projeção ortogonal (Euclidiana) sobre X. O multiplicador μ e o parâmetro de penalização ρ são fixos, e ficarão claros no contexto. Como Ω é compacto e convexo, a projeção $P_\Omega(x - \nabla_x L_\rho(x,\mu))$ é única. A utilização do gradiente projetado contínuo se justifica pelos seguintes motivos:

- A projeção ortogonal de um vetor $z \in \mathbb{R}^n$ na caixa (1.5) é calculada de forma direta por
$$[P_\Omega(z)]_i = \min\{u_i, \max\{\ell_i, z_i\}\} = \begin{cases} \ell_i & \text{se } z_i \leqslant \ell_i \\ z_i & \text{se } \ell_i < z_i < u_i \\ u_i & \text{se } z_i \geqslant u_i \end{cases};$$

- Se $\mathcal{G}_P(x^*) = 0$ então x^* é ponto KKT do subproblema (1.4). De fato, $\mathcal{G}_P(x^*) = 0$ implica
$$\begin{cases} x_i = \ell_i & \text{se} \quad [\nabla_x L_\rho(x,\mu)]_i \geqslant 0 \\ [\nabla_x L_\rho(x,\mu)]_i = 0 & \text{se} \quad [\nabla_x L_\rho(x,\mu)]_i \in (x_i - u_i, x_i - \ell_i) \\ x_i = u_i & \text{se} \quad [\nabla_x L_\rho(x,\mu)]_i \leqslant 0 \end{cases}. \tag{1.6}$$
Tomando
$$\mu^\ell = \left(\nabla_x L_\rho(x,\mu)\right)_+ \quad \text{e} \quad \mu^u = \left(-\nabla_x L_\rho(x,\mu)\right)_+ \tag{1.7}$$
vemos que $\mu^\ell, \mu^u \geqslant 0$ e
$$\mu_i^\ell(\ell_i - x_i) = \mu_i^u(x_i - u_i) = 0 \quad \text{para todo } i = 1, \ldots, n. \tag{1.8}$$
Isto é, as equações (1.3) do sistema KKT (Definição 1.1, página 19) se cumprem para as restrições de caixa $\ell - x \leqslant 0$ e $x - u \leqslant 0$ com os multiplicadores μ^ℓ e μ^u, respectivamente. Ademais, (1.6)–(1.8) fornecem, para cada $i = 1, \ldots, n$,
$$\left[\nabla_x L(x,\mu) - \sum_{j=1}^n \mu_j^\ell e_j + \sum_{j=1}^n \mu_j^u e_j\right]_i$$

$$= [\nabla_x L(x,\mu)]_i - ([\nabla_x L(x,\mu)]_i)_+ + ([-\nabla_x L(x,\mu)]_i)_+ = 0.$$

Ou seja, o gradiente da função Lagrangiano do subproblema (1.4) se anula com os multiplicadores μ^ℓ e μ^u. Desta forma, x^* é ponto KKT do subproblema (1.4).

Portanto, a escrita $\mathcal{G}_P(x^*) = 0$ é adequada na descrição de minimizadores de (1.4), como se faz em [5]. Notemos que neste subproblema vale a condição de qualificação de linearidade das restrições;

- Como a projeção ortogonal sobre convexos é contínua, $\mathcal{G}_P$ é uma função contínua. Assim, $\|\mathcal{G}_P(x)\|$ serve como medida de estacionariedade/otimalidade para o subproblema (1.4). Ou seja, "$\|\mathcal{G}_P(x)\| \approx 0$" é um critério de parada razoável para a iteração interna do método de Lagrangiano aumentado de primeira ordem (veja a expressão (1.9) no Algoritmo 1).

Dado $x \in \Omega$, definimos

$$\mathcal{F}(x) = \left\{ z \in \Omega \;\middle|\; \begin{array}{ll} z_i = \ell_i, & \text{se } x_i = \ell_i, \\ z_i = u_i, & \text{se } x_i = u_i, \\ \ell_i < z_i < u_i, & \text{caso contrário} \end{array} \right\}.$$

Este conjunto corresponde à menor face da caixa Ω que contém x (veja a Figura 4). Variáveis fixadas em um de seus limites são ditas *fixas*, e as não fixadas, *livres*. Seja $\mathcal{V}(x)$ a variedade afim de menor dimensão que contém $\mathcal{F}(x)$, e $\mathcal{S}(x)$ o subespaço paralelo a $\mathcal{V}(x)$. Diremos que a *dimensão de* $\mathcal{F}(x)$, denotada por $\dim \mathcal{F}(x)$, é a dimensão de $\mathcal{S}(x)$, que por sua vez coincide com o número de variáveis livres na face $\mathcal{F}(x)$. Definimos assim o *gradiente projetado interno* como

$$\mathcal{G}_I(x) = P_{\mathcal{S}(x)}\left(\mathcal{G}_P(x)\right).$$

De maneira análoga ao gradiente projetado contínuo, se $\mathcal{G}_I(x^*) = 0$ então x^* é ponto KKT do problema (1.4) restrito à face $\mathcal{F}(x^*)$,

$$\min_x \; L_\rho(x,\mu) \quad \text{s.a} \quad x \in \mathcal{F}(x^*).$$

Ou seja, a expressão $\mathcal{G}_I(x^*) = 0$ é adequada para descrever minimizadores dos subproblemas (1.4) *dentro* de uma face da caixa Ω. Ademais, $\mathcal{G}_I$ é contínua. A Figura 4 ilustra a geometria dos gradientes $\mathcal{G}_P$ e $\mathcal{G}_I$.

Algencan é apresentado no Algoritmo 1. Cabe ressaltar que no passo 1, apesar do uso específico do Algoritmo 2 para resolver subproblemas, temos flexibilidade para incorporar diferentes métodos de minimização em caixas. Portanto, o Algoritmo 1 poderá ter sua performance melhorada na medida em que melhores resolvedores sobre caixas sejam desenvolvidos (uma comparação numérica entre algoritmos de minimização em caixas é feita

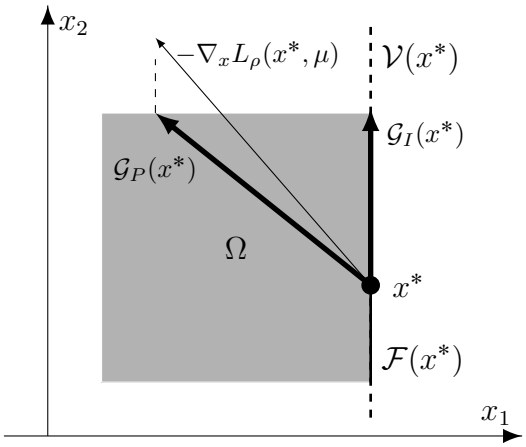

Figura 4 – Geometria dos gradientes $\mathcal{G}_P$ e $\mathcal{G}_I$. A região viável do subproblema (1.4) é a caixa Ω, enquanto que a face $\mathcal{F}(x^*)$ é formada pela borda direita. Se $\mathcal{G}_P(x^*) \neq 0$ então $\mathcal{G}_P(x^*)$ é uma direção de descida para $L_\rho(\cdot, \mu)$ a partir de x^*, viável para Ω. Por outro lado, $\mathcal{G}_I(x^*) \neq 0$ é uma direção de descida viável em relação à face $\mathcal{F}(x^*)$.

em [34]). Neste trabalho, estamos particularmente interessados no método Gencan [36] (Algoritmo 2), um resolvedor sobre caixas baseado em estratégias de restrições ativas e passos de gradiente espectral projetado (SPG[7]). O principal motivo para seu uso específico é que o método de LA de segunda ordem é definido modificando diretamente Gencan. Não por acaso, são utilizados os termos "Algencan" e "Algencan-second" em referência aos métodos de LA baseados em Gencan.[8]

Algoritmo 1 Algencan

Sejam $\mu^G_{\max} > 0$, $\mu^H_{\min} < \mu^H_{\max}$, $\gamma > 1$, $0 < \tau < 1$ e $\{\varepsilon^{\text{grad}}_k\}$ uma sequência de escalares positivos com $\lim_{k\to\infty} \varepsilon^{\text{grad}}_k = 0$. Sejam ainda

$$\mu^{G,1}_i \in [0, \mu^G_{\max}], \quad i = 1, \ldots, s,$$
$$\mu^{H,1}_i \in [\mu^H_{\min}, \mu^H_{\max}], \quad i = 1, \ldots, q$$

e $\rho_1 > 0$. Inicialize $k \leftarrow 1$.

Passo 1. Se possível, calcule um minimizador aproximado x^k do subproblema

$$\min_{x\in\Omega} L_{\rho_k}(x, \mu^k)$$

que satisfaça a condição de otimalidade aproximada

$$\left\|\mathcal{G}_P(x^k)\right\| \leqslant \varepsilon^{\text{grad}}_k, \tag{1.9}$$

utilizando Gencan (Algoritmo 2) com $\overline{\rho} = \rho_k$, $\overline{\mu} = \mu^k$ e $\varepsilon^{\text{grad}} = \varepsilon^{\text{grad}}_k$. Se não for possível, pare a execução declarando falha.

[7] Do inglês *Spectral Projected Gradient*.

[8] O leitor não deve confundir a indexação das sequências geradas por Gencan e Algencan, apesar de usarmos a mesma notação "x^k".

Passo 2. Defina

$$V_i^k = \max\left\{G_i(x^k), -\frac{\mu_i^{G,k}}{\rho_k}\right\}, \quad i = 1, \ldots, s.$$

Se $k > 1$ e

$$\max\{\|H(x^k)\|_\infty, \|V^k\|_\infty\} \leqslant \tau \max\{\|H(x^{k-1})\|_\infty, \|V^{k-1}\|_\infty\},$$

defina $\rho_{k+1} = \rho_k$. Caso contrário, defina $\rho_{k+1} = \gamma\rho_k$.

Passo 3. Calcule

$$\mu_i^{G,k+1} \in [0, \mu_{\max}^G], \quad i = 1, \ldots, s \quad \text{e}$$
$$\mu_i^{H,k+1} \in [\mu_{\min}^H, \mu_{\max}^H], \quad i = 1, \ldots, q.$$

Faça $k \leftarrow k + 1$ e vá para o Passo 1.

Algoritmo 2 GENCAN

Seja $x^1 \in \Omega$ uma solução inicial aproximada do problema

$$\min_{x\in\Omega} L_{\overline{\rho}}(x, \overline{\mu})$$

($\overline{\mu}$ e $\overline{\rho}$ são fixos aqui). Assuma que $\eta \in (0, 1)$ e $\varepsilon^{\text{grad}} > 0$. Inicialize $k \leftarrow 1$.

Passo 1. (Critério de parada) Se

$$\|\mathcal{G}_P(x^k)\| \leqslant \varepsilon^{\text{grad}}$$

pare declarando convergência.

Passo 2. (Decisão de permanência ou abandono da face corrente) Se

$$\|\mathcal{G}_I(x^k)\| \leqslant \eta\|\mathcal{G}_P(x^k)\|$$

calcule $x^{k+1} \in \Omega$ usando o método SPG [40] (iteração de abandono). Caso contrário, calcule x^{k+1} usando a estratégia de busca linear de [36, Algoritmo 2.1] (iteração interna).

Passo 3. Faça $k \leftarrow k + 1$ e vá para o Passo 1.

Antes de prosseguirmos, faremos uma observação interessante acerca do Algoritmo 1. No passo 3, os multiplicadores $\mu^{G,k+1}$ e $\mu^{H,k+1}$ da iteração seguinte são calculados em compactos, mantendo-se portanto limitados durante todo o processo. Ou seja, se as estimativas "reais" para estes multiplicadores dadas pelo gradiente da função Lagrangiano aumentado, a saber, $(\mu^{G,k} + \rho_k G(x^k))_+$ e $\mu^{H,k} + \rho_k H(x^k)$, tendem a crescer exageradamente, truncamo-las para a próxima iteração (*salvaguarda*). Pode ocorrer ainda do ponto limite não ser KKT, e logo nem sequer existam multiplicadores (é possível mostrar que, neste

caso, estimativas dadas via $\nabla_x L_{\rho_k}(x^k, \mu^k)$ vão em módulo a infinito). Neste ponto, o leitor pode se perguntar: ainda que as estimativas "reais" explodam, por que usar salvaguardas? De fato, o método de LA clássico, que remonta aos trabalhos de Hestenes [78] e Powell [125], não as utiliza. Como justificativa podemos dizer o seguinte: em última análise, qualquer implementação prática trabalha com multiplicadores truncados, pois no computador podemos representar apenas uma faixa finita de números. Portanto o uso de salvaguardas simula o que já ocorre no computador, e do qual não se pode escapar. Evidentemente, o truncamento não deve ser muito restritivo, sob pena de forçar o crescimento do parâmetro de penalização ρ desnecessariamente. Em outras palavras, $\mu^G_{\max}$, $\mu^H_{\min}$ e $\mu^H_{\max}$ devem ser suficientemente grandes em módulo de modo que multiplicadores para o ponto limite, quando existirem, estejam contidos nos intervalos do passo 3 (veja a discussão feita em [5]). Por outro lado, e como já dito, quando o ponto limite não é KKT, necessariamente algumas estimativas "reais" vão em módulo a infinito. Neste caso há truncamento de qualquer forma, e ρ aumenta. Recaímos assim, de certa forma, na estratégia de penalização externa pura, que possui excelentes resultados teóricos de convergência (lembre-se da comparação feita no início deste capítulo). Esta conexão, devida à salvaguarda, é fortemente explorada em [9]. Em [97], os autores comparam os métodos de LA com e sem salvaguardas, concluindo que elas são benéficas na prática numérica.

No que segue, analisamos os passos do Algoritmo 2. Como já dissemos, Gencan emprega uma estratégia de restrições ativas. Em nosso caso, a caixa Ω é dividida em faces disjuntas. A minimização ocorre em duas etapas: uma dentro da face corrente (iteração interna), e outra para fora da face (iteração de abandono). A ideia fundamental é que exploremos a face corrente ao máximo e só a abandonemos quando for vantajoso. O teste do passo 2

$$\|\mathcal{G}_I(x)\| \leqslant \eta \|\mathcal{G}_P(x)\|, \tag{1.10}$$

$\eta \in (0, 1)$, busca esse equilíbrio. Quando (1.10) ocorre, a medida $\|\mathcal{G}_I(x)\|$ de estacionariedade dentro da face é uma fração pequena da medida $\|\mathcal{G}_P(x)\|$ de estacionariedade em relação à caixa Ω. Neste caso, julgamos que a face corrente já foi suficientemente explorada, e é melhor abandoná-la.

No passo 2 do Algoritmo 2, temos a liberdade de escolha da estratégia de minimização interna. De fato, em princípio qualquer método de minimização irrestrita pode ser utilizado. Inicialmente, Gencan foi concebido com uma estratégia baseada em busca linear com *backtracking*, cujas direções são calculadas pelo método de gradientes conjugados (Newton Truncado) [36]. Ao longo do tempo outras estratégias foram usadas, resultando em diferentes versões de Gencan (a saber, com quase-Newton [37] e regiões de confiança [21]). A grosso modo, o cálculo de direções por métodos quase-Newton não se diferencia de Newton Truncando porque ambos visam resolver um sistema Newtoniano. Portanto em ambos, *backtracking* é empregado. Mas, por sua natureza, a estratégia

baseada em regiões de confiança não realiza *backtracking*. Isso torna a versão com regiões de confiança incompatível com a adaptação de GENCAN desenvolvida em [4] para lidar com direções de segunda ordem. De qualquer forma, não consideramos outras opções além de Newton Truncado.

Finalizamos esta seção com o principal resultado sobre convergência teórica do Algoritmo 1. Originalmente, a teoria de convergência de ALGENCAN foi estabelecida sob CPLD [5, Teorema 4.1]. No entanto, com o advento das condições sequenciais de otimalidade, sabemos hoje que ALGENCAN gera pontos limite *Aproximadamente KKT* (AKKT) [39, Teorema 6.2]. Como todo ponto AKKT sob CCP é KKT [15, Teorema 3.2], a convergência é automaticamente estabelecida com a condição CCP.

Teorema 1.5. *Seja $\{x^k\}$ uma sequência gerada pelo Algoritmo 1. Então*

- *$\{x^k\}$ admite pelo menos um ponto limite e todo limite x^* é um ponto KKT do problema de inviabilidade*

$$\min_x \|G(x)_+\|^2 + \|H(x)\|^2 \quad \text{s.a} \quad x \in \Omega.$$

- *Se a sequência dos parâmetros de penalidade $\{\rho_k\}$ for limitada, então todo ponto limite x^* é viável para (P).*
- *Todo ponto limite viável x^* que satisfaz CCP é ponto KKT de (P).*

1.3 Lagrangiano aumentado de segunda ordem

Na seção anterior, vimos que pontos limite viáveis de sequências geradas por ALGENCAN são pontos KKT mediante a condição de qualificação de cone continuidade (Teorema 1.5). Em outras palavras, ALGENCAN atinge pontos estacionários de primeira ordem.

ALGENCAN-SECOND [4] é um método de Lagrangiano aumentado que converge a pontos estacionários de segunda ordem de (P). A diferença básica para o método de primeira ordem está na resolução do subproblema (1.4). Aqui, GENCAN (Algoritmo 2) é adaptado para garantir que o iterando x^k satisfaça condições de segunda ordem para o subproblema. Essa adaptação, chamada GENCAN-SECOND, emprega uma busca linear, o cálculo de direções de curvatura negativa, e um critério que decide qual direção, de primeira ou segunda ordens, é escolhida. O objetivo desta seção é apresentar ALGENCAN-SECOND e GENCAN-SECOND, bem como resultados de convergência.

Além dos gradientes $\mathcal{G}_P(x)$ e $\mathcal{G}_I(x)$ da seção anterior, definimos, para todo $x \in \Omega$, o *gradiente interno* como

$$\mathcal{G}_{\mathcal{F}}(x) = P_{\mathcal{S}(x)}\left(\nabla_x L_\rho(x, \mu)\right)$$

(novamente μ e ρ ficarão claros no contexto). O papel do gradiente interno é essencialmente o mesmo que o desempenhado pelo gradiente projetado interno $\mathcal{G}_I(x^*)$. A diferença é que $\mathcal{G}_\mathcal{F}$ não é projetado na caixa Ω, o que implica $\|\mathcal{G}_\mathcal{F}(x^*)\| \geqslant \|\mathcal{G}_I(x^*)\|$ (essa desigualdade pode ser estrita, veja a Figura 4). Em especial, $\mathcal{G}_\mathcal{F}(x^*) \neq 0$ é uma direção de descida para o subproblema (1.4) restrito apenas à variedade afim $\mathcal{V}(x^*)$,

$$\min_x L_\rho(x,\mu) \quad \text{s.a} \quad x \in \mathcal{V}(x^*),$$

ou seja, o subproblema está restrito à face $\mathcal{F}(x^*)$ mas sem os limitantes nas variáveis livres.

Retomemos a expressão (1.1) da função Lagrangiano aumentado PHR do início deste capítulo:

$$L_\rho(x,\mu) = F(x) + \frac{\rho}{2}\left[\left\|\left(\frac{\mu^G}{\rho} + G(x)\right)_+\right\|^2 + \left\|\frac{\mu^H}{\rho} + H(x)\right\|^2\right].$$

Essa função é continuamente diferenciável em relação a x e

$$\nabla_x L_\rho(x,\mu) = \nabla F(x) + \nabla G(x)\left(\mu^G + \rho G(x)\right)_+ + \nabla H(x)\left(\mu^H + \rho H(x)\right).$$

No método de segunda ordem, é natural que busquemos trabalhar com informações de segunda ordem dos subproblemas. No entanto, a Hessiana de L_ρ pode não estar definida se $(\mu_i^G + \rho G_i(x))_+ = 0$ para algum i. Definimos então, para cada $x \in \Omega$ e $\varepsilon > 0$, a *ε-Hessiana aproximada* de L_ρ (com respeito a x) como

$$\begin{aligned}\nabla^2_\varepsilon L_\rho(x,\mu) = \nabla^2 F(x) &+ \sum_{i=1}^{s}\left(\mu_i^G + \rho G_i(x)\right)_+ \nabla^2 G_i(x) + \rho \sum_{i\in I_\varepsilon(x)} \nabla G_i(x)\nabla G_i(x)^T \\ &+ \sum_{i=1}^{q}\left(\mu_i^H + \rho H_i(x)\right)\nabla^2 H_i(x) + \rho\sum_{i=1}^{q}\nabla H_i(x)\nabla H_i(x)^T\end{aligned}$$

onde

$$I_\varepsilon(x) = \left\{ j \in \{1,\ldots,s\} \;\middle|\; \frac{1}{\sqrt{\rho}}\left(\mu_j^G + \rho G_j(x)\right) \geqslant -\varepsilon \right\}.$$

Observamos que $\nabla^2_0 L_\rho(x,\mu)$ é a Hessiana verdadeira de L_ρ onde ela está definida. Nesses pontos, os autovalores de $\nabla^2_\varepsilon L_\rho$ fornecem limitantes superiores para os autovalores da Hessiana verdadeira. Mais ainda, se $\nabla^2_0 L_\rho$ é semidefinida positiva então $\nabla^2_\varepsilon L_\rho$ também é.

Agora, seja $\mathcal{F}$ uma face e $\varepsilon > 0$. Para cada $x \in \mathcal{F}$, $\rho > 0$ e $\mu \in \mathbb{R}^s_+ \times \mathbb{R}^q$, definimos a *$\varepsilon$-Hessiana reduzida* $\mathcal{H}_{[\mathcal{F},\varepsilon,\rho]}(x,\mu)$ pondo

$$\left[\mathcal{H}_{[\mathcal{F},\varepsilon,\rho]}(x,\mu)\right]_{ij} = \begin{cases} \left[\nabla^2_\varepsilon L_\rho(x,\mu)\right]_{ij} & \text{se } x_i \text{ e } x_j \text{ forem livres em } \mathcal{F} \\ \delta_{ij} & \text{caso contrário} \end{cases},$$

onde δ_{ij} é o delta de Kronecker. A ideia por detrás da ε-Hessiana reduzida é a seguinte: assim como Gencan, sua adaptação para segunda ordem Gencan-second baseia-se em uma

estratégia de restrições ativas. Na iteração interna, buscamos resolver o subproblema (1.4) restrito à face corrente $\mathcal{F}$. Em especial, GENCAN usa o gradiente projetado interno $\mathcal{G}_I(x)$ (página 28) para decidir otimalidade de primeira ordem dentro da face. Em GENCAN-SECOND, a matriz $\mathcal{H}_{[\mathcal{F},\varepsilon,\rho]}(x,\mu)$ tem o papel de medir otimalidade de segunda ordem restrita a $\mathcal{F}$. De fato, se x_i é fixa em $\mathcal{F}$ então as direções que saem da face $\mathcal{F}$ têm componente i não nula (reflita sobre a Figura 4, página 29). Para simplificar, tomemos a direção do vetor canônico e_i de $\mathbb{R}^n$. Pela definição,

$$(\alpha e_i)^T \mathcal{H}_{[\mathcal{F},\varepsilon,\rho]}(x,\mu)(\alpha e_i) = \alpha^2 \geqslant 0,$$

ou seja, a direção αe_i não influencia na decisão de segunda ordem. Esse argumento pode ser estendido para uma direção qualquer decompondo-a na base canônica de $\mathbb{R}^n$. Em outras palavras, $\mathcal{H}_{[\mathcal{F},\varepsilon,\rho]}(x,\mu)$ está relacionada a WSONC dentro da face corrente $\mathcal{F}$.

ALGENCAN-SECOND é apresentado no Algoritmo 3 e GENCAN-SECOND no Algoritmo 4. Apresentamos ainda o algoritmo da iteração interna de GENCAN-SECOND (Algoritmo 5), que contém o critério de decisão para escolha entre direções de primeira e segunda ordens. Cabe ressaltar que, assim como em GENCAN, usamos por simplicidade a mesma notação "x^k" para indexação da sequência gerada por GENCAN-SECOND e ALGENCAN-SECOND. O leitor não deve confundi-las.

Algoritmo 3 ALGENCAN-SECOND

Sejam $\mu^G_{\max} > 0$, $\mu^H_{\min} < \mu^H_{\max}$, $\gamma > 1$ e $0 < \tau < 1$. Sejam $\{\varepsilon^{\text{fun}}_k\}$, $\{\varepsilon^{\text{grad}}_k\}$ e $\{\varepsilon^{\text{hess}}_k\}$ sequências de escalares positivos tais que

$$\lim_{k\to\infty} \varepsilon^{\text{fun}}_k = \lim_{k\to\infty} \varepsilon^{\text{grad}}_k = \lim_{k\to\infty} \varepsilon^{\text{hess}}_k = 0.$$

Sejam ainda

$$\mu^{G,1}_i \in [0, \mu^G_{\max}], \quad i = 1,\ldots,s,$$
$$\mu^{H,1}_i \in [\mu^H_{\min}, \mu^H_{\max}], \quad i = 1,\ldots,q$$

e $\rho_1 > 0$. Inicialize $k \leftarrow 1$.

Passo 1. Se possível, calcule um minimizador aproximado x^k do problema

$$\min_{x\in\Omega} L_{\rho_k}(x,\mu^k)$$

usando GENCAN-SECOND (Algoritmo 4) com $\overline{\rho} = \rho_k$, $\overline{\mu} = \mu^k$, $\varepsilon^{\text{fun}} = \varepsilon^{\text{fun}}_k$, $\varepsilon^{\text{grad}} = \varepsilon^{\text{grad}}_k$ e $\varepsilon^{\text{hess}} = \varepsilon^{\text{hess}}_k$. O iterando x^k satisfaz

$$\left\|\mathcal{G}_P(x^k)\right\| \leqslant \varepsilon^{\text{grad}}_k$$

e

$$\dim \mathcal{F}(x^k) = 0 \quad \text{ou} \quad \lambda_{\min}\left(\mathcal{H}_{[\mathcal{F}(x^k),\varepsilon^{\text{fun}}_k,\rho_k]}(x^k,\mu^k)\right) \geqslant -\varepsilon^{\text{hess}}_k.$$

Se não for possível, pare declarando falha.

Passo 2. Defina

$$V_i^k = \max\left\{G_i(x^k), -\frac{\mu_i^{G,k}}{\rho_k}\right\}, \quad i = 1, \ldots, s.$$

Se $k > 1$ e

$$\max\{\|H(x^k)\|_\infty, \|V^k\|_\infty\} \leqslant \tau \max\{\|H(x^{k-1})\|_\infty, \|V^{k-1}\|_\infty\},$$

defina $\rho_{k+1} = \rho_k$. Caso contrário, defina $\rho_{k+1} = \gamma\rho_k$.

Passo 3. Calcule

$$\mu_i^{G,k+1} \in [0, \mu_{\max}^G], \quad i = 1, \ldots, s \quad \text{e}$$
$$\mu_i^{H,k+1} \in [\mu_{\min}^H, \mu_{\max}^H], \quad i = 1, \ldots, q.$$

Faça $k \leftarrow k + 1$ e vá para o Passo 1.

Algoritmo 4 GENCAN-SECOND

Seja $x^1 \in \Omega$ uma solução inicial aproximada do problema

$$\min_{x\in\Omega} L_{\overline{\rho}}(x, \overline{\mu})$$

($\overline{\mu}$ e $\overline{\rho}$ são fixos aqui). Assuma que $\eta \in (0, 1)$, $\varepsilon^{\text{fun}}, \varepsilon^{\text{grad}}, \varepsilon^{\text{hess}} > 0$ e $\varepsilon^{\text{curv}} \in (0, \varepsilon^{\text{hess}})$. Inicialize $k \leftarrow 1$.

Passo 1. (Critério de parada) Se

$$\|\mathcal{G}_P(x^k)\| \leqslant \varepsilon^{\text{grad}}$$

e

$$\dim \mathcal{F}(x^k) = 0 \quad \text{ou} \quad \lambda_{\min}\left(\mathcal{H}_{[\mathcal{F}(x^k), \varepsilon_k^{\text{fun}}, \overline{\rho}]}(x^k, \overline{\mu})\right) \geqslant -\varepsilon_k^{\text{hess}},$$

pare declarando convergência.

Passo 2. (Decisão de permanência ou abandono da face corrente) Se

$$\|\mathcal{G}_I(x^k)\| \leqslant \eta\|\mathcal{G}_P(x^k)\|$$

e

$$\dim \mathcal{F}(x^k) = 0 \quad \text{ou} \quad \lambda_{\min}\left(\mathcal{H}_{[\mathcal{F}(x^k), \varepsilon^{\text{fun}}, \overline{\rho}]}(x^k, \overline{\mu})\right) \geqslant -\varepsilon^{\text{curv}},$$

calcule $x^{k+1} \in \Omega$ usando o método SPG [40] (iteração de abandono). Caso contrário, calcule x^{k+1} pelo Algoritmo 5 usando os parâmetros $\overline{\rho}$, $\overline{\mu}$, ε^{fun} e $\varepsilon^{\text{curv}}$ (iteração interna).

Passo 3. Faça $k \leftarrow k + 1$ e vá para o Passo 1.

Algoritmo 5 Iteração interna de GENCAN-SECOND

Seja $x^1 \in \Omega$ uma solução inicial aproximada do problema

$$\min_{x\in\Omega} L_{\overline{\rho}}(x, \overline{\mu})$$

($\overline{\mu}$ e $\overline{\rho}$ são fixos aqui). Assuma que $\alpha \in (0, 1/2)$, $\beta > 0$, $\theta \in (0, 1)$, $\kappa > 0$, $\varepsilon_0 > 0$ e $\varepsilon^{\text{fun}}, \varepsilon^{\text{curv}} > 0$. Inicialize $k \leftarrow 1$.

Passo 1. (Direção de primeira ordem) Se $\|\mathcal{G}_{\mathcal{F}}(x^k)\| \neq 0$, calcule $d^{k,1} \in \mathcal{S}(x^k)$ satisfazendo a "condição do ângulo"

$$\mathcal{G}_{\mathcal{F}}(x^k)^T d^{k,1} \leqslant -\theta \|\mathcal{G}_{\mathcal{F}}(x^k)\| \|d^{k,1}\|$$

e a "condição *beta*"

$$\|d^{k,1}\| \geqslant \beta \|\mathcal{G}_{\mathcal{F}}(x^k)\|.$$

Caso contrário, defina $d^{k,1} = 0$.

Passo 2. (Direção de segunda ordem) Se

$$\|\mathcal{G}_{\mathcal{F}}(x^k)\| \geqslant \varepsilon_0 \quad \text{ou} \quad \lambda_{\min}\left(\mathcal{H}_{[\mathcal{F}(x^k),\varepsilon^{\text{fun}},\overline{\rho}]}(x^k, \overline{\mu})\right) \geqslant -\varepsilon^{\text{curv}}, \tag{1.11}$$

defina $d^k = d^{k,1}$ e vá para o Passo 4. Caso contrário, calcule $d^{k,2} \in \mathcal{S}(x^k)$ unitária satisfazendo

$$\mathcal{G}_{\mathcal{F}}(x^k)^T d^{k,2} \leqslant 0 \quad \text{e} \quad (d^{k,2})^T \mathcal{H}_{[\mathcal{F}(x^k),\varepsilon^{\text{fun}},\overline{\rho}]}(x^k, \overline{\mu}) d^{k,2} < -\varepsilon^{\text{curv}}. \tag{1.12}$$

Passo 3. (Decisão entre as direções de primeira e segunda ordens) Se $\|d^{k,1}\| \neq 0$ e

$$\frac{\mathcal{G}_{\mathcal{F}}(x^k)^T d^{k,1}}{\|d^{k,1}\|} \leqslant \kappa \left(\frac{\mathcal{G}_{\mathcal{F}}(x^k)^T d^{k,2}}{\|d^{k,2}\|} + \frac{1}{2} \cdot \frac{(d^{k,2})^T \mathcal{H}_{[\mathcal{F}(x^k,\overline{\mu}),\varepsilon^{\text{fun}},\overline{\rho}]}(x^k, \overline{\mu}) d^{k,2}}{\|d^{k,2}\|^2} \right) \tag{1.13}$$

faça $d^k = d^{k,1}$. Caso contrário, faça $d^k = d^{k,2}$.

Passo 4. (Tamanho do passo) Calcule o máximo tamanho de passo

$$t_k^{\max} = \max\left\{ t \geqslant 0 \mid x^k + s d^k \in \Omega,\ \forall s \in [0, t] \right\}. \tag{1.14}$$

Passo 5. (Calcula x^{k+1} na borda de $\mathcal{F}(x^k)$) Se

$$t_k^{\max} < 1 \quad \text{e} \quad L_{\overline{\rho}}(x^k + t_k^{\max} d^k, \overline{\mu}) < L_{\overline{\rho}}(x^k, \overline{\mu})$$

então calcule x^{k+1} na borda de $\mathcal{F}(x^k)$ tal que $L_{\overline{\rho}}(x^{k+1}, \overline{\mu}) \leqslant L_{\overline{\rho}}(x^k + t_k^{\max} d^k, \overline{\mu})$ e finalize a execução.

Passo 6. (Escolha da função de descenso) Se $d^k = d^{k,2}$ defina $\gamma(t) = -t^2 \varepsilon^{\text{curv}}$. Caso contrário, defina $\gamma(t) = t(d^k)^T \nabla_x L_{\overline{\rho}}(x^k, \overline{\mu})$.

Passo 7. (*Backtracking*) Faça $t \leftarrow \min\{1, t_k^{\max}\}$. Se a condição de descenso

$$L_{\overline{\rho}}(x^k + td^k, \overline{\mu}) \leqslant L_{\overline{\rho}}(x^k, \overline{\mu}) + \alpha\gamma(t) \tag{1.15}$$

é satisfeita, faça $t_k = t$, calcule x^{k+1} na borda de $\mathcal{F}(x^k)$ tal que $L_{\overline{\rho}}(x^{k+1}, \overline{\mu}) \leqslant L_{\overline{\rho}}(x^k + t_k d^k, \overline{\mu})$ e finalize a execução. Caso contrário, escolha $t^{\text{novo}} \in [0.1t, 0.9t]$, faça $t \leftarrow t^{\text{novo}}$ e repita o teste (1.15).

Basicamente, as mesmas justificativas usadas para Gencan (Algoritmo 2, página 30) podem ser empregadas para Gencan-second, com as devidas exigências de segunda ordem (compare os Algoritmos 2 e 4 e veja as semelhanças). Fazemos a seguir alguns comentários sobre a iteração interna de Gencan-second (Algoritmo 5).

- Como vimos, se $\mathcal{G}_{\mathcal{F}}(x^k) = 0$ então x^k é ponto estacionário de primeira ordem (KKT) do subproblema (1.4) dentro da variedade afim $\mathcal{V}(x^k)$, em particular, restrito à face $\mathcal{F}(x^k)$. Portanto, só faz sentido calcular direções de primeira ordem se $\mathcal{G}_{\mathcal{F}}(x^k) \neq 0$ (passo 1). As condições do ângulo e *beta* do passo 1 são clássicas em otimização, e procuram evitar passos pequenos ou pouco efetivos: a primeira evita direções de descida "quase ortogonais" ao gradiente; a segunda garante que a direção seja grande em relação à norma do gradiente. Veja por exemplo [109];

- Se a condição (1.11) do passo 2 vale, a estacionariedade de primeira ordem não foi atingida e a ε-Hessiana reduzida é quase semidefinida positiva. Neste caso, uma direção de segunda ordem não é útil, e portanto não é calculada. No caso contrário, (1.12) garante que a direção $d^{k,2}$ não seja de subida dentro da face, e que seja de curvatura negativa. Mais ainda, como neste ponto da execução temos $\lambda_{\min}\left(\mathcal{H}_{[\mathcal{F}(x^k),\varepsilon^{\text{fun}},\overline{\rho}]}(x^k, \overline{\mu})\right) < -\varepsilon^{\text{curv}}$, uma direção como esta existe;

- O critério do passo 3 só pode ser verificado se ambas direções de primeira e segunda ordens são calculadas. Ambos os lados da desigualdade (1.13) representam decréscimos esperados de $L_{\overline{\rho}}(\cdot, \overline{\mu})$ a partir de x^k. Em relação à segunda ordem, a estimativa obedece um modelo quadrático. Já para a primeira ordem, o modelo linear é suficiente. O parâmetro κ pode ser visto como o peso que a primeira ordem tem no processo de escolha (nos testes do Capítulo 4 tomamos $\kappa = 1$). Veja a Observação 3 de [4] para uma discussão detalhada;

- Uma escolha admissível para x^{k+1} no passo 5 é $x^k + t_k^{\max} d^k$, que inclusive adotamos em nossos testes numéricos do Capítulo 4. Notemos que pela expressão (1.14), este ponto atinge a fronteira de $\mathcal{F}(x^k)$. Analogamente, adotamos $x^{k+1} = x^k + t_k d^k$ no passo 7;

- Se a direção de primeira ordem é escolhida, então o descenso suficiente é do tipo Armijo (veja os passos 6 e 7). Caso a direção de segunda ordem seja escolhida, é

exigido um decréscimo quadrático que independe do gradiente de $L_{\overline{\rho}}$. Notemos que a direção $d^{k,2}$ não necessariamente satisfaz a condição do ângulo, apenas não é direção de subida. O termo $-t^2\varepsilon^{\text{curv}}$ é inspirado na expansão de Taylor de ordem 2 da função $\phi(t) = L_{\overline{\rho}}(x^k + td^{k,2}, \overline{\mu})$ em $t = 0$, e do fato de $(d^{k,2})^T \mathcal{H}_{[\mathcal{F}(x^k),\varepsilon^{\text{fun}},\overline{\rho}]}(x^k, \overline{\mu})d^{k,2} < -\varepsilon^{\text{curv}}$ (veja a expressão (1.12)).

No que segue, enunciamos o principal resultado de convergência do Algoritmo 3. Evidentemente, o Teorema 1.5 (página 32) vale para Algencan-second pois a sequência gerada pelo algoritmo de segunda ordem cumpre as condições exigidas em Algencan. Adicionalmente, sabemos que Algencan-second atende à condição sequencial de segunda ordem conhecida como AKKT2 [11], que por sua vez garante WSONC sob CCP2 [11, Teorema 4.16]. O resultado seguinte é portanto uma atualização do Teorema 2 de [4].

Teorema 1.6. *Seja $\{x^k\}$ uma sequência gerada pelo Algoritmo 3. Então*

- *$\{x^k\}$ admite pelo menos um ponto limite e todo limite x^* é um ponto KKT do problema de inviabilidade*

$$\min_x \|G(x)_+\|^2 + \|H(x)\|^2 \quad \text{s.a} \quad x \in \Omega.$$

- *Se a sequência dos parâmetros de penalidade $\{\rho_k\}$ for limitada, então todo ponto limite x^* é viável.*
- *Todo ponto limite viável x^* que satisfaz CCP é um ponto KKT de (P).*
- *Todo ponto limite viável x^* que satisfaz CCP2 cumpre WSONC (em particular, é KKT).*

1.4 Melhoramentos no cálculo de autopares

Em métodos de segunda ordem como Algencan-second, um enorme custo computacional está associado ao cálculo de autopares. Aqui apresentamos melhorias visando uma implementação mais econômica. A ideia é realizar computações somente quando realmente necessárias. As estratégias adotadas são as seguintes:

(a) Como vimos na seção anterior, em Algencan-second a matriz $\mathcal{H}_{[\mathcal{F}(x^k),\varepsilon^{\text{fun}},\overline{\rho}]}(x^k, \overline{\mu})$ tem linhas e colunas da identidade para índices de variáveis fixas em alguns dos limites da caixa Ω. O método Gencan para subproblemas implementado em Algencan traz naturalmente a identificação de variáveis fixas, que usamos também para Gencan-second. É natural em Gencan permutar as entradas de direções de descida e gradientes a fim de representar suas projeções na caixa. Especificamente, as entradas de variáveis livres são colocadas nas primeiras posições. Identificadas as variáveis

livres, procedemos de maneira análoga permutando linhas e colunas para escrever $\mathcal{H}_{[\mathcal{F}(x^k),\varepsilon^{\text{fun}},\overline{\rho}]}(x^k,\overline{\mu})$ na forma

$$\begin{bmatrix} R & 0 \\ 0 & I \end{bmatrix}.$$

Como $\lambda_{\min}\left(\mathcal{H}_{[\mathcal{F}(x^k),\varepsilon^{\text{fun}},\overline{\rho}]}(x^k,\overline{\mu})\right) = \min\{1,\lambda_{\min}(R)\}$, o cálculo do menor autovalor em GENCAN-SECOND fica reduzido à matriz de menor ordem R, isto é, devemos resolver apenas $Rv = \gamma v$. Neste caso, um autovetor associado é $(v,0)$, onde as últimas entradas nulas são relativas às variáveis fixas. Finalmente permutamos o vetor $(v,0)$ para representação no espaço de dimensão maior n. Em resumo, todo o custo computacional da segunda ordem fica restrito à submatriz R. Ademais, essa estratégia minimiza problemas com escalamento, dado que as únicas entradas consideradas no cálculo de autovalores são entradas da ε-Hessiana aproximada $\nabla^2_{\varepsilon^{\text{fun}}} L_{\overline{\rho}}(x^k,\overline{\mu})$, uma matriz totalmente proveniente dos dados do problema e dos parâmetros do algoritmo. Infelizmente, a experiência numérica indica que há poucas variáveis fixas, ou seja, R tem quase dimensão total. No entanto, como o custo computacional da permutação é baixo, a realizamos;

(b) Utilizamos o método iterativo de gradientes conjugados precondicionado com correções de Jacobi (JPCG[9]) desenvolvido em [120] (veja ainda [121]) para cálculo de autopares, e adequado a matrizes esparsas. Estimativas de autovetores de uma iteração de GENCAN-SECOND são usadas em iterações posteriores como aproximações de uma base inicial do subespaço de iteração em JPCG. Além disso, discos de Gershgorin são usados para fornecer uma estimativa inicial para o autovalor, ou até mesmo declarar que não há autovalores suficientemente negativos (caso em que nenhuma iteração de JPCG é feita). Apesar do uso de pré-condicionadores em JPCG poder acelerar o método, não utilizamos/desenvolvemos algum;

(c) Observamos com certa frequência durante a prática numérica que iterações consecutivas de GENCAN-SECOND são de passos curtos, isto é, $\|x^{k+1} - x^k\| \approx 0$. Logo, uma análise de sensibilidade do autovalor da Hessiana pode ser útil no seguinte sentido: se o passo foi curto, $\lambda_{\min}$ menos um limitante superior para sua perturbação pode ser maior do que ou igual a $-\varepsilon^{\text{curv}}$. Neste caso, não necessitamos de um novo cálculo de autovalor. Evidentemente isso só faz sentido se a estimativa para perturbação de $\lambda_{\min}$ for computacionalmente barata.

Neste contexto, lançamos mão do seguinte resultado [139, Corolário 4.13, pg 205]: se A e E são matrizes simétricas de mesma dimensão n e $\lambda_0(Z) = \lambda_{\min}(Z) \leqslant \cdots \leqslant$

[9] Do inglês *Jacobi Preconditioned Conjugate Gradients*.

$\lambda_{n-1}(Z)$ $(Z = A, E)$ são seus autovalores, então

$$\sqrt{\sum_{i=0}^{n-1} \left(\lambda_i(A+E) - \lambda_i(A)\right)^2} \leqslant \|E\|_F, \tag{1.16}$$

onde

$$\|E\|_F = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} E_{ij}^2}$$

é a norma de Frobenius da matriz E. Em nosso caso, consideramos as ε-Hessianas reduzidas das iterações k e $k+r$, $A = \mathcal{H}^k$ e $A + E = \mathcal{H}^{k+r}$ respectivamente. De (1.16), segue que

$$\lambda_{\min}(\mathcal{H}^{k+r}) \geqslant \lambda_{\min}(\mathcal{H}^k) - \|\mathcal{H}^{k+r} - \mathcal{H}^k\|_F. \tag{1.17}$$

Na presença de continuidade, a matriz $E = \mathcal{H}^{k+r} - \mathcal{H}^k$ pode ser vista como uma pequena perturbação quando $x^{k+r} \approx x^k$. Assim, é possível que na iteração $k+r$, a desigualdade (1.17) forneça $\lambda_{\min}(\mathcal{H}^{k+r}) \geqslant -\varepsilon^{\text{curv}}$ se anteriormente $\lambda_{\min}(\mathcal{H}^k) > -\varepsilon^{\text{curv}}$. Notemos que a computação de $\|\mathcal{H}^{k+r} - \mathcal{H}^k\|_F$ tem custo computacional baixo em relação ao cálculo de autopares (comparável a uma iteração de JPCG, cujo custo é linear em relação ao número de elementos não nulos da Hessiana). Especificamente, fazemos o seguinte:

- calculamos o autopar quando a primeira requisição é feita;
- em qualquer requisição futura, ainda que em outra iteração externa de ALGENCAN-SECOND, se o último autovalor calculado for maior que $-\varepsilon^{\text{curv}}$, verificamos se (1.17) ocorre. Em caso afirmativo, não calculamos novo autopar.

Por fim, não faremos uma análise do efeito da estratégia (c) em nossos testes numéricos do Capítulo 4, pois, em princípio, não vemos razões para acreditarmos que ela interfira na qualidade do ponto limite obtido pelo método. Nos limitaremos a dizer que constatamos ser benéfica para redução do tempo computacional.

Capítulo 2

Problemas com restrições de complementariedade

Neste capítulo apresentamos propriedades fundamentais dos problemas de programação matemática com restrições de complementariedade. Retomemos o modelo dado na introdução:

$$\begin{array}{ll} \min\limits_{x} & f(x) \\ \text{s.a} & G(x) \leqslant 0, \quad H(x) = 0 \\ & g(x) \geqslant 0, \quad h(x) \geqslant 0, \quad g(x)^T h(x) \leqslant 0, \end{array} \tag{MPCC}$$

onde $f : \mathbb{R}^n \to \mathbb{R}$, $G : \mathbb{R}^n \to \mathbb{R}^s$, $H : \mathbb{R}^n \to \mathbb{R}^q$, e $g, h : \mathbb{R}^n \to \mathbb{R}^m$ são funções continuamente diferenciáveis.

Como comentamos na introdução, (MPCC) não satisfaz a maioria das condições de qualificação para programação não linear geral, apresentadas na Seção 1.1 (página 18). Somente a condição mais fraca, Guignard, pode ser esperada com certa frequência [56]. Isso nos leva a considerar condições de qualificação específicas, menos exigentes que Guignard. Sob tais CQs não há garantia de pontos KKT, e logo conceitos de estacionariedade menos exigentes foram desenvolvidos. A Seção 2.1 dedica-se a estas questões. Na Seção 2.2, discutimos relações entre condições sequenciais de otimalidade conhecidas na literatura e (MPCC), justificando o porquê de não as considerarmos em nossa análise de convergência.

2.1 Condições de qualificação e estacionariedade para MPCC

Em relação a (MPCC), a função Lagrangiano usual definida no Capítulo 1 toma a forma

$$L(x, \mu) = f(x) + (\mu^G)^T G(x) + (\mu^H)^T H(x) - (\mu^g)^T g(x) - (\mu^h)^T h(x) + \mu^0 g(x)^T h(x), \tag{2.1}$$

onde $\mu = (\mu^G, \mu^H, \mu^g, \mu^h, \mu^0) \in \mathbb{R}^s_+ \times \mathbb{R}^q \times \mathbb{R}^{2m+1}_+$. Como vimos, $L(x, \mu)$ é usada para definir as condições KKT (Definição 1.1, página 19). Por outro lado, KKT é um conceito de

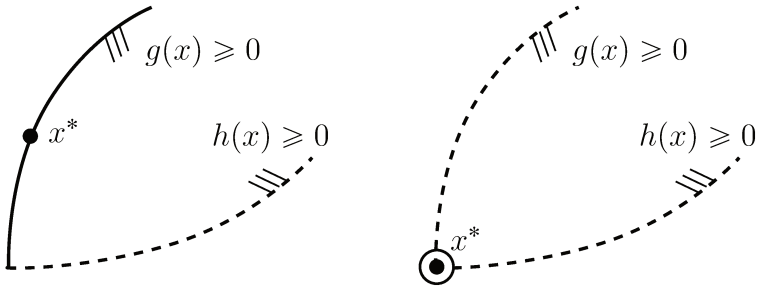

Figura 5 – Região viável do problema (TNLP). As restrições $g(x) \geqslant 0$ e $h(x) \geqslant 0$ compõem a complementariedade $g(x)^T h(x) \leqslant 0$. À esquerda, apenas a restrição $g(x) \geqslant 0$ é ativa em x^*, e logo a menor face que contém x^* é definida por $g(x) = 0$, $h(x) \geqslant 0$. À direita, ambas restrições são ativas em x^*, e a menor face é formada apenas pelo próprio ponto x^*.

estacionariedade muito exigente para (MPCC) como veremos nesta seção. Definimos assim, para cada x^* viável, o *problema não linear apertado*

$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.a} \quad & G(x) \leqslant 0, \quad H(x) = 0 \\ & g_i(x) = 0, \quad i \in I_g(x^*) \\ & h_i(x) = 0, \quad i \in I_h(x^*) \\ & g_i(x) \geqslant 0, \quad i \in I_h(x^*) \backslash I_g(x^*) \\ & h_i(x) \geqslant 0, \quad i \in I_g(x^*) \backslash I_h(x^*), \end{aligned} \tag{TNLP}$$

onde

$$I_g(x^*) = \{i \mid g_i(x^*) = 0\} \quad \text{e} \quad I_h(x^*) = \{i \mid h_i(x^*) = 0\}$$

são conjuntos de índices de restrições ativas. Notemos que, sendo x^* viável para (MPCC), $I_g(x^*) \cup I_h(x^*) = \{1, \ldots, m\}$ e logo

$$i \in I_h(x^*) \backslash I_g(x^*) \;\Leftrightarrow\; i \notin I_g(x^*)$$

(o mesmo ocorre para $I_g(x^*) \backslash I_h(x^*)$). Portanto o modelo (TNLP)[1] é obtido precisamente pela fixação das restrições $g_i(x) \geqslant 0$ e $h_i(x) \geqslant 0$ ativas em x^*, descartando a complementariedade. Geometricamente, a região viável de (TNLP) é a interseção da menor face do conjunto

$$\{x \in \mathbb{R}^n \mid g(x) \geqslant 0,\ h(x) \geqslant 0\}$$

que contém x^* com o conjunto $\{x \in \mathbb{R}^n \mid G(x) \leqslant 0,\ H(x) = 0\}$ (veja a Figura 5).

Como (TNLP) é um problema sem restrições de complementariedade, podemos esperar que valham as condições de qualificação padrão apresentadas na Seção 1.1. Mais ainda, todo minimizador local x^* de (MPCC) é minimizador de (TNLP) pois a região

[1] Do inglês *Tightened NonLinear Program*.

viável do segundo problema está contida na região viável do primeiro. Portanto, uma condição necessária de otimalidade para (TNLP) é também uma CNO para (MPCC). Logo, é conveniente considerar as condições KKT para (TNLP), como faremos a seguir.

Consideremos a função Lagrangiano usual relativa a (TNLP)

$$\mathcal{L}(x, \lambda) = f(x) + (\lambda^G)^T G(x) + (\lambda^H)^T H(x) - (\lambda^g)^T g(x) - (\lambda^h)^T h(x).$$

Nos referimos a $\mathcal{L}$ como a função *Lagrangiano-MPCC* (durante todo este trabalho, μ e λ denotarão os multiplicadores relativos a L e $\mathcal{L}$, respectivamente). Os sinais dos multiplicadores em $\mathcal{L}$ coincidem com aqueles na função Lagrangiano usual (2.1) em relação às *restrições ordinárias* $G(x) \leqslant 0$ e $H(x) = 0$, mas se diferencia em relação às *restrições complementares* $g(x) \geqslant 0$ e $h(x) \geqslant 0$. De fato, como em (TNLP) as restrições complementares ativas em x^* são restrições de igualdade, o sinal dos multiplicadores correspondentes não está definido. Por outro lado, KKT para (TNLP) mantém a nulidade dos multiplicadores das restrições inativas. Isso nos leva ao seguinte conceito de estacionariedade: diremos que x^* é *fracamente estacionário* para (MPCC) se x^* é ponto KKT para (TNLP). Mais especificamente:

Definição 2.1. *Dizemos que um ponto viável x^* de* (MPCC) *é* fracamente estacionário *(W-estacionário)*[2] *se existe um vetor de multiplicadores $\lambda = (\lambda^G, \lambda^H, \lambda^g, \lambda^h) \in \mathbb{R}^s_+ \times \mathbb{R}^q \times \mathbb{R}^{2m}$ tal que*

$$\nabla_x \mathcal{L}(x^*, \lambda) = 0, \quad \lambda^g_{I_h(x^*) \setminus I_g(x^*)} = 0 \quad e \quad \lambda^h_{I_g(x^*) \setminus I_h(x^*)} = 0.$$

Como já dito, todo minimizador local de (MPCC) é também minimizador local de (TNLP). Logo,

"W-estacionário ou não-(CQ para TNLP)"

é uma condição necessária de otimalidade para (MPCC). Qualquer CQ padrão dada na Seção 1.1 (página 18) gera automaticamente uma condição necessária dessa forma. Por isso, é comum nos referirmos às condições de qualificação específicas para (MPCC) pela sigla "MPCC-CQ", onde "CQ" é uma condição de qualificação padrão. Algumas MPCC-CQs são definidas justamente exigindo a CQ correspondente sobre (TNLP), como por exemplo MPCC-LICQ (relativa à LICQ, item 1 da Definição 1.2, página 20), MPCC-MFCQ [135], MPCC-CPLD [82] e MPCC-CRCQ [138]. Destacamos a seguir a condição de independência linear, que será de nosso interesse na análise de convergência do Capítulo 3.

Definição 2.2. *Dizemos que um ponto viável x^* satisfaz a condição de qualificação de* independência linear dos gradientes das restrições ativas para MPCC *(MPCC-LICQ) se*

[2] A letra "W" refere-se à palavra inglesa *weakly*.

os gradientes

$$\nabla G_i(x^*), i \in I_G(x^*), \quad \nabla H_i(x^*), i = 1, \ldots, q, \quad \nabla g_i(x^*), i \in I_g(x^*), \quad \nabla h_i(x^*), i \in I_h(x^*)$$

são linearmente independentes.

De maneira análoga à LICQ, a condição MPCC-LICQ implica a unicidade do vetor de multiplicadores λ em um ponto fracamente estacionário. Esse resultado é trivial pois, como $\lambda^g_{I_h(x^*)\setminus I_g(x^*)} = 0$ e $\lambda^h_{I_g(x^*)\setminus I_h(x^*)} = 0$, na equação $\nabla_x \mathcal{L}(x^*, \lambda) = 0$ só há gradientes de restrições ativas linearmente independentes.

Teorema 2.1. *Seja x^* um ponto W-estacionário com vetor de multiplicadores λ associado. Se x^* satisfaz MPCC-LICQ então λ é único.*

Veremos adiante que MPCC-LICQ, assim como LICQ em programação não linear geral, é uma condição de qualificação muito exigente quando comparada às outras definidas na literatura. No entanto, mesmo sob MPCC-LICQ o conceito de W-estacionariedade pode ser muito fraco, como ilustra o próximo exemplo.

Exemplo 2.1. *Consideremos o MPCC bidimensional*

$$\min_x f(x) \quad \text{s.a} \quad x_1 \geqslant 0, \quad x_2 \geqslant 0, \quad x_1 x_2 \leqslant 0,$$

com $f : \mathbb{R}^2 \to \mathbb{R}$ diferenciável na origem, onde vale MPCC-LICQ. Fixada qualquer função objetivo f, a origem é ponto W-estacionário com multiplicadores $\lambda^{x_1} = f_{x_1}(0,0)$ e $\lambda^{x_2} = f_{x_2}(0,0)$. Em particular, para a escolha $f(x) = -x_1 - x_2$ a origem é maximizador global, e para $f(x) = x_1 - x_2$ um ponto de sela.

A fim de obtermos condições de estacionariedade mais fortes, restringimos os sinais dos multiplicadores λ^g_i e λ^h_i com índices no conjunto

$$I_0(x^*) = I_g(x^*) \cap I_h(x^*),$$

ou seja, índices onde as restrições complementares $g_i(x) \geqslant 0$ e $h_i(x) \geqslant 0$ são simultaneamente ativas em x^*. Dentre os conceitos estabelecidos na literatura, trabalharemos com outros três além da W-estacionariedade.

Definição 2.3. *Seja x^* um ponto W-estacionário com multiplicador associado $\lambda = (\lambda^G, \lambda^H, \lambda^g, \lambda^h) \in \mathbb{R}^s_+ \times \mathbb{R}^q \times \mathbb{R}^{2m}$.*

- *[135] Se $\lambda^g_i \cdot \lambda^h_i \geqslant 0$ para todo $i \in I_0(x^*)$ então dizemos que x^* é* Clarke estacionário *(C-estacionário);*
- *[118] Se para todo $i \in I_0(x^*)$ temos $\lambda^g_i \cdot \lambda^h_i = 0$ ou ($\lambda^g_i > 0$ e $\lambda^h_i > 0$) então dizemos que x^* é* Mordukhovich estacionário *(M-estacionário);*

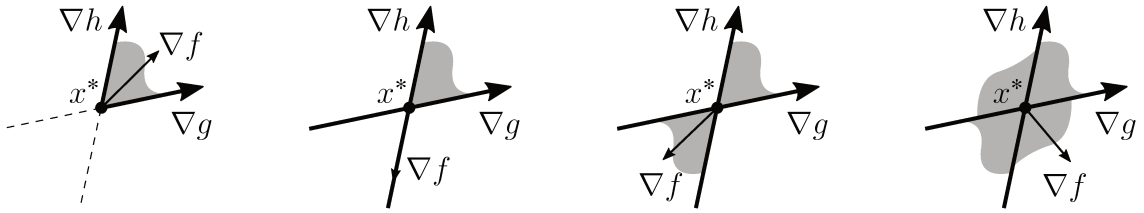

Figura 6 – Geometria dos conceitos de estacionariedade para (MPCC). As restrições $g(x), h(x) \geqslant 0$ são ativas em x^*, e a equação $\nabla_x \mathcal{L}(x^*, \lambda) = 0$ é equivalente a $\nabla f(x^*) = \lambda^g \nabla g(x^*) + \lambda^h \nabla h(x^*)$. Os sinais dos multiplicadores λ^g e λ^h determinam as regiões onde $\nabla f(x^*)$ pode pertencer. Conceitos mais exigentes resultam em regiões menores.

- *[135] Se $\lambda_i^g \geqslant 0$ e $\lambda_i^h \geqslant 0$ para todo $i \in I_0(x^*)$ então dizemos que x^* é* fortemente estacionário *(S-estacionário)*[3].

A Figura 6 ilustra a geometria dos diferentes conceitos de estacionariedade. Evidentemente,

$$\text{S-estac.} \Rightarrow \text{M-estac.} \Rightarrow \text{C-estac.} \Rightarrow \text{W-estac.}$$

As recíprocas não valem, como mostra o próximo exemplo.

Exemplo 2.2. *Consideremos o MPCC do Exemplo 2.1. A origem é ponto W, mas não C-estacionário se $f(x) = x_1 - x_2$; é C, mas não M-estacionário se $f(x) = -x_1 - x_2$; é M, mas não S-estacionário se $f(x) = -x_2$; e S-estacionário se $f(x) = x_1 + x_2$. O leitor pode facilmente verificar tais afirmações algebricamente, ou por ilustrações como as da Figura 6.*

As noções de estacionariedade apresentadas diferenciam-se apenas pelos sinais dos multiplicadores λ_i^g e λ_i^h com índices no conjunto $I_0(x^*)$. Como comentamos, MPCC-CQs podem ser obtidas pela simples exigência de uma CQ tradicional ao problema (TNLP). No entanto, é comum na literatura que os sinais dos multiplicadores com índices em $I_0(x^*)$ na definição das condições de qualificação sigam, quando possível, o padrão do conceito de estacionariedade. O objetivo é conceber MPCC-CQs menos exigentes, mas que ainda garantam que minimizadores sejam pontos estacionários de interesse. É o caso, por exemplo, de MPCC-RCPLD, definida em [70, 71]. Essa condição é inspirada na CQ de dependência linear positiva constante relaxada (RCPLD) (item 2 da Definição 1.2, página 20), e será de nosso interesse.

[3] A letra "S" refere-se à palavra inglesa *strongly*.

Definição 2.4 ([70]). *Sejam* $\mathcal{I}_H \subset \{1,\dots,q\}$, $\mathcal{I}_g \subset I_g(x^*)\backslash I_h(x^*)$ *e* $\mathcal{I}_h \subset I_h(x^*)\backslash I_g(x^*)$ *tais que*

$$\nabla H_i(x^*),\ i\in\mathcal{I}_H,\quad \nabla g_i(x^*),\ i\in\mathcal{I}_g,\quad \nabla h_i(x^*),\ i\in\mathcal{I}_h$$

formem uma base para o espaço

$$\operatorname{span}\{\nabla H_k(x^*),\nabla g_i(x^*),\nabla h_j(x^*)\mid k\in\{1,\dots,q\},\ i\in I_g(x^*)\backslash I_h(x^*),\ j\in I_h(x^*)\backslash I_g(x^*)\}.$$

Dizemos que um ponto viável x^* *satisfaz* condição de qualificação de dependência linear positiva constante relaxada para MPCC *(MPCC-RCPLD) se existe uma vizinhança* $\mathcal{U}(x^*)$ *de* x^* *tal que*

- *O conjunto*

  $$\operatorname{span}\{\nabla H_k(x),\nabla g_i(x),\nabla h_j(x)\mid k\in\{1,\dots,q\},\ i\in I_g(x^*)\backslash I_h(x^*),\ j\in I_h(x^*)\backslash I_g(x^*)\}.$$

  tem o mesmo posto para todo $x\in\mathcal{U}(x^*)$*;*

- *para cada* $\mathcal{I}_G\subset I_G(x^*)$ *e* $\mathcal{I}_g^0,\mathcal{I}_h^0\subset I_0(x^*)$*, sempre que existirem multiplicadores* $\lambda^G,\lambda^H,\lambda^g,\lambda^h$*, não todos nulos, satisfazendo*

  - ✓ $\lambda_i^G\geqslant 0$ *para todo* $i\in\mathcal{I}_G$*,*
  - ✓ $\lambda_i^h\lambda_i^g=0$ *ou* $\lambda_i^h,\lambda_i^g>0$ *para todo* $i\in I_0(x^*)$ *e*
  - ✓ $\displaystyle\sum_{i\in\mathcal{I}_G}\lambda_i^G\nabla G_i(x^*)+\sum_{i\in\mathcal{I}_H}\lambda_i^H\nabla H_i(x^*)-\sum_{i\in\mathcal{I}_g\cup\mathcal{I}_g^0}\lambda_i^g\nabla g_i(x^*)-\sum_{i\in\mathcal{I}_h\cup\mathcal{I}_h^0}\lambda_i^h\nabla h_i(x^*)=0,$

  então

  $$\nabla G_i(x),\ i\in\mathcal{I}_G,\quad \nabla H_i(x),\ i\in\mathcal{I}_H,\quad \nabla g_i(x),\ i\in\mathcal{I}_g\cup\mathcal{I}_g^0,\quad \nabla h_i(x),\ i\in\mathcal{I}_h\cup\mathcal{I}_h^0$$

  são linearmente dependentes para todo $x\in\mathcal{U}(x^*)$*.*

Como fizemos com as CQs da Seção 1.1, listamos a seguir algumas MPCC-CQs da literatura, indicando referências para o leitor interessado. As siglas seguem as CQs para programação não linear geral: MPCC-Linearidade [144]; MPCC-MFCQ [135], MPCC-CPLD [82, 145]; MPCC-CRCQ [138], MPCC-RCRCQ [71]; MPCC-Abadie [54]; MPCC-Guignard [58] e MPCC-CCP [130]. A Figura 7 traz a hierarquia entre as condições mencionadas. O leitor pode comparar com a Figura 1 (página 21) e observar que muitas das implicações são mantidas. É possível, no entanto, que hajam implicações em falta na Figura 7, pois até o momento elas não foram totalmente esclarecidas. Por exemplo, não sabemos se MPCC-RCPLD implica MPCC-Abadie. Cabe notarmos que MPCC-CPLD no sentido de [82] implica MPCC-Abadie, o que não está representado.

Na introdução deste trabalho, argumentamos que nenhum ponto viável de (MPCC) cumpre a CQ usual LICQ, ou mesmo CCP (Definição 1.2, página 20). Dissemos ainda que nem mesmo a CQ de Abadie pode ser esperada com frequência [56]. Para ilustrarmos, discutimos nos dois exemplos a seguir como Abadie e CCP podem falhar.

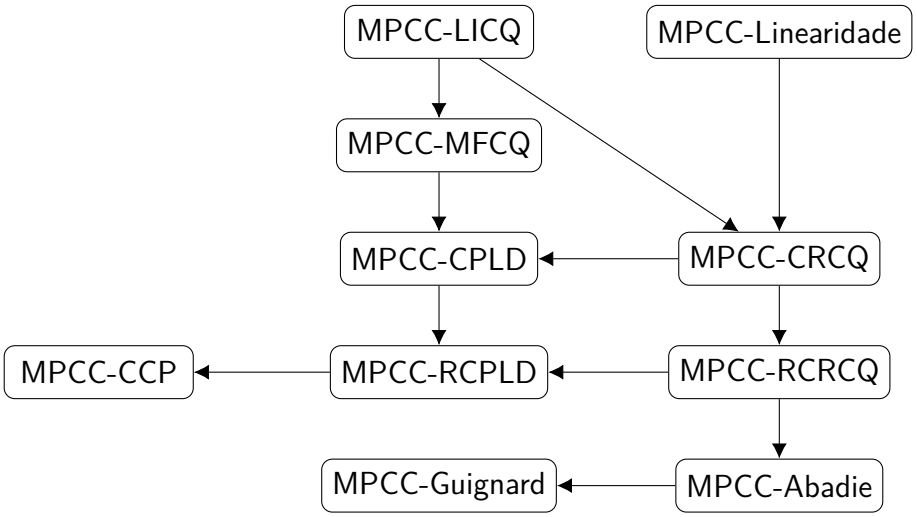

Figura 7 – Relações entre CQs para MPCC. As setas indicam implicações lógicas.

Exemplo 2.3. *O objetivo deste exemplo é ilustrar que a CQ de Abadie pode falhar em MPCCs simples. Retomemos o problema geral de programação não linear dado início do Capítulo 1:*

$$\min_x \; F(x) \quad \text{s.a} \quad G(x) \leqslant 0, \quad H(x) = 0. \tag{P}$$

Como discutido naquele capítulo, podemos escrever as condições KKT (Definição 1.1, página 19) na forma geométrica

$$-\nabla F(x^*) \in \mathcal{T}^{\text{lin}}(x^*)^\circ,$$

onde $\mathcal{T}^{\text{lin}}(x^*)^\circ$ é o polar do cone

$$\mathcal{T}^{\text{lin}}(x^*) = \left\{ d \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} d^T \nabla G_i(x^*) \leqslant 0, & \forall i \in I_G(x^*) \\ d^T \nabla H_i(x^*) = 0, & i = 1, \ldots, q \end{array} \right\}$$

(ou seja, $\mathcal{T}^{\text{lin}}(x^*)^\circ = \{z \in \mathbb{R}^n \mid z^T d \leqslant 0,\ \forall d \in \mathcal{T}^{\text{lin}}(x^*)\}$). O cone de direções tangentes ao conjunto viável de (P) no ponto x^* é composto pelas direções secantes limites partindo de x^*, cuja descrição algébrica é dada por

$$\mathcal{T}(x^*) = \left\{ d \in \mathbb{R}^n \;\middle|\; \exists x^k \text{ viável tal que } x^k \to x^* \text{ e } \frac{x^k - x^*}{\|x^k - x^*\|} \to \frac{d}{\|d\|} \right\} \cup \{0\}. \tag{2.2}$$

Finalmente, x^* cumpre a CQ de Abadie se [20, 28]

$$\mathcal{T}(x^*) = \mathcal{T}^{\text{lin}}(x^*).$$

Agora, consideremos o conjunto simples

$$X = \left\{ (x_1, x_2) \in \mathbb{R}^2 \mid x_1 \geqslant 0,\ x_2 \geqslant 0,\ x_1 x_2 \leqslant 0 \right\}, \tag{2.3}$$

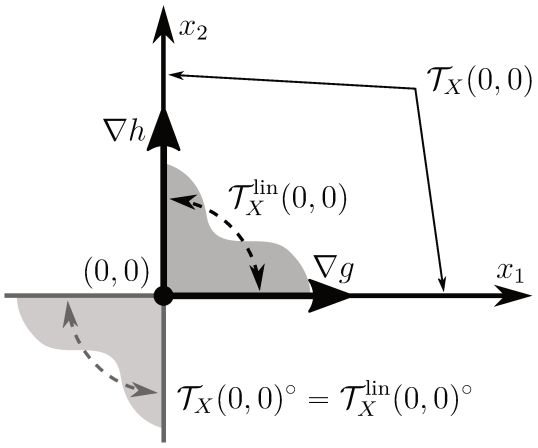

Figura 8 – Geometria para o Exemplo 2.3. As direções tangentes ao conjunto X partindo da origem ($\mathcal{T}_X(0,0)$) são aquelas na direção dos semieixos, enquanto que a linearização $\mathcal{T}_X^{\text{lin}}(0,0)$ é composta pelas direções em todo o primeiro quadrante. Em contrapartida, os cones polares de $\mathcal{T}_X(0,0)$ e $\mathcal{T}_X^{\text{lin}}(0,0)$ coincidem.

onde duas as inequações $x_1 \geqslant 0$ e $x_2 \geqslant 0$ são complementares entre si. Neste caso, temos

$$\mathcal{T}_X(0,0) = \{d \in \mathbb{R}^2_+ \mid d_1 d_2 = 0\} \quad \neq \quad \mathbb{R}^2_+ = \{d \in \mathbb{R}^2_+ \mid d_1 \cdot 0 + d_2 \cdot 0 \leqslant 0\} = \mathcal{T}_X^{\text{lin}}(0,0),$$

isto é, não vale a CQ de Abadie na origem. A Figura 8 ilustra esse fato geometricamente. Por outro lado, cabe comentar que o conjuntos polares de $\mathcal{T}_X(0,0)$ e $\mathcal{T}_X^{\text{lin}}(0,0)$ são iguais, a saber, ao conjunto $\{d \in \mathbb{R}^2 \mid d \leqslant 0\}$, exatamente no que consiste a CQ de Guignard [20]. Ou seja, Guignard é a única CQ usual que vale na origem.

Exemplo 2.4. *Neste exemplo mostramos, como dissemos na introdução, o fato da condição de qualificação CCP não valer em ponto algum do conjunto X dado por (2.3). O cone da Definição 1.2, item 3, torna-se*

$$\mathcal{K}(x_1,x_2) = \left\{ -\mu_1^G \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \mu_2^G \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \mu^0 \begin{bmatrix} x_2 \\ x_1 \end{bmatrix} \;\middle|\; \begin{array}{l} \mu \geqslant 0,\ \mu^0(x_1x_2) = 0, \\ \mu_1^G x_1 = \mu_2^G x_2 = 0 \end{array} \right\}.$$

Em particular, quando $x_1 = 0$ e $x_2 > 0$ temos

$$\mathcal{K}(0,x_2) = \left\{(-\mu_1^G + \mu^0 x_2, 0) \in \mathbb{R}^2 \mid \mu_1^G, \mu^0 \geqslant 0\right\} = \mathbb{R} \times \{0\},$$

e para $x_1 \neq 0$ e $x_2 > 0$,

$$\mathcal{K}(x_1,x_2) = \left\{(\mu^0 x_2, \mu^0 x_1) \in \mathbb{R}^2 \mid \mu^0 \geqslant 0\right\}.$$

Assim, nos pontos $(0, x_2^*) \in X$ com $x_2^* > 0$ temos

$$\limsup_{(x_1,x_2)\to(0,x_2^*)} \mathcal{K}(x_1,x_2) = \mathbb{R}_+ \times \mathbb{R} \quad \not\subset \quad \mathbb{R} \times \{0\} = \mathcal{K}(0,x_2^*),$$

e não vale a condição de qualificação CCP. De maneira análoga, CCP não vale nos pontos $(x_1^*, 0) \in X$ com $x_1^* > 0$. Finalmente, como a CQ de Abadie não vale em $(0,0)$

(Exemplo 2.3), CCP também não se verifica dado que é mais restritiva (veja a Figura 1, página 21).

Vimos no Capítulo 1 que as CQs tradicionais em programação não linear garantem que minimizadores locais sejam pontos KKT. Dado que a CQ de Abadie pode não ocorrer mesmo com restrições complementares muito simples, uma pergunta torna-se pertinente: quais noções de estacionariedade as CQs específicas para (MPCC) garantem que minimizadores locais atendam? No que segue, mostraremos que em certas situações, pontos KKT não podem ser esperados. Isso esclarece ainda mais o porquê de trabalharmos com conceitos de estacionariedade menos exigentes.

Em primeiro lugar, sabemos que S-estacionariedade é equivalente a KKT, como enunciado a seguir.

Teorema 2.2 ([56, Proposição 4.2], veja também [86, Proposição 1.1])**.** *Um ponto x^* é S-estacionário se, e somente se, é ponto KKT para* (MPCC)*.*

A condição de qualificação MPCC-LICQ (Definição 2.2, página 43) atesta que minimizadores sejam pontos S-estacionários (ou equivalentemente, pontos KKT). O teorema a seguir e seu corolário justificam tal afirmação.

Teorema 2.3 ([56, Teorema 4.6])**.** *Se um ponto x^*, viável para* (MPCC)*, satisfaz MPCC-LICQ então também satisfaz a CQ de Guignard.*

Corolário 2.1. *Seja x^* um minimizador local de* (MPCC)*. Se x^* satisfaz MPCC-LICQ então x^* é ponto S-estacionário.*

Demonstração. O resultado é consequência dos Teoremas 2.2 e 2.3 (lembremos que a CQ de Guignard garante que minimizadores locais de um problema de programação não linear qualquer sejam pontos KKT, veja a Seção 1.1, página 18). □

Infelizmente, nem a condição um pouco menos exigente MPCC-MCFQ ou mesmo MPCC-Linearidade (veja a Figura 7, página 47) garantem que minimizadores de (MPCC) sejam pontos S-estacionários[4], como mostra o exemplo a seguir.

Exemplo 2.5 ([135])**.** *Consideremos o MPCC "linear"*

$$\begin{aligned} \min_x \quad & f(x) = x_1 + x_2 - x_3 \\ \text{s.a} \quad & G_1(x) = -4x_1 + x_3 \leqslant 0, \quad G_2(x) = -4x_2 + x_3 \leqslant 0, \\ & g(x) = x_1 \geqslant 0, \quad h(x) = x_2 \geqslant 0, \quad x_1 x_2 \leqslant 0. \end{aligned}$$

[4] Há uma outra condição de qualificação, não citada neste trabalho, conhecida como MPCC-SMFCQ (Mangasarian-Fromovitz estrita para MPCC) que também garante S-estacionariedade em minimizadores. Ela é menos exigente que MPCC-LICQ, porém mais exigente que MPCC-MFCQ. Veja [135] para detalhes.

O ponto $x^* = (0,0,0)$ *é o único minimizador deste problema. De fato, na viabilidade temos* $x_1 = 0$ *ou* $x_2 = 0$*, e logo* $x_3 \leqslant 0$*. Assim, a função objetivo atinge seu mínimo em* x^**, onde vale zero. Além disso, como* $x_1 \geqslant 0$ *e* $x_2 \geqslant 0$*, o único ponto onde* f *se anula é* x^**.*

Como todas as restrições são lineares (exceto, claro, a complementariedade), vale MPCC-Linearidade. Ademais, MPCC-MFCQ consiste na exigência de MFCQ sobre as restrições do problema apertado TNLP [135] que, na origem, são dadas por

$$G_1(x) = -4x_1 + x_3 \leqslant 0, \quad G_2(x) = -4x_2 + x_3 \leqslant 0,$$
$$g(x) = x_1 = 0, \quad h(x) = x_2 = 0$$

(compare com o modelo (TNLP)*, página 42). MFCQ se verifica na origem* x^* *em relação a estas últimas restrições se*

$$\mu_1^G \nabla G_1(x^*) + \mu_2^G \nabla G_2(x^*) + \mu^g \nabla g(x^*) + \mu^h \nabla h(x^*) = 0, \quad \mu_1^G \geqslant 0, \quad \mu_2^G \geqslant 0$$

admitir somente a solução nula $\mu = 0$ *(consulte [85] para definição da condição de Mangasarian-Fromovitz). A expressão anterior fornece*

$$-4\mu_1^G + \mu^g = 0, \quad -4\mu_2^G + \mu^h = 0, \quad \mu_1^G + \mu_2^G = 0 \quad e \quad \mu_1^G, \mu_2^G \geqslant 0,$$

o que implica $\mu = 0$*. Ou seja,* x^* *cumpre MPCC-MFCQ em relação ao conjunto viável do MPCC linear dado. No entanto,* x^* *não é ponto S-estacionário: o sistema*

$$\mathcal{L}(x^*, \lambda) = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} + \lambda_1^G \begin{bmatrix} -4 \\ 0 \\ 1 \end{bmatrix} + \lambda_2^G \begin{bmatrix} 0 \\ -4 \\ 1 \end{bmatrix} - \lambda^g \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \lambda^h \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = 0, \quad \lambda^G \geqslant 0$$

fornece $\lambda^g = 1 - 4\lambda_1^G$ *(da* 1^a *equação),* $\lambda^h = 1 - 4\lambda_2^G$ *(da* 2^a *equação) e* $\lambda_1^G + \lambda_2^G = 1$ *(da* 3^a *equação). Somando as duas primeiras expressões e considerando a terceira, obtemos* $\lambda^g + \lambda^h = -2$*. Logo não pode ser* $\lambda^g \geqslant 0$ *e* $\lambda^h \geqslant 0$*.*

A discussão anterior mostra que S-estacionariedade (ou equivalentemente, KKT) é um conceito muito exigente em MPCCs, só pode ser esperado em geral sob MPCC-LICQ. Em particular, o Teorema 2.3 não pode ser estendido a nenhuma outra MPCC-CQ mencionada neste capítulo. Isso justifica e abre margem para considerarmos os conceitos de estacionariedade menos exigentes já enunciados. Em particular, pontos M-estacionários têm seu papel destacado. Como veremos nos próximos parágrafos, este conceito faz um papel próximo ao de KKT para programação não linear padrão, pois ocorre em minimizadores mediante MPCC-CQs pouco exigentes, além de ser um dos conceitos mais fortes conhecidos na literatura com tal propriedade.

A fim de analisarmos o papel da M-estacionariedade em MPCCs, pontuamos em primeiro lugar que, como era de se esperar, este conceito não constitui uma condição necessária de otimalidade legítima. O exemplo a seguir ilustra tal fato.

Exemplo 2.6. *Consideremos o MPCC*

$$\min_x x_1 + x_2 \quad \text{s.a} \quad x_1^3 \geqslant 0, \quad x_2^3 \geqslant 0, \quad x_1^3 x_2^3 \leqslant 0.$$

O ponto $x^* = (0, 0)$ *é minimizador global, mas não é ponto W-estacionário. De fato, a equação* $\mathcal{L}(x^*, \lambda) = 0$ *não é satisfeita para* λ *algum, dado que os gradientes das restrições são nulos na origem, mas o gradiente da função objetivo não. Em particular,* x^* *não é ponto C ou M-estacionário, e logo C e M-estacionariedade não são condições necessárias de otimalidade.*

Ou seja, assim como ocorre com KKT em programação não linear geral, são necessárias condições de qualificação das restrições para que minimizadores de (MPCC) sejam pontos W, C ou M-estacionários. Por outro lado, minimizadores são pontos M-estacionários mesmo sob as condições menos exigentes MPCC-Guignard e MPCC-CCP (veja a Figura 7, página 47). Cabe ressaltar que Guignard é a condição de qualificação menos exigente possível para KKT [67], e que MPCC-Guignard é uma adaptação sua a MPCCs. Por sua vez, MPCC-Guignard é a condição menos exigente para o chamado conceito de B-estacionariedade [70], mais rigoroso que os anteriormente citados.

Teorema 2.4 ([57, Teorema 3.1], [130, Corolário 3.8])**.** *Seja* x^* *um minimizador local de* (MPCC)*. Se* x^* *satisfaz MPCC-Guignard ou MPCC-CCP (e logo qualquer outra MPCC-CQ mencionada) então* x^* *é ponto M-estacionário. Ou seja,*

"M-estacionário ou não-(MPCC-CQ)"

é uma CNO para (MPCC)*.*

Como todo ponto M-estacionário também é W e C-estacionário, o resultado a seguir é direto.

Corolário 2.2. *Seja* x^* *um minimizador local de* (MPCC)*. Se* x^* *satisfaz MPCC-Guignard ou MPCC-CCP então* x^* *é ponto W/C-estacionário.*

O leitor pode verificar rapidamente que todos os conceitos de estacionariedade para (MPCC) são equivalentes se $I_0(x^*)$ é vazio (veja a Definição 2.3, página 44). Esta propriedade é conhecida como *complementariedade estrita do nível inferior*, ou simplesmente *complementariedade estrita*. Neste caso, o Teorema 2.2 (página 49) diz que todo ponto W, C, M ou S-estacionário é KKT, e logo não se justifica um estudo específico como fazemos neste trabalho. Nosso interesse é analisar a convergência de métodos de Lagrangiano aumentado do ponto de vista mais geral. Portanto, não levamos em consideração a hipótese de complementariedade estrita na teoria desenvolvida no Capítulo 3. Cabe notarmos que

sob esta hipótese, a convergência de ALGENCAN (Algoritmo 1, página 29) foi recentemente generalizada em [9].

Para finalizarmos esta seção, observamos que os conceitos de C e M-estacionariedade têm origem em condições de otimalidade obtidas com auxílio dos cálculos diferenciais de Clarke e Mordukhovich, respectivamente. A abordagem via sinais dos multiplicadores apresentada na Definição 2.3 (página 44) é adequada à análise de convergência de algoritmos, e é amplamente utilizada na literatura. A título de curiosidade, fornecemos a seguir uma justificativa geométrica para o MPCC bidimensional

$$\min_x f(x) \quad \text{s.a.} \quad g(x) = x_1 \geqslant 0, \quad h(x) = x_2 \geqslant 0, \quad g(x)h(x) = x_1x_2 \leqslant 0, \tag{2.4}$$

onde f é continuamente diferenciável (de certa forma, estas simples restrições contemplam o caso geral mediante inserção de folgas).

Seja X o conjunto viável do problema (2.4). O cone de direções tangentes a X em um ponto viável x, tal como em (2.2) (página 47), é dado por

$$\mathcal{T}_X(x) = \left\{ d \in \mathbb{R}^2 \;\middle|\; \begin{array}{ll} d_1 \in \mathbb{R}, \;\; d_2 = 0 & \text{se } x_1 > 0, \; x_2 = 0 \\ d_1 = 0, \;\; d_2 \in \mathbb{R} & \text{se } x_1 = 0, \; x_2 > 0 \\ d \geqslant 0, \;\; d_1 d_2 \leqslant 0 & \text{se } x_1 = x_2 = 0 \end{array} \right\}.$$

O cone de direções normais, ou *cone normal*, a X em um ponto $x \in X$ é definido por

$$\mathcal{N}_X(x) = \mathcal{T}_X(x)^\circ = \left\{ d \in \mathbb{R}^2 \;\middle|\; \begin{array}{ll} d_1 = 0, \;\; d_2 \in \mathbb{R} & \text{se } x_1 > 0, \; x_2 = 0 \\ d_1 \in \mathbb{R}, \;\; d_2 = 0 & \text{se } x_1 = 0, \; x_2 > 0 \\ d \leqslant 0 & \text{se } x_1 = x_2 = 0 \end{array} \right\}.$$

A Figura 9(a) ilustra geometricamente os cones $\mathcal{T}_X(x)$ e $\mathcal{N}_X(x)$. É possível mostrar que todo minimizador local x^* de (2.4) satisfaz

$$-\nabla f(x^*) \in \mathcal{N}_X(x^*). \tag{2.5}$$

Utilizando o cálculo de Mordukhovich [116], obtemos uma condição necessária de otimalidade de primeira ordem na forma (2.5) expandindo o cone normal $\mathcal{N}_X(x^*)$ a

$$\mathcal{N}_X^M(x^*) = \limsup_{x \to x^*} \mathcal{N}_X(x).$$

A grosso modo, o *cone normal de Mordukhovich* $\mathcal{N}_X^M(x^*)$ (também chamado *cone normal limite*) contém os limites de direções contidas nos cones *ao redor* de x^*. Em particular, $\mathcal{N}_X^M(0,0)$ contém os eixos cartesianos, obtidos como limites de direções contidas nos cones normais $\mathcal{N}_X(x_1, 0)$ e $\mathcal{N}_X(0, x_2)$ (veja a Figura 9(b)). Algebricamente, temos

$$\mathcal{N}_X^M(x) = \left\{ d \in \mathbb{R}^2 \;\middle|\; \begin{array}{ll} d_1 = 0, \;\; d_2 \in \mathbb{R} & \text{se } x_1 > 0, \; x_2 = 0 \\ d_1 \in \mathbb{R}, \;\; d_2 = 0 & \text{se } x_1 = 0, \; x_2 > 0 \\ d < 0 \text{ ou } d_1 d_2 = 0 & \text{se } x_1 = x_2 = 0 \end{array} \right\}.$$

Analogamente, o *cone normal de Clarke* é dado por

$$\mathcal{N}_X^W(x^*) = \operatorname{cl}\operatorname{conv}\left(\mathcal{N}_X^M(x^*)\right) = \operatorname{cl}\operatorname{conv}\left(\limsup_{x\to x^*}\mathcal{N}_X(x)\right)$$

(veja [133]), onde $\operatorname{cl}\operatorname{conv}(D)$ denota o fecho convexo do conjunto D. Temos

$$\mathcal{N}_X^W(x) = \left\{ d\in\mathbb{R}^2 \;\middle|\; \begin{array}{lll} d_1 = 0, & d_2\in\mathbb{R} & \text{se } x_1>0,\ x_2=0\\ d_1\in\mathbb{R}, & d_2 = 0 & \text{se } x_1=0,\ x_2>0\\ d_1\in\mathbb{R}, & d_2\in\mathbb{R} & \text{se } x_1=x_2=0 \end{array} \right\}$$

(o leitor pode imaginar o fecho convexo de $\mathcal{N}_X^M(x)$ pela Figura 9(b)). Scheel e Scholtes [135] mostraram, como consequência direta de uma condição de otimalidade devida ao próprio Clarke [46, Teorema 1], que em todo minimizador local x^* de (MPCC) a pertinência $-\nabla f(x^*)\in\mathcal{N}_X^W(x^*)$ é válida se ainda exigirmos $d_1d_2\geqslant 0$ na descrição de $\mathcal{N}_X^W(x^*)$. Isso nos leva a considerar o cone

$$\mathcal{N}_X^C(x) = \left\{ d\in\mathbb{R}^2 \;\middle|\; \begin{array}{lll} d_1 = 0, & d_2\in\mathbb{R} & \text{se } x_1>0,\ x_2=0\\ d_1\in\mathbb{R}, & d_2 = 0 & \text{se } x_1=0,\ x_2>0\\ d_1d_2\geqslant 0 & & \text{se } x_1=x_2=0 \end{array} \right\}.$$

Imediatamente,

$$\mathcal{N}_X(x)\subset\mathcal{N}_X^M(x)\subset\mathcal{N}_X^C(x)\subset\mathcal{N}_X^W(x)$$

para todo $x\in X$. Assim, (2.5) fornece as condições necessárias de otimalidade

$$-\nabla f(x^*)\in\mathcal{N}_X^W(x^*),\qquad -\nabla f(x^*)\in\mathcal{N}_X^C(x^*)\qquad\text{e}\qquad -\nabla f(x^*)\in\mathcal{N}_X^M(x^*). \tag{2.6}$$

Todas implicam a existência de um $d\in\mathbb{R}^2$ tal que

$$\nabla f(x^*) - (-d_1)\nabla g(x^*) - (-d_1)\nabla h(x^*) = 0.$$

Quando $x^* = (0,0)$, a primeira condição em (2.6) não impõe limitações em d; já a segunda condição diz que $(-d_1)(-d_2)\geqslant 0$, enquanto a terceira, $-d>0$ ou $(-d_1)(-d_2)=0$. Isso coincide com os conceitos de W, C e M-estacionariedade das Definições 2.1 e 2.3 (páginas 43 e 44). O leitor interessado neste ponto de vista pode consultar [41, 55, 116, 118, 133, 135].

É curioso que a chamada C-estacionariedade não proceda diretamente da computação do cone normal de Clarke, parte dos fundamentos de seu cálculo diferencial, assim como ocorre entre a M-estacionariedade e os cones normais de Mordukhovich. Aliás, os cones de Clarke fornecem, sim, pontos W-estacionários. O termo "C-estacionário" foi cunhado por Scheel e Scholtes [135], e justifica-se pelo uso de um resultado posterior do próprio Clarke, como mencionado.

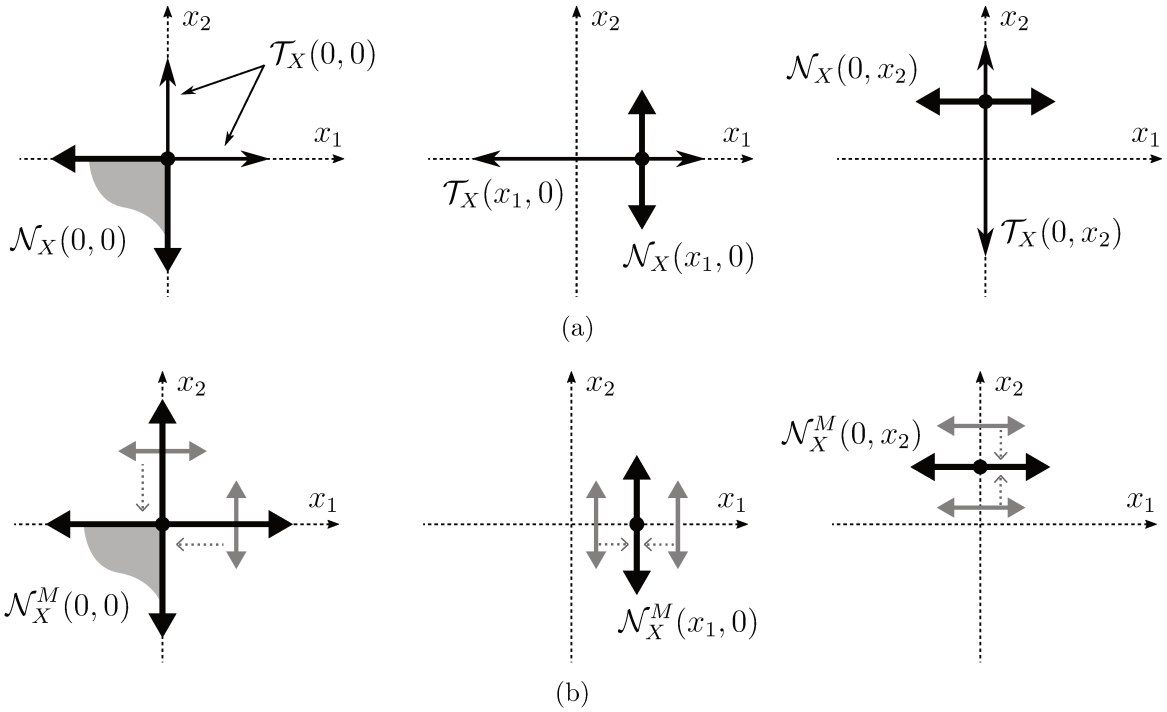

Figura 9 – Geometria do conceito de M-estacionariedade para o problema (2.4). Em (a), os cones tangente $\mathcal{T}_X(x)$ e normal $\mathcal{N}_X(x)$. Em (b), o cone $\mathcal{N}_X^M(0,0)$ contém os eixos cartesianos, limites de direções contidas nos cones normais ao redor da origem. Compare com a Figura 6 (página 45).

2.2 Notas sobre condições sequenciais de otimalidade

Na Seção 1.1 tratamos das condições de otimalidade para programação não linear padrão, onde KKT teve um papel central. Requeremos a existência de um vetor de multiplicadores μ^* que satisfaça (1.2) e (1.3) (Definição 1.1, página 19). Por outro lado, a existência de multiplicadores *para o ponto* x^* não é garantida: lembremos que no problema

$$\min_x x \quad \text{s.a} \quad x^2 = 0,$$

a origem é o único minimizador, e não é ponto KKT. Mas o que ocorre *próximo* à origem? Vejamos: a equação (1.2) relativa a este pequeno problema toma a forma

$$1 + 2\mu x = 0,$$

cuja solução em μ é $\mu = -1/(2x)$ se $x \neq 0$. Ou seja, em qualquer ponto arbitrariamente próximo à origem há multiplicador, que naturalmente explode quando $x \to 0$. Em outras palavras, dada uma sequência $\{x^k\} \subset \mathbb{R}^n \backslash \{0\}$ convergindo a zero, há uma sequência de multiplicadores $\{\mu^k\}$ que satisfazem KKT aproximadamente. Isso é possível em qualquer problema de programação matemática, e motiva a definição das chamadas *condições sequenciais de otimalidade.*

Condições sequenciais de otimalidade estão naturalmente associadas a algoritmos iterativos, e logo não fazem sentido sem eles. Em particular, Algencan (Algoritmo 1,

página 29) e ALGENCAN-SECOND (Algoritmo 3, página 34) geram sequências *Aproximadamente KKT* (AKKT) [10, 39] e *Aproximadamente KKT de segunda ordem* (AKKT2) [11]. Sob hipóteses bastante fracas, ALGENCAN gera ainda uma sequência mais exigente que AKKT, conhecida como CAKKT[5] [18]. Dizemos que o ponto viável x^* é CAKKT para o problema de programação não linear geral (P) (página 17) se existirem sequências $\{x^k\} \subset \mathbb{R}^n$, $\{\mu^{G,k}\} \subset \mathbb{R}^s_+$ e $\{\mu^{H,k}\} \subset \mathbb{R}^q$ tais que

$$\lim_{k\to\infty} x^k = x^*, \tag{2.7}$$

$$\lim_{k\to\infty} \|\nabla F(x^k) + \nabla G(x^k)\mu^{G,k} + \nabla H(x^k)\mu^{H,k}\| = 0, \tag{2.8}$$

$$\lim_{k\to\infty} \mu_i^{G,k} G_i(x^k) = 0,\ i = 1, \ldots, s, \tag{2.9}$$

$$\lim_{k\to\infty} \mu_i^{H,k} H_i(x^k) = 0,\ i = 1, \ldots, q. \tag{2.10}$$

As expressões (2.8) e (2.9) estão diretamente relacionadas às equações KKT da Definição 1.1. Já a expressão (2.10) é uma exigência extra de complementariedade para restrições de igualdade.

Sabemos que todo minimizador local de (P) é ponto CAKKT, ou seja, CAKKT é uma condição necessária de otimalidade independentemente de condições de qualificação [18, Teorema 3.3]. Sabemos também que CAKKT é uma condição sequencial forte, no sentido que todo ponto CAKKT atende a outras condições sequenciais de otimalidade de primeira ordem (a saber, AKKT e AGP [108]), e que sob a CQ CCP (Definição 1.2, página 20), garante pontos KKT (consulte [15, 18]). Nesse sentido, condições sequenciais de otimalidade já são realidade no estudo da convergência de algoritmos. No entanto, como problemas com restrições de complementariedade não satisfazem a maioria das CQs padrão, devemos nos questionar quais garantias as condições sequenciais tradicionais fornecem em MPCCs. Infelizmente, nem mesmo CAKKT garante mais do que W-estacionariedade, ainda sob MPCC-LICQ, como mostra o próximo exemplo.

Exemplo 2.7. *Consideremos o MPCC bidimensional*

$$\min_x\ x_1 - x_2 \quad \text{s.a} \quad x_1 \geqslant 0, \quad x_2 \geqslant 0, \quad x_1 x_2 \leqslant 0.$$

O ponto $x^* = (0, 0)$ é apenas W-estacionário e satisfaz MPCC-LICQ. O gradiente da função Lagrangiano usual $L(x, \mu)$ se anula quando

$$\begin{bmatrix} 1 \\ -1 \end{bmatrix} - \mu^{x_1,k} \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \mu^{x_2,k} \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \mu^{0,k} \begin{bmatrix} x_2^k \\ x_1^k \end{bmatrix} = 0.$$

As sequências definidas por $x^k = (1/k, -1/k)$, $\mu^{x_1,k} = \mu^{x_2,k} = 0$ e $\mu^{0,k} = k$ satisfazem a igualdade acima para todo $k \geqslant 1$. Ademais, satisfazem as expressões (2.7)–(2.10), isto é, x^* é CAKKT.

[5] Do inglês *Complementary Approximate KKT*.

No próximo capítulo mostraremos que Algencan converge a pontos C-estacionários sob MPCC-LICQ (na verdade, estendemos o resultado de [87]), evitando a origem no exemplo acima. Ou seja, apesar do método gerar sequências CAKKT neste exemplo, seus pontos limite atendem a propriedades adicionais. Essa lacuna entre uma sequência CAKKT qualquer e aquela gerada por Algencan motiva o estudo de condições sequenciais específicas para (MPCC), sempre, é claro, associadas a algoritmos. Como comentado na introdução, este é um terreno inexplorado até o momento, mas que não é abordado neste trabalho.

Em menor quantidade, condições sequenciais de segunda ordem foram propostas na literatura. Além da já mencionada AKKT2 [11], há ainda uma recente adaptação de CAKKT à segunda ordem denominada CAKKT2 [74]. No contexto de MPCCs, as mesmas questões levantadas nos parágrafos anteriores se colocam. É possível mostrar que no exemplo desta seção, a origem é ponto CAKKT2 (e logo AKKT2) usando as mesmas sequências. Assim, todas as condições sequenciais de nosso conhecimento não garantem mais do que W-estacionariedade. Por outro lado, no próximo capítulo mostraremos que Algencan-second converge a pontos M-estacionários sob MPCC-LICQ.

A especificidade de (MPCC), portanto, justifica o porquê de não tratarmos condições sequenciais neste trabalho. Em especial, apenas tangenciamos o tema nos resultados de convergência dos métodos de Lagrangiano aumentado do Capítulo 1, sem prejuízo ao entendimento do principal objetivo desta pesquisa, apresentado no próximo capítulo.

Capítulo 3

Convergência de métodos de Lagrangiano aumentado em MPCCs

Neste capítulo apresentamos os resultados de convergência dos métodos de Lagrangiano aumentado de primeira e segunda ordens, descritos no Capítulo 1, quando aplicados a problemas com restrições de complementariedade. Os conceitos trabalhados nos capítulos anteriores serão utilizados. Em particular, durante todo este capítulo citamos frequentemente Algencan (Algoritmo 1, página 29) e Algencan-second (Algoritmo 3, página 34).

Por simplicidade, omitiremos do modelo (MPCC) (página 41) as restrições ordinárias $G(x) \leqslant 0$ e $H(x) = 0$. Portanto, o problema analisado aqui é

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{s.a} \quad & g(x) \geqslant 0, \quad h(x) \geqslant 0, \\ & g(x)^T h(x) \leqslant 0. \end{aligned} \tag{MPCC}$$

Neste capítulo, vamos supor f, h e g duas vezes continuamente diferenciáveis. Como comentamos na Seção 1.3 (página 32), Algencan e Algencan-second tratam restrições de caixa separadamente, sem penalizá-las. Quando a caixa Ω está presente, as ε-Hessianas reduzida $\mathcal{H}_{[\mathcal{F},\varepsilon,\rho]}(x,\mu)$ e aproximada $\nabla^2_\varepsilon L_\rho(x,\mu)$ têm entradas relativas à variáveis fixas na face corrente diferentes. Assim como no capítulo anterior, omitiremos a caixa Ω por simplicidade, e logo $\mathcal{H}_{[\mathcal{F},\varepsilon,\rho]}(x,\mu) = \nabla^2_\varepsilon L_\rho(x,\mu)$. Isso facilitará as provas de convergência. Tanto a presença das restrições $G(x) \leqslant 0$ e $H(x) = 0$ quanto da caixa Ω não oferecem maiores dificuldades, e serão tratadas em separado nas Seções 3.1 e 3.2.

Izmailov, Solodov e Uskov [87] mostraram que, sob MPCC-LICQ, métodos de Lagrangiano aumentado de primeira ordem convergem a pontos pelo menos C-estacionários. Mais ainda, os autores mostraram que quando certa sequência dual associada à complementariedade é limitada, então há convergência a pontos S-estacionários sob MPCC-LICQ. Especificamente, os autores mostraram o seguinte:

Teorema 3.1 ([87, Teorema 3.2]). *Seja $\{x^k\}$ uma sequência gerada por* Algencan *e x^* um ponto limite viável seu, digamos, $\lim_{k\in K} x^k = x^*$. Suponhamos que x^* satisfaça MPCC-LICQ. Se a sequência de multiplicadores da complementariedade é tal que $\liminf_{k\in K}(\mu^{0,k} + \rho_k g(x^k)^T h(x^k)) < \infty$ então x^* é S-estacionário. Caso contrário, este ponto limite é pelo menos C-estacionário.*

No que segue, estendemos o Teorema 3.1 no primeiro caso, onde alguma subsequência de multiplicadores da complementariedade é limitada, mostrando que o mesmo vale sob a CQ menos exigente MPCC-RCPLD. Para tanto, aplicamos a mesma técnica utilizada em [12] para lidar com RCPLD em programação não linear geral. O próximo lema pode ser visto como consequência do conhecido Lema de Carathéodory.

Lema 3.1 ([12, Lema 1]). *Sejam I e J conjuntos finitos de índices e v^i, $i \in I \cup J$, vetores de $\mathbb{R}^n$. Suponhamos que $\{v^i\}_{i\in I}$ seja linearmente independente. Se*

$$w = \sum_{i\in I} \alpha_i v^i + \sum_{i\in J} \alpha_i v^i$$

com $\alpha_i \neq 0$ para todo $i \in J$, então existem um subconjunto $\hat{J} \subset J$ e escalares $\hat{\alpha}_i$, $i \in I \cup \hat{J}$, tais que

$$w = \sum_{i\in I} \hat{\alpha}_i v^i + \sum_{i\in \hat{J}} \hat{\alpha}_i v^i,$$

$\alpha_i \cdot \hat{\alpha}_i > 0$ para todo $i \in \hat{J}$, e $\{v^i\}_{i\in I\cup\hat{J}}$ é linearmente independente.

Lembramos que a função Lagrangiano usual para o problema (MPCC) deste capítulo toma a forma

$$L(x,\mu) = f(x) - (\mu^g)^T g(x) - (\mu^h)^T h(x) + \mu^0 g(x)^T h(x),$$

onde $\mu = (\mu^g, \mu^h, \mu^0) \in \mathbb{R}^{2m+1}_+$ (veja a expressão (2.1), página 41). Denotaremos as estimativas dos multiplicadores geradas pelos métodos por

$$\mu^k = (\mu^{g,k}, \mu^{h,k}, \mu^{0,k}).$$

Por simplicidade, denotaremos ainda

$$I_g = I_g(x^*), \quad I_h = I_h(x^*) \quad \text{e} \quad I_0 = I_0(x^*)$$

quando x^* estiver claro no contexto.

Teorema 3.2. *Seja $\{x^k\}$ uma sequência gerada por* Algencan *e x^* um ponto limite viável seu, digamos, $\lim_{k\in K} x^k = x^*$. Se a sequência de multiplicadores da complementariedade é tal que $\liminf_{k\in K}(\mu^{0,k} + \rho_k g(x^k)^T h(x^k)) < \infty$ e vale MPCC-RCPLD em x^*, então x^* é S-estacionário. Caso contrário, se x^* satisfaz MPCC-LICQ então este ponto limite é pelo menos C-estacionário.*

Demonstração. Vamos demonstrar apenas a primeira afirmação, a segunda segue do Teorema 3.1. Portanto assumimos que MPCC-RCPLD vale em x^* e que $\liminf_{k\in K}(\mu^{0,k} + \rho_k g(x^k)^T h(x^k)) < \infty$.

A sequência $\{x^k, \mu^k, \rho_k\}$ gerada por ALGENCAN satisfaz

$$\nabla_x L_{\rho_k}(x^k, \mu^k) = \nabla f(x^k) - \sum_{i=1}^{m} \left(\mu_i^{g,k} - \rho_k g_i(x^k)\right)_+ \nabla g_i(x^k) - \sum_{i=1}^{m} \left(\mu_i^{h,k} - \rho_k h_i(x^k)\right)_+ \nabla h_i(x^k) + \left(\mu^{0,k} + \rho_k g(x^k)^T h(x^k)\right)_+ v^k \to 0, \quad (3.1)$$

onde

$$v^k = \sum_{i=1}^{m} \left(\nabla g_i(x^k) h_i(x^k) + \nabla h_i(x^k) g_i(x^k)\right).$$

De maneira compacta, podemos escrever

$$\nabla f(x^k) - \sum_{i=1}^{m} \lambda_i^{g,k} \nabla g_i(x^k) - \sum_{i=1}^{m} \lambda_i^{h,k} \nabla h_i(x^k) \to 0, \quad (3.2)$$

onde

$$\lambda_i^{g,k} = \left(\mu_i^{g,k} - \rho_k g_i(x^k)\right)_+ - \lambda^{0,k} h_i(x^k), \quad (3.3)$$

$$\lambda_i^{h,k} = \left(\mu_i^{h,k} - \rho_k h_i(x^k)\right)_+ - \lambda^{0,k} g_i(x^k), \quad (3.4)$$

e $\lambda^{0,k} = (\mu^{0,k} + \rho_k g(x^k)^T h(x^k))_+$. Desta forma, a sequência dos multiplicadores da complementariedade em ALGENCAN é $\{\lambda^{0,k}\}$.

Se a sequência não decrescente dos parâmetros de penalização $\{\rho_k\}$ for limitada, então as sequências

$$\left\{\left(\mu^{g,k} - \rho_k g(x^k)\right)_+\right\}_{k\in K}, \left\{\left(\mu^{h,k} - \rho_k h(x^k)\right)_+\right\}_{k\in K} \text{ e } \left\{\left(\mu^{0,k} - \rho_k g(x^k)^T h(x^k)\right)_+\right\}_{k\in K}$$

também são limitadas (lembremos que $\{\mu^k\}$ pertence a um compacto em ALGENCAN, veja o Algoritmo 1, página 29). Assim, estas sequências possuem pontos limite $\mu^{g,*}$, $\mu^{h,*}$ e $\mu^{0,*}$ não negativos satisfazendo, por (3.1),

$$\nabla f(x^*) - \sum_{i=1}^{m} \mu_i^{g,*} \nabla g_i(x^*) - \sum_{i=1}^{m} \mu_i^{h,*} \nabla h_i(x^*) + \mu^{0,*} \sum_{i=1}^{m} \left(\nabla g_i(x^*) h_i(x^*) + \nabla h_i(x^*) g_i(x^*)\right) = 0.$$

Isto é, x^* é ponto KKT para (MPCC), e logo um ponto S-estacionário (Teorema 2.2, página 49).

Supomos daqui em diante que $\rho_k \to \infty$. Como $\liminf_{k\in K} \lambda^{0,k} < \infty$, existe um conjunto infinito de índices $K_1 \subset K$ tal que $\{\lambda^{0,k}\}_{k\in K_1}$ é limitada. Logo, (3.3) e (3.4)

implicam $\lim\limits_{k\in K_1} \lambda^{g,k}_{I_h\setminus I_g} = 0$ e $\lim\limits_{k\in K_1} \lambda^{h,k}_{I_g\setminus I_h} = 0$, e de (3.2) segue que

$$\lim_{k\in K_1} \left[\nabla f(x^k) - \sum_{i\in I_g} \lambda_i^{g,k}\nabla g_i(x^k) - \sum_{i\in I_h} \lambda_i^{h,k}\nabla h_i(x^k)\right] = 0. \tag{3.5}$$

Novamente por (3.3) e (3.4) temos

$$\lim_{k\in K_1} \lambda^{g,k}_{I_0} \geqslant 0 \quad \text{e} \quad \lim_{k\in K_1} \lambda^{h,k}_{I_0} \geqslant 0. \tag{3.6}$$

Essas expressões nos permitem perturbar $\lambda^{g,k}_{I_0}$ e $\lambda^{h,k}_{I_0}$ a fim de obtermos (3.5) com $\lambda^{g,k}_{I_0} \geqslant 0$ e $\lambda^{h,k}_{I_0} \geqslant 0$ para todo $k \in K_1$ suficientemente grande. Consideremos $\mathcal{I}_g \subset I_g\setminus I_h$ e $\mathcal{I}_h \subset I_h\setminus I_g$ como na definição de MPCC-RCPLD (Definição 2.4, página 45), isto é, tais que

$$\mathcal{B}(x^*) = \{\nabla g_i(x^*), \nabla h_j(x^*) \mid i \in \mathcal{I}_g,\ j \in \mathcal{I}_h\}$$

seja uma base para o espaço

$$\mathcal{S}(x^*) = \text{span}\,\{\nabla g_i(x^*), \nabla h_j(x^*) \mid i \in I_g\setminus I_h,\ j \in I_h\setminus I_g\}.$$

A primeira condição da definição de MPCC-RCPLD garante que

$$\mathcal{S}(x^k) = \left\{\nabla g_i(x^k), \nabla h_j(x^k) \mid i \in I_g\setminus I_h,\ j \in I_h\setminus I_g\right\}$$

tem o mesmo posto que $\mathcal{S}(x^*)$ para todo $k \in K_1$ suficientemente grande, digamos, para todo $k \in K_2 \subset K_1$. Assim, $\mathcal{B}(x^k)$ é uma base de $\mathcal{S}(x^k)$ para todo $k \in K_2$. Existem então sequências $\{\hat\lambda^{g,k}_{\mathcal{I}_g}\}_{k\in K_2}$ e $\{\hat\lambda^{h,k}_{\mathcal{I}_h}\}_{k\in K_2}$ tais que

$$\sum_{i\in I_g\setminus I_h} \lambda_i^{g,k}\nabla g_i(x^k) + \sum_{i\in I_h\setminus I_g} \lambda_i^{h,k}\nabla h_i(x^k) = \sum_{i\in\mathcal{I}_g} \hat\lambda_i^{g,k}\nabla g_i(x^k) + \sum_{i\in\mathcal{I}_h} \hat\lambda_i^{h,k}\nabla h_i(x^k),$$

para todo $k \in K_2$, e (3.5) fornece

$$\lim_{k\in K_2} \left[\nabla f(x^k) - \sum_{i\in\mathcal{I}_g} \hat\lambda_i^{g,k}\nabla g_i(x^k) - \sum_{i\in\mathcal{I}_h} \hat\lambda_i^{h,k}\nabla h_i(x^k) - \sum_{i\in I_0} \lambda_i^{g,k}\nabla g_i(x^k) - \sum_{i\in I_0} \lambda_i^{h,k}\nabla h_i(x^k)\right] = 0.$$

Os gradientes das duas primeiras somas são linearmente independentes e os multiplicadores das duas últimas somas são não negativos. Pelo Lema 3.1 (página 58), para cada $k \in K_2$ existem um conjunto $I_0^k \subset I_0$ e vetores $\hat\lambda^{g,k}_{I_0^k} \geqslant 0$, $\hat\lambda^{h,k}_{I_0^k} \geqslant 0$ tais que

$$\lim_{k\in K_2} \left[\nabla f(x^k) - \sum_{i\in\mathcal{I}_g} \hat\lambda_i^{g,k}\nabla g_i(x^k) - \sum_{i\in\mathcal{I}_h} \hat\lambda_i^{h,k}\nabla h_i(x^k) - \sum_{i\in I_0^k} \hat\lambda_i^{g,k}\nabla g_i(x^k) - \sum_{i\in I_0^k} \hat\lambda_i^{h,k}\nabla h_i(x^k)\right] = 0,$$

onde os gradientes das restrições envolvidas são linearmente independentes. Como há apenas um número finito de conjuntos I_0^k, existe $K_3 \subset K_2$ infinito e $\mathcal{I}_0 \subset I_0$ tais que

$$\lim_{k\in K_3}\left[\nabla f(x^k) - \sum_{i\in\mathcal{I}_g\cup\mathcal{I}_0}\hat{\lambda}_i^{g,k}\nabla g_i(x^k) - \sum_{i\in\mathcal{I}_h\cup\mathcal{I}_0}\hat{\lambda}_i^{h,k}\nabla h_i(x^k)\right] = 0, \tag{3.7}$$

onde todos os gradientes de restrições permanecem linearmente independentes. Mais ainda, $\hat{\lambda}_{\mathcal{I}_0}^{g,k} \geqslant 0$ e $\hat{\lambda}_{\mathcal{I}_0}^{h,k} \geqslant 0$ para todo $k \in K_3$.

Seja

$$S_k = \max\left\{\|\hat{\lambda}_{\mathcal{I}_g\cup\mathcal{I}_0}^{g,k}\|_\infty, \|\hat{\lambda}_{\mathcal{I}_h\cup\mathcal{I}_0}^{h,k}\|_\infty\right\}.$$

Se $\liminf_{k\in K_3} S_k < \infty$ então há um conjunto infinito $K_4 \subset K_3$ tal que

$$\lim_{k\in K_4}\hat{\lambda}_{\mathcal{I}_g\cup\mathcal{I}_0}^{g,k} = \hat{\lambda}_{\mathcal{I}_g\cup\mathcal{I}_0}^{g} \quad \text{e} \quad \lim_{k\in K_4}\hat{\lambda}_{\mathcal{I}_h\cup\mathcal{I}_0}^{h,k} = \hat{\lambda}_{\mathcal{I}_h\cup\mathcal{I}_0}^{h}.$$

Como $\hat{\lambda}_{\mathcal{I}_0}^{g,k} \geqslant 0$ e $\hat{\lambda}_{\mathcal{I}_0}^{h,k} \geqslant 0$ para todo $k \in K_4$, tomando $\hat{\lambda}_{I_0\setminus\mathcal{I}_0}^{g,k} = \hat{\lambda}_{I_0\setminus\mathcal{I}_0}^{h,k} = 0$ concluímos que x^* é S-estacionário.

Supomos agora que $\lim_{k\in K_3} S_k = \infty$. Dividindo (3.7) por S_k obtemos

$$\lim_{k\in K_3}\left[\sum_{i\in\mathcal{I}_g\cup\mathcal{I}_0}\frac{\hat{\lambda}_i^{g,k}}{S_k}\nabla g_i(x^k) + \sum_{i\in\mathcal{I}_h\cup\mathcal{I}_0}\frac{\hat{\lambda}_i^{h,k}}{S_k}\nabla h_i(x^k)\right] = 0.$$

Todas as sequências $\hat{\lambda}_i^{g,k}/S_k$ e $\hat{\lambda}_i^{h,k}/S_k$ são limitadas, e portanto admitem subsequências convergentes com índices em um conjunto infinito $K_5 \subset K_3$. Pela definição de S_k, um desses pontos limite é igual a 1, e assim os gradientes

$$\nabla g_i(x^*), i \in \mathcal{I}_g \cup \mathcal{I}_0, \quad \nabla h_i(x^*), i \in \mathcal{I}_h \cup \mathcal{I}_0,$$

são linearmente dependentes. Ademais, $\hat{\lambda}_{\mathcal{I}_0}^{g,k}/S_k \geqslant 0$ e $\hat{\lambda}_{\mathcal{I}_0}^{h,k}/S_k \geqslant 0$ para todo $k \in K_5$. Tomando $\lambda_{I_0\setminus\mathcal{I}_0}^{g} = \lambda_{I_0\setminus\mathcal{I}_0}^{h} = 0$ na definição MPCC-RCPLD, contrariamos seu segundo item, o que completa a prova. □

Observação 3.1. *Restrições ordinárias e de caixa não oferecem maiores dificuldades. A adaptação do Teorema 3.2 ainda segue a estratégia adotada em [12].*

O resultado anterior não pode ser estendido a M-estacionários sem hipóteses adicionais. De fato, o próximo exemplo mostra que Algencan pode convergir a pontos não M-estacionários mesmo sob MPCC-LICQ.

Exemplo 3.1. *Consideremos o MPCC bidimensional*

$$\min_x -x_1 - x_2 \quad \text{s.a} \quad x_1 \geqslant 0, \quad x_2 \geqslant 0, \quad x_1x_2 \leqslant 0.$$

Mostremos que Algencan *pode convergir à origem, um ponto C, mas não M-estacionário, e que cumpre MPCC-LICQ. De fato, consideremos as sequências* $\{x^k = (1/k, 1/k)\}$ *e* $\{\rho_k = k^3\}$. *Temos*

$$\nabla_x L_{\rho_k}(x^k, \mu^k) = \begin{bmatrix} -1 \\ -1 \end{bmatrix} - (\mu^{x_1,k} - k^2)_+ \begin{bmatrix} 1 \\ 0 \end{bmatrix} - (\mu^{x_2,k} - k^2)_+ \begin{bmatrix} 0 \\ 1 \end{bmatrix} + (\mu^{0,k} + k)_+ \begin{bmatrix} 1/k \\ 1/k \end{bmatrix}.$$

Pela limitação da sequência $\{\mu^k\}$ *em* Algencan, *temos* $(\mu^{x_1,k} - k^2)_+ = (\mu^{x_2,k} - k^2)_+ = 0$ *e* $(\mu^{0,k} + k)_+ = \mu^{0,k} + k$ *para todo* k *suficientemente grande. Assim, para estes índices* k,

$$\nabla_x L_{\rho_k}(x^k, \mu^k) = \frac{\mu^{0,k}}{k} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \to 0.$$

Consideramos, sem prejuízo, que a caixa $\Omega = [0, 1]^2$ *está presente para concluir que*

$$\mathcal{G}_P(x^k) = P_\Omega(x^k - \nabla_x L_{\rho_k}(x^k, \mu^k)) - x^k = -\nabla_x L_{\rho_k}(x^k, \mu^k) \to 0,$$

ou seja, Algencan *pode convergir ao ponto* $(0, 0)$.

Prosseguimos agora com a análise da convergência de Algencan-second em (MPCC). No método, consideramos o subproblema

$$\min_x \ f(x) + \frac{\rho_k}{2} \left\{ \left\| \left(\frac{\mu^{g,k}}{\rho_k} - g(x) \right)_+ \right\|^2 + \left\| \left(\frac{\mu^{h,k}}{\rho_k} - h(x) \right)_+ \right\|^2 + \left[\left(\frac{\mu^{0,k}}{\rho_k} + g(x)^T h(x) \right)_+ \right]^2 \right\}.$$

Algencan-second gera sequências $\{x^k\} \subset \mathbb{R}^n$, $\left\{\left(\mu^{g,k}, \mu^{h,k}, \mu^{0,k}\right)\right\} \subset \mathbb{R}^m_+ \times \mathbb{R}^m_+ \times \mathbb{R}_+$, $\{\rho_k\} \subset \mathbb{R}_+$ e $\{\varepsilon_k^{\text{fun}}\} \subset \mathbb{R}_+$ satisfazendo as seguintes condições de primeira e segunda ordens (por simplicidade omitiremos o termo "(x^k)" quando conveniente):

Condição de primeira ordem:

$$\begin{aligned} \nabla_x L_{\rho_k}(x^k, \mu^k) = \ \nabla f - \sum_{i=1}^m \left(\mu_i^{g,k} - \rho_k g_i \right)_+ \nabla g_i - \sum_{i=1}^m \left(\mu_i^{h,k} - \rho_k h_i \right)_+ \nabla h_i \\ + (\mu^{0,k} + \rho_k g^T h)_+ \sum_{i=1}^m (g_i \nabla h_i + h_i \nabla g_i) \to 0. \end{aligned} \tag{3.8}$$

Condição de segunda ordem:

A sequência dos menores autovalores das matrizes $\mathcal{H}^k = \mathcal{H}_{[\mathcal{F}(x^k), \varepsilon_k^{\text{fun}}, \rho_k]}$ tende a ficar não negativa. A ε-Hessiana aproximada toma a forma

$$\begin{aligned} \nabla^2_{\varepsilon_k^{\text{fun}}} L_{\rho_k}(x^k, \mu^k) = \nabla^2 f \ - \sum_{i=1}^m \left[(\mu_i^{g,k} - \rho_k g_i)_+ - (\mu^{0,k} + \rho_k g^T h)_+ h_i \right] \nabla^2 g_i \\ - \sum_{i=1}^m \left[(\mu_i^{h,k} - \rho_k h_i)_+ - (\mu^{0,k} + \rho_k g^T h)_+ g_i \right] \nabla^2 h_i \end{aligned} \tag{3.9}$$

$$
\begin{aligned}
&+ \rho_k \sum_{i \in I_g^{\varepsilon,k}} \nabla g_i \nabla g_i^T + \rho_k \sum_{i \in I_h^{\varepsilon,k}} \nabla h_i \nabla h_i^T \\
&+ (\mu^{0,k} + \rho_k g^T h)_+ \left[\sum_{i=1}^{m} (\nabla h_i \nabla g_i^T + \nabla g_i \nabla h_i^T) \right] \\
&+ z_k \rho_k \left[\sum_{i=1}^{m} (g_i \nabla h_i + h_i \nabla g_i) \right] \left[\sum_{i=1}^{m} (g_i \nabla h_i + h_i \nabla g_i) \right]^T .
\end{aligned}
$$

onde

$$
I_r^{\varepsilon,k} = \left\{ i \;\middle|\; \frac{1}{\sqrt{\rho_k}} \left(\mu_i^{r,k} - \rho_k r_i(x^k) \right) \geqslant -\varepsilon_k^{\text{fun}} \right\}
$$

para $r = g, h$, e

$$
z_k = \begin{cases} 1 & \text{se } \dfrac{1}{\sqrt{\rho_k}} \left(\mu^{0,k} + \rho_k g(x^k)^T h(x^k) \right) \geqslant -\varepsilon_k^{\text{fun}} \\ 0 & \text{se } \dfrac{1}{\sqrt{\rho_k}} \left(\mu^{0,k} + \rho_k g(x^k)^T h(x^k) \right) < -\varepsilon_k^{\text{fun}} \end{cases} .
$$

Lembrando que, como estamos supondo que não há restrições de caixa, segue que $\mathcal{H}^k = \nabla^2_{\varepsilon_k^{\text{fun}}} L_{\rho_k}(x^k, \mu^k)$.

O teorema a seguir é o principal resultado de nosso trabalho, e trata da convergência de Algencan-second em MPCCs. Nele utilizamos a notação de "grande O", comum em complexidade computacional. Especificamente, sejam $\{a_k\}$ e $\{b_k\}$ sequências de números reais não negativos. Escrevemos $a_k = O(b_k)$ para indicar que a sequência $\{b_k\}$ limita superiormente o crescimento de a_k por um fator linear. Em outras palavras, $a_k = O(b_k)$ significa que existe uma constante $M > 0$ tal que $a_k \leqslant M \cdot b_k$ para todo k suficientemente grande.

Teorema 3.3. *Seja $\{x^k\}$ uma sequência gerada por* Algencan-second *e x^* um ponto limite viável seu, digamos, $\lim_{k \in K} x^k = x^*$. Se a sequência de multiplicadores da complementariedade é tal que $\liminf_{k \in K} (\mu^{0,k} + \rho_k g(x^k)^T h(x^k)) < \infty$ e vale MPCC-RCPLD em x^*, então x^* é S-estacionário. Caso contrário, se x^* satisfaz MPCC-LICQ e $\epsilon_k^{fun} = O(1/\sqrt{\rho_k})$ então este ponto limite é pelo menos M-estacionário.*

Demonstração. A primeira afirmação é consequência direta do Teorema 3.2. Mostremos a segunda afirmação. Portanto, ao longo da demonstração estamos supondo que x^* satisfaz MPCC-LICQ e que

$$
\lim_{k \in K} (\mu^{0,k} + \rho_k g(x^k)^T h(x^k)) = \infty. \tag{3.10}
$$

Em particular, como a sequência $\{\mu^{0,k}\}$ computada por Algencan-second é limitada, temos $\rho_k \to \infty$.

Os iterandos de Algencan-second satisfazem

$$
\lim_{k \in K} \left[\nabla f(x^k) - \sum_{i=1}^{m} \lambda_i^{g,k} \nabla g_i(x^k) - \sum_{i=1}^{m} \lambda_i^{h,k} \nabla h_i(x^k) \right] = 0 \tag{3.11}
$$

onde

$$\begin{aligned}\lambda_i^{g,k} &= \left(\mu_i^{g,k} - \rho_k g_i(x^k)\right)_+ - \lambda^{0,k} h_i(x^k),\\ \lambda_i^{h,k} &= \left(\mu_i^{h,k} - \rho_k h_i(x^k)\right)_+ - \lambda^{0,k} g_i(x^k) \text{ e}\\ \lambda^{0,k} &= \left(\mu^{0,k} + \rho_k g(x^k)^T h(x^k)\right)_+.\end{aligned}$$

Seguindo a prova do Teorema 3.1 fornecida em [87], e passando a uma subsequência se necessário, podemos assumir que $\{\lambda^{g,k}\}_{k\in K}$ e $\{\lambda^{h,k}\}_{k\in K}$ convergem a $\overline{\lambda}^g$ e $\overline{\lambda}^h$ com $\overline{\lambda}^g_{I_h\setminus I_g} = 0$, $\overline{\lambda}^h_{I_g\setminus I_h} = 0$ e $\overline{\lambda}^g_i\overline{\lambda}^h_i \geqslant 0$ para todo $i \in I_0$. Isto é, x^* é ponto C-estacionário com

$$\nabla f(x^*) - \sum_{i\in I_g} \overline{\lambda}^g_i \nabla g_i(x^*) - \sum_{i\in I_h} \overline{\lambda}^h_i \nabla h_i(x^*) = 0. \tag{3.12}$$

MPCC-LICQ em x^* garante a unicidade de $\overline{\lambda}^g_{I_g}$ e $\overline{\lambda}^h_{I_h}$ na expressão anterior (Teorema 2.1, página 44). A prova até aqui depende somente da primeira ordem. Nosso objetivo agora é mostrar que as propriedades de segunda ordem de Algencan-second garantem que x^* é um ponto M-estacionário.

Suponhamos por contradição que x^* não seja M-estacionário. Portanto, existe um índice $j \in I_0$ tal que $\overline{\lambda}^g_j\overline{\lambda}^h_j \neq 0$ e ($\overline{\lambda}^g_j < 0$ ou $\overline{\lambda}^h_j < 0$). Como $\overline{\lambda}^g_j\overline{\lambda}^h_j \geqslant 0$, temos $\overline{\lambda}^g_j < 0$ e $\overline{\lambda}^h_j < 0$. Daí, para todo $k \in K$ suficientemente grande, temos

$$\lambda_j^{g,k} = (\mu_j^{g,k} - \rho_k g_j)_+ - (\mu^{0,k} + \rho_k g^T h)_+ h_j \leqslant -\delta \tag{3.13}$$

e

$$\lambda_j^{h,k} = (\mu_j^{h,k} - \rho_k h_j)_+ - (\mu^{0,k} + \rho_k g^T h)_+ g_j \leqslant -\delta \tag{3.14}$$

para um certo $\delta > 0$ fixo.

Afirmamos que

$$\mu_j^{g,k} - \rho_k g_j \to -\infty \quad \text{e} \quad \mu_j^{h,k} - \rho_k h_j \to -\infty.$$

De fato, se $\mu_j^{g,k} - \rho_k g_j \to \infty$ então $g_j < 0$ e (3.13) fornece

$$(\mu^{0,k} + \rho_k g^T h)_+ h_j \to \infty.$$

Por (3.14) só poderia ser $(\mu^{0,k} + \rho_k g^T h)_+ \to -\infty$, o que é impossível. Agora, se $\mu_j^{g,k} - \rho_k g_j \to a \in \mathbb{R}$ então

$$(\mu^{0,k} + \rho_k g^T h)_+ h_j \to b = (a)_+ - \overline{\lambda}^g_j > 0.$$

Como $\mu^{0,k} h_j \to 0$ obtemos $(g^T h)\rho_k h_j \to b > 0$, que implica $|\rho_k h_j| \to \infty$ pois $g^T h \to 0$. No caso em que $\rho_k h_j \to -\infty$ temos $g^T h < 0$, donde segue a limitação de $\{(\mu^{0,k} + \rho_k g^T h)_+\}$, contradizendo (3.10). Por outro lado, se $\rho_k h_j \to \infty$ então $g^T h > 0$ e $(\mu_j^{h,k} - \rho_k h_j)_+ \to 0$. Por (3.14) segue que

$$(\mu^{0,k} + \rho_k g^T h) g_j = (\mu^{0,k} + \rho_k g^T h)_+ g_j \to -\overline{\lambda}^h_j > 0,$$

que fornece $\rho_k g_j \to \infty$. Então $\mu_j^{g,k} - \rho_k g_j \to -\infty$, contradizendo a convergência ao escalar a. Concluímos portanto que $\mu_j^{g,k} - \rho_k g_j \to -\infty$, como queríamos. Esta análise é válida para todo $k \in K$ suficientemente grande, tomando sucessivas subsequências se necessário. De maneira análoga provamos também que $\mu_j^{h,k} - \rho_k h_j \to -\infty$.

Escrevemos

$$v^k = \sum_{i=1}^{m} \left(g_i(x^k)\nabla h_i(x^k) + h_i(x^k)\nabla g_i(x^k) \right)$$

e para cada $k \in K$ tomamos um vetor unitário arbitrário

$$d^k \in \text{span}\{v^k, \nabla g_{I_g\setminus\{j\}}(x^k), \nabla h_{I_h\setminus\{j\}}(x^k)\}^{\perp}. \tag{3.15}$$

Tal d^k existe pois x^* cumpre MPCC-LICQ e portanto não há mais do que $n-1$ vetores no conjunto de geradores deste subespaço. Multiplicando (3.8) (página 62) por d^k obtemos

$$\nabla f^T d^k - \sum_{i=1}^{m} \left(\mu_i^{g,k} - \rho_k g_i \right)_+ \nabla g_i^T d^k - \sum_{i=1}^{m} \left(\mu_i^{h,k} - \rho_k h_i \right)_+ \nabla h_i^T d^k \to 0. \tag{3.16}$$

Para todo $k \in K$ suficientemente grande temos

- $\nabla g_i(x^k)^T d^k = 0, \forall i \in I_g\setminus\{j\}$, pela definição de d^k;
- $\left(\mu_i^{g,k} - \rho_k g_i(x^k) \right)_+ = 0, \forall i \notin I_g$, pela viabilidade de x^*;
- $\left(\mu_j^{g,k} - \rho_k g_j(x^k) \right)_+ = 0$ pois $\mu_j^{g,k} - \rho_k g_j(x^k) \to -\infty$.

O mesmo ocorre para h, e portanto (3.16) fornece $\nabla f(x^k)^T d^k \to 0$. Como $\{d^k\}_{k\in K}$ é uma sequência de vetores unitários, existe um conjunto infinito de índices $K_1 \subset K$ tal que $\lim\limits_{k\in K_1} d^k = d$. Pela continuidade dos gradientes de f, g e h, segue que

$$d \in S = \text{span}\{\nabla f(x^*), \nabla g_{I_g\setminus\{j\}}(x^*), \nabla h_{I_h\setminus\{j\}}(x^*)\}^{\perp}.$$

Notemos que, fixado *a priori* $d \in S$, podemos escolher uma sequência $\{d^k\}_{k\in K_1}$ convergindo a d, e que ainda satisfaça (3.15) para todo $k \in K_1$.

Como x^* satisfaz MPCC-LICQ, podemos supor que $\nabla h_j(x^*)^T d \neq 0$. De fato, se $\nabla h_j(x^*)^T d = 0$ para todo $d \in S$, isto é, se

$$\nabla h_j(x^*) \in S^{\perp} = \text{span}\{\nabla f(x^*), \nabla g_{I_g\setminus\{j\}}(x^*), \nabla h_{I_h\setminus\{j\}}(x^*)\},$$

então podemos escrever

$$\nabla h_j(x^*) = \alpha\nabla f(x^*) + \sum_{i\in I_g\setminus\{j\}} \alpha_i^g \nabla g_i(x^*) + \sum_{i\in I_h\setminus\{j\}} \alpha_i^h \nabla h_i(x^*) + 0\cdot\nabla g_j(x^*).$$

Temos $\alpha \neq 0$ pois, como vale MPCC-LICQ em x^*, os gradientes das restrições envolvidas são linearmente independentes. Dividindo esta expressão por α obtemos $\nabla f(x^*)$ como combinação linear dos gradientes de restrições ativas, onde o fator de $\nabla g_j(x^*)$ é nulo. Mas isto contraria a unicidade dos multiplicadores na escrita (3.12) (página 64), dado que naquela expressão $\overline{\lambda}_j^g < 0$.

Multiplicando (3.12) por d obtemos

$$\nabla g_j(x^*)^T d = -\frac{\overline{\lambda}_j^h}{\overline{\lambda}_j^g} \nabla h_j(x^*)^T d \neq 0.$$

Como $\overline{\lambda}_j^g \leqslant -\delta$ e $\overline{\lambda}_j^h \leqslant -\delta$, temos $\overline{\lambda}_j^h / \overline{\lambda}_j^g > 0$, e logo

$$d^T \left[\nabla h_j(x^*) \nabla g_j(x^*)^T + \nabla g_j(x^*) \nabla h_j(x^*)^T \right] d = 2(\nabla g_j(x^*)^T d)(\nabla h_j(x^*)^T d) < 0. \quad (3.17)$$

Multiplicando (3.9) (página 62) por d_k, para $k \in K_1$, e lembrando que $I_g \cup I_h = \{1, \ldots, m\}$, obtemos

$$\begin{aligned} d_k^T \nabla^2_{\varepsilon_k} L_{\rho_k} d_k = & \underbrace{d_k^T \left[\nabla^2 f(x^k) - \sum_{i=1}^m \lambda_i^{g,k} \nabla^2 g_i(x^k) - \sum_{i=1}^m \lambda_i^{h,k} \nabla^2 h_i(x^k) \right] d_k}_{A} \\ & \underbrace{+ \rho_k \sum_{i \in I_g^{\varepsilon,k}} \left(\nabla g_i(x^k)^T d_k \right)^2 + \rho_k \sum_{i \in I_h^{\varepsilon,k}} \left(\nabla h_i(x^k)^T d_k \right)^2}_{B} \\ & + \underbrace{\left(\mu^{0,k} + \rho_k g(x^k)^T h(x^k) \right)_+ \left[d_k^T \left(\nabla h_j(x^k) \nabla g_j(x^k)^T + \nabla g_j(x^k) \nabla h_j(x^k)^T \right) d_k \right]}_{C}. \end{aligned} \quad (3.18)$$

O termo A é limitado pois $\lim_{k \in K_1} \lambda^g = \overline{\lambda}^g$ e $\lim_{k \in K_1} \lambda^h = \overline{\lambda}^h$. Como $\lim_{k \in K_1} (\mu_j^{g,k} - \rho_k g_j) = \lim_{k \in K_1} (\mu_j^{h,k} - \rho_k h_j) = -\infty$, concluímos que $j \notin I_g^{\varepsilon,k} \cup I_h^{\varepsilon,k}$ para todo $k \in K_1$, desde de que ϵ_k^{fun} seja pequeno o suficiente para garantir que

$$0 < \varepsilon_k^{\text{fun}} < -\max \left\{ \frac{1}{\sqrt{\rho_k}} \left(\mu_j^{g,k} - \rho_k g_j(x^k) \right), \frac{1}{\sqrt{\rho_k}} \left(\mu_j^{h,k} - \rho_k h_j(x^k) \right) \right\}. \quad (3.19)$$

Ademais, temos, para todo $k \in K_1$ suficientemente grande e $\gamma > 0$ fixo, $\mu_j^{g,k} - \rho_k g_j \leqslant -\gamma$ e $\mu_j^{h,k} - \rho_k h_j \leqslant -\gamma$. Portanto, a condição algorítmica $\varepsilon_k^{\text{fun}} = O(1/\sqrt{\rho_k})$ é suficiente para garantir a validade de (3.19). Agora, para todo $i \notin I_g$, se $k \in K_1$ é suficientemente grande então $g_j(x^k)$ é maior que um certo patamar positivo fixado (por exemplo, $g_j(x^*)/2$). Logo, como $\rho_k \to \infty$, segue que

$$\lim_{k \in K_1} \frac{1}{\sqrt{\rho_k}} \left(\mu_j^{g,k} - \rho_k g_j(x^k) \right) = \lim_{k \in K_1} \left(\frac{\mu_j^{g,k}}{\sqrt{\rho_k}} - \sqrt{\rho_k} g_j(x^k) \right) = -\infty.$$

Também, $(I_h \backslash I_g) \cap I_g^{\varepsilon,k} = \varnothing$ para estes índices k. Analogamente, $(I_g \backslash I_h) \cap I_h^{\varepsilon,k} = \varnothing$ para todo $k \in K_1$ suficientemente grande. Concluímos portanto que o termo B em (3.18) não

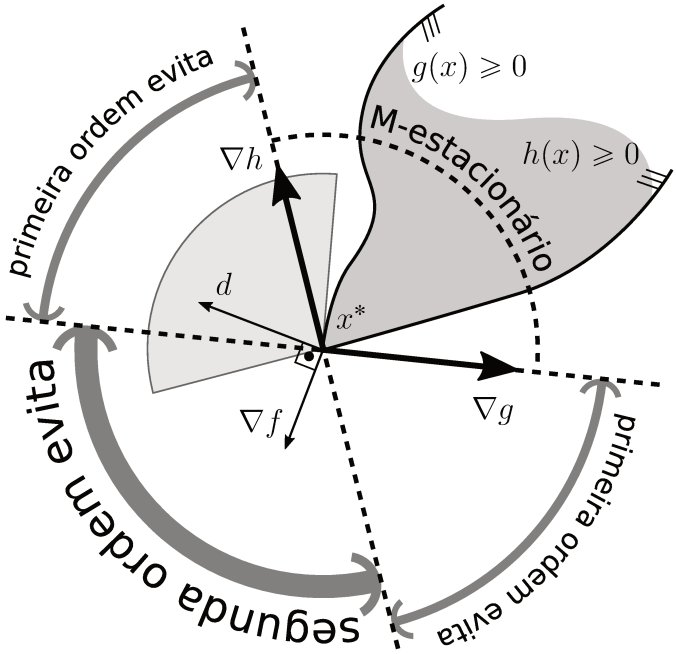

Figura 10 – Convergência de Algencan-second. As duas restrições complementares $g(x) \geqslant 0$ e $h(x) \geqslant 0$ são ativas em x^*. Quando $\nabla f(x^*)$ está nas regiões delimitadas pelos arcos sólidos finos, x^* não é C-estacionário, enquanto o arco sólido mais grosso indica a região para a qual x^* é C, mas não M-estacionário. Quando $\nabla f(x^*)$ aponta para as regiões indicadas pelas linhas pontilhadas, x^* é M-estacionário. Algencan não converge a x^* quando $\nabla f(x^*)$ aponta para os arcos finos, enquanto que Algencan-second ainda evita o caso em que $\nabla f(x^*)$ está na região delimitada pelo arco grosso. Neste último caso, um vetor d ortogonal a $\nabla f(x^*)$ é tal que $d^T(\nabla h \nabla g^T + \nabla g \nabla h^T)d = 2(\nabla h^T d)(\nabla g^T d) < 0$.

aparece. Finalmente, por (3.10) (página 63) e (3.17), o termo C em (3.18) tende a $-\infty$. Isso mostra que

$$\lim_{k \in K_1} d_k^T \nabla^2_{\varepsilon_k^{\text{fun}}} L_{\rho_k}(x^k, \mu^k) d_k = -\infty,$$

contradizendo o critério de segunda ordem (3.9) (página 62) em Algencan-second. Logo, x^* é um ponto M-estacionário, o que completa a prova. □

A Figura 10 fornece uma interpretação geométrica da convergência de Algencan-second.

O teorema anterior garante que Algencan-second converge a pontos M-estacionários de (MPCC) mediante MPCC-LICQ. Por outro lado, não podemos esperar convergência a S-estacionários, como mostra o exemplo a seguir.

Exemplo 3.2. *No Exemplo 3.1 (página 61), mostramos que* Algencan *pode convergir à origem, um ponto C-estacionário, mas não M-estacionário. Naquele exemplo,* Algencan-second *evita este inconveniente. Já no MPCC*

$$\min_x \frac{1}{2}\left[(x_1 - 1)^2 + 4x_2^2\right] \quad \text{s.a} \quad x_1 \geqslant 0, \quad x_2 \geqslant 0, \quad x_1 x_2 \leqslant 0,$$

ALGENCAN-SECOND *pode convergir à origem com as sequências definidas por* $x_k = (1/k^2, 1/k)$ *e* $\rho_k = k^4$*, caso em que* $\mu^{0,k} + \rho_k x_1^k x_2^k \to \infty$*. Neste problema, a origem é M, mas não S-estacionário.*

Por fim, notemos que informações de segunda ordem foram usadas anteriormente na literatura para obter convergência a pontos M-estacionários. Scholtes [137] a usou em seu método de regularização; Anitescu [23] mostrou que sua estratégia de penalização "elástica" converge a pontos M-estacionários sob MPCC-LICQ (veja ainda [24]); em [25], um algoritmo deste tipo que permite cálculos inexatos foi apresentado; informações de segunda ordem foram consideradas ainda em estratégias de Lagrangiano aumentado parcial [105]. Diferentemente deste último trabalho, estabelecemos resultados para um método de Lagrangiano aumentado de segunda ordem "puro", onde todas as restrições podem ser penalizadas. Esta situação é adequada especialmente à restrições não lineares.

3.1 Restrições de caixa

Nesta seção discutimos a adaptação da demonstração do Teorema 3.3 (página 63) para o caso em que restrições de caixa

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

são presentes. O caso de interesse é quando a sequência dual $\{\mu^{0,k} + \rho_k g(x^k)^T h(x^k)\}_{k\in K}$ é ilimitada, e vale MPCC-LICQ em x^*.

Naturalmente, a suposição de MPCC-LICQ em x^* diz respeito também às restrições de caixa ativas em x^*. Podemos portanto adaptar a expressão (3.15) (página 65) para

$$d^k \in \operatorname{span}\{v^k, \nabla g_{I_g\setminus\{j\}}(x^k), \nabla h_{I_h\setminus\{j\}}(x^k), e_F\}^\perp,$$

onde e_F é a matriz cujas colunas são os vetores canônicos e_i's de $\mathbb{R}^n$ com índices relativos a variáveis fixas na face corrente $\mathcal{F}(x^*)$. Ou seja, aqui d^k será ortogonal também às colunas de e_F. Multiplicando a expressão de primeira ordem por d^k obtemos

$$\nabla f^T d^k - \sum_{i=1}^m \left(\mu_i^{g,k} - \rho_k g_i\right)_+ \nabla g_i^T d^k - \sum_{i=1}^m \left(\mu_i^{h,k} - \rho_k h_i\right)_+ \nabla h_i^T d^k + \sum_{i=1}^n \left(\mu_i^{u,k} + \rho_k(x_i^k - u_i)\right)_+ e_i^T d^k - \sum_{i=1}^n \left(\mu_i^{\ell,k} - \rho_k(x_i^k - \ell_i)\right)_+ e_i^T d^k \to 0$$

(omitimos o termo "(x^k)" por simplicidade), onde $\mu^{u,k}$ e $\mu^{\ell,k}$ são as estimativas do método para os multiplicadores das restrições de caixa $x \leqslant u$ e $x \geqslant \ell$, respectivamente. Temos $e_i^T d^k = 0$ para todo índice i de variável fixa, e quando x_i é variável livre na face $\mathcal{F}(x^*)$, os termos entre parênteses nas duas últimas somas são nulos para todo $k \in K$ suficientemente

grande. Assim, obtemos novamente $\nabla f(x^k)^T d^k \to 0$, e de maneira análoga concluímos que

$$\lim_{k\in K_1} (d^k)^T \nabla^2_{\varepsilon_k^{\text{fun}}} L_{\rho_k}(x^k, \mu^k) d^k \to -\infty,$$

onde o conjunto K_1 é obtido como anteriormente. Agora, como d^k é ortogonal às colunas de e_F, podemos escrever

$$d^k = \sum_{i;\, x_i \text{livre}} d_i^k e_i$$

e logo

$$(d^k)^T \mathcal{H}^k d^k = \sum_{\substack{(i,j);\\ x_i, x_j \text{livres}}} d_i^k d_j^k \mathcal{H}_{ij}^k = \sum_{\substack{(i,j);\\ x_i, x_j \text{livres}}} d_i^k d_j^k \left(\nabla^2_{\varepsilon_k^{\text{fun}}} L_{\rho_k}\right)_{ij} = (d^k)^T \nabla^2_{\varepsilon_k^{\text{fun}}} L_{\rho_k} d^k \to -\infty,$$

completando o argumento.

3.2 Restrições gerais

Nesta seção mostramos brevemente que o Teorema 3.3 é válido na presença das restrições ordinárias $G(x) \leqslant 0$ e $H(x) = 0$, que não participam da complementariedade. Restrições de caixa são negligenciadas por simplicidade, pois os argumentos da seção anterior podem ser aplicados no presente caso. Naturalmente, o caso de interesse é quando a sequência dual $\{\mu^{0,k} + \rho_k g(x^k)^T h(x^k)\}_{k\in K}$ é ilimitada, e vale MPCC-LICQ em x^*. Vale ressaltar que MPCC-LICQ cobre as restrições $G_i(x) \leqslant 0$ e $H_i(x) = 0$ ativas em x^* (veja a Definição 2.2, página 43). Ao longo desta seção omitiremos o termo "(x^k)" sempre que conveniente.

Retomemos o subproblema de Algencan-second:

$$\begin{aligned}\min\ f(x) + \frac{\rho_k}{2}\Bigg\{ &\left\| \left(\frac{\mu^{G,k}}{\rho_k} + G(x) \right)_+ \right\|^2 + \left\| \frac{\mu^{H,k}}{\rho_k} + H(x) \right\|^2 \\ &+ \left\| \left(\frac{\mu^{g,k}}{\rho_k} - g(x) \right)_+ \right\|^2 + \left\| \left(\frac{\mu^{h,k}}{\rho_k} - h(x) \right)_+ \right\|^2 + \left[\left(\frac{\mu^{0,k}}{\rho_k} + g(x)^T h(x) \right)_+ \right]^2 \Bigg\}.\end{aligned}$$

O algoritmo gera sequências $\{x^k\}$, $\{\mu^k\}$ e $\{\rho_k\}$ satisfazendo as seguintes condições de primeira e segunda ordens:

Condição de primeira ordem:

$$\begin{aligned}\nabla_x L_{\rho_k} = \nabla f &+ \sum_{i=1}^{s} (\mu_i^{G,k} + \rho_k G_i)_+ \nabla G_i + \sum_{i=1}^{q} (\mu_i^{H,k} + \rho_k H_i) \nabla H_i \\ &- \sum_{i=1}^{m} (\mu_i^{g,k} - \rho_k g_i)_+ \nabla g_i - \sum_{i=1}^{m} (\mu_i^{h,k} - \rho_k h_i)_+ \nabla h_i \\ &+ (\mu^{0,k} + \rho_k g^T h)_+ \sum_{i=1}^{m} (\nabla h_i g_i + \nabla g_i h_i) \to 0.\end{aligned} \tag{3.20}$$

Condição de segunda ordem:

O menor autovalor da ε-Hessiana aproximada $\nabla^2_{\varepsilon_k} L_{\rho_k}(x^k, \mu^k)$ tende a ficar não negativo. Neste caso, tal matriz é dada por:

$$\begin{aligned}\nabla^2_{\varepsilon_k} L_{\rho_k} = \nabla^2 f &+ \sum_{i=1}^{s}(\mu_i^{G,k} + \rho_k G_i)_+ \nabla^2 G_i + \sum_{i=1}^{q}(\mu_i^{H,k} + \rho_k H_i)\nabla^2 H_i \\ &+ \rho_k \sum_{i\in I_G^{\varepsilon,k}} \nabla G_i \nabla G_i^T + \rho_k \sum_{i=1}^{q} \nabla H_i \nabla H_i^T \\ &- \sum_{i=1}^{m}[(\mu_i^{g,k} - \rho_k g_i)_+ - (\mu^{0,k} + \rho_k g^T h)_+ h_i]\nabla^2 g_i \\ &- \sum_{i=1}^{m}[(\mu_i^{h,k} - \rho_k h_i)_+ - (\mu^{0,k} + \rho_k g^T h)_+ g_i]\nabla^2 h_i \\ &+ \rho_k \sum_{i\in I_g^{\varepsilon,k}} \nabla g_i \nabla g_i^T + \rho_k \sum_{i\in I_h^{\varepsilon,k}} \nabla h_i \nabla h_i^T \\ &+ (\mu^{0,k} + \rho_k g^T h)_+ \left[\sum_{i=1}^{m}(\nabla h_i \nabla g_i^T + \nabla g_i \nabla h_i^T)\right] \\ &+ z_k \rho_k \left[\sum_{i=1}^{m}(\nabla h_i g_i + \nabla g_i h_i)\right]\left[\sum_{i=1}^{m}(\nabla h_i g_i + \nabla g_i h_i)\right]^T.\end{aligned}$$

onde

$$z_k = \begin{cases} 1 & \text{se } \dfrac{1}{\sqrt{\rho_k}}\left(\mu^{0,k} + \rho_k g(x^k)^T h(x^k)\right) \geqslant -\varepsilon_k^{\text{fun}} \\ 0 & \text{se } \dfrac{1}{\sqrt{\rho_k}}\left(\mu^{0,k} + \rho_k g(x^k)^T h(x^k)\right) < -\varepsilon_k^{\text{fun}} \end{cases}.$$

e

$$I_G^{\varepsilon,k} = \left\{ i \subset \{1, \ldots, p\} \;\middle|\; \frac{1}{\sqrt{\rho_k}}\left(\mu_i^{G,k} + \rho_k G_i(x^k)\right) \geqslant -\varepsilon_k^{\text{fun}} \right\}.$$

Consideramos ainda o conjunto adicional de índices $I_G(x) = \{i \mid G_i(x) = 0\}$ e, como anteriormente, denotamos $I_G = I_G(x^*)$, $I_g = I_g(x^*)$ e $I_h = I_h(x^*)$.

Considerando apenas a primeira ordem, podemos afirmar que Algencan-second converge a pontos C-estacionários, e logo as sequências de multiplicadores (3.13), (3.14) (página 64),

$$\lambda^{G,k} = \left(\mu^{G,k} + G(x^k)\right)_+ \quad \text{e} \quad \lambda^{H,k} = \left(\mu^{H,k} + H(x^k)\right)$$

admitem um vetor limite $\overline{\lambda}$ tal que

$$\nabla f(x^*) + \sum_{i\in I_G} \overline{\lambda}_i^G \nabla G_i(x^*) + \sum_{i=1}^{q} \overline{\lambda}_i^H \nabla H_i(x^*) - \sum_{i\in I_g} \overline{\lambda}_i^g \nabla g_i(x^*) - \sum_{i\in I_h} \overline{\lambda}_i^h \nabla h_i(x^*) = 0. \quad (3.21)$$

Procedemos assumindo por contradição que x^* não é M-estacionário, e logo existe um índice $j \in I_0$ tal como na demonstração do Teorema 3.3. Como MPCC-LICQ vale em x^*,

os gradientes

$$\nabla G_{I_G}(x^*), \quad \nabla H(x^*), \quad \nabla g_{I_g}(x^*) \quad \text{e} \quad \nabla h_{I_h}(x^*)$$

são linearmente independentes. Portanto, podemos tomar

$$d_k \in \operatorname{span}\left\{v^k, \nabla G_{I_G}(x^k), \nabla H(x^k), \nabla g_{I_g\setminus\{j\}}(x^k), \nabla h_{I_h\setminus\{j\}}(x^k)\right\}^{\perp}.$$

Multiplicando a expressão de primeira ordem (3.20) por d_k obtemos

$$\nabla f^T d_k + \sum_{i=1}^{s}(\mu_i^{G,k} + \rho_k G_i)_+ \nabla G_i^T d_k + \sum_{i=1}^{q}(\mu_i^{H,k} + \rho_k H_i)\nabla H_i^T d_k \\ - \sum_{i=1}^{m}(\mu_i^{g,k} - \rho_k g_i)_+ \nabla g_i^T d_k - \sum_{i=1}^{m}(\mu_i^{h,k} - \rho_k h_i)_+ \nabla h_i^T d_k \to 0,$$

e analogamente $\nabla f(x^k)^T d_k \to 0$. O vetor limite $d = \lim_{k\in K_1} d_k$ satisfaz

$$d \in \operatorname{span}\{\nabla f(x^*), \nabla G_{I_G}(x^*), \nabla H(x^*), \nabla g_{I_g\setminus\{j\}}(x^*), \nabla h_{I_h\setminus\{j\}}(x^*)\}^{\perp}.$$

Com argumentos similares sobre (3.21) obtemos $\nabla g_j(x^*)^T d = -(\overline{\lambda}_j^h/\overline{\lambda}_j^g)\nabla h_j(x^*)^T d \neq 0$, e (3.17) (página 66) ocorre. Multiplicando a expressão de segunda ordem por d_k, $k \in K_1$, obtemos

$$d_k^T \nabla^2_{\varepsilon_k} L_{\rho_k} d_k = d_k^T \left[\nabla^2 f(x^k) + \sum_{i=1}^{s}\left(\mu_i^{G,k} + \rho_k G_i(x^k)\right)_+ \nabla^2 G_i(x^k) \right. \\ \left. + \sum_{i=1}^{q}\left(\mu_i^{H,k} + \rho_k H_i(x^k)\right)\nabla^2 H_i(x^k) - \sum_{i=1}^{m}\lambda_i^{g,k}\nabla^2 g_i(x^k) - \sum_{i=1}^{m}\lambda_i^{h,k}\nabla^2 h_i(x^k)\right] d_k \\ \underbrace{+ \rho_k \sum_{i\in I^G_{\varepsilon_k}(x^k)}\left(\nabla G_i(x^k)^T d_k\right)^2 + \rho_k \sum_{i\in I^g_{\varepsilon_k}(x^k)}\left(\nabla g_i(x^k)^T d_k\right)^2 + \rho_k \sum_{i\in I^h_{\varepsilon_k}(x^k)}\left(\nabla h_i(x^k)^T d_k\right)^2}_{B} \\ + \underbrace{\left(\mu^{0,k} + \rho_k g(x^k)^T h(x^k)\right)_+ \left[d_k^T\left(\nabla h_j(x^k)\nabla g_j(x^k)^T + \nabla g_j(x^k)\nabla h_j(x^k)^T\right) d_k\right]}_{C}.$$

O termo entre colchetes é limitado e C tende a $-\infty$. Em B, as duas primeiras somas não aparecem quando ε_k é suficientemente pequeno. Agora, se $i \notin I_G(x^*)$ então, para todo $k \in K_1$ suficientemente grande, $G_i(x^k) \leqslant G_i(x^*)/2 < 0$, e logo

$$\lim_{k\in K_1}\frac{1}{\sqrt{\rho_k}}\left(\mu_i^{G,k} + \rho_k G_i(x^k)\right) \leqslant \lim_{k\in K_1}\left(\frac{\mu_i^{G,k}}{\sqrt{\rho_k}} - \sqrt{\rho_k}\frac{G_i(x^*)}{2}\right) = -\infty.$$

Daí, $i \notin I^G_{\varepsilon_k}(x^k)$, e logo a terceira soma em B não aparece. Isto completa a adaptação da demonstração do Teorema 3.3.

Capítulo 4

Resultados Numéricos

Neste capítulo reportamos e discutimos testes numéricos. Nossa implementação de Algencan-second (Algoritmo 3, página 34) foi baseada na versão 3.0.0 beta do pacote Algencan disponibilizado pelo projeto TANGO [140][1]. Analogamente, nossa implementação de Gencan-second (Algoritmo 4, página 35) baseou-se na versão de Gencan que acompanha o pacote Algencan. Como explicitado na Seção 1.4 (página 38), utilizamos para o cálculo de autovalores o método iterativo de gradientes conjugados precondicionado com correções de Jacobi (JPCG) [120, 121] implementado no pacote HSL EA19 [83] (versão 1.4.0). A fim de forçar Algencan a realizar somente passos de Lagrangiano aumentado, não utilizamos sua opção de aceleração [38]. A aceleração é acionada próximo ao método convergir. Ela consiste na aplicação de uma estratégia Newtoniana ao sistema KKT obtido pela fixação de restrições ativas como igualdade, tentando usufruir da boa convergência local de Newton.

Em relação à nossa implementação de Algencan-second, destacamos:

- Gencan emprega um *backtracking* com interpolação quadrática. Nós mantemos a interpolação se a direção de primeira ordem é escolhida. Quando uma direção de segunda ordem é escolhida, realizamos o *backtracking* simples $t \leftarrow 0.9 \times t$. Observamos que o critério de decréscimo suficiente para segunda ordem em Algencan-second não é do tipo Armijo clássico (veja os passos 6 e 7 do Algoritmo 5, página 36);

- A implementação de Gencan traz consigo passos de extrapolação [36] que melhoram o desempenho global de Algencan. Uma extrapolação ocorre quando, após um passo $t \in (0, 1]$ satisfazer a condição de Armijo, aumenta-se t exigindo que a função objetivo não cresça, mesmo que o novo t ultrapasse 1. Por padrão, a extrapolação em Gencan é feita dobrando o tamanho do passo anterior por 100 vezes ou até que a função objetivo cresça. Em Gencan-second mantemos essa estratégia;

[1] Os nomes dos métodos serão usados neste capítulo para referenciar também às suas implementações computacionais.

- Como enunciamos no Teorema 3.3 (página 63), a tolerância ε^{fun} para a ε-Hessiana aproximada deve ser da ordem de $O(1/\sqrt{\rho})$. Em nossa implementação, tomamos a sequência não crescente

$$\varepsilon_k^{\text{fun}} = \max\left\{10^{-14}, \min\{\varepsilon_{k-1}^{\text{fun}}, 1/\sqrt{\rho_k}\}\right\};$$

- A cada iteração de GENCAN, o decréscimo da função objetivo, o decréscimo da norma do gradiente projetado e o tamanho do passo são monitorados. GENCAN para por falta de progresso se por `maxinnitnp` iterações consecutivas pelo menos `itnplevel` desses valores não melhorarem. Em GENCAN-SECOND monitoramos adicionalmente o menor autovalor da ε-Hessiana reduzida. Especificamente, paramos a execução de GENCAN-SECOND se durante `maxinnitnp` iterações consecutivas pelo menos `itnplevel` das medidas originais não melhorarem e, adicionalmente, se o menor autovalor for menor que $-\varepsilon_k^{\text{curv}}$ sem melhora. A intenção é declarar falha somente se também não houve progresso da segunda ordem quando não satisfeita. Usamos os valores padrão `maxinnitnp` $= 3$ e `itnplevel` $= 2$;

- Além das melhorias da Seção 1.4 (página 38), GENCAN-SECOND inicia a busca por direções de curvatura negativa somente próximo a pontos estacionários de primeira ordem. Mais especificamente, autovalores são calculados quando ambas as condições ocorrem:

  – a norma do gradiente projetado é menor do que ou igual a $\varepsilon^{\text{grad}}$ (passo 1 do Algoritmo 4, página 34);

  – a norma do gradiente é menor do que ou igual a um patamar $\varepsilon^0 > 0$ (passo 2 do Algoritmo 5, página 36).

  Além disso, o teste por falta de progresso do menor autovalor que trata o item anterior só é realizado quando já não há progresso na primeira ordem;

- Na iteração interna de GENCAN-SECOND (Algoritmo 5, página 36), uma direção de primeira ordem $d^{k,1}$ é preferida a uma unitária de segunda ordem $d^{k,2}$ quando $\|d^{k,1}\| \neq 0$ e

$$\frac{\mathcal{G}_{\mathcal{F}}(x^k)^T d^{k,1}}{\|d^{k,1}\|} \leqslant \kappa\left(\frac{\mathcal{G}_{\mathcal{F}}(x^k)^T d^{k,2}}{\|d^{k,2}\|} + \frac{1}{2}\cdot\frac{(d^{k,2})^T \mathcal{H}_{[\mathcal{F}(x^k),\varepsilon^{\text{fun}},\overline{\rho}]}(x^k,\overline{\mu})d^{k,2}}{\|d^{k,2}\|^2}\right).$$

  Em nossa implementação, adotamos o teste *para aceitação da direção de segunda ordem*

$$\|d^{k,1}\| \leqslant \varepsilon^{\text{opt}} \quad \text{ou} \quad \kappa\left(\frac{\mathcal{G}_{\mathcal{F}}(x^k)^T d^{k,2}}{\|d^{k,2}\|} + \frac{\lambda_{\min}}{2}\right) < \frac{\mathcal{G}_{\mathcal{F}}(x^k)^T d^{k,1}}{\|d^{k,1}\|},$$

  onde ε^{opt} é a precisão para otimalidade e $\lambda_{\min}$ é o menor autovalor da ε-Hessiana reduzida. Cabe observarmos que neste ponto da execução, $\lambda_{\min} < 0$ e a direção de

segunda ordem $d^{k,2}$ não nula já foi calculada (veja os passos 2, 3 e 4 do Algoritmo 5). De acordo com o exposto na Seção 1.4 (página 38), o autopar $(\lambda_{\min}, d^{k,2})$ é calculado pelo método JPCG, e logo temos em mãos $\lambda_{\min}$;

- A direção de segunda ordem $d^{k,2}$ na iteração interna de GENCAN-SECOND não pode ser de subida dentro da face corrente (inequações (1.12) do Algoritmo 5, página 36). Assim, caso o método iterativo JPCG retorne uma direção de subida d, consideramos a direção $-d$;

- Em complemento ao item anterior, a direção $d^{k,2}$ deve pertencer a $\mathcal{S}(x^*)$, subespaço paralelo à variedade afim de menor dimensão contendo a face corrente. Como comentado na Seção 1.3 (página 32), a matriz $\mathcal{H}_{[\mathcal{F}(x^k),\varepsilon^{\text{fun}},\overline{\rho}]}(x^k, \overline{\mu})$ tem colunas/linhas da identidade para os índices relativos a variáveis fixas, o que a faz ser uma medida adequada de estacionariedade de segunda ordem dentro da face. Essas entradas não são consideradas no cálculo de direções de segunda ordem, pois permutamos linhas e colunas de modo a trabalhar somente com o bloco relativo às variáveis livres (veja o item (a) da Seção 1.4, página 38). Assim, o cálculo de direções de segunda ordem pertencentes a $\mathcal{S}(x^*)$ fica facilitado, e pode ser feito da seguinte forma: digamos que J_ℓ seja o conjunto dos índices das variáveis livres na face corrente. Calculamos assim um autovetor $\tilde{d} \in \mathbb{R}^{|J_\ell|}$ associado ao menor autovalor da submatriz $|J_\ell| \times |J_\ell|$ das variáveis livres, como explicitado na Seção 1.4. Claramente, o vetor definido por $d_i = \tilde{d}_i$ se $i \in J_\ell$ e $d_i = 0$ se $i \notin J_\ell$ é uma direção de curvatura negativa para $\mathcal{H}_{[\mathcal{F}(x^k),\varepsilon^{\text{fun}},\overline{\rho}]}(x^k, \overline{\mu})$, e pertencente a $\mathcal{S}(x^*)$. Finalmente, tomamos $d^{k,2} = d$ ou $d^{k,2} = -d$ de acordo com o item anterior;

- Usualmente iniciamos métodos de Lagrangiano aumentado com multiplicadores arbitrários, ou mesmo nulos como é o caso em nossos testes. Somente após algumas iterações as estimativas para os multiplicadores começam fornecer informações reais do problema. Portanto, só faz sentido iterações de segunda ordem após certo amadurecimento. Em nosso caso, aplicamos GENCAN (Algoritmo 2, página 30) no início do processo e decidimos quando é pertinente mudar para GENCAN-SECOND (Algoritmo 4, página 35). As seguintes estratégias são adotadas:

  - Uma iteração de GENCAN-SECOND nunca é executada na primeira iteração externa de ALGENCAN-SECOND;
  - Mudamos para GENCAN-SECOND se as estimativas para os multiplicadores λ^k "estabilizarem", isto é, se $\|\lambda^k - \lambda^{k-1}\|_\infty \leqslant \varepsilon^\lambda$ por N_λ iterações consecutivas;
  - Mudamos para GENCAN-SECOND também se os critérios de convergência para a primeira ordem em ALGENCAN-SECOND são satisfeitos com sua tolerância multiplicada por 10. Ou seja, se a complementariedade, a viabilidade e a norma do gradiente projetado são no máximo 10 vezes a precisão requerida;

  - Após a primeira execução de GENCAN-SECOND, todas as futuras iterações de ALGENCAN-SECOND são de segunda ordem. Isto garante que realizemos um número "infinito" de iterações de segunda ordem, caso em que o Teorema 3.3 (página 63) se aplica;
  - Não declaramos convergência em ALGENCAN-SECOND antes de executar uma iteração de segunda ordem. Logo, evitamos os casos em que a convergência a pontos estacionários ocorre rapidamente somente com direções de primeira ordem, antes da execução de pelo menos uma iteração de segunda ordem. Com isso temos um método de segunda ordem legítimo.

- Por outro lado, iniciamos o uso de GENCAN-SECOND sempre que detectarmos uma eminente falha de ALGENCAN-SECOND com GENCAN:

  - ALGENCAN-SECOND pode falhar declarando que um ponto estacionário da inviabilidade foi encontrado. Portanto, mudamos para GENCAN-SECOND se o gradiente da medida de inviabilidade no ponto corrente é menor do que duas vezes a tolerância que ALGENCAN-SECOND usa para declarar falha;
  - Uma falha pode ocorrer se o parâmetro de penalização ρ_k cresce muito. Portanto, mudamos para GENCAN-SECOND se ρ_k é maior que um parâmetro fixo $\rho_{2\text{nd}}$;
  - Mudamos ainda para GENCAN-SECOND se 2/3 do número máximo de iterações de ALGENCAN-SECOND for alcançado.

Os testes computacionais foram realizados com problemas da coletânea MacMPEC, disponível no endereço `http://www.mcs.anl.gov/~leyffer/MacMPEC` em linguagem de modelagem AMPL [62]. Dos 193 problemas testes, 19 foram descartados por serem multiobjetivo (`ralph1`), inviáveis (`pack-rig2-16/32`,`pack-rig2c-16/32`), terem variáveis inteiras (`ex9.1.2`) ou restrições de complementariedade mistas (`bilevel1m`, `bilevel2m`, `gnash10m`–`19m`, `taxmcp`). Outros 15 problemas teste não foram considerados pois causam erros fatais desconhecidos na interface AMPL de ALGENCAN (`gnash14`, `incid-set-1-8/16/32`, `incid-set-1c-8/16/32`, `incid-set-2-8/16/32`, `incid-set-2c-8/16/32`, `water-FL`, `water-net`). O problema `pack-rig3c-32` causou estouro de memória nos cálculos de segunda ordem, enquanto `bem-milanc30-s`, `tollmpec`, e `tollmpec1` excederam cinco horas de execução. Estes problemas também foram ignorados. Cabe notarmos que as instâncias originais usam a diretiva `complements` para representar restrições de complementariedade, mas essas diretivas não são entendidas pela interface AMPL de ALGENCAN. Portanto as reescrevemos explicitamente. A opção `presolve` do AMPL foi desligada a fim de preservar a estrutura original dos problemas. A única modificação usada é o escalamento dos dados do problema, padrão em ALGENCAN.

Um computador equipado com processador Intel i7 2.6 GHz, 8 GB de memória RAM e sistema operacional GNU/Linux 64 bits (Ubuntu 16.04) foi usado. Para ALGENCAN

e ALGENCAN-SECOND usamos $\varepsilon^{\text{opt}} = \varepsilon^{\text{feas}} = 10^{-6}$ e o número máximo de iterações externas igual a 100. Em relação aos parâmetros exclusivos do método de segunda ordem, fixamos $\varepsilon^0 = 10^{-4}$, $\varepsilon^{\text{curv}} = 0.99 \times 10^{-6}$, $\varepsilon^{\text{hess}} = 10^{-6}$, $\kappa = 1$, $\eta = 0.1$, $\varepsilon_0^{\text{fun}} = 10^{-6}$, $\varepsilon^{\lambda} = \sqrt{\varepsilon^{\text{feas}}} = 10^{-3}$, $N_\lambda = 3$ e $\rho_{2\text{nd}} = 10^8$. Inicializamos as estimativas dos multiplicadores todas em zero. Quando não fornecido pela instância do problema, tomamos o ponto primal inicial como sendo a origem.

O uso de variáveis de folga em ALGENCAN não é recomendado em geral pois o método pode se beneficiar das restrições de desigualdade [39]. Logo o uso de folgas é desabilitado por padrão em ALGENCAN. No entanto, a experiência numérica com métodos de programação quadrática sequencial aplicados a MPCCs sugerem que é melhor escrevermos a complementariedade com folga, na forma $g(x)^T h(x) + s = 0$, $s \geqslant 0$ [61]. Izmailov, Solodov e Uskov [87] adotam esta forma em seus testes computacionais com ALGENCAN. A fim de comparar o comportamento de ALGENCAN e ALGENCAN-SECOND nessas duas situações, realizamos testes em ambas as formas $g(x)^T h(x) \leqslant 0$ e $g(x)^T h(x) + s = 0, s \geqslant 0$.

Os teoremas de convergência 3.2 e 3.3 (páginas 58 e 63, respectivamente) garantem que os métodos de Lagrangiano aumentado atingem pontos S-estacionários se a sequência de multiplicadores da complementariedade é limitada. Logo, se apenas pontos C ou M-estacionários forem atingidos, esta sequência necessariamente é ilimitada. Em particular, o parâmetro de penalização ρ explode. Evidentemente, este caso não pode ser considerado uma falha do método. Portanto, consideramos que ALGENCAN e ALGENCAN-SECOND convergem se uma das duas situações ocorrer:

- o algoritmo declara convergência a um ponto estacionário viável;
- o algoritmo para com um grande ρ (maior que 10^8, o valor padrão em ALGENCAN), mas com viabilidade e com valor funcional final não pior que o melhor valor conhecido na literatura, como reportado na coletânea MacMPEC. Esta é a situação típica de convergência a pontos somente C ou M-estacionários.

Se nenhuma dessas condições se verificarem, consideramos que o método falhou.

Resultados para complementariedade sem folga

Neste caso, a complementariedade toma a forma $g(x)^T h(x) \leqslant 0$. Ambos os métodos falharam em 15 dos 155 problemas considerados (9,68%). A Tabela 1 traz os problemas onde ocorreu convergência em pelo menos um dos métodos, mas diferentes pontos limites foram "atingidos". A coluna "Melhor FO" contém o melhor valor conhecido para a função objetivo, como descrito na coletânea MacMPEC. As colunas "FO" e "Inviab" contêm, respectivamente, o valor final da função objetivo e a medida de inviabilidade [39]

		Algencan			Algencan-second		
Problema	Melhor FO	FO	Saída	Inviab	FO	Saída	Inviab
`pack-comp1c-32`	6.61e-01	6.01e-01	$\rho \gg 1$	1.14e-03	6.61e-01	conv./S	3.47e-08
`scale4`	1.00e+00	1.71e+00	conv./C	5.57e-07	1.00e+00	conv./S	1.58e-08
`scale5`	1.00e+02	2.00e+02	conv./C	5.56e-07	1.00e+02	conv./S	3.00e-09
`scholtes3`	5.00e-01	9.99e-01	conv./C	5.56e-07	5.00e-01	conv./S	3.00e-10

Tabela 1 – Testes computacionais, complementariedade na forma $g(x)^T h(x) \leqslant 0$. Diferenças entre Algencan e Algencan-second.

dada por

$$\max\left\{\|G(x^k)_+\|_\infty,\ \|H(x^k)\|_\infty,\ \|\left(-g(x^k)\right)_+\|_\infty,\ \|\left(-h(x^k)\right)_+\|_\infty,\ \left(g(x^k)^T h(x^k)\right)_+\right\}. \quad (4.1)$$

Na coluna "Saída", "$\rho \gg 1$" indica que o algoritmo falhou com parâmetro de penalização grande. A entrada "conv./C" indica que o algoritmo declarou convergência, mas pelo menos um par de multiplicadores da função Lagrangiano-MPCC associados à restrições complementares pareceu convergir a valores negativos. Ou seja, o ponto encontrado é aproximadamente C-estacionário apenas. De maneira análoga, a entrada "conv./S" indica convergência a pontos aproximadamente S-estacionários, isto é, quando todos multiplicadores parecem convergir a valores não negativos.

Observamos que no problema `pack-comp1c-32`, Algencan-second alcançou viabilidade, enquanto que Algencan não atingiu a precisão requerida. Nos outros três problemas da Tabela 1, Algencan-second convergiu a um ponto com valor funcional menor que o atingido pelo método de primeira ordem. É possível verificar em cada caso que Algencan convergiu a um ponto apenas C-estacionário, enquanto Algencan-second alcançou a melhor solução conhecida, um ponto S-estacionário. Isso corrobora nossa teoria. Nos outros 135 problemas, ambos os algoritmos declararam convergência com o mesmo valor de função objetivo. Nestes problemas, o valor funcional foi igual ao reportado pela coletânea MacMPEC, com exceção dos problemas `ex9.2.5`, `qpecgen100-4`, `qpecgen200-3` e `qpecgen200-4`, onde um ponto aparentemente S-estacionário com melhor valor funcional foi atingido, e `bard3`, `ex9.2.3`, `pack-comp1p-8`, `qpecgen100-1`, `qpecgen100-3`, `qpecgen200-2` e `TrafficSignalCycle-1`–`11/13`, onde os algoritmos convergiram a S-estacionários aproximados com pior valor funcional.

Resultados para complementariedade com folga

Quando consideramos a complementariedade na forma $g(x)^T h(x) + s = 0$, $s \geqslant 0$, o comportamento geral dos algoritmos foi muito similar ao caso anterior. Aqui, o problema `hakonsen` apresentou um erro na interface AMPL de Algencan, e foi desconsiderado. Dos 154 problemas considerados, novamente ambos os algoritmos falharam em 15 deles (9,74%). Como anteriormente, evidenciamos na Tabela 2 apenas os problemas onde a viabilidade foi alcançada por algum dos algoritmos e onde pontos limite diferentes foram

		Algencan			Algencan-second		
Problema	Melhor FO	FO	Saída	Inviab	FO	Saída	Inviab
`pack-comp1c-32`	6.61e-01	8.25e-01	$\rho \gg 1$	9.32e+06	6.61e-01	conv./S	6.46e-07
`pack-comp1p-8`	6.00e-01	6.61e-01	conv./S	2.11e-07	6.01e-01	conv./S	1.50e-09
`scale4`	1.00e+00	1.71e+00	conv./C	5.57e-07	1.00e+00	conv./S	1.58e-08
`scale5`	1.00e+02	2.00e+02	conv./C	5.56e-07	1.00e+02	conv./S	3.00e-10
`scholtes3`	5.00e-01	9.99e-01	conv./C	5.56e-07	5.00e-01	conv./S	3.00e-10

Tabela 2 – Testes computacionais, complementariedade na forma $g(x)^T h(x) + s = 0$, $s \geqslant 0$. Diferenças entre Algencan e Algencan-second.

"atingidos". No caso anterior, Algencan e Algencan-second alcançaram o mesmo valor funcional em `pack-comp1p-8`, porém maior que o melhor valor conhecido. Agora, ambos os algoritmos declararam convergência, mas Algencan-second obteve um valor funcional melhor.

Notamos que Algencan comportou-se de maneira muito similar nas duas formas de escrita da complementariedade. Logo, não podemos afirmar que uma forma é melhor que a outra. Em [87] os autores fazem a mesma observação. O mesmo comentário é válido para Algencan-second.

Nos testes computacionais realizados em [87], os autores consideraram 161 problemas da coletânea MacMPEC e reportaram que Algencan convergiu em cerca de 97% deles. Em nossos testes, essa porcentagem girou em torno de 90%. Provavelmente, esta diferença foi causada pelas diferentes opções adotadas, e pelas diferentes versões de Algencan utilizadas em cada bateria de testes. Em particular, não usamos os resolvedores de sistemas lineares esparsos da biblioteca HSL [83], cujo uso é recomendado na documentação de Algencan. Quando um resolvedor destes está disponível, Algencan emprega uma estratégia Newtoniana com regiões de confiança para resolução dos subproblemas [39], que não são compatíveis com a busca linear usada em Gencan-second. O leitor pode consultar novamente a Seção 1.2, quando da apresentação do Algoritmo 2 (página 30), para uma discussão mais detalhada sobre esta opção. Por outro lado, como já comentado na introdução, a maior contribuição deste trabalho é na teoria de convergência de métodos de Lagrangiano aumentado em MPCCs. Os testes numéricos aqui apresentados servem como confirmação da teoria desenvolvida no Capítulo 3, indicando que há casos em que a segunda ordem de fato recupera pontos estacionários melhores quando comparados à primeira ordem. A fim de fazermos esta comparação, Gencan, tal como descrito no Algoritmo 2, foi usado em todos os testes. De qualquer forma, tudo indica que Algencan atinge minimizadores de (MPCC) com frequência. Logo, não consideramos aqui os métodos de primeira e segunda ordens competidores entre si, mas sim, a informação de segunda ordem como complemento ao método de primeira ordem. Este é o espírito de Algencan-second: só recorremos à segunda ordem quando a primeira tende a se esgotar. De fato, observe que o critério (1.11) (página 36) somente é reprovado se a norma do gradiente

interno $\mathcal{G}_{\mathcal{F}}(x^k)$ é pequena. Cabe ressaltarmos que, ao contrário de Algencan-second, outros métodos empregam estratégias que conjugam direções de primeira e segunda ordens simultaneamente, e durante todo o processo de minimização. Veja por exemplo [44, 65, 114] e suas referências. Naturalmente, isso é razoável quando informações de segunda ordem puderem ser computadas a um custo relativamente baixo.

Por fim, lembramos que a análise feita nesta seção pontua apenas as diferenças observadas entre os métodos de primeira e segunda ordens. No Anexo A relatamos a lista completa dos testes realizados.

4.1 Uma estratégia alternativa

Nos testes computacionais relatados na seção anterior, verificamos que Algencan tem alta taxa de sucesso, e que nossa implementação de Algencan-second apresenta resultados muito parecidos com o método de primeira ordem. Em particular, pudemos observar que o método de segunda ordem atinge pontos estacionários qualitativamente melhores apenas em muito poucos casos (Tabelas 1 e 2). Nesta situação, é evidente que o método de segunda ordem, que realiza computações explícitas de autopares, tenha maior custo computacional frente ao de primeira ordem. A Figura 11 traz um comparativo dos tempos de execução entre os dois algoritmos. Foram considerados problemas com complementariedade sem folga, e para os quais ambos os métodos declararam convergência a um ponto viável ou terminaram com parâmetro de penalização ρ grande, mas com medida de inviabilidade menor do que ou igual a $10 \times \varepsilon^{\text{feas}} = 10^{-5}$ (esta é a típica situação de convergência a C ou M-estacionários apenas). Por outro lado, foram descartados problemas em que, para pelo menos um dos algoritmos, as funções internas do AMPL retornaram `NaN` ("*Not A Number*") ou `Inf` ("Infinito") em algum momento da minimização. Essas respostas atendem a padrões estabelecidos pela IEEE para representação de pontos flutuantes, e ocorrem quando operações ilegais são realizadas (tais como divisão por zero ou raiz quadrada de números negativos), ou quando a capacidade do tipo numérico é extrapolada. O fato é que nessas situações, Algencan tenta "sair do iterando problemático e continuar a minimização", muitas vezes com sucesso, mas a um custo computacional alto. Julgamos que este custo não pode ser comparado, e portanto descartamos o problema. Descartamos ainda problemas em que os algoritmos alcançaram valores funcionais diferentes, pois é justificável tempos diferentes para soluções diferentes.

De maneira global, Algencan-second foi 76% mais lento que Algencan. Em particular, o tempo de execução do primeiro método nos problemas com $n \in [200, 246]$ ultrapassou mais de cinco vezes o tempo do segundo (veja a Figura 11). Isso ocorreu porque Algencan-second realizou muitas iterações de segunda ordem até parar com ρ grande. Cabe ressaltar que o multiplicador do tempo é a média geométrica das razões entre

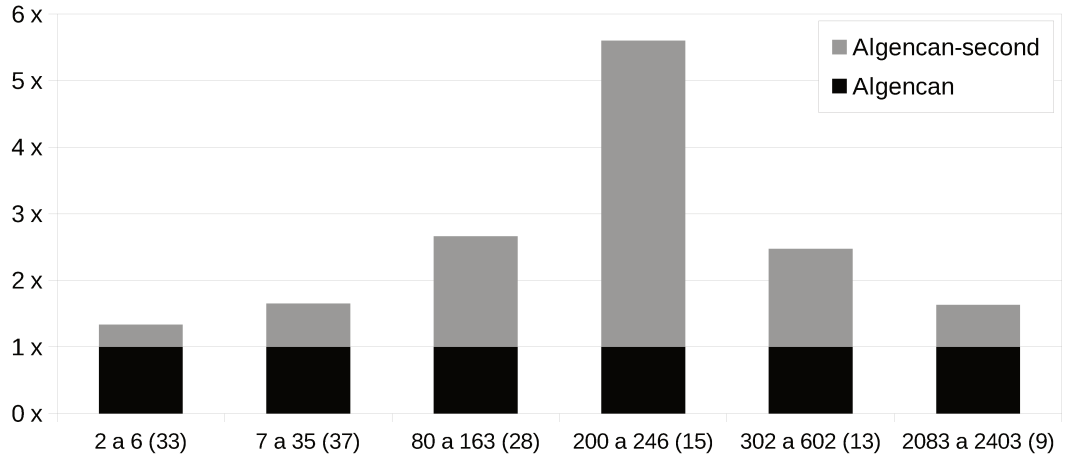

Figura 11 – Tempos de execução de ALGENCAN-SECOND em relação a ALGENCAN para problemas com complementariedade sem folga, e quando ambos os métodos atingiram viabilidade. Na horizontal, os intervalos para o número n de variáveis, e entre parênteses o número de problemas no intervalo. Na vertical, o número de vezes do tempo de execução de ALGENCAN.

o tempo T_2 de ALGENCAN-SECOND e o tempo T_1 de ALGENCAN de cada problema. Assim, ALGENCAN é a referência. Por sua vez, os tempos de execução T_1 e T_2 foram obtidos pela média aritmética das rodadas necessárias para um tempo total mínimo de 15 segundos. Isso torna a medição mais confiável em problemas pequenos.

Diante desta discussão, podemos nos indagar se devemos pagar o custo computacional de iterações de segunda ordem sem efeito prático. Na intenção de evitá-las, consideramos a seguinte estratégia:

1. executamos ALGENCAN até parar, digamos, na iteração k;
2. se ALGENCAN atingiu um ponto "suficientemente viável", analisamos se o limite "aparenta" ser M-estacionário. Em caso negativo, executamos uma iteração de segunda ordem.

   Caso ALGENCAN declarou falha com um ponto inviável, encerramos o processo declarando falha;
3. repetimos este processo até parar, ou por um número máximo de vezes.

Os melhoramentos no cálculo de autopares feitos na Seção 1.4 (página 38) e neste capítulo são de propósito geral, servem em princípio a qualquer problema. Ademais, garantem que o método resultante seja legitimamente de segunda ordem. Ao contrário, a estratégia acima descrita não garante pontos estacionários de segunda ordem, dado que iterações de segunda ordem são sujeitas à critérios de estacionariedade específicos para MPCCs. No entanto, isso não fere nossa teoria, pois procuramos apenas pontos M-estacionários. Assim, é importante salientar que encaramos aqui a segunda ordem não

como fim, mas como um recurso adicional, tendo em mente a manutenção da eficiência computacional global do método de primeira ordem. Isso, aliado ao comportamento observado em nossos testes numéricos, justifica a nova estratégia.

Passamos a especificar o procedimento adotado. No item 3, fixamos em 10 o número máximo de vezes para aplicação da estratégia. Se o procedimento iniciar após uma falha ALGENCAN, mas com viabilidade suficiente e em um ponto aparentemente não M-estacionário, adicionamos uma única vez 10 iterações ao número máximo permitido e reiniciamos ρ pelo procedimento padrão de ALGENCAN. Quanto ao item 2, consideramos que o iterando x^k de ALGENCAN é suficientemente viável se a medida de inviabilidade (4.1) (página 77) for menor do que ou igual a $10 \times \varepsilon^{\text{feas}} = 10^{-5}$. A fim de verificar a estacionariedade de x^k, devemos calcular estimativas para os multiplicadores λ_i^k da função Lagrangiano-MPCC com índices em $I_0(x^k)$. Vimos, no Capítulo 3, que a função Lagrangiano aumentado usada por ALGENCAN fornece as estimativas

$$\lambda_i^{g,k} = \left(\mu_i^{g,k} - \rho_k g_i(x^k)\right)_+ - \left(\mu^{0,k} + \rho_k g(x^k)^T h(x^k)\right)_+ h_i(x^k) \text{ e} \tag{4.2}$$

$$\lambda_i^{h,k} = \left(\mu_i^{h,k} - \rho_k h_i(x^k)\right)_+ - \left(\mu^{0,k} + \rho_k g(x^k)^T h(x^k)\right)_+ g_i(x^k), \tag{4.3}$$

onde μ^k são os multiplicadores calculados pelo algoritmo na iteração k (veja as expressões (3.1), (3.3) e (3.4), página 59). Testar se x^k é ponto M-estacionário aproximado envolve verificar, ainda que aproximadamente, se

$$\lambda_i^{g,k}\lambda_i^{h,k} = 0 \quad \text{ou} \quad \lambda_i^{g,k}, \lambda_i^{h,k} \geqslant 0, \quad \forall i \in I_0(x^k) \tag{4.4}$$

(veja a Definição 2.3, página 44). No entanto, destacamos dois possíveis inconvenientes do mundo numérico:

- a identificação de restrições ativas em programação não linear pode errar. Em particular, ao errar na identificação de atividade das restrições $g_i(x) \geqslant 0$ e $h_i(x) \geqslant 0$ em x^k, erramos na descrição de $I_0(x^k)$ (aliás, o desejo é descrever o conjunto $I_0(x^*)$ no provável limite x^*);
- o produto $\lambda_i^{g,k}\lambda_i^{h,k}$ tem escala diferente de $\lambda_i^{g,k}$ e $\lambda_i^{h,k}$. É razoável levar isso em consideração no teste (4.4).

Se identificarmos erroneamente $g_i(x) \geqslant 0$ e $h_i(x) \geqslant 0$ como restrições ativas, faremos um teste do tipo (4.4) a mais. Não julgamos isso uma falha grave, pois no máximo realizaremos uma iteração de segunda ordem sem necessidade. Agora, se não identificarmos tais restrições como ativas quando realmente forem, estaremos potencialmente deixando de obter um ponto estacionário melhor. Em outras palavras, é melhor errar por excesso, identificar um conjunto de índices que contenha $I_0(x^k)$. Nossa estratégia é então declarar $i \in I_0(x^k)$ quando

$$|g_i(x^k)| \leqslant \delta \quad \text{e} \quad |h_i(x^k)| \leqslant \delta$$

para $\delta > 0$ fixo.

Para a identificação acima, escolhemos

$$\delta = \begin{cases} \sqrt{\varepsilon^{\text{feas}}} = \sqrt{10^{-6}} & \text{se ALGENCAN declarou convergência} \\ \sqrt{10 \times \varepsilon^{\text{feas}}} = \sqrt{10^{-5}} & \text{se ALGENCAN só atingiu viabilidade suficiente} \end{cases}.$$

Separamos em dois casos pois consideramos que houve convergência se de fato o algoritmo a declarou (viabilidade com precisão $\varepsilon^{\text{feas}}$), ou se não declarou convergência, mas alcançou uma viabilidade satisfatória (cuja precisão estabelecemos em $10 \times \varepsilon^{\text{feas}}$). A motivação para tomar a raiz quadrada é a seguinte: suponhamos que ALGENCAN atingiu viabilidade com precisão $\varepsilon > 0$ para as restrições $g(x) = x_1 \geqslant 0$ e $h(x) = x_2 \geqslant 0$. Isto é,

$$\max\left\{ |(-x_1^k)_+|,\ |(-x_2^k)_+|,\ |(x_1^k x_2^k)_+| \right\} \leqslant \varepsilon$$

(veja a expressão (4.1), página 77). O algoritmo pode ter parado com $x_1^k = x_2^k = \sqrt{\varepsilon}$ pois, neste caso, este máximo será igual a $|(x_1^k x_2^k)_+| = \varepsilon$. No entanto, é possível que se o método continuasse, os iterandos se aproximariam cada vez mais da origem (isso de fato ocorre no problema do Exemplo 3.1, página 61). Na dúvida, deveríamos identificar como ativas ambas as restrições $x_1 \geqslant 0$ e $x_2 \geqslant 0$, e logo devemos tomar, no mínimo, $\delta = \sqrt{\varepsilon}$. Este argumento é razoável no caso geral, dado que a complementariedade é formada por produtos $g_i(x)h_i(x)$.

Finalmente, resolvemos a incompatibilidade de escala entre o produto $\lambda_i^{g,k}\lambda_i^{h,k}$ e os multiplicadores $\lambda_i^{g,k}$ e $\lambda_i^{h,k}$ substituindo (4.4) por

$$\min\left\{ \max\{\lambda_i^{g,k}, -\lambda_i^{h,k}\},\ \max\{-\lambda_i^{g,k}, \lambda_i^{h,k}\},\ \max\{\lambda_i^{g,k}, \lambda_i^{h,k}\} \right\} \geqslant -\delta, \quad \forall i \in I_0(x^k). \quad (4.5)$$

O critério (4.4) descarta ambos multiplicadores negativos ou com sinais trocados, exatamente o que faz (aproximadamente) o critério acima[2].

Resultados e análise da nova estratégia

Comentamos sucintamente o comportamento da nova estratégia para complementariedade sem folga, isto é, escrita como $g(x)^T h(x) \leqslant 0$.

Observamos que em quase todos os problemas, a nova estratégia comportou-se como o método de primeira ordem. Este comportamento é esperado, já que ALGENCAN parece convergir a pontos S-estacionários com frequência. As exceções foram os problemas `scale4`, `scale5` e `scholtes3`, únicos casos onde ALGENCAN-SECOND alcançou um ponto qualitativamente melhor que ALGENCAN (veja a Tabela 1, página 77). Nestes problemas,

[2] Anitescu [23] discute a identificação de pontos apenas M-estacionários em aritmética de ponto flutuante. O autor conclui que verificações numéricas sobre sinais de multiplicadores não são suficientes para predizer se o ponto é M ou S-estacionário. Mas em nosso caso, ambos conceitos de estacionariedade são desejáveis. Pelo menos de nosso conhecimento, não há na literatura algo análogo envolvendo C-estacionariedade.

a nova estratégia também atingiu pontos S-estacionários. Isso indica que o teste (4.5) funcionou, identificando corretamente pontos M-estacionários. Ademais, a proximidade entre a nova estratégia e o método de primeira ordem se refletiu no tempo computacional: nas duas primeiras faixas ($n \in [2, 6]$ e $n \in [7, 35]$), o primeiro algoritmo foi, respectivamente, apenas 6% e 11% mais custoso que o segundo, enquanto que nas demais faixas, das maiores dimensões, o tempo de execução foi o mesmo. Globalmente, o custo adicional foi de 4%.

Ressaltamos que ALGENCAN e a nova estratégia pararam com viabilidade insuficiente para o problema `pack-comp1c-32`. Neste caso, o teste de M-estacionariedade (4.5) não foi aplicado em momento algum. No entanto, ALGENCAN-SECOND alcançou viabilidade (veja a Tabela 1). Este problema ilustra que aqui encaramos a segunda ordem como um recurso adicional, não temos em mãos um método de segunda ordem legítimo. Ou seja, quando a primeira ordem falha na inviabilidade, encerramos declarando falha na resolução. Cabe ressaltar que, como já comentado neste capítulo, os casos onde ALGENCAN não atinge viabilidade, mas ALGENCAN-SECOND sim, são raros, pelo menos considerando nossa implementação e a coletânea de problemas utilizada.

Capítulo 5

Considerações Finais

O objetivo central desta tese é analisar a convergência de métodos de Lagrangiano aumentado aplicados a *problemas de programação matemática com restrições de complementariedade* (MPCCs). Tais métodos constituem uma importante classe de algoritmos destinados a resolver problemas gerais de otimização não linear. Remontam aos trabalhos seminais de Hestenes [78] e Powell [125], do fim dos anos 60, tendo nos dias atuais uma vasta literatura dedicada. Possuem boas propriedades de convergência e bom desempenho numérico, em particular, aqueles que utilizam a função Lagrangiano aumentado obtida via penalização quadrática. Esta é a variante considerada neste trabalho, a qual descrevemos em detalhes no Capítulo 1.

Problemas com restrições de complementariedade surgem em variados contextos, tais como otimização em dois níveis, otimização de portfólio, máquinas de suporte vetorial, otimização discreta, controle ótimo; e em aplicações como controle de tráfego urbano, economia, problemas do setor elétrico, dentre outras. A introdução traz algumas referências. Do ponto de vista teórico, têm alto grau de degeneração, no sentido que não satisfazem a maioria das condições de qualificação para programação não linear geral estabelecidas na literatura. Assim, conceitos de estacionariedade, diferentes das condições KKT, e condições de qualificação adequadas foram definidas na literatura. O Capítulo 2 traz uma revisão dos principais conceitos relacionados a MPCCs.

A degeneração presente nos problemas com restrições de complementariedade impõe desafios na análise de convergência de algoritmos. Não por acaso, há uma expressiva literatura dedicada ao tema. Em particular, esquemas de penalização e de Lagrangiano aumentado receberam especial atenção [23–25, 84, 87, 105, 106, 142, 143], pois suas boas propriedades de convergência se refletem também em MPCCs. Neste trabalho, estendemos os resultados de convergência de métodos de Lagrangiano aumentado de primeira ordem desenvolvidos em [87]. Mais especificamente, provamos que quando certa sequência de multiplicadores da complementariedade é limitada, o método converge a pontos KKT com a condição de qualificação MPCC-RCPLD, que é menos restritiva que MPCC-

LICQ, usada no resultado anterior. Cabe ressaltar que a nova hipótese engloba o caso de restrições lineares, antes descoberto. Ademais, provamos um novo resultado sobre a convergência do método de Lagrangiano aumentado de segunda ordem conhecido como Algencan-second, desenvolvido em [4]. Mostramos que este método é capaz de atingir melhores pontos estacionários que seu par de primeira ordem, com as mesmas condições de qualificação. Os resultados de convergência são discutidos no Capítulo 3.

Do ponto de vista da resolução numérica de MPCCs, métodos tradicionais em otimização são aplicados com certo sucesso. A introdução traz várias referências. Em particular, esquemas de Lagrangiano aumentado têm bom comportamento frente a outros métodos [87]. No Capítulo 4, apresentamos propostas para uma implementação prática de Algencan-second, bem como uma alternativa específica para MPCCs que minimiza o tempo computacional gasto com computações de segunda ordem. Realizamos testes numéricos com problemas da tradicional coletânea MacMPEC, mantida por Sven Leyffer, e exibimos casos em que Algencan-second converge a pontos estacionários qualitativamente melhores que aqueles obtidos pelo método de primeira ordem. Ademais, os testes corroboram o bom desempenho dos esquemas de Lagrangiano aumentado verificado anteriormente na literatura.

5.1 Trabalhos futuros

Condições sequenciais de otimalidade para MPCCs

A convergência de métodos em otimização não linear contínua pode ser tratada por meio das chamadas *condições sequenciais de otimalidade* (veja [9–11, 16, 18, 74–77, 108]). Argumentamos na Seção 2.2 (página 54) que as principais condições sequenciais conhecidas para programação não linear padrão não garantem mais do que W-estacionariedade, uma caracterização muito pouco exigente dos minimizadores de MPCCs. No entanto, métodos como Lagrangiano aumentado possuem convergência mais forte. Logo, um tema de pesquisa ainda quase intocado é o estudo de condições sequenciais de otimalidade específicas para MPCCs, e a consequente unificação dos resultados de convergência dos algoritmos. A única proposta de nosso conhecimento é o recente trabalho, ainda não publicado, de Ramos [130], no qual uma condição sequencial relacionada à M-estacionariedade é estabelecida. Ademais, há propostas de condições sequenciais para outros contextos específicos, como programação multiobjetivo [66]. Condições sequenciais para problemas "híbridos", como multiobjetivo com restrições de complementariedade, podem também ser objeto de estudo.

MPCCs e estacionariedade segundo Bouligand

Restrições de complementariedade têm intrinsecamente características combinatoriais. De fato, ao exigirmos $g_i(x) \geqslant 0$, $h_i(x) \geqslant 0$ e $g_i(x)h_i(x) \leqslant 0$, naturalmente surgem

três possibilidades: $g_i(x) > 0$ e $h_i(x) = 0$; $g_i(x) = 0$ e $h_i(x) > 0$; ou $g_i(x) = h_i(x) = 0$. Dado um ponto viável x^*, essas escolhas nos levam a considerar os problemas resultantes do MPCC original fixando as restrições ativas $g_i(x) \geqslant 0$ ou $h_i(x) \geqslant 0$ em x^* como igualdade. Como exemplo, para as restrições $x_1, x_2 \geqslant 0$ e $x_1 x_2 \leqslant 0$, e o ponto $x^* = (0, 0)$, as três alternativas correspondentes são $x_1 \geqslant x_2 = 0$, $0 = x_1 \leqslant x_2$ e $x_1 = x_2 = 0$ (em tempo, o problema não linear apertado (TNLP), página 42, refere-se à última escolha). O conceito de estacionariedade segundo Bouligand (B-estacionariedade) [135] captura essa estrutura no seguinte sentido: um ponto viável x^* para MPCC é *B-estacionário* quando é ponto KKT para cada um dos problemas acima descritos[1]. Observe que há $2^{|I_0(x^*)|}$ desses problemas, onde $I_0(x^*) = \{i \mid g_i(x^*) = h_i(x^*) = 0\}$. Ou seja, verificar que x^* é B-estacionário envolve, em princípio, verificar a otimalidade em $2^{|I_0(x^*)|}$ problemas. O interesse na B-estacionariedade justifica-se por ser uma propriedade forte, muito próxima à S-estacionariedade, mas, ao contrário desta, satisfeita por minimizadores sob hipóteses pouco exigentes. Em particular, em um ponto B-estacionário não há direções de descida de primeira ordem, enquanto que a W, C e M-estacionariedade não as eliminam completamente (imagine a direção $-\nabla f$ na Figura 6, página 45). Um desafio é estabelecer métodos com convergência a B-estacionários, especialmente sem a necessidade de enumerações. A dificuldade mantém-se mesmo quando a complementariedade é formada por restrições lineares, veja [90] e suas referências.

Esquemas de regularização

Esquemas de regularização são métodos desenvolvidos especificamente para MPCCs (consulte a introdução para referências). Sabemos que tais métodos têm boas propriedades de convergência, mas as perdem se, relativo aos seus subproblemas, obtemos pontos estacionários apenas aproximados [96]. Esta é uma questão pertinente pois, na prática numérica, aproximações são a realidade, especialmente mediante não linearidades. Somente muito recentemente foi proposto um método de regularização que mantém bons resultados teóricos de convergência sob inexatidão [110] (artigo ainda não publicado). De qualquer forma, pelo menos de nosso conhecimento todos esses métodos, quando sujeitos à inexatidão, necessitam de alguma limitação de multiplicadores, usualmente requerendo uma hipótese do tipo Mangasarian Fromovitz – MPCC-MFCQ. De fato, o grande desafio para métodos que trabalham com sequências duais parece ser justamente controlar a velocidade de explosão de seus multiplicadores. Um desafio, portanto, é flexibilizar hipóteses de limitação mantendo convergência.

[1] Em otimização não linear padrão, é comum estabelecer a condição necessária de otimalidade de primeira ordem sobre o chamado cone tangente de Bouligand (veja [85]). Esse exato cone é considerado no contexto de MPCCs, por exemplo em [107]. O termo "B-estacionário" usado aqui, devido a Scheel e Scholtes [135], não faz referência a este cone, mas sim a uma linearização sua específica para MPCCs. Isso pode causar confusão. Em [54], o termo "B-estacionário linearizado" é utilizado como referência ao conceito considerado aqui. Preferimos a definição de Scheel e Scholtes, pois é adequada ao ponto de vista algorítmico e amplamente utilizada.

Mathematical problems with vanishing constraints (MPVCs)

MPVCs são problemas muito semelhantes a MPCCs, e foram introduzidos por Achtziger e Kanzow em 2008 [1]. Enquanto que em um MPCC temos tipicamente as restrições $g(x) \geqslant 0$, $h(x) \geqslant 0$ e as complementariedades $g_i(x)h_i(x) \leqslant 0$, em um MPVC a restrição $g(x) \geqslant 0$ é desconsiderada. Assim, quando $h_i(x) = 0$, a função g_i não restringe o domínio do problema, dado que a complementariedade $g_i(x)h_i(x) \leqslant 0$ sempre é satisfeita. Ou seja, g_i é simplesmente ignorada (daí vem o termo "*vanishing*"). Como ocorre com MPCCs, MPVCs também sofrem de certo grau de degeneração, sendo consideradas, portanto, condições de estacionariedade e CQs específicas. No entanto, a simples ausência de uma das restrições envolvidas na complementariedade tornam MPVCs "mais tratáveis" que um MPCC. Isso justifica a profusão de artigos dedicados ao problema nos últimos anos. Acerca dos MPVCs, citamos possíveis temas de pesquisa: novas condições de qualificação; convergência de algoritmos; novos métodos de resolução; estabilidade de pontos estacionários; condições sequenciais de otimalidade específicas; e aplicações.

Referências

[1] W. ACHTZIGER e C. KANZOW. Mathematical programs with vanishing constraints: optimality conditions and constraint qualifications. *Mathematical Programming*, 114(1):69–99, 2008. doi:10.1007/s10107-006-0083-3.

[2] S. ALBRECHT e M. ULBRICH. Mathematical programs with complementarity constraints in the context of inverse optimal control for locomotion. *Optimization Methods and Software*, páginas 1–29, 2016. doi:10.1080/10556788.2016.1225212.

[3] R. ANDREANI, R. BEHLING, G. HAESER, e P. J. S. SILVA. On second-order optimality conditions in nonlinear optimization. *Optimization Methods and Software*, 32(1):22–38, 2016. doi:10.1080/10556788.2016.1188926.

[4] 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(2):209–236, 2010. doi:10.1007/s10589-009-9240-y.

[5] 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(4):1286–1309, 2007. doi:10.1137/060654797.

[6] R. ANDREANI, E. G. BIRGIN, J. M. MARTÍNEZ, e J. YUAN. Spectral projected gradient and variable metric methods for optimization with linear inequalities. *IMA Journal of Numerical Analysis*, 25(2):221–252, 2005. doi:10.1093/imanum/drh020.

[7] R. ANDREANI, C. DUNDER, e J. M. MARTÍNEZ. Nonlinear-programming reformulation of the order-value optimization problem. *Mathematical Methods of Operations Research*, 61(3):365–384, 2005. doi:10.1007/s001860400410.

[8] R. ANDREANI, C. E. ECHAGÜE, e M. L. SCHUVERDT. Constant-rank condition and second-order constraint qualification. *Journal of Optimization Theory and Applications*, 146(2):255–266, 2010. doi:10.1007/s10957-010-9671-8.

[9] R. ANDREANI, N. S. FAZZIO, M. L. SCHUVERDT, e L. D. SECCHIN. A sequential optimality condition related to the quasinormality constraint qualification and its

algorithmic consequences. Relatório técnico, 2017. Último acesso em 09/12/17. URL: http://www.optimization-online.org/DB_HTML/2017/09/6194.html.

[10] R. ANDREANI, G. HAESER, e J. M. MARTÍNEZ. On sequential optimality conditions for smooth constrained optimization. *Optimization*, 60(5):627–641, 2011. doi:10.1080/02331930903578700.

[11] R. ANDREANI, G. HAESER, A. RAMOS, e P. J. S. SILVA. A second-order sequential optimality condition associated to the convergence of optimization algorithms. *IMA Journal of Numerical Analysis*, 37(4):1902–1929, 2017. doi:10.1093/imanum/drw064.

[12] R. ANDREANI, G. HAESER, M. L. SCHUVERDT, e P. J. S. SILVA. A relaxed constant positive linear dependence constraint qualification and applications. *Mathematical Programming*, 135(1):255–273, 2012. doi:10.1007/s10107-011-0456-0.

[13] R. ANDREANI, G. HAESER, M. L. SCHUVERDT, e P. J. S. SILVA. Two new weak constraint qualifications and applications. *SIAM Journal on Optimization*, 22(3):1109–1135, 2012. doi:10.1137/110843939.

[14] R. ANDREANI, J. M. MARTINEZ, e M. L. SCHUVERDT. On the relation between constant positive linear dependence condition and quasinormality constraint qualification. *Journal of Optimization Theory and Applications*, 125(2):473–483, 2005. doi:10.1007/s10957-004-1861-9.

[15] R. ANDREANI, J. M. MARTÍNEZ, A. RAMOS, e P. J. S. SILVA. A cone-continuity constraint qualification and algorithmic consequences. *SIAM Journal on Optimization*, 26(1):96–110, 2016. doi:10.1137/15M1008488.

[16] R. ANDREANI, J. M. MARTÍNEZ, A. RAMOS, e P. J. S. SILVA. Strict constraint qualifications and sequential optimality conditions for constrained optimization. *Mathematics of Operations Research*, páginas 1–25, 2018. doi:10.1287/moor.2017.0879.

[17] R. ANDREANI, J. M. MARTÍNEZ, e M. L. SCHUVERDT. On second-order optimality conditions for nonlinear programming. *Optimization*, 56(5–6):529–542, 2007. doi:10.1080/02331930701618617.

[18] R. ANDREANI, J. M. MARTÍNEZ, e B. F. SVAITER. A new sequential optimality condition for constrained optimization and algorithmic consequences. *SIAM Journal on Optimization*, 20(6):3533–3554, 2010. doi:10.1137/090777189.

[19] R. ANDREANI e L. D. SECCHIN. A note on the convergence of an augmented Lagrangian algorithm to second-order stationary points. In *Proceeding Series of the*

Brazilian Society of Computational and Applied Mathematics, volume 6. Sociedade Brasileira de Matemática Aplicada e Computacional, 2018. doi:10.5540/03.2018.006.01.0303.

[20] R. ANDREANI e P. J. S. SILVA. Constant rank constraint qualifications: a geometric introduction. *Pesquisa Operacional*, 34(3):481–494, 2014. doi:10.1590/0101-7438.2014.034.03.0481.

[21] M. ANDRETTA, E. G. BIRGIN, e J. M. MARTÍNEZ. Practical active-set Euclidian trust-region method with spectral projected gradients for bound-constrained minimization. *Optimization*, 54(3):305–325, 2005. doi:10.1080/02331930500100270.

[22] M. ANITESCU. Degenerate nonlinear programming with a quadratic growth condition. *SIAM Journal on Optimization*, 10(4):1116–1135, 2000. doi:10.1137/S1052623499359178.

[23] M. ANITESCU. Global convergence of an elastic mode approach for a class of mathematical programs with complementarity constraints. *SIAM Journal on Optimization*, 16(1):120–145, 2005. doi:10.1137/040606855.

[24] M. ANITESCU. On using the elastic mode in nonlinear programming approaches to mathematical programs with complementarity constraints. *SIAM Journal on Optimization*, 15(4):1203–1236, 2005. doi:10.1137/S1052623402401221.

[25] M. ANITESCU, P. TSENG, e S. J. WRIGHT. Elastic-mode algorithms for mathematical programs with equilibrium constraints: global convergence and stationarity properties. *Mathematical Programming*, 110(2):337–371, 2007. doi:10.1007/s10107-006-0005-4.

[26] A. BACCARI. On the classical necessary second-order optimality conditions. *Journal of Optimization Theory and Applications*, 123(1):213–221, 2004. doi:10.1023/B:JOTA.0000043998.04008.e6.

[27] A. BACCARI e A. TRAD. On the classical necessary second-order optimality conditions in the presence of equality and inequality constraints. *SIAM Journal on Optimization*, 15(2):394–408, 2005. doi:10.1137/S105262340342122X.

[28] M. S. BAZARAA, H. D. SHERALI, e C. M. SHETTY. *Nonlinear Programming. Theory and Algorithms*. John Wiley & Sons, New Jersey, 3ª edição, 2006.

[29] D. P. BERTSEKAS. *Convex Optimization Algorithms*. Athena Scientific, Belmont, Massachusetts, 2015.

[30] D. P. BERTSEKAS e A. E. OZDAGLAR. Pseudonormality and a Lagrange multiplier theory for constrained optimization. *Journal of Optimization Theory and Applications*, 114(2):287–343, 2002. doi:10.1023/A:1016083601322.

[31] E. G. BIRGIN, L. F. BUENO, e J. M. MARTÍNEZ. Sequential equality-constrained optimization for nonlinear programming. *Computational Optimization and Applications*, 65(3):699–721, 2016. doi:10.1007/s10589-016-9849-6.

[32] E. G. BIRGIN, R. A. CASTILLO, e J. M. MARTÍNEZ. Numerical comparison of augmented Lagrangian algorithms for nonconvex problems. *Computational Optimization and Applications*, 31(1):31–55, 2005. doi:10.1007/s10589-005-1066-7.

[33] E. G. BIRGIN, C. A. FLOUDAS, e J. M. MARTÍNEZ. Global minimization using an augmented Lagrangian method with variable lower-level constraints. *Mathematical Programming*, 125(1):139–162, 2010. doi:10.1007/s10107-009-0264-y.

[34] E. G. BIRGIN e J. M. GENTIL. Evaluating bound-constrained minimization software. *Computational Optimization and Applications*, 53(2):347–373, 2012. doi:10.1007/s10589-012-9466-y.

[35] E. G. BIRGIN, G. HAESER, e A. RAMOS. Augmented Lagrangians with constrained subproblems and convergence to second-order stationary points. *Computational Optimization and Applications*, 2017. doi:10.1007/s10589-017-9937-2.

[36] E. G. BIRGIN e J. M. MARTÍNEZ. Large-scale active-set box-constrained optimization method with spectral projected gradients. *Computational Optimization and Applications*, 23(1):101–125, 2002. doi:10.1023/A:1019928808826.

[37] E. G. BIRGIN e J. M. MARTÍNEZ. Structured minimal-memory inexact quasi-Newton method and secant preconditioners for augmented Lagrangian optimization. *Computational Optimization and Applications*, 39(1):1–16, 2008. doi:10.1007/s10589-007-9050-z.

[38] E. G. BIRGIN e J. M. MARTÍNEZ. Improving ultimate convergence of an augmented Lagrangian method. *Optimization Methods and Software*, 23(2):177–195, 2008. doi:10.1080/10556780701577730.

[39] E. G. BIRGIN e J. M. MARTÍNEZ. *Practical Augmented Lagrangian Methods for Constrained Optimization*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2014. doi:10.1137/1.9781611973365.

[40] E. G. BIRGIN, J. M. MARTÍNEZ, e M. RAYDAN. Inexact spectral projected gradient methods on convex sets. *IMA Journal of Numerical Analysis*, 23(4):539–559, 2003. doi:10.1093/imanum/23.4.539.

[41] M. BOUNKHEL. *Regularity Concepts in Nonsmooth Analysis. Theory and Applications*, volume 59 de *Springer Optimization and Its Applications*. Springer, New York, 2012. doi:10.1007/978-1-4614-1019-5.

[42] M. BRANDA, M. BUCHER, M. ČERVINKA, e A. SCHWARTZ. Convergence of a Scholtes-type regularization method for cardinality-constrained optimization problems with an application in sparse robust portfolio optimization. Relatório técnico, 2017. Último acesso em 09/12/17. URL: https://arxiv.org/abs/1703.10637.

[43] O. P. BURDAKOV, C. KANZOW, e A. SCHWARTZ. Mathematical programs with cardinality constraints: reformulation by complementarity-type conditions and a regularization method. *SIAM Journal on Optimization*, 26(1):397–425, 2016. doi:10.1137/140978077.

[44] J. CANO, J. M. MOGUERZA, e F. J. PRIETO. Using improved directions of negative curvature for the solution of bound-constrained nonconvex problems. *Journal of Optimization Theory and Applications*, 2017. doi:10.1007/s10957-017-1137-9.

[45] N. H. CHIEU e G. M. LEE. Constraint qualifications for mathematical programs with equilibrium constraints and their local preservation property. *Journal of Optimization Theory and Applications*, 163(3):755–776, 2014. doi:10.1007/s10957-014-0546-2.

[46] F. H. CLARKE. A new approach to Lagrange multipliers. *Mathematics of Operations Research*, 1(2):165–174, 1976. doi:10.1287/moor.1.2.165.

[47] S. DEMPE. *Foundations of Bilevel Programming*, volume 61 de *Nonconvex Optimization and Its Applications*. Springer US, New York, 1ª edição, 2002. doi:10.1007/b101970.

[48] S. DEMPE. Annotated bibliography on bilevel programming and mathematical programs with equilibrium constraints. *Optimization*, 52(3):333–359, 2003. doi:10.1080/0233193031000149894.

[49] N. ECHEBEST, M. D. SÁNCHEZ, e M. L. SCHUVERDT. Convergence results of an augmented Lagrangian method using the exponential penalty function. *Journal of Optimization Theory and Applications*, 168(1):92–108, 2016. doi:10.1007/s10957-015-0735-7.

[50] A. EHRENMANN e K. NEUHOFF. A comparison of electricity market designs in networks. *Operations Research*, 57(2):274–286, 2009. doi:10.1287/opre.1080.0624.

[51] F. FACCHINEI, H. JIANG, e L. QI. A smoothing method for mathematical programs with equilibrium constraints. *Mathematical Programming*, 85(1):107–134, 1999. doi:10.1007/s10107990015a.

[52] M. C. FERRIS, S. P. DIRKSE, e A. MEERAUS. Mathematical programs with equilibrium constraints: automatic reformulation and solution via constrained optimization. Relatório técnico, Oxford University Numerical Analysis Group Research Report NA 02-11, 2002. URL: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.9.7017.

[53] M. C. FERRIS e J. S. PANG. Engineering and economic applications of complementarity problems. *SIAM Review*, 39(4):669–713, 1997. doi:10.1137/S0036144595285963.

[54] M. L. FLEGEL e C. KANZOW. Abadie-type constraint qualification for mathematical programs with equilibrium constraints. *Journal of Optimization Theory and Applications*, 124(3):595–614, 2005. doi:10.1007/s10957-004-1176-x.

[55] M. L. FLEGEL e C. KANZOW. On M-stationary points for mathematical programs with equilibrium constraints. *Journal of Mathematical Analysis and Applications*, 310(1):286–302, 2005. doi:10.1016/j.jmaa.2005.02.011.

[56] M. L. FLEGEL e C. KANZOW. On the Guignard constraint qualification for mathematical programs with equilibrium constraints. *Optimization*, 54(6):517–534, 2005. doi:10.1080/02331930500342591.

[57] M. L. FLEGEL e C. KANZOW. A direct proof for M-stationarity under MPEC-GCQ for mathematical programs with equilibrium constraints. In S. DEMPE e V. KALASHNIKOV, editors, *Optimization with Multivalued Mappings: Theory, Applications, and Algorithms*, volume 2 de *Springer Optimization and Its Applications*, páginas 111–122. Springer US, Boston, MA, 2006. doi:10.1007/0-387-34221-4_6.

[58] M. L. FLEGEL, C. KANZOW, e J. V. OUTRATA. Optimality conditions for disjunctive programs with application to mathematical programs with equilibrium constraints. *Set-Valued Analysis*, 15(2):139–162, 2007. doi:10.1007/s11228-006-0033-5.

[59] R. FLETCHER e S. LEYFFER. Numerical experience with solving MPECs as NLPs. Relatório técnico, University of Dundee Report NA 210, 2002.

[60] R. FLETCHER e S. LEYFFER. Solving mathematical programs with complementarity constraints as nonlinear programs. *Optimization Methods and Software*, 19(1):15–40, 2004. doi:10.1080/10556780410001654241.

[61] R. FLETCHER, S. LEYFFER, D. RALPH, e S. SCHOLTES. Local convergence of SQP methods for mathematical programs with equilibrium constraints. *SIAM Journal on Optimization*, 17(1):259–286, 2006. doi:10.1137/S1052623402407382.

[62] R. FOURER, D. M. GAY, e B. W. KERNIGHAN. *AMPL: A Modeling Language for Mathematical Programming*. Duxbury Resource Center, 2ª edição, 2003. Disponível em http://ampl.com/resources/the-ampl-book/chapter-downloads. Acesso em Novembro de 2016.

[63] M. FUKUSHIMA, Z.-Q. LUO, e J.-S. PANG. A globally convergent sequential quadratic programming algorithm for mathematical programs with linear complementarity constraints. *Computational Optimization and Applications*, 10(1):5–34, 1998. doi:10.1023/A:1018359900133.

[64] H. GFRERER. Optimality conditions for disjunctive programs based on generalized differentiation with application to mathematical programs with equilibrium constraints. *SIAM Journal on Optimization*, 24(2):898–931, 2014. doi:10.1137/130914449.

[65] P. E. GILL, V. KUNGURTSEV, e D. P. ROBINSON. A stabilized SQP method: global convergence. *IMA Journal of Numerical Analysis*, 37(1):407–443, 2017. doi:10.1093/imanum/drw004.

[66] G. GIORGI, B. JIMÉNEZ, e V. NOVO. Approximate Karush-Kuhn-Tucker condition in multiobjective optimization. *Journal of Optimization Theory and Applications*, 171(1):70–89, 2016. doi:10.1007/s10957-016-0986-y.

[67] F. J. GOULD e J. W. TOLLE. A necessary and sufficient qualification for constrained optimization. *SIAM Journal on Applied Mathematics*, 20(2):164–172, 1971. doi:10.1137/0120021.

[68] N. I. M. GOULD e P. L. TOINT. A note on the convergence of barrier algorithms to second-order necessary points. *Mathematical Programming*, 85(2):433–438, 1999. doi:10.1007/s101070050066.

[69] M. GUIGNARD. Generalized Kuhn-Tucker conditions for mathematical programming problems in a Banach space. *SIAM Journal on Control*, 7(2):232–241, 1969. doi:10.1137/0307016.

[70] L. GUO e G.-H. LIN. Notes on some constraint qualifications for mathematical programs with equilibrium constraints. *Journal of Optimization Theory and Applications*, 156(3):600–616, 2013. doi:10.1007/s10957-012-0084-8.

[71] L. GUO, G.-H. LIN, e J. J. YE. Second-order optimality conditions for mathematical programs with equilibrium constraints. *Journal of Optimization Theory and Applications*, 158(1):33–64, 2013. doi:10.1007/s10957-012-0228-x.

[72] L. GUO, G.-H. LIN, e J. J. YE. Solving mathematical programs with equilibrium constraints. *Journal of Optimization Theory and Applications*, 166(1):234–256, 2015. doi:10.1007/s10957-014-0699-z.

[73] L. GUO, G.-H. LIN, D. ZHANG, e D. ZHU. An MPEC reformulation of an EPEC model for electricity markets. *Operations Research Letters*, 43(3):262–267, 2015. doi:10.1016/j.orl.2015.03.001.

[74] G. HAESER. A second-order optimality condition with first and second-order complementarity associated to global convergence of algorithms. Relatório técnico, 2016. Último acesso em 09/12/17. URL: http://www.optimization-online.org/DB_HTML/2016/10/5690.html.

[75] G. HAESER e V. V. DE MELO. Convergence detection for optimization algorithms: Approximate-KKT stopping criterion when lagrange multipliers are not available. *Operations Research Letters*, 43(5):484–488, 2015. doi:https://doi.org/10.1016/j.orl.2015.06.009.

[76] G. HAESER, H. LIU, e Y. YE. Optimality condition and complexity analysis for linearly-constrained optimization without differentiability on the boundary. Relatório técnico, 2017. Último acesso em 09/12/17. URL: http://www.optimization-online.org/DB_HTML/2017/02/5861.html.

[77] G. HAESER e M. L. SCHUVERDT. On approximate KKT condition and its extension to continuous variational inequalities. *Journal of Optimization Theory and Applications*, 149(3):528–539, 2011. doi:10.1007/s10957-011-9802-x.

[78] M. R. HESTENES. Multiplier and gradient methods. *Journal of Optimization Theory and Applications*, 4(5):303–320, 1969. doi:10.1007/BF00927673.

[79] M. R. HESTENES. *Optimization Theory: The Finite Dimensional Case*. John Wiley & Sons, New York, 1975.

[80] T. HOHEISEL, C. KANZOW, e A. SCHWARTZ. Convergence of a local regularization approach for mathematical programmes with complementarity or vanishing constraints. *Optimization Methods and Software*, 27(3):483–512, 2012. doi:10.1080/10556788.2010.535170.

[81] T. HOHEISEL, C. KANZOW, e A. SCHWARTZ. Mathematical programs with vanishing constraints: a new regularization approach with strong convergence properties. *Optimization*, 61(6):619–636, 2012. doi:10.1080/02331934.2011.608164.

[82] T. HOHEISEL, C. KANZOW, e A. SCHWARTZ. Theoretical and numerical comparison of relaxation methods for mathematical programs with complementarity constraints. *Mathematical Programming*, 137(1):257–288, 2013. doi:10.1007/s10107-011-0488-5.

[83] HSL. A collection of fortran codes for large scale scientific computation. Acesso em 10/11/2016. URL: http://www.hsl.rl.ac.uk.

[84] X. M. HU e D. RALPH. Convergence of a penalty method for mathematical programming with complementarity constraints. *Journal of Optimization Theory and Applications*, 123(2):365–390, 2004. doi:10.1007/s10957-004-5154-0.

[85] A. F. IZMAILOV e M. V. SOLODOV. *Otimização, vol 1. Condições de Otimalidade, Elementos de Análise Convexa e de Dualidade*. IMPA, Rio de Janeiro, 2005.

[86] A. F. IZMAILOV e M. V. SOLODOV. An active-set Newton method for mathematical programs with complementarity constraints. *SIAM Journal on Optimization*, 19(3):1003–1027, 2008. doi:10.1137/070690882.

[87] A. F. IZMAILOV, M. V. SOLODOV, e E. I. USKOV. Global convergence of augmented Lagrangian methods applied to optimization problems with degenerate constraints, including problems with complementarity constraints. *SIAM Journal on Optimization*, 22(4):1579–1606, 2012. doi:10.1137/120868359.

[88] R. JANIN. Directional derivative of the marginal function in nonlinear programming. In A. V. FIACCO, editor, *Sensitivity, Stability and Parametric Analysis*, volume 21 de *Mathematical Programming Studies*, páginas 110–126. Springer Berlin Heidelberg, Berlin, Heidelberg, 1984. doi:10.1007/BFb0121214.

[89] H. JIANG e D. RALPH. Smooth SQP methods for mathematical programs with nonlinear complementarity constraints. *SIAM Journal on Optimization*, 10(3):779–808, 2000. doi:10.1137/S1052623497332329.

[90] J. J. JÚDICE. Algorithms for linear programming with linear complementarity constraints. *TOP*, 20(1):4–25, 2012. doi:10.1007/s11750-011-0228-2.

[91] A. KADRANI, J.-P. DUSSAULT, e A. BENCHAKROUN. A new regularization scheme for mathematical programs with complementarity constraints. *SIAM Journal on Optimization*, 20(1):78–103, 2009. doi:10.1137/070705490.

[92] A. KADRANI, J. P. DUSSAULT, e A. BENCHAKROUN. A globally convergent algorithm for MPCC. *EURO Journal on Computational Optimization*, 3(3):263–296, 2015. doi:10.1007/s13675-015-0044-9.

[93] C. KANZOW e A. SCHWARTZ. Mathematical programs with equilibrium constraints: enhanced Fritz John-conditions, new constraint qualifications, and improved exact penalty results. *SIAM Journal on Optimization*, 20(5):2730–2753, 2010. doi:10.1137/090774975.

[94] C. KANZOW e A. SCHWARTZ. A new regularization method for mathematical programs with complementarity constraints with strong convergence properties. *SIAM Journal on Optimization*, 23(2):770–798, 2013. doi:10.1137/100802487.

[95] C. KANZOW e A. SCHWARTZ. Convergence properties of the inexact Lin-Fukushima relaxation method for mathematical programs with complementarity constraints. *Computational Optimization and Applications*, 59(1):249–262, 2014. doi:10.1007/s10589-013-9575-2.

[96] C. KANZOW e A. SCHWARTZ. The price of inexactness: convergence properties of relaxation methods for mathematical programs with complementarity constraints revisited. *Mathematics of Operations Research*, 40(2):253–275, 2015. doi:10.1287/moor.2014.0667.

[97] C. KANZOW e D. STECK. An example comparing the standard and safeguarded augmented Lagrangian methods. *Operations Research Letters*, 45(6):598–603, 2017. doi:10.1016/j.orl.2017.09.005.

[98] Y.-C. LEE, J.-S. PANG, e J. E. MITCHELL. Global resolution of the support vector machine regression parameters selection problem with LPCC. *EURO Journal on Computational Optimization*, 3(3):197–261, 2015. doi:10.1007/s13675-015-0041-z.

[99] S. LEYFFER. Complementarity constraints as nonlinear equations: theory and numerical experience. In S. DEMPE e V. KALASHNIKOV, editors, *Optimization with Multivalued Mappings: Theory, Applications, and Algorithms*, volume 2 de *Springer Optimization and Its Applications*, páginas 169–208. Springer US, Boston, MA, 2006. doi:10.1007/0-387-34221-4_9.

[100] S. LEYFFER, G. LÓPEZ-CALVA, e J. NOCEDAL. Interior methods for mathematical programs with complementarity constraints. *SIAM Journal on Optimization*, 17(1):52–77, 2006. doi:10.1137/040621065.

[101] S. LEYFFER e T. S. MUNSON. A globally convergent filter method for MPECs. Relatório técnico, Argonne National Laboratory, 2007. URL: http://www.mcs.anl.gov/~tmunson/papers/slpec.pdf.

[102] C. LI e L. GUO. A single-level reformulation of mixed integer bilevel programming problems. *Operations Research Letters*, 45(1):1–5, 2017. doi:10.1016/j.orl.2016.10.013.

[103] G.-H. LIN e M. FUKUSHIMA. A modified relaxation scheme for mathematical programs with complementarity constraints. *Annals of Operations Research*, 133(1):63–84, 2005. doi:10.1007/s10479-004-5024-z.

[104] X. LIU e J. SUN. Generalized stationary points and an interior-point method for mathematical programs with equilibrium constraints. *Mathematical Programming*, 101(1):231–261, 2004. doi:10.1007/s10107-004-0543-6.

[105] H. Z. LUO, X. L. SUN, e Y. F. XU. Convergence properties of modified and partially-augmented Lagrangian methods for mathematical programs with complementarity constraints. *Journal of Optimization Theory and Applications*, 145(3):489–506, 2010. doi:10.1007/s10957-009-9642-0.

[106] H. Z. LUO, X. L. SUN, Y. F. XU, e H. X. WU. On the convergence properties of modified augmented Lagrangian methods for mathematical programming with complementarity constraints. *Journal of Global Optimization*, 46(2):217–232, 2010. doi:10.1007/s10898-009-9419-x.

[107] Z.-Q. LUO, J.-S. PANG, e D. RALPH. *Mathematical Programs with Equilibrium Constraints*. Cambridge University Press, United Kingdom, 1996. doi:10.1017/CBO9780511983658.

[108] J. M. MARTÍNEZ e B. F. SVAITER. A practical optimality condition without constraint qualifications for nonlinear programming. *Journal of Optimization Theory and Applications*, 118(1):117–133, 2003. doi:10.1023/A:1024791525441.

[109] J. M. MARTÍNEZ. *Otimização prática usando o Lagrangiano aumentado*. Departamento de Matemática Aplicada IMECC/Unicamp, 2009. URL: http://www.ime.unicamp.br/~martinez/lagraum.pdf.

[110] T. MIGOT, J.-P. DUSSAULT, M. HADDOU, e A. KADRANI. How to compute a local minimum of the MPCC. Relatório técnico, 2017. Último acesso em 09/12/17. URL: https://hal.archives-ouvertes.fr/hal-01525402.

[111] T. MIGOT, M. HADDOU, e J.-P. DUSSAULT. The new butterfly relaxation methods for mathematical program with complementarity constraints. Relatório técnico, 2016. Último acesso em 09/12/17. URL: https://hal.archives-ouvertes.fr/hal-01525399.

[112] L. MINCHENKO e S. STAKHOVSKI. On relaxed constant rank regularity condition in mathematical programming. *Optimization*, 60(4):429–440, 2011. doi:10.1080/02331930902971377.

[113] L. MINCHENKO e S. STAKHOVSKI. Parametric nonlinear programming problems under the relaxed constant rank condition. *SIAM Journal on Optimization*, 21(1):314–332, 2011. doi:10.1137/090761318.

[114] J. M. MOGUERZA e F. J. PRIETO. An augmented Lagrangian interior-point method using directions of negative curvature. *Mathematical Programming*, 95(3):573–616, 2003. doi:10.1007/s10107-002-0360-8.

[115] M. T. T. MONTEIRO e H. S. RODRIGUES. Combining the regularization strategy and the SQP to solve MPCC – a MATLAB implementation. *Journal of Computational and Applied Mathematics*, 235(18):5348–5356, 2011. doi:10.1016/j.cam.2010.05.008.

[116] B. S. MORDUKHOVICH. *Variational Analysis and Generalized Differentiation, vol 1. Basic Theory*. Grundlehren der mathematischen Wissenschaften. Springer-Verlag Berlin Heidelberg, New York, 2006. doi:10.1007/3-540-31247-1.

[117] K. G. MURTY e S. N. KABADI. Some NP-complete problems in quadratic and nonlinear programming. *Mathematical Programming*, 39(2):117–129, 1987. doi:10.1007/BF02592948.

[118] J. V. OUTRATA. Optimality conditions for a class of mathematical programs with equilibrium constraints: strongly regular case. *Kybernetika*, 35(2):177–193, 1999.

[119] J. V. OUTRATA, M. KOCVARA, e J. ZOWE. *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints*, volume 28 de *Nonconvex Optimization and Its Applications*. Springer US, New York, 1ª edição, 1998. doi:10.1007/978-1-4757-2825-5.

[120] E. E. OVTCHINNIKOV. Jacobi correction equation, line search, and conjugate gradients in Hermitian eigenvalue computation I: computing an extreme eigenvalue. *SIAM Journal on Numerical Analysis*, 46(5):2567–2592, 2008. doi:10.1137/070688742.

[121] E. E. OVTCHINNIKOV e J. K. REID. A preconditioned block conjugate gradient algorithm for computing extreme eigenpairsexistence of augmented Lagrange multipliers: reduction to exact penalty functions and localization principle of symmetric and Hermitian problems. Relatório Técnico RAL-TR-2010-019, 2010. URL: ftp://ftp.numerical.rl.ac.uk/pub/reports/orRAL2010019.pdf.

[122] A. E. OZDAGLAR e D. P. BERTSEKAS. The relation between pseudonormality and quasiregularity in constrained optimization. *Optimization Methods and Software*, 19(5):493–506, 2004. doi:10.1080/10556780410001709420.

[123] J.-S. PANG e M. FUKUSHIMA. Complementarity constraint qualifications and simplified B-stationarity conditions for mathematical programs with equilibrium constraints. *Computational Optimization and Applications*, 13(1):111–136, 1999. doi:10.1023/A:1008656806889.

[124] J.-S. PANG e S. A. GABRIEL. NE/SQP: A robust algorithm for the nonlinear complementarity problem. *Mathematical Programming*, 60(1):295–337, 1993. doi:10.1007/BF01580617.

[125] M. J. D. POWELL. A method for nonlinear constraints in minimization problems. In R. FLETCHER, editor, *Optimization*, páginas 283–298. Academic Press, New York, 1969.

[126] L. QI e Z. WEI. On the constant positive linear dependence condition and its application to SQP methods. *SIAM Journal on Optimization*, 10(4):963–981, 2000. doi:10.1137/S1052623497326629.

[127] A. U. RAGHUNATHAN e L. T. BIEGLER. Mathematical programs with equilibrium constraints (MPECs) in process engineering. *Computers & Chemical Engineering*, 27(10):1381–1392, 2003. doi:10.1016/S0098-1354(03)00092-9.

[128] A. U. RAGHUNATHAN e L. T. BIEGLER. An interior point method for mathematical programs with complementarity constraints (MPCCs). *SIAM Journal on Optimization*, 15(3):720–750, 2005. doi:10.1137/S1052623403429081.

[129] D. RALPH e S. J. WRIGHT. Some properties of regularization and penalization schemes for MPECs. *Optimization Methods and Software*, 19(5):527–556, 2004. doi:10.1080/10556780410001709439.

[130] A. RAMOS. Mathematical programs with equilibrium constraints: a sequential optimality condition, new constraint qualifications and algorithmic consequences. Relatório técnico, 2016. Último acesso em 09/12/17. URL: http://www.optimization-online.org/DB_HTML/2016/04/5423.html.

[131] A. RAMOS. Two new weak constraint qualifications for mathematical programs with equilibrium constraints annd applications. Relatório técnico, 2017. Último acesso em 07/02/18. URL: http://www.optimization-online.org/DB_HTML/2017/05/6043.html.

[132] R. T. ROCKAFELLAR. Augmented Lagrange multiplier functions and duality in nonconvex programming. *SIAM Journal on Control*, 12(2):268–285, 1974. doi:10.1137/0312021.

[133] R. T. ROCKAFELLAR e R. J.-B. WETS. *Variational Analysis*, volume 317 de *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag Berlin Heidelberg, 1ª edição, 1998. doi:10.1007/978-3-642-02431-3.

[134] A. RUBINOV e X. YANG. *Lagrange-type Functions in Constrained Non-Convex Optimization*, volume 85 de *Applied Optimization*. Springer US, 2003. doi:10.1007/978-1-4419-9172-0.

[135] H. SCHEEL e S. SCHOLTES. Mathematical programs with complementarity constraints: stationarity, optimality, and sensitivity. *Mathematics of Operations Research*, 25(1):1–22, 2000. doi:10.1287/moor.25.1.1.15213.

[136] L. SCHEWE e M. SCHMIDT. Computing feasible points for MINLPs with MPECs. Relatório técnico, 2016. Último acesso em 09/12/17. URL: http://www.optimization-online.org/DB_HTML/2016/12/5778.html.

[137] S. SCHOLTES. Convergence properties of a regularization scheme for mathematical programs with complementarity constraints. *SIAM Journal on Optimization*, 11(4):918–936, 2001. doi:10.1137/S1052623499361233.

[138] S. STEFFENSEN e M. ULBRICH. A new relaxation scheme for mathematical programs with equilibrium constraints. *SIAM Journal on Optimization*, 20(5):2504–2539, 2010. doi:10.1137/090748883.

[139] G. W. STEWART e J.-G. SUN. *Matriz Pertubation Theory*. Computer Science and Scientific Computing. Academic Press, New York, 1990.

[140] TANGO. Trustable algorithms for nonlinear general optimization. Acesso em 10/11/2016. URL: http://www.ime.usp.br/~egbirgin/tango/.

[141] H. XU, J.-S. PANG, F. ORDÓÑEZ, e M. DESSOUKY. Complementarity models for traffic equilibrium with ridesharing. *Transportation Research Part B: Methodological*, 81, Part 1:161–182, 2015. doi:10.1016/j.trb.2015.08.013.

[142] X. Q. YANG e X. X. HUANG. Lower-order penalty methods for mathematical programs with complementarity constraints. *Optimization Methods and Software*, 19(6):693–720, 2004. doi:10.1080/10556780410001697659.

[143] X. Q. YANG e X. X. HUANG. Convergence analysis of an augmented Lagrangian method for mathematical programs with complementarity constraints. *Nonlinear*

Analysis: Theory, Methods & Applications, 63(5–7):e2247–e2256, 2005. doi:10.1016/j.na.2005.03.027.

[144] J. J. YE. Necessary and sufficient optimality conditions for mathematical programs with equilibrium constraints. *Journal of Mathematical Analysis and Applications*, 307(1):350–369, 2005. doi:10.1016/j.jmaa.2004.10.032.

[145] J. J. YE e J. ZHANG. Enhanced Karush-Kuhn-Tucker conditions for mathematical programs with equilibrium constraints. *Journal of Optimization Theory and Applications*, 163(3):777–794, 2014. doi:10.1007/s10957-013-0493-3.

Anexos

ANEXO A

Resultados numéricos detalhados

Reproduzimos a seguir os testes numéricos realizados com os problemas da coletânea MacMPEC. Reportamos os testes em duas tabelas: na primeira, a complementariedade é tratada como desigualdade; na segunda tabela, é tratada como igualdade via inserção de folgas. Veja a discussão do Capítulo 4 (página 72) para detalhes.

A coluna "Problema" traz a instância com mesma nomenclatura dada em MacMPEC; as colunas "FO", "Saída" e "Inviab" trazem, respectivamente, o valor da função objetivo no término da execução do algoritmo, o motivo de parada do algoritmo ("conv" se o algoritmo declarou convergência ou "$\rho \gg 1$" se parou com parâmetro penalizador muito grande), e a medida de inviabilidade. Finalmente, a coluna "Melhor FO" traz o melhor valor funcional conhecido para o problema, como reportado em MacMPEC.

		Algencan			Algencan-second		
Problema	Melhor FO	FO	Saída	Inviab	FO	Saída	Inviab
bar-truss	1,02E+04	1,26E+01	$\rho \gg 1$	5,54E-01	1,16E+01	$\rho \gg 1$	5,04E-01
bard1	1,70E+01	1,70E+01	conv.	7,97E-08	1,70E+01	conv.	7,97E-08
bard1m	1,70E+01	1,70E+01	conv.	7,97E-08	1,70E+01	conv.	7,98E-08
bard2	-6,60E+03	-6,60E+03	conv.	2,85E-07	-6,60E+03	conv.	2,85E-07
bard2m	-6,60E+03	-6,60E+03	conv.	1,59E-07	-6,60E+03	conv.	1,59E-07
bard3	-1,27E+01	-1,04E+01	conv.	9,94E-08	-1,04E+01	conv.	9,94E-08
bard3m	-1,27E+01	-1,27E+01	conv.	6,37E-07	-1,27E+01	conv.	6,37E-07
bilevel1	0,00E+00	5,61E-08	conv.	1,15E-08	5,61E-08	conv.	1,15E-08
bilevel2	-6,60E+03	-6,60E+03	conv.	6,60E-09	-6,60E+03	conv.	6,60E-09
bilevel3	-1,27E+01	-1,27E+01	conv.	7,41E-07	-1,27E+01	conv.	7,41E-07
bilin	-1,84E+01	-1,84E+01	conv.	2,09E-07	-1,84E+01	conv.	2,58E-07
dempe	2,83E+01	2,83E+01	conv.	1,65E-08	2,83E+01	conv.	1,65E-08
design-cent-1	-1,86E+00	-1,86E+00	conv.	4,15E-07	-1,86E+00	conv.	4,15E-07
design-cent-2	-3,48E+00	-3,14E+40	$\rho \gg 1$	4,62E+08	-3,14E+40	$\rho \gg 1$	4,62E+08
design-cent-21	-3,48E+00	-6,35E-02	$\rho \gg 1$	2,00E-06	-8,43E-02	$\rho \gg 1$	3,22E-06
design-cent-3	-3,72E+00	-6,28E+40	$\rho \gg 1$	1,00E+00	-6,28E+40	$\rho \gg 1$	1,00E+00
design-cent-31	-3,72E+00	-3,72E+00	conv.	8,17E-07	-3,72E+00	conv.	8,17E-07
design-cent-4	-3,08E+00	-4,47E+30	$\rho \gg 1$	8,32E+10	-4,47E+30	$\rho \gg 1$	8,32E+10
desilva	-1,00E+00	-1,00E+00	conv.	5,14E-08	-1,00E+00	conv.	5,14E-08
df1	0,00E+00	0,00E+00	conv.	0,00E+00	0,00E+00	conv.	0,00E+00

ex9.1.1	-1,30E+01	-1,30E+01	conv.	4,15E-07	-1,30E+01	conv.	1,51E-07
ex9.1.10	-3,25E+00	-3,25E+00	conv.	0,00E+00	-3,25E+00	conv.	0,00E+00
ex9.1.3	-2,92E+01	-2,92E+01	conv.	1,90E-09	-2,92E+01	conv.	4,00E-10
ex9.1.4	-3,70E+01	-3,70E+01	conv.	1,38E-07	-3,70E+01	conv.	1,38E-07
ex9.1.5	-1,00E+00	-1,00E+00	conv.	1,89E-07	-1,00E+00	conv.	7,00E-10
ex9.1.6	-4,90E+01	-4,90E+01	conv.	2,25E-07	-4,90E+01	conv.	2,27E-07
ex9.1.7	-2,60E+01	-2,60E+01	conv.	1,16E-08	-2,60E+01	conv.	1,16E-08
ex9.1.8	-3,25E+00	-3,25E+00	conv.	1,00E-10	-3,25E+00	conv.	0,00E+00
ex9.1.9	3,11E+00	3,11E+00	conv.	5,85E-07	3,11E+00	conv.	5,85E-07
ex9.2.1	1,70E+01	1,70E+01	conv.	1,41E-07	1,70E+01	conv.	1,41E-07
ex9.2.2	1,00E+02	1,00E+02	conv.	7,11E-07	1,00E+02	conv.	7,11E-07
ex9.2.3	-5,50E+01	5,00E+00	conv.	5,72E-07	5,00E+00	conv.	2,11E-08
ex9.2.4	5,00E-01	5,00E-01	conv.	7,04E-07	5,00E-01	conv.	7,04E-07
ex9.2.5	6,00E+00	5,00E+00	conv.	2,79E-07	5,00E+00	conv.	2,79E-07
ex9.2.6	-1,00E+00	-1,00E+00	conv.	7,65E-07	-1,00E+00	conv.	7,65E-07
ex9.2.7	1,70E+01	1,70E+01	conv.	1,41E-07	1,70E+01	conv.	1,41E-07
ex9.2.8	1,50E+00	1,50E+00	conv.	3,97E-08	1,50E+00	conv.	3,97E-08
ex9.2.9	2,00E+00	2,00E+00	conv.	1,00E-10	2,00E+00	conv.	1,00E-10
flp2	0,00E+00	0,00E+00	conv.	2,10E-08	0,00E+00	conv.	2,10E-08
flp4-1	0,00E+00	-6,31E-08	conv.	4,50E-09	-6,31E-08	conv.	4,50E-09
flp4-2	0,00E+00	3,27E-08	conv.	8,98E-08	3,27E-08	conv.	8,98E-08
flp4-3	0,00E+00	-4,30E-09	conv.	1,00E-10	-4,30E-09	conv.	1,00E-10
flp4-4	0,00E+00	-1,30E-09	conv.	0,00E+00	-1,30E-09	conv.	0,00E+00
gauvin	2,00E+01	2,00E+01	conv.	5,61E-08	2,00E+01	conv.	5,61E-08
gnash10	-2,31E+02	-2,31E+02	conv.	6,30E-07	-2,31E+02	conv.	6,30E-07
gnash11	-1,30E+02	-1,30E+02	conv.	1,73E-07	-1,30E+02	conv.	1,73E-07
gnash12	-3,69E+01	-3,69E+01	conv.	2,25E-07	-3,69E+01	conv.	1,22E-08
gnash13	-7,06E+00	-7,06E+00	conv.	1,64E-07	-7,06E+00	conv.	1,64E-07
gnash15	-3,55E+02	-3,55E+02	conv.	7,00E-07	-3,55E+02	conv.	7,00E-07
gnash16	-2,41E+02	-2,41E+02	conv.	5,42E-08	-2,41E+02	conv.	5,42E-08
gnash17	-9,07E+01	-9,07E+01	conv.	3,24E-08	-9,07E+01	conv.	3,24E-08
gnash18	-2,57E+01	-2,57E+01	conv.	6,75E-07	-2,57E+01	conv.	6,75E-07
gnash19	-6,12E+00	-6,12E+00	conv.	8,68E-07	-6,12E+00	conv.	8,68E-07
hakonsen	2,44E+01	-1,49E-01	$\rho \gg 1$	1,20E-02	-1,49E-01	$\rho \gg 1$	1,20E-02
hs044-i	1,56E+01	1,84E+01	$\rho \gg 1$	3,39E-06	8,47E+00	$\rho \gg 1$	6,08E+00
jr1	5,00E-01	5,00E-01	conv.	2,45E-07	5,00E-01	conv.	2,45E-07
jr2	5,00E-01	5,00E-01	conv.	8,65E-07	5,00E-01	conv.	8,65E-07
kth1	0,00E+00	-1,78E-07	conv.	8,88E-08	-1,78E-07	conv.	8,88E-08
kth2	0,00E+00	0,00E+00	conv.	0,00E+00	0,00E+00	conv.	0,00E+00
kth3	5,00E-01	5,00E-01	conv.	1,11E-07	5,00E-01	conv.	1,11E-07
liswet1-100	1,37E+01	1,19E+10	$\rho \gg 1$	3,94E+03	5,04E+09	$\rho \gg 1$	2,41E+03
liswet1-200	1,70E+01	1,61E-02	$\rho \gg 1$	2,34E-06	1,61E-02	$\rho \gg 1$	2,34E-06
liswet1-50	1,40E+01	1,40E-02	conv.	7,45E-07	1,40E-02	conv.	6,91E-07
monteiro	3,75E+01	1,00E+07	$\rho \gg 1$	1,32E+01	2,48E+08	$\rho \gg 1$	4,63E+01
monteirob	8,28E+02	-4,93E+02	$\rho \gg 1$	4,45E-03	-4,77E+02	$\rho \gg 1$	4,95E-03
nash1a	7,89E-30	0,00E+00	conv.	3,09E-07	0,00E+00	conv.	8,90E-09
nash1b	7,89E-30	0,00E+00	conv.	4,28E-08	0,00E+00	conv.	1,00E-10
nash1c	7,89E-30	0,00E+00	conv.	0,00E+00	0,00E+00	conv.	0,00E+00
nash1d	7,89E-30	1,00E-10	conv.	1,44E-07	0,00E+00	conv.	9,18E-08
nash1e	7,89E-30	0,00E+00	conv.	9,55E-07	0,00E+00	conv.	9,55E-07

outrata31	3,21E+00	3,21E+00	conv.	6,05E-07	3,21E+00	conv.	1,23E-08
outrata32	3,45E+00	3,45E+00	conv.	3,03E-07	3,45E+00	conv.	1,45E-08
outrata33	4,60E+00	4,60E+00	conv.	2,74E-07	4,60E+00	conv.	2,74E-07
outrata34	6,59E+00	6,59E+00	conv.	1,37E-07	6,59E+00	conv.	1,37E-07
pack-comp1-16	6,17E-01	6,17E-01	conv.	5,30E-07	6,17E-01	conv.	5,30E-07
pack-comp1-32	6,53E-01	6,53E-01	conv.	8,33E-08	6,61E-01	conv.	3,47E-08
pack-comp1-8	6,00E-01	6,00E-01	conv.	8,89E-08	6,00E-01	conv.	2,80E-09
pack-comp1c-16	6,23E-01	6,23E-01	conv.	8,86E-07	6,23E-01	conv.	7,08E-07
pack-comp1c-32	6,61E-01	6,01E-01	$\rho \gg 1$	1,14E-03	6,61E-01	conv.	3,47E-08
pack-comp1c-8	6,00E-01	6,00E-01	conv.	8,18E-08	6,00E-01	conv.	8,18E-08
pack-comp1p-16	6,17E-01	6,18E-01	conv.	1,27E-07	6,17E-01	conv.	3,00E-10
pack-comp1p-32	6,53E-01	6,53E-01	conv.	5,61E-08	6,53E-01	conv.	5,61E-08
pack-comp1p-8	6,00E-01	1,43E+00	conv.	2,20E-07	1,43E+00	conv.	2,20E-07
pack-comp2-16	7,27E-01	7,27E-01	conv.	2,13E-07	7,27E-01	conv.	2,13E-07
pack-comp2-32	7,83E-01	7,83E-01	conv.	4,92E-07	7,83E-01	conv.	1,12E-07
pack-comp2-8	6,73E-01	6,73E-01	conv.	2,69E-07	6,73E-01	conv.	2,69E-07
pack-comp2c-16	7,27E-01	7,27E-01	conv.	9,95E-07	7,27E-01	conv.	9,95E-07
pack-comp2c-32	7,83E-01	7,83E-01	conv.	5,68E-07	7,83E-01	conv.	2,65E-08
pack-comp2c-8	6,73E-01	6,73E-01	conv.	4,44E-07	6,73E-01	conv.	4,44E-07
pack-comp2p-16	7,27E-01	7,36E-01	conv.	8,54E-07	7,27E-01	conv.	3,20E-09
pack-comp2p-32	7,83E-01	7,83E-01	conv.	5,52E-08	7,83E-01	conv.	5,52E-08
pack-comp2p-8	6,73E-01	6,76E-01	conv.	3,00E-09	6,76E-01	conv.	3,00E-09
pack-rig1-16	8,26E-01	8,26E-01	conv.	4,04E-07	8,26E-01	conv.	4,04E-07
pack-rig1-32	8,51E-01	8,51E-01	conv.	6,44E-07	8,51E-01	conv.	6,44E-07
pack-rig1-8	7,88E-01	7,88E-01	conv.	1,22E-07	7,88E-01	conv.	1,22E-07
pack-rig1c-16	8,26E-01	8,26E-01	conv.	4,82E-07	8,26E-01	conv.	8,81E-07
pack-rig1c-32	8,52E-01	8,52E-01	conv.	5,40E-08	8,52E-01	conv.	2,00E-07
pack-rig1c-8	7,88E-01	7,88E-01	conv.	6,11E-07	7,88E-01	conv.	1,13E-07
pack-rig1p-16	8,26E-01	8,26E-01	conv.	2,91E-07	8,26E-01	conv.	2,91E-07
pack-rig1p-32	8,51E-01	8,51E-01	conv.	7,08E-07	8,51E-01	conv.	6,63E-08
pack-rig1p-8	7,88E-01	7,88E-01	conv.	9,00E-07	7,88E-01	conv.	9,00E-07
pack-rig2-8	7,80E-01	7,80E-01	conv.	3,83E-07	7,80E-01	conv.	3,83E-07
pack-rig2c-8	7,99E-01	7,99E-01	conv.	1,00E-10	7,99E-01	conv.	1,00E-10
pack-rig2p-16	1,09E+00	1,09E+00	conv.	7,07E-07	1,09E+00	conv.	0,00E+00
pack-rig2p-32	1,14E+00	1,14E+00	conv.	1,00E-10	1,14E+00	conv.	1,00E-10
pack-rig2p-8	7,80E-01	7,80E-01	conv.	5,77E-07	7,80E-01	conv.	5,77E-07
pack-rig3-16	8,00E-01	8,00E-01	conv.	3,50E-07	8,00E-01	conv.	3,50E-07
pack-rig3-32	1,00E+20	8,86E-01	conv.	3,84E-08	8,86E-01	conv.	4,87E-07
pack-rig3-8	7,35E-01	7,35E-01	conv.	5,28E-07	7,35E-01	conv.	5,28E-07
pack-rig3c-16	8,19E-01	8,19E-01	conv.	5,26E-07	8,19E-01	conv.	5,26E-07
pack-rig3c-8	7,53E-01	7,53E-01	conv.	4,53E-07	7,53E-01	conv.	4,53E-07
portfl-1	1,50E-05	1,47E-05	conv.	3,18E-07	1,47E-05	conv.	3,18E-07
portfl-2	1,46E-05	1,43E-05	conv.	4,67E-07	1,43E-05	conv.	4,67E-07
portfl-3	6,27E-06	6,13E-06	conv.	6,46E-07	6,13E-06	conv.	6,46E-07
portfl-4	2,18E-06	2,14E-06	conv.	2,18E-07	2,14E-06	conv.	2,18E-07
portfl-6	2,36E-06	2,29E-06	conv.	5,95E-07	2,29E-06	conv.	5,95E-07
qpec1	8,00E+01	8,00E+01	conv.	1,45E-07	8,00E+01	conv.	4,00E-10
qpec2	4,50E+01	4,50E+01	conv.	8,12E-07	4,50E+01	conv.	8,12E-07
qpecgen100-1	9,90E-02	2,41E-01	conv.	4,38E-07	2,41E-01	conv.	4,38E-07
qpecgen100-2	-6,59E+00	-6,59E+00	conv.	6,90E-08	-6,59E+00	conv.	6,90E-08

qpecgen100-3	-5,48E+00	-5,45E+00	conv.	1,77E-07	-5,45E+00	conv.	1,77E-07
qpecgen100-4	-3,98E+00	-4,09E+00	conv.	3,27E-07	-4,09E+00	conv.	3,27E-07
qpecgen200-1	-1,93E+00	-1,94E+00	conv.	9,12E-07	-1,94E+00	conv.	9,12E-07
qpecgen200-2	-2,40E+01	-2,41E+01	conv.	9,36E-07	-2,41E+01	conv.	9,36E-07
qpecgen200-3	-1,95E+00	-1,95E+00	conv.	7,58E-07	-1,95E+00	conv.	7,58E-07
qpecgen200-4	-6,19E+00	-6,22E+00	conv.	7,54E-07	-6,22E+00	conv.	7,54E-07
ralph2	0,00E+00	-0,00E+00	conv.	0,00E+00	-0,00E+00	conv.	0,00E+00
ralphmod	-6,83E+02	-6,83E+02	conv.	8,34E-07	-6,83E+02	conv.	3,10E-09
scale1	1,00E+00	1,00E+00	conv.	4,00E-09	1,00E+00	conv.	4,00E-10
scale2	1,00E+00	1,00E+00	conv.	9,98E-07	1,00E+00	conv.	1,00E-09
scale3	1,00E+00	1,00E+00	conv.	2,75E-07	1,00E+00	conv.	2,75E-07
scale4	1,00E+00	1,71E+00	conv.	5,57E-07	1,00E+00	conv.	1,58E-08
scale5	1,00E+02	2,00E+02	conv.	5,56E-07	1,00E+02	conv.	3,00E-10
scholtes1	2,00E+00	2,00E+00	conv.	1,30E-07	2,00E+00	conv.	1,30E-07
scholtes2	1,50E+01	1,50E+01	conv.	9,18E-08	1,50E+01	conv.	9,18E-08
scholtes3	5,00E-01	9,99E-01	conv.	5,56E-07	5,00E-01	conv.	3,00E-10
scholtes4	-3,07E-07	-1,24E-03	conv.	3,86E-07	-1,24E-03	conv.	3,86E-07
scholtes5	1,00E+00	1,00E+00	conv.	2,85E-07	1,00E+00	conv.	6,90E-09
sl1	1,00E-04	1,00E-04	conv.	1,20E-09	1,00E-04	conv.	1,20E-09
stackelberg1	-3,27E+03	-3,27E+03	conv.	1,30E-08	-3,27E+03	conv.	1,30E-08
tap-09	1,09E+02	1,11E+02	$\rho \gg 1$	2,00E-03	1,11E+02	$\rho \gg 1$	1,57E-03
tap-15	1,84E+02	1,84E+02	$\rho \gg 1$	1,84E-04	1,84E+02	$\rho \gg 1$	1,26E-04
trafficsignalcycle-1	4,52E+01	5,50E+01	conv.	3,33E-07	5,50E+01	conv.	8,60E-08
trafficsignalcycle-10	4,87E+01	5,48E+01	conv.	5,52E-07	5,48E+01	conv.	5,52E-07
trafficsignalcycle-11	9,04E+01	1,02E+02	conv.	1,05E-07	1,02E+02	conv.	1,05E-07
trafficsignalcycle-12	1,31E+02	2,02E+02	$\rho \gg 1$	7,71E-03	2,02E+02	$\rho \gg 1$	7,71E-03
trafficsignalcycle-13	8,36E+01	8,64E+01	conv.	4,50E-09	8,64E+01	conv.	4,50E-09
trafficsignalcycle-2	4,21E+01	5,26E+01	conv.	2,58E-08	5,26E+01	conv.	2,58E-08
trafficsignalcycle-3	8,06E+01	8,71E+01	conv.	4,38E-08	8,71E+01	conv.	4,38E-08
trafficsignalcycle-4	7,50E+01	7,90E+01	conv.	6,01E-08	7,90E+01	conv.	3,55E-08
trafficsignalcycle-5	8,81E+01	1,01E+02	conv.	7,40E-09	1,01E+02	conv.	7,40E-09
trafficsignalcycle-6	8,98E+01	1,02E+02	conv.	3,50E-09	1,02E+02	conv.	3,50E-09
trafficsignalcycle-7	1,73E+02	1,94E+02	$\rho \gg 1$	7,95E-01	1,94E+02	$\rho \gg 1$	7,95E-01
trafficsignalcycle-8	1,78E+02	1,99E+02	$\rho \gg 1$	7,73E-01	1,99E+02	$\rho \gg 1$	7,73E-01
trafficsignalcycle-9	4,63E+01	5,32E+01	conv.	6,00E-09	5,32E+01	conv.	6,00E-09

Tabela 3 – Testes computacionais, complementariedade na forma $g(x)^T h(x) \leqslant 0$.

		Algencan			Algencan-second		
Problema	Melhor FO	FO	Saída	Inviab	FO	Saída	Inviab
bar-truss	1,02E+04	8,27E-01	$\rho \gg 1$	5,07E-02	8,27E-01	$\rho \gg 1$	5,07E-02
bard1	1,70E+01	1,70E+01	conv.	7,97E-08	1,70E+01	conv.	7,97E-08
bard1m	1,70E+01	1,70E+01	conv.	7,97E-08	1,70E+01	conv.	7,98E-08
bard2	-6,60E+03	-6,60E+03	conv.	3,00E-09	-6,60E+03	conv.	3,00E-09
bard2m	-6,60E+03	-6,60E+03	conv.	2,87E-07	-6,60E+03	conv.	2,87E-07
bard3	-1,27E+01	-1,27E+01	conv.	3,64E-08	-1,27E+01	conv.	3,64E-08
bard3m	-1,27E+01	-1,27E+01	conv.	5,49E-07	-1,27E+01	conv.	5,49E-07
bilevel1	0,00E+00	5,75E-08	conv.	3,79E-08	5,75E-08	conv.	4,02E-08
bilevel2	-6,60E+03	-6,60E+03	conv.	7,01E-07	-6,60E+03	conv.	7,01E-07

bilevel3	-1,27E+01	-1,27E+01	conv.	8,38E-07	-1,27E+01	conv.	8,38E-07
bilin	-1,84E+01	-1,84E+01	conv.	1,09E-07	-1,84E+01	conv.	1,09E-07
dempe	2,83E+01	2,83E+01	conv.	0,00E+00	2,83E+01	conv.	0,00E+00
design-cent-1	-1,86E+00	-1,86E+00	conv.	4,21E-07	-1,86E+00	conv.	4,21E-07
design-cent-2	-3,48E+00	-1,99E+27	$\rho \gg 1$	6,81E+05	-1,99E+27	$\rho \gg 1$	6,81E+05
design-cent-21	-3,48E+00	-7,00E-02	$\rho \gg 1$	2,50E-06	-7,10E-02	$\rho \gg 1$	2,46E-06
design-cent-3	-3,72E+00	-1,28E+21	$\rho \gg 1$	3,08E+05	-1,28E+21	$\rho \gg 1$	3,08E+05
design-cent-31	-3,72E+00	-3,72E+00	conv.	2,45E-07	-3,72E+00	conv.	2,45E-07
design-cent-4	-3,08E+00	-2,52E+16	$\rho \gg 1$	3,16E-01	-2,52E+16	$\rho \gg 1$	3,20E-01
desilva	-1,00E+00	-1,00E+00	conv.	5,08E-08	-1,00E+00	conv.	5,08E-08
df1	0,00E+00	0,00E+00	conv.	0,00E+00	0,00E+00	conv.	0,00E+00
ex9.1.1	-1,30E+01	-1,30E+01	conv.	3,67E-08	-1,30E+01	conv.	3,67E-08
ex9.1.10	-3,25E+00	-3,25E+00	conv.	8,00E-09	-3,25E+00	conv.	8,00E-09
ex9.1.3	-2,92E+01	-2,92E+01	conv.	3,99E-07	-2,92E+01	conv.	3,80E-08
ex9.1.4	-3,70E+01	-3,70E+01	conv.	2,33E-07	-3,70E+01	conv.	2,33E-07
ex9.1.5	-1,00E+00	-1,00E+00	conv.	2,07E-08	-1,00E+00	conv.	2,07E-08
ex9.1.6	-4,90E+01	-4,90E+01	conv.	7,43E-07	-4,90E+01	conv.	1,36E-07
ex9.1.7	-2,60E+01	-2,30E+01	conv.	2,36E-07	-2,30E+01	conv.	2,36E-07
ex9.1.8	-3,25E+00	-3,25E+00	conv.	8,00E-09	-3,25E+00	conv.	8,00E-09
ex9.1.9	3,11E+00	3,11E+00	conv.	5,85E-07	3,11E+00	conv.	5,85E-07
ex9.2.1	1,70E+01	1,70E+01	conv.	6,43E-07	1,70E+01	conv.	2,08E-08
ex9.2.2	1,00E+02	1,00E+02	conv.	4,33E-07	1,00E+02	conv.	4,33E-07
ex9.2.3	-5,50E+01	5,00E+00	conv.	2,99E-08	5,00E+00	conv.	2,99E-08
ex9.2.4	5,00E-01	5,00E-01	conv.	2,53E-07	5,00E-01	conv.	2,53E-07
ex9.2.5	6,00E+00	5,00E+00	conv.	2,88E-07	5,00E+00	conv.	2,88E-07
ex9.2.6	-1,00E+00	-1,00E+00	conv.	7,79E-07	-1,00E+00	conv.	7,79E-07
ex9.2.7	1,70E+01	1,70E+01	conv.	6,43E-07	1,70E+01	conv.	2,08E-08
ex9.2.8	1,50E+00	1,50E+00	conv.	3,97E-08	1,50E+00	conv.	3,97E-08
ex9.2.9	2,00E+00	2,00E+00	conv.	4,19E-08	2,00E+00	conv.	7,60E-09
flp2	0,00E+00	0,00E+00	conv.	1,05E-07	0,00E+00	conv.	1,05E-07
flp4-1	0,00E+00	0,00E+00	conv.	0,00E+00	0,00E+00	conv.	0,00E+00
flp4-2	0,00E+00	1,00E-10	conv.	0,00E+00	1,00E-10	conv.	0,00E+00
flp4-3	0,00E+00	7,79E-07	conv.	9,59E-07	-6,00E-10	conv.	1,70E-09
flp4-4	0,00E+00	-6,00E-08	conv.	1,56E-07	-6,00E-08	conv.	1,56E-07
gauvin	2,00E+01	2,00E+01	conv.	5,61E-08	2,00E+01	conv.	5,61E-08
gnash10	-2,31E+02	-2,31E+02	conv.	5,37E-07	-2,31E+02	conv.	5,37E-07
gnash11	-1,30E+02	-1,30E+02	conv.	1,71E-07	-1,30E+02	conv.	1,71E-07
gnash12	-3,69E+01	-3,69E+01	conv.	3,09E-07	-3,69E+01	conv.	3,09E-07
gnash13	-7,06E+00	-7,06E+00	conv.	7,65E-07	-7,06E+00	conv.	7,65E-07
gnash15	-3,55E+02	-3,55E+02	conv.	7,06E-07	-3,55E+02	conv.	7,06E-07
gnash16	-2,41E+02	-2,41E+02	conv.	5,35E-08	-2,41E+02	conv.	5,35E-08
gnash17	-9,07E+01	-9,07E+01	conv.	3,17E-08	-9,07E+01	conv.	3,17E-08
gnash18	-2,57E+01	-2,57E+01	conv.	9,95E-07	-2,57E+01	conv.	9,95E-07
gnash19	-6,12E+00	-6,12E+00	conv.	9,34E-07	-6,12E+00	conv.	9,34E-07
hs044-i	1,56E+01	1,36E+00	$\rho \gg 1$	5,78E-05	1,20E+00	$\rho \gg 1$	5,87E-05
jr1	5,00E-01	5,00E-01	conv.	2,45E-07	5,00E-01	conv.	2,45E-07
jr2	5,00E-01	5,00E-01	conv.	8,65E-07	5,00E-01	conv.	8,65E-07
kth1	0,00E+00	-1,78E-07	conv.	8,92E-08	-1,78E-07	conv.	8,92E-08
kth2	0,00E+00	-2,14E-07	conv.	2,14E-07	-0,00E+00	conv.	0,00E+00
kth3	5,00E-01	5,00E-01	conv.	1,14E-07	5,00E-01	conv.	1,14E-07

liswet1-100	1,37E+01	4,41E+11	$\rho \gg 1$	2,67E+05	4,37E+07	$\rho \gg 1$	4,55E+01
liswet1-200	1,70E+01	2,89E-01	$\rho \gg 1$	3,27E-05	2,89E-01	$\rho \gg 1$	3,27E-05
liswet1-50	1,40E+01	1,40E-02	conv.	6,88E-07	1,40E-02	conv.	6,99E-07
monteiro	3,75E+01	-3,44E+01	$\rho \gg 1$	3,09E-03	1,89E+13	$\rho \gg 1$	3,66E+05
monteirob	8,28E+02	-8,77E+02	$\rho \gg 1$	3,81E-01	-8,78E+02	$\rho \gg 1$	3,81E-01
nash1a	7,89E-30	0,00E+00	conv.	2,99E-08	0,00E+00	conv.	2,99E-08
nash1b	7,89E-30	0,00E+00	conv.	1,38E-07	0,00E+00	conv.	7,77E-08
nash1c	7,89E-30	0,00E+00	conv.	2,29E-08	0,00E+00	conv.	2,29E-08
nash1d	7,89E-30	0,00E+00	conv.	3,34E-07	0,00E+00	conv.	2,60E-09
nash1e	7,89E-30	0,00E+00	conv.	1,87E-07	0,00E+00	conv.	1,94E-08
outrata31	3,21E+00	3,21E+00	conv.	3,54E-07	3,21E+00	conv.	1,26E-08
outrata32	3,45E+00	3,45E+00	conv.	3,21E-07	3,45E+00	conv.	1,56E-08
outrata33	4,60E+00	4,60E+00	conv.	2,58E-07	4,60E+00	conv.	2,58E-07
outrata34	6,59E+00	6,59E+00	conv.	1,37E-07	6,59E+00	conv.	1,37E-07
pack-comp1-16	6,17E-01	6,17E-01	conv.	5,30E-07	6,17E-01	conv.	5,30E-07
pack-comp1-32	6,53E-01	6,53E-01	conv.	1,10E-08	6,53E-01	conv.	1,10E-08
pack-comp1-8	6,00E-01	6,00E-01	conv.	1,30E-07	6,00E-01	conv.	1,00E-10
pack-comp1c-16	6,23E-01	6,23E-01	conv.	4,97E-08	6,23E-01	conv.	7,07E-07
pack-comp1c-32	6,61E-01	8,25E-01	$\rho \gg 1$	9,32E+06	6,61E-01	conv.	6,46E-07
pack-comp1c-8	6,00E-01	6,00E-01	conv.	9,70E-09	6,00E-01	conv.	3,00E-10
pack-comp1p-16	6,17E-01	6,17E-01	conv.	9,40E-09	6,17E-01	conv.	9,40E-09
pack-comp1p-32	6,53E-01	6,53E-01	conv.	3,09E-07	6,53E-01	conv.	4,90E-09
pack-comp1p-8	6,00E-01	6,61E-01	conv.	2,11E-07	6,01E-01	conv.	1,50E-09
pack-comp2-16	7,27E-01	7,27E-01	conv.	2,13E-07	7,27E-01	conv.	2,13E-07
pack-comp2-32	7,83E-01	7,83E-01	conv.	6,24E-07	7,83E-01	conv.	6,33E-07
pack-comp2-8	6,73E-01	6,73E-01	conv.	2,69E-07	6,73E-01	conv.	2,69E-07
pack-comp2c-16	7,27E-01	7,27E-01	conv.	9,95E-07	7,27E-01	conv.	9,95E-07
pack-comp2c-32	7,83E-01	7,83E-01	conv.	1,59E-07	7,83E-01	conv.	2,34E-08
pack-comp2c-8	6,73E-01	6,73E-01	conv.	4,43E-07	6,73E-01	conv.	4,43E-07
pack-comp2p-16	7,27E-01	7,24E-01	conv.	1,34E-07	7,24E-01	conv.	1,34E-07
pack-comp2p-32	7,83E-01	7,83E-01	conv.	2,27E-07	7,83E-01	conv.	2,27E-07
pack-comp2p-8	6,73E-01	6,63E-01	conv.	1,83E-07	6,63E-01	conv.	1,83E-07
pack-rig1-16	8,26E-01	8,26E-01	conv.	2,18E-08	8,26E-01	conv.	3,76E-07
pack-rig1-32	8,51E-01	8,51E-01	conv.	1,58E-07	8,51E-01	conv.	1,58E-07
pack-rig1-8	7,88E-01	7,88E-01	conv.	1,22E-07	7,88E-01	conv.	1,22E-07
pack-rig1c-16	8,26E-01	8,26E-01	conv.	3,30E-07	8,26E-01	conv.	8,99E-07
pack-rig1c-32	8,52E-01	8,52E-01	conv.	8,07E-08	8,52E-01	conv.	2,04E-07
pack-rig1c-8	7,88E-01	7,88E-01	conv.	1,13E-07	7,88E-01	conv.	1,13E-07
pack-rig1p-16	8,26E-01	8,26E-01	conv.	2,93E-07	8,26E-01	conv.	2,93E-07
pack-rig1p-32	8,51E-01	8,51E-01	conv.	5,48E-07	8,51E-01	conv.	2,69E-07
pack-rig1p-8	7,88E-01	7,88E-01	conv.	7,03E-07	7,88E-01	conv.	7,03E-07
pack-rig2-8	7,80E-01	7,80E-01	conv.	3,83E-07	7,80E-01	conv.	3,83E-07
pack-rig2c-8	7,99E-01	7,99E-01	conv.	1,30E-08	7,99E-01	conv.	1,30E-08
pack-rig2p-16	1,09E+00	1,09E+00	conv.	5,00E-10	1,09E+00	conv.	1,00E-10
pack-rig2p-32	1,14E+00	1,14E+00	conv.	3,77E-07	1,14E+00	conv.	1,00E-10
pack-rig2p-8	7,80E-01	7,80E-01	conv.	2,08E-07	7,80E-01	conv.	2,08E-07
pack-rig3-16	8,00E-01	8,00E-01	conv.	3,48E-07	8,00E-01	conv.	3,48E-07
pack-rig3-32	1,00E+20	8,86E-01	conv.	3,75E-08	8,86E-01	conv.	4,41E-07
pack-rig3-8	7,35E-01	7,35E-01	conv.	5,27E-07	7,35E-01	conv.	5,27E-07
pack-rig3c-16	8,19E-01	8,19E-01	conv.	4,37E-07	8,19E-01	conv.	4,37E-07

pack-rig3c-8	7,53E-01	7,53E-01	conv.	4,68E-07	7,53E-01	conv.	7,67E-08
portfl-1	1,50E-05	1,47E-05	conv.	3,18E-07	1,47E-05	conv.	3,18E-07
portfl-2	1,46E-05	1,43E-05	conv.	4,67E-07	1,43E-05	conv.	4,67E-07
portfl-3	6,27E-06	6,13E-06	conv.	6,46E-07	6,13E-06	conv.	6,46E-07
portfl-4	2,18E-06	2,14E-06	conv.	2,27E-07	2,14E-06	conv.	2,27E-07
portfl-6	2,36E-06	2,29E-06	conv.	5,95E-07	2,29E-06	conv.	5,95E-07
qpec1	8,00E+01	8,00E+01	conv.	3,30E-09	8,00E+01	conv.	3,30E-09
qpec2	4,50E+01	4,50E+01	conv.	8,12E-07	4,50E+01	conv.	8,12E-07
qpecgen100-1	9,90E-02	2,41E-01	conv.	5,49E-07	2,41E-01	conv.	5,49E-07
qpecgen100-2	-6,59E+00	-6,59E+00	conv.	7,01E-07	-6,59E+00	conv.	7,01E-07
qpecgen100-3	-5,48E+00	-5,45E+00	conv.	1,49E-07	-5,45E+00	conv.	1,49E-07
qpecgen100-4	-3,98E+00	-4,09E+00	conv.	3,62E-07	-4,09E+00	conv.	3,62E-07
qpecgen200-1	-1,93E+00	-1,94E+00	conv.	9,13E-07	-1,94E+00	conv.	9,13E-07
qpecgen200-2	-2,40E+01	-2,41E+01	conv.	1,05E-07	-2,41E+01	conv.	3,59E-07
qpecgen200-3	-1,95E+00	-1,95E+00	conv.	5,49E-07	-1,95E+00	conv.	5,49E-07
qpecgen200-4	-6,19E+00	-6,22E+00	conv.	1,78E-07	-6,22E+00	conv.	1,78E-07
ralph2	0,00E+00	-0,00E+00	conv.	0,00E+00	-0,00E+00	conv.	0,00E+00
ralphmod	-6,83E+02	-6,83E+02	conv.	3,05E-07	-6,83E+02	conv.	1,00E-09
scale1	1,00E+00	1,00E+00	conv.	2,40E-09	1,00E+00	conv.	2,00E-10
scale2	1,00E+00	1,00E+00	conv.	6,09E-08	1,00E+00	conv.	6,09E-08
scale3	1,00E+00	1,00E+00	conv.	2,75E-07	1,00E+00	conv.	2,75E-07
scale4	1,00E+00	1,71E+00	conv.	5,57E-07	1,00E+00	conv.	1,58E-08
scale5	1,00E+02	2,00E+02	conv.	5,56E-07	1,00E+02	conv.	3,00E-10
scholtes1	2,00E+00	2,00E+00	conv.	4,70E-08	2,00E+00	conv.	2,00E-10
scholtes2	1,50E+01	1,50E+01	conv.	1,00E-07	1,50E+01	conv.	1,00E-07
scholtes3	5,00E-01	9,99E-01	conv.	5,56E-07	5,00E-01	conv.	3,00E-10
scholtes4	-3,07E-07	-1,24E-03	conv.	3,86E-07	-1,24E-03	conv.	3,86E-07
scholtes5	1,00E+00	1,00E+00	conv.	5,07E-07	1,00E+00	conv.	5,90E-09
sl1	1,00E-04	1,00E-04	conv.	3,48E-08	1,00E-04	conv.	3,48E-08
stackelberg1	-3,27E+03	-3,27E+03	conv.	5,60E-09	-3,27E+03	conv.	5,60E-09
tap-09	1,09E+02	1,10E+02	$\rho \gg 1$	5,32E-02	1,11E+02	$\rho \gg 1$	1,04E-01
tap-15	1,84E+02	1,87E+02	$\rho \gg 1$	2,05E-03	1,87E+02	$\rho \gg 1$	1,21E-03
trafficsignalcycle-1	4,52E+01	5,50E+01	conv.	3,00E-09	5,50E+01	conv.	3,00E-09
trafficsignalcycle-10	4,87E+01	5,48E+01	conv.	5,10E-09	5,48E+01	conv.	5,10E-09
trafficsignalcycle-11	9,04E+01	1,02E+02	conv.	1,19E-07	1,02E+02	conv.	2,00E-10
trafficsignalcycle-12	1,31E+02	2,02E+02	$\rho \gg 1$	7,71E-03	2,02E+02	$\rho \gg 1$	7,71E-03
trafficsignalcycle-13	8,36E+01	8,64E+01	conv.	1,92E-07	8,64E+01	conv.	8,10E-09
trafficsignalcycle-2	4,21E+01	5,26E+01	conv.	8,80E-09	5,26E+01	conv.	8,80E-09
trafficsignalcycle-3	8,06E+01	8,71E+01	conv.	4,00E-10	8,71E+01	conv.	4,00E-10
trafficsignalcycle-4	7,50E+01	7,90E+01	conv.	6,48E-08	7,90E+01	conv.	3,45E-08
trafficsignalcycle-5	8,81E+01	1,01E+02	conv.	2,53E-08	1,01E+02	conv.	1,70E-09
trafficsignalcycle-6	8,98E+01	1,02E+02	conv.	4,46E-07	1,02E+02	conv.	4,00E-09
trafficsignalcycle-7	1,73E+02	1,94E+02	$\rho \gg 1$	7,95E-01	1,94E+02	$\rho \gg 1$	7,95E-01
trafficsignalcycle-8	1,78E+02	1,99E+02	$\rho \gg 1$	7,73E-01	1,99E+02	$\rho \gg 1$	7,73E-01
trafficsignalcycle-9	4,63E+01	5,32E+01	conv.	5,98E-08	5,32E+01	conv.	5,98E-08

Tabela 4 – Testes computacionais, complementariedade na forma $g(x)^T h(x) + s = 0$, $s \geqslant 0$.