Universidade Estadual de Campinas

Intituto de Matemática Estatística e Computação Científica

Departamento de Matemática Aplicada

Resolução de sistema KKT por método de tipo Newton não diferenciável

Renaud Gilles Gaujoux
Mestrado em Matemática Aplicada

Orientador: Prof. Dr. Roberto Andreani

Resolução de sistema KKT por método de tipo Newton não diferenciável

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por **Renaud Gilles Gaujoux** e aprovada pela comissão julgadora.

Campinas, 25 de fevereiro de 2005

Banca examinadora:

Prof. Dr. Sandra Augusta Santos

Prof. Dr. Roberto Andreani

Prof. Dr. Paulo Silva e Silva

Suplentes:

Prof. Dr. Vera Lúcia Rocha Lopes

Prof. Dr. Geraldo Silva

Prof. Dr. **Roberto Andreani**
Orientador

Dissertação apresentada ao Instituto de Matemática, Estatística e Computação Científica, UNICAMP, como requisito parcial para obtenção do Título de **MESTRE** em **Matemática Aplicada**

CHAMADA TZ UNICAMP
G235r
EX
TOMBO BC/ 64300
PROC. 16/P-36/05
C ☐ D ☒
PREÇO R$ 11,00
DATA 13/6/05
CPD
BIB-ID 352800

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO IMECC DA UNICAMP

Bibliotecário: Maria Júlia Milani Rodrigues – CRB8a 2116

Gaujoux, Renaud Gilles

G235r Resolução de sistema KKT por método tipo Newton não diferenciável / Renaud Gilles Gaujoux -- Campinas, [S.P. :s.n.], 2005.

Orientador : Roberto Andreani

Dissertação (mestrado) - Universidade Estadual de Campinas, *Instituto de Matemática, Estatística e Computação Científica.*

1. Otimização matemática. 2. Programação não-linear. 3. Métodos iterativos (Matemática). 4. Algoritmos. I. Andreani, Roberto. II. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. III. Título.

Título em inglês: Resolution of KKT system by generalized Newton type method

Palavras-chave em inglês (Keywords): 1. Mathematical optimization. 2. Nonlinear programming. 3. Iterative methods (Mathematics). 4. Algorithms.

Área de concentração: ~~Otimização~~ Matemática Aplicada.

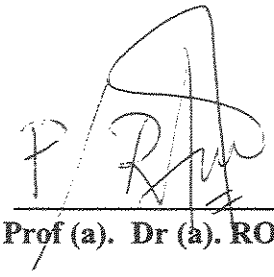
Titulação: Mestre em Matemática Aplicada

Banca examinadora: Prof. Dr. Roberto Andreani (UNICAMP)
Profa. Dra. Sandra Augusta Santos (UNICAMP)
Prof. Dr. Paulo Silva e Silva (USP)

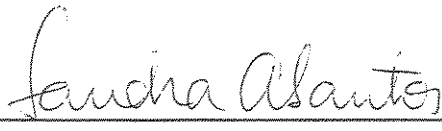
Data da defesa: 16/02/2005

Dissertação de Mestrado defendida em 16 de fevereiro de 2005

e aprovada pela Banca Examinadora composta pelos Profs. Drs.



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



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

Prof (a). Dr (a). PAULO JOSÉ DA SILVA E SILVA

Resumo

Esta dissertação trata da aplicação de um método de tipo Newton generalizado aos sistemas KKT. Graças às funções chamadas de NCP, o sistema KKT pode ser reformulado como uma equação do tipo $H(z) = 0$, onde H é uma função semi-suave.

Nos preliminares teóricos apresentamos os conceitos importantes para a análise desse tipo de sistema quando a função involvida não é diferenciável. Trata-se de subdiferencial, semi-suavidade, semi-derivada. Então, usando um ponto de vista global, descrevemos de uma vez só as diferentes generalizações do método de Newton, apresentando as condições suficientes de convergência local. Uma versão globalizada do método é também detalhada.

Com o fim de aplicar o algoritmo à reformulação semi-suave do sistema KKT, estudamos as propriedades da função H, primeiro independentemente da função NCP usada. Então analisamos o caso de três funções NCP particulares: a função do Mínimo, a função de Fischer-Burmeister, a função de Fischer-Burmeister Penalizada.

Apresentamos os resultados de testes numéricos que comparam o desempenho do algoritmo quando usa as diferentes funções NCP acima.

Abstract

This work deals with the use of generalized Newton type method to solve KKT systems. By the mean of so called NCP functions, any KKT system can be writen as an equation of type $H(z) = 0$, where H is a semismooth function.

In a teorical preliminaries part, we present some key notions for the analysis of such a type of system, whose the involved function is not differentiable. It deals with subdifferential, semismoothness, semiderivative. Then, tackling the problem with a very general point of view, we make a unified description of different generalizations of Newton method, giving sufficient local convergence conditions. More over, we detail a possible globalization of such methods.

In order to use this global algorithm to solve semismooth form of KKT systems, we study some of the H function's properties, first without specifying any underlying NCP function, and then in the case of three known NCP functions: the minimum function, the Fischer-Burmeister function and the penalized Fischer-Burmeister function.

Finally, we give the results of numerical tests, which compare the algorithm's performance for each of these three NCP functions.

À Papex©, mon professeur,

sapiance qui nous a quitté trop tôt.

Agradecimentos

Agradeço,

ao meu orientador Roberto Andreani pelo o seu suporte e a sua confiança constantes, os seus conselhos e sua simpatia;

às Professoras Sandra Augusta Santos e Vera Lúcia Rocha Lopes pelos seus conselhos e encorajamentos;

à María Laura pela sua ajuda na correção e o seus encorajamentos, e ao Flávio por ter-me iniciado ao MATLAB;

à Tânia e o pessoal do secretaria da pós, que me guiaram nos labirintos administrativos;

ao meu Professor de matemática tão especiais, Jean-Marie Exbrayat, "notre Papex©", por todas essas aulas que ele professou do seu jeito, pela formação que ele deu, pelo gosto da matemática que ele transmitiu;

aos meus pais e ao meu irmão que conseguem me dar força e assistência, mesmo tão longe de mim;

às minhas outras famílias do coração, Bénavent, Albejante, Mazon e Marim;

a todos os meus amigos tanto daqui quanto de lá, que participaram cada um à sua maneira à concretização e à finalização deste trabalho;

a todos de quem esqueci de citar o nome mas que sabem que não foi de propósito;

...

Conteúdo

Introdução **11**

1 Sistemas de Karush-Kuhn-Tucker (KKT) **13**
1.1 Problemas relacionados . . . 14
1.1.1 Programação não linear (NLP) . . . 14
1.1.2 Complementaridade não linear (NCP) . . . 15
1.1.3 Desigualdade variacional (VIP) . . . 15
1.2 Abordagem clássica do problema . . . 18
1.2.1 Equação não linear: método de Newton . . . 19
1.2.2 Minimização da norma: método do gradiente . . . 21
1.3 Abordagem não diferenciável do problema . . . 22
1.3.1 Transformação do sistema: funções NCP . . . 22
1.3.2 As funções NCP estudadas . . . 23

2 Preliminares teóricos **27**
2.1 Definições e teoremas gerais . . . 27
2.2 Funções ponto-conjunto . . . 32
2.3 Derivada e gradiente generalizados . . . 38
2.3.1 Derivada direcional generalizada . . . 38
2.3.2 Gradiente generalizado . . . 40
2.3.3 Regras de cálculo do gradiente generalizado . . . 49
2.4 Jacobianos generalizados . . . 50
2.4.1 B-subdiferencial: $\partial_B f$ . . . 50
2.4.2 Subdiferencial de Clarke: ∂f . . . 53
2.4.3 C-subdiferencial: $\partial_C f$ . . . 56
2.5 Semi-suavidade . . . 57
2.6 Semi-derivada . . . 60

3 Algoritmos **65**
3.1 Método de Newton clássico . . . 65
3.2 Métodos de Newton não diferenciáveis . . . 66
3.2.1 Método de Newton generalizado local . . . 67
3.2.2 Método de Newton generalizado global . . . 71

4 Aplicação aos sistemas KKT **81**
4.1 Considerações gerais . . . 82
4.1.1 Semi-suavidade . . . 82
4.1.2 Regularidade na solução . . . 83
4.1.3 Caráter C^1 da função de mérito . . . 87
4.1.4 Escolha de um elemento em $\partial_B H(z)$ . . . 88
4.2 Estudo das funções NCP caso a caso . . . 90
4.2.1 A função do Mínimo . . . 91
4.2.2 A função de Fisher-Burmeister (FB) . . . 95
4.2.3 A função de Fischer-Burmeister Penalizada (FBP) . . . 101

5 Resultados numéricos **109**
5.1 Detalhes e parâmetros da implementação . . . 110
5.1.1 Busca linear não monótona . . . 110
5.1.2 Modificações para a implementação . . . 111
5.1.3 Critérios de parada . . . 112
5.1.4 Pontos iniciais . . . 112
5.1.5 Parâmetros e legenda do resultados . . . 113
5.2 Problemas testes . . . 113
5.3 Resultados . . . 114

Conclusão **149**

Notações

$\mathbb{R}^*$	Números reais não nulos
$\mathbb{R}_+$	Números reais positivos ou nulos
$\mathbb{R}_+^*$	Números reais estritamente positivos
$\mathbb{R}_+ +$	Números reais estritamente positivos $(= \mathbb{R}_+^*)$
$]a, b[$	Intervalo real aberto: $\{x \in \mathbb{R} \mid a < x < b\}$
$\mathcal{M}_{n,p}(\mathbb{R})$	Conjunto das matrizes reais de n linhas e p colunas
$\mathcal{M}_n(\mathbb{R})$	$\mathcal{M}_{n,n}(\mathbb{R})$
$\mathbb{GL}_n(\mathbb{R})$	Subgrupo de $\mathcal{M}_{n,n}(\mathbb{R})$ das matrizes reais inversíveis
$L(X, Y)$	Conjunto das aplicações lineares contínuas de X em Y
$Id_{\mathbb{R}^n}$	Operador identidade de $\mathbb{R}^n$ em $\mathbb{R}^n$
$\langle x, y \rangle$	Produto escalar entre os vetores x e y
$\langle \xi, x \rangle$	Avaliação do funcional $\xi : X \longrightarrow \mathbb{R}$ no ponto x $(= \xi(x))$
$f\|_A$	Restrição de uma função $f : X \longrightarrow Y$ a um subconjunto $A \subset X$
$\nabla f(x)$	Gradiente de $f : \mathbb{R}^n \longmapsto \mathbb{R}$ no ponto x
$f'(x)$	Matriz Jacobiana de $f : \mathbb{R}^n \longmapsto \mathbb{R}^p$ no ponto x $(\in \mathcal{M}_{p,n}(\mathbb{R}))$
$f'_d(x; d)$	Derivada direcional de f em x na direção d
$f'_s(x)$	Derivada estrita de f em x
$f^\circ(x; d)$	Derivada direcional generalizada de f em x, na direção d
$\partial_B f(x)$	B-subdiferencial de f no ponto x
$\partial f(x)$	Subdiferencial de Clarke de f no ponto x
$\partial_C f(x)$	C-subdiferencial de f no ponto x
$\partial_\bullet f(x)$	Notação genérica do subdiferencial de f no ponto x (pode ser substituido por $\partial_B f(x)$, $\partial f(x)$ ou $\partial_C f(x)$)

Introdução

Sejam três funções F, h e g com a seguinte regularidade:

- $F : \mathbb{R}^n \longmapsto \mathbb{R}^n$, contínua
- $h : \mathbb{R}^n \longmapsto \mathbb{R}^p$, e $g : \mathbb{R}^n \longmapsto \mathbb{R}^q$, de classe C^1

O problema de Karush-Kuhn-Tucker geral, denotado $KKT(F, h, g)$, consiste em achar uma tripla (x, λ, μ) em $\mathbb{R}^{n+p+q}$ que verifique as seguintes equações:

$$\begin{array}{rr} F(x) + h'(x)^T\lambda - g'(x)^T\mu = 0 & \\ h(x) = 0 & \\ \text{para } 1 \leq i \leq q \qquad \mu_i g_i(x) = 0 & \\ g(x) \geq 0 & \\ \mu \geq 0 & \end{array} \tag{1}$$

O interesse pelos sistemas KKT vem do fato que, escolhendo as funções F, h e g de maneira adequada, vários problemas clássicos da matemática podem ser relacionados a um sistema KKT particular.

Em geral, estes sistemas são constituídos de equações-inequações não lineares, o que torna a sua resolução analítica mais difícil, e até impossível. Por isto, os problemas desse tipo se resolvem geralmente por métodos numéricos.

Nesta dissertação, apresentamos um *método numérico de tipo Newton*, baseado em uma reformulação do problema original em um sistema onde só aparecem equações. Essa reformulação usa funções chamadas de NCP (Nonlinear Complementary Problem), que no caso possuem pontos de não diferenciabilidade. A abordagem que descrevemos, entra então no quadro não diferenciável, e apela a uma teoria de cálculo diferencial estendido, cujos fundamentos são devidos a Clarke [5].

O nosso estudo trata de três funções NCP particulares: a função do mínimo, a função de Fischer-Burmeister devida a Fischer (1992), e a função de Fischer-Burmeister Penalizada devida a Chen, Chen e Kanzow (1997).

Essas três funções foram bastante estudadas, especialmente a função de Fischer-Burmeister, isso na resolução de problemas de complementaridade não linear [6, 8, 15, 21, 31] ou de desigualdade variacional [7, 10]. Quanto à função de Fischer-Burmeister penalizada, que nós saibamos, essa só foi estudada no quadro dos problemas de complementaridade [3], para os quais apresentou resultados promissores.

Um dos aportes desta dissertação é, em particular, fornecer um estudo tanto teórico quanto numérico da função de Fischer-Burmeister penalizada aplicada à resolução dos sistemas KKT. Ao mesmo tempo, fornecemos uma revisão teórica importante e bastante geral dos conceitos da análise não diferenciável. Esta pode ser usada em contextos diferentes, na resolução de outros problemas. Também, apresentamos resultados de *testes comparativos* entre diferentes funções NCP. Em efeito, exceto em Qi [35], trabalhos análogos fornecem resultados de testes para uma função NCP apenas.

A dissertação se organiza da seguinte maneira. No capítulo 1, tratamos dos sistemas KKT em geral, apresentando três problemas relacionados aos sistemas KKT: o problema de programação não linear (NLP), o problema de complementaridade não linear (NCP) e o problema de desigualdade variacional (VIP). Neste mesmo capítulo, detalhamos a reformulação de um sistema KKT na sua forma (1), em um sistema composto somente de equações.

O capítulo 2 fornece a base teórica necessária ao entendimento dos capítulos seguintes. Em particular, são desenvolvidos as noções de *subdiferencial* e de *semi-derivadas*, usadas no capítulo 3 para a descrição e as provas de convergência dos métodos de Newton no quadro não diferenciável.

O capítulo 4 é consagrado à aplicação desses métodos aos sistemas KKT. Primeiro, estudamos as propriedades gerais dos sistemas reformulados, independentemente da função NCP escolhida para efetuar a reformulação. Segundo, tratamos separadamente cada uma das três funções NCP estudadas: a função do mínimo, a função de Fischer-Burmeister e a função de Fischer-Burmeister penalizada.

Cada uma dessas funções tem propriedades diferentes (pontos de não diferenciabilidade, curvas de nível, ...), levando a resultados numéricos diferentes, e portanto interessantes. Comparações de comportamento do método de resolução relativamente à função NCP usada são apresentadas no último capítulo (5).

Capítulo 1

Sistemas de Karush-Kuhn-Tucker (KKT)

Neste capítulo, fazemos um estudo preliminar dos sistemas KKT. Apresentamos primeiro três diferentes classes de problemas matemáticos clássicos que lhe são relacionadas: os problemas de programação não linear, os problemas de complementaridade não linear e os problemas de desigualdade variacional. Segundo, tratamos brevemente de como poderia ser abordada a questão de resolver um sistema KKT usando métodos básicos em um quadro diferenciável. Por fim, tratamos da abordagem não diferenciável, que consiste em reformular o sistema KKT em uma equação não linear não diferenciável via um certo tipo de funções: as funções NCP.

Lembramos a forma do sistema $KKT(F, h, g)$ geral dada na introdução:

$$\begin{array}{rrcl} & F(x) + h'(x)^T\lambda - g'(x)^T\mu & = & 0 \\ & h(x) & = & 0 \\ \text{para } 1 \leq i \leq q \qquad & \mu_i g_i(x) & = & 0 \\ & g(x) & \geq & 0 \\ & \mu & \geq & 0. \end{array}$$

Por analogia com os problemas de otimização (cf. próxima seção), chamaremos a função h (resp. g) de restrição de igualdade (resp. desigualdade).

1.1 Problemas relacionados

1.1.1 Programação não linear (NLP)

Seja o problema de otimização geral na sua seguinte forma padrão:

$$\begin{array}{l} \min f(x) \\ s.a. \left\{ \begin{array}{l} h(x) = 0 \\ g(x) \leq 0 \end{array} \right. \end{array} \tag{1.1}$$

onde $h : \mathbb{R}^n \longrightarrow \mathbb{R}^p$ e $g : \mathbb{R}^n \longrightarrow \mathbb{R}^q$ são funções C^1, que definem as restrições do domínio em que procura-se o mínimo de f.

Suponhamos que x^* é solução local do problema (1.1), em que se verifica uma condição de qualificação das restrições (constraint qualification) como a de independência linear, de Mangasarian-Fromovitz, do posto constante, CPLD, quase-normalidade, etc. [1, 2, 23]

Então, pela teoria clássica da otimização, sabemos que x^* verifica as equações chamadas de *equações de Karush-Kuhn-Tucker*. Essas estipulam que em x^*, o gradiente ∇f da função objetivo é combinação linear dos gradientes das restrições ativas, ou seja:

existe (λ^*, μ^*) em $\mathbb{R}^p \times \mathbb{R}^q$ tal que o vetor (x^*, λ^*, μ^*) seja solução do sistema:

$$\left\{ \begin{array}{l} \nabla f(x) + h'(x)^T \lambda + g'(x)^T \mu = 0 \\ h(x) = 0 \\ g(x) \leq 0 \\ \mu \geq 0 \\ \mu^T g(x) = 0 \end{array} \right. \tag{1.2}$$

x^* é então chamado de ponto KKT do problema de otimização ou de solução primal de (1.2). O par (λ, μ) é chamado de solução dual de (1.2).

Assim, é imediato ver que o sistema (1.2) é exatamente o sistema $KKT(\nabla f, h, -g)$. Portanto, sob as condições enunciadas acima, uma solução local do problema de otimização, também é solução primal de um sistema KKT especial.

É bem conhecido no entanto que, em geral, os problemas (1.2) e (1.1) não são equivalentes. Com efeito, uma solução primal do sistema KKT (1.2) pode não ser solução local do problema (1.1), e reciprocamente, uma solução local do problema (1.1) pode não verificar as condições KKT associadas.

Entretanto, no caso onde f é convexa, h afim e g convexa (problema dito convexo), sabemos que qualquer solução primal de (1.2) também é solução local do problema de minimização (1.1).

Apesar de não existir equivalência estrita entre os dois problema, é freqüente que o problema de minimização seja tal que qualquer solução local verifique uma das condições de qualificação das restrições. Neste caso, o conjunto das soluções locais do problema de otimização está incluído no das soluções primais do sistema $KKT(\nabla f, h, -g)$. Logo, este último conjunto, a priori mais fácil de caracterizar, é constituído por pontos candidatos à resolução do problema original. Obviamente, o caso mais favorável é quando o sistema KKT associado possui uma solução única.

1.1.2 Complementaridade não linear (NCP)

Este é um outro problema matemático clássico que aparece quando se trata de equilíbrios em economia ou de problemas de engenharia.

Dado $F : \mathbb{R}^n \longmapsto \mathbb{R}^n$, o problema de complementaridade consiste em achar x em $\mathbb{R}^n$ tal que:

$$x \geq 0, \quad F(x) \geq 0, \quad x^T F(x) = 0. \tag{1.3}$$

Adicionando uma variável $\mu \in \mathbb{R}^n$ ao problema, podemos reescrevê-lo na seguinte forma:

$$\begin{aligned} F(x) - \mu &= 0 && (1.4)\\ \text{para } 1 \leq i \leq n \qquad \mu_i x_i &= 0 && (1.5)\\ x &\geq 0 && (1.6)\\ \mu &\geq 0 && (1.7)\\ & && (1.8) \end{aligned}$$

Aqui também reconhecemos um sistema KKT especial: $KKT(F, \emptyset, Id_{\mathbb{R}^n})$, onde o símbolo $\emptyset$ significa que não tem restrição de igualdade e $Id_{\mathbb{R}^n}$ é a identidade de $\mathbb{R}^n$. Este sistema é claramente estritamente equivalente ao problema NCP original. Portanto, o problema NCP é um caso particular dos sistemas KKT, que pode ser resolvido usando diretamente os métodos descritos no capítulo 3 que são desenvolvidos para o caso geral.

1.1.3 Desigualdade variacional (VIP)

Os problemas de desigualdade variacional são muito importantes pois são muito gerais, e vários problemas em economia, teoria dos jogos, etc, ver [9, 11, 19], podem ser formulados como um VIP particular.

Sejam $F : \mathbb{R}^n \longmapsto \mathbb{R}^n$ e $\mathcal{C} \subset \mathbb{R}^n$ um conjunto fechado não vazio.

O problema VIP consiste em achar x em $\mathcal{C}$ tal que:

$$\text{para todo } y \text{ em } \mathcal{C}, \ \langle F(x), y - x \rangle \geq 0. \tag{1.9}$$

Dependendo do conjunto $\mathcal{C}$ e da função F envolvidos, o problema VIP pode assumir formas diversas, inclusive reduzir-se a outros problemas de formulação mais simples:

• **Caso $\mathcal{C} = \mathbb{R}^n_+$: o problema de complementaridade** $NCP(F)$.

Seja x^* uma solução do problema $VIP(F, \mathbb{R}^n_+)$.
Então, escolhendo sucessivamente $y = 0$, $y = 2x^*$ e $y = x^* + e_i$ para cada e_i vetor da base canônica de $\mathbb{R}^n$, obtemos que:

$$x^* \geq 0,\ F(x^*) \geq 0 \quad \text{e} \quad \langle F(x), x\rangle = 0.$$

Ou seja: x^* é solução do $NCP(F)$.
Reciprocamente, seja agora x^* uma solução do problema $NCP(F)$.
Então x^* e $F(x^*) \geq 0$ pertencem a $\mathbb{R}^n_+$, e para cada y em $\mathbb{R}^n_+$ temos:

$$\langle F(x), y - x^*\rangle = \underbrace{\langle F(x^*), y\rangle}_{\geq 0} - \underbrace{\langle F(x^*), x^*\rangle}_{=0} \geq 0.$$

Portanto x^* é solução do $VIP(F, \mathbb{R}^n_+)$.

• **Caso $\mathcal{C} = \mathbb{R}^n$: a equação $F(x) = 0$.**

Se x^* é solução de $VIP(F, \mathbb{R}^n)$, então para todo y em $\mathbb{R}^n$ temos:

$$\langle F(x^*), y\rangle \geq 0.$$

Com $y = -F(x^*)$, isto implica $F(x^*) = 0$. A recíproca é trivial.

• **Caso $F = \nabla f$, $\mathcal{C}$ convexo: condição necessária de otimalidade.**

Consideremos o seguinte problema de minimização:

$$\begin{array}{l} \min f(x) \\ \text{s.a } x \in \mathcal{C} \end{array}$$

Vamos ver que o $VIP(F, \mathcal{C})$ é equivalente à condição necessária de otimalidade de primeira ordem deste problema.

Seja x em $\mathcal{C}$.
A condição necessária de otimalidade de primeira ordem do problema em x diz que para qualquer curva $\gamma : [0, 1] \longrightarrow \mathcal{C}$ de classe C^1 partindo de x temos:

$$\langle \nabla f(x), \gamma'(0)\rangle \geq 0. \tag{1.10}$$

Como $\mathcal{C}$ é convexo, para qualquer ponto y em $\mathcal{C}$, dispomos da curva:

$$\gamma_y(t) = x + t(y - x),$$

que pertence a $\mathcal{C}$, é C^1 e verifica:

$$\gamma_y'(0) = y - x.$$

Portanto, se x verifica a condição (1.10), então para qualquer y em $\mathcal{C}$ temos:

$$\langle \nabla f(x), y - x \rangle \geq 0.$$

Ou seja: x é solução do problema $VIP(\nabla f, \mathcal{C})$

Reciprocamente, suponhamos que x^* é solução do $VIP(\nabla f, \mathcal{C})$.
Seja γ uma curva em $\mathcal{C}$ partindo de x^*. Tem-se:

$$\gamma'(0) = \lim_{t \to 0^+} \frac{\gamma(t) - \gamma(0)}{t} = \lim_{t \to 0^+} \frac{\gamma(t) - x^*}{t}$$

Portanto, por continuidade do produto escalar:

$$\begin{aligned} \langle \nabla f(x^*), \gamma'(0) \rangle &= \lim_{t \to 0^+} \langle \nabla f(x^*), \tfrac{\gamma(t) - x^*}{t} \rangle \\ &= \lim_{t \to 0^+} \tfrac{1}{t} \underbrace{\langle \nabla f(x^*), \gamma(t) - x^* \rangle}_{\geq 0} \geq 0 \end{aligned}$$

Ou seja: x^* verifica a condição (1.10).

- **Caso $\mathcal{C}$ é definido por restrições diferenciáveis: um sistema KKT**

Consideramos um conjunto $\mathcal{C}$ da forma:

$$\mathcal{C} = \mathcal{C}_{h,g} = \{x \in \mathbb{R}^n | h(x) = 0 \text{ e } g(x) \geq 0\},$$

onde h e g são funções C^1 como na definição do sistema KKT geral. Observemos que aqui as restrições de desigualdade têm um sinal oposto ao das restrições definidas para o problema de programação não linear.

Dessa vez, não vamos conseguir uma equivalência estrita entre o VIP e um sistema KKT particular. Em efeito, a relação entre eles resulta da reformulação do VIP como um problema de minimização. Mas como já vimos, este último é relacionado a um sistema KKT sem lhe ser equivalente.

Todo o raciocínio é baseado no fato de que para qualquer escolha da função F e do conjunto $\mathcal{C}$, as duas asserções seguintes são equivalentes:

(a) x^* é solução do $VIP(F, \mathcal{C})$

(b) x^* é minimizador global do problema:

$$\begin{array}{l} \min f(y) = \langle F(x^*), y - x^* \rangle \\ \text{s.a. } y \in \mathcal{C} \end{array} \tag{1.11}$$

É claro que se x^* é solução solução do VIP, o valor mínimo do problema (1.11) tem de ser positivo ou nulo. Já que $f(x^*) = 0$, o valor ótimo de f é então 0 e x^* é minimizador global.

Reciprocamente, se x^* for minimizador global de (1.11), como $f(x^*) = 0$, então tem-se $f(y) \geq 0$ para todo y em $\mathcal{C}$. Ou seja, x^* é solução do $VIP(F, \mathcal{C})$.

Logo, se verifica-se uma condição de qualificação das restrições em um ponto x^*, então este é solução primal do sistema KKT associado ao problema (1.11):

$$\begin{array}{rr} F(x^*) + h'(y)^T\lambda - g'(y)^T\mu = 0 & \\ h(y) = 0 & \\ \text{para } 1 \leq i \leq q \qquad \mu_i g_i(y) = 0 & \\ g(y) \geq 0, \mu \geq 0 & \end{array} \tag{1.12}$$

Esse é o sistema $KKT(F(x^*), h, g)$ na incógnita (y, λ, μ), onde $F(x^*)$ representa a função constante $y \longmapsto F(x^*)$. Obviamente, o sistema (1.12) não pode ser usado diretamente para achar x^* pois ele já supõe conhecido o ponto x^*, que é a solução que procuramos.

No entanto, sabemos que x^* é solução deste sistema. Portanto temos:

$$\left.\begin{array}{rr} F(x^*) + h'(x^*)^T\lambda^* - g'(x^*)^T\mu^* = 0 & \\ h(x^*) = 0 & \\ \text{para } 1 \leq i \leq q \qquad \mu_i^* g_i(x^*) = 0 & \\ g(x^*) \geq 0, \mu^* \geq 0 & \end{array}\right\} = KKT(F, h, g). \tag{1.13}$$

Assim, aparece um outro sistema KKT de qual x^* é solução primal. Desta vez ele é explorável para a resolução do problema VIP, pois é independente de qualquer solução.

Reciprocamente, se uma tripla x^* é solução primal do sistema $KKT(F, h, g)$, então ela é solução primal do sistema $KKT(F(x^*), h, g)$. Porém, a função objetivo do problema (1.11) é convexa, pois linear. Logo, se h for afim e g côncava, sabemos que x^* é um minimizador global de (1.11), e por isto uma solução do problema $VIP(F, h, g)$ original.

1.2 Abordagem clássica do problema

A fim de simplificar as notações para esta secção, denotamos $z = (x, \lambda, \mu)$ e escrevemos o sistema $KKT(F, h, g)$ sob a forma mais compacta seguinte:

$$\begin{cases} M(z) &= 0 \\ c(z) &\geq 0, \end{cases} \tag{1.14}$$

onde:

$$M(z) = \begin{pmatrix} A(z) \\ \mu_1 g_1(x) \\ \vdots \\ \mu_q g_q(x) \end{pmatrix}, \; A(z) = \begin{pmatrix} F(x) + h'(x)^T\lambda - g'(x)^T\mu \\ h(x) \end{pmatrix} \text{ e } \quad c(z) = \begin{pmatrix} g(x) \\ \mu_1 \\ \vdots \\ \mu_q \end{pmatrix}.$$

Entre os diversos métodos para tratar um problema do tipo (1.14), teremos interesse nas duas abordagens seguintes:

- Considerar o problema diretamente como uma equação não linear, resolvendo-o com um *método de Newton*, adaptado para levar em conta as restrições de desigualdade.
- Considerar o problema de minimização com restrições não lineares seguinte:

$$\begin{array}{l} \min \|M(z)\|^2 \\ \text{s.a. } c(z) \geq 0 \end{array}$$

As soluções de (1.14) são exatamente os minimizadores globais deste problema de programação não-linear que pode ser resolvido por um *método do tipo gradiente*.

1.2.1 Equação não linear: método de Newton

O método de Newton clássico segue o esquema de atualização seguinte:

$$x_{k+1} = x_k - M'(x_k)^{-1}M(x_k).$$

Tal e qual, este método se aplica a sistemas de equações. Porém, o sistema (1.14) contém as inequações $c(z) \geq 0$. Assim, neste caso, tem-se que contar com elas, e usar algum processo que garanta que essas restrições de desigualdade se cumpram, pelo menos nos pontos limites.

Enquanto isso, seja qual for esse processo de viabilização, certas condições primordiais à convergência local do método de Newton têm de ser verificadas para que o método dê resultados.

Em particular, se z^* é solução de (1.14), a matriz jacobiana $M'(z^*)$ tem que ser *não singular*. Isso é imperativo, pois garante a inversibilidade da jacobiana em torno desta solução, e portanto a validade da atualização acima, pelo menos localmente.

Porém, a matriz jacobiana de M em um ponto z qualquer é:

$$M'(z) = \begin{pmatrix} \frac{\partial A}{\partial x}(z) & \frac{\partial A}{\partial \lambda}(z) & & \frac{\partial A}{\partial \mu}(z) & \\ \mu_1 g_1'(x) & 0 & g_1(x) & & \\ \vdots & \vdots & & \ddots & \\ \mu_q g_q'(x) & 0 & & & g_q(x) \end{pmatrix}.$$

Assim, caso, para algum $1 \leq i_0 \leq q$, tenha-se:

$$|\mu_{i_0}| + |g_{i_0}(x)| = 0,$$

ou seja, caso μ_{i_0} e $g_{i_0}(x)$ sejam nulos ao mesmo tempo, então a $(n+i_0)$-ésima linha de $M'(z)$ será nula, fazendo com que $M'(z)$ seja trivialmente singular.

Este fenômeno de anulação simultânea de um coeficiente μ_i junto com a restrição de desigualdade g_i é chamado de *não complementaridade estrita*.

Resulta disso que se não se cumprir a complementaridade estrita na solução, o método de Newton é *estruturalmente inaplicável*, pois a matriz jacobiana falha em ser regular neste ponto.

Logo, muitos métodos para resolver problemas onde aparece considerações de complementaridade são válidos sob a hipótese de complementaridade estrita na solução. O inconveniente é que a ausência de complimentaridade estrita não é nada rara e acontece com problemas muito simples.

Exemplo 1.1

Consideremos o seguinte programa de minimização:

$$\begin{array}{l} \min \frac{1}{2}x^2 \\ \text{s.a. } x \geq 0 \end{array}$$

Este problema é de muito simples: a função objetiva é quadrática, convexa, com uma única restrição de desigualdade, que é linear. Portanto não apresenta nenhuma dificuldade de resolução teórica.
Porém, do ponto de vista da resolução numérica, ele não pode ser resolvido pelo método de Newton, pois não cumpre a condição de convergência. Com efeito, o seu sistema KKT associado é:

$$\begin{cases} x - \mu = 0 \\ \mu x = 0 \\ x \geq 0, \mu \geq 0. \end{cases}$$

Como é óbvio, a única solução deste sistema é o ponto $(x, \mu) = (0, 0)$. Ou seja, não vale a complementaridade estrita na solução, o que leva à singularidade da jacobiana na solução.

Efetivamente, a jacobiana do sistema é:

$$M'(z) = M'(x,\mu) = \begin{pmatrix} 1 & -1 \\ \mu & x \end{pmatrix}.$$

Dai, na solução, obtemos a matriz singular anunciada:

$$M'(0,0) = \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix}.$$

□

1.2.2 Minimização da norma: método do gradiente

Uma outra maneira de abordar o problema consiste em considerar um problema de minimização equivalente, no sentido que os minimizadores globais desse problema são as soluções do sistema KKT original. Por exemplo, consideremos o problema de minimização da função de mérito natural:

$$\begin{array}{l} \min \|M(z)\|^2 \\ \text{s.a. } c(z) \geq 0. \end{array} \tag{1.15}$$

Este tipo de abordagem tem alguns inconvenientes. A primeira observação é que, elevando o sistema ao quadrado, aumentamos um eventual mal condicionamento do problema. Segundo, sem nenhuma hipótese suplementar sobre a função M (monotonicidade, P-função), os métodos do tipo gradiente fornecem pontos estacionários de $\|M\|^2$, que podem ser só minimizadores locais, e não serem solução do problema KKT original. É a situação ilustrada pela Figura 1.1. Por fim, esses métodos não apresentam a convergência local rápida que pode ter um método tipo Newton.

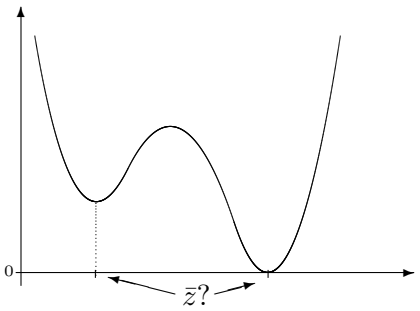


Figura 1.1: Mínimo local ou mínimo global ?

1.3 Abordagem não diferenciável do problema

De uma maneira geral [12, 30], o primeiro passo da abordagem não diferenciável para resolver o sistema KKT consiste em transformar o sistema original em um sistema equivalente da forma

$$H(x) = 0.$$

Essa transformação se apóia no fato de que um sistema KKT inclui nas suas equações um *problema de complementaridade não linear*. Então, usando funções ditas NCP, pode-se substituir as desigualdades por igualdades, e obter um sistema da forma desejada.

Ao longo deste trabalho, nós vamos focalizar três funções NCP: a função do mínimo, a função de Fischer-Burmeister e a função de Fischer-Burmeister penalizada. Nesta secção, só vamos definir essas três funções e enfatizar os problemas de não diferenciabilidade de cada uma, justificando assim o uso de algoritmos de resolução especiais. Deixamos para a Seção 4.2 um estudo mais profundo dessas funções e das propriedades necessárias à aplicação do método numérico descrito no Capítulo 3.

1.3.1 Transformação do sistema: funções NCP

O sistema $KKT(F, h, g)$ geral é:

$$\begin{array}{rcl} F(x) + h'(x)^T\lambda - g'(x)^T\mu & = & 0 \\ h(x) & = & 0 \end{array}$$

$$\boxed{\begin{array}{lrcl} \text{para } 1 \le i \le q & \mu_i g_i(x) & = & 0 \\ & g(x) & \ge & 0 \\ & \mu & \ge & 0 \end{array}} \longrightarrow \begin{array}{c} \text{Problema} \\ \text{de} \\ \text{Complementaridade} \end{array} \tag{1.16}$$

Nas três últimas equações e inequações reconhece-se um problema de complementaridade, que pode ser transformado em um sistema só com equações, por meio das funções chamadas de NCP.

Definição 1.1 (função NCP)
Uma função NCP é uma função $\phi : \mathbb{R}^2 \longrightarrow \mathbb{R}$ que se anula exatamente sobre o conjunto formado pelos dois semi-eixos positivos. Ou seja, a variedade dos zeros de ϕ é:

$$v(\phi) = \mathbb{R}_+ \times \{0\} \cup \{0\} \times \mathbb{R}_+$$

Ou, colocado de uma outra maneira, ϕ é tal que:

$$\phi(x, y) = 0 \Leftrightarrow x \ge 0, y \ge 0 \text{ e } xy = 0.$$

No que segue, denotamos:

$$N_0 = \mathbb{R}_+ \times \{0\} \cup \{0\} \times \mathbb{R}_+.$$

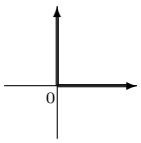


Figura 1.2: Zeros de uma função NCP

Com esta notação, uma função de $\mathbb{R}^2$ em $\mathbb{R}$ é NCP se e somente se o seu conjunto de zero é exatamente N_0.

Dada uma função NCP qualquer ϕ, os vetores (x, λ, μ) de $\mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^q$ que verificam as três equações-inequações de complementaridade em (1.16) são exatamente os vetores tais que:

$$\phi(g_i(x), \mu_i) = 0 \qquad \text{, para } 1 \leq i \leq q.$$

Logo, pode-se reescrever o sistema KKT geral sob a forma:

$$H_\phi(x, \lambda, \mu) = \begin{pmatrix} F(x) + h'(x)^T\lambda + g'(x)^T\mu \\ h(x) \\ [\phi(g_i(x), \mu_i)]_{1 \leq i \leq q} \end{pmatrix} = 0. \tag{1.17}$$

Dependendo da função ϕ escolhida, pode-se esperar obter sistemas transformados mais ou menos complexos, e mais ou menos adequados à aplicação de métodos numéricos de resolução.

1.3.2 As funções NCP estudadas

Aqui só apresentamos as definições das funções que serão estudadas mais em detalhe na Seção 4.2. Em particular, deixamos para essa secção a prova de que cada uma é de fato uma função NCP mesmo.

A função do mínimo

A função dita do mínimo é definida como:

$$\begin{array}{rccc} \phi_* : & \mathbb{R}^2 & \longrightarrow & \mathbb{R} \\ & (a,b) & \longmapsto & \min(a,b) \end{array}$$

Notemos que a função ϕ_* possui outras expressões tais como:

$$\phi_*(a,b) = \frac{1}{2}(a+b-|a-b|) = a - \max\{0, a-b\}$$

Da expressão com o valor absoluto, aparece claro que ϕ_* não é diferenciável na primeira bissetriz onde $a = b$.

A função de Fisher-Burmeister (FB)

A função de Fischer-Burmeister foi introduzida por Fischer [12], que a atribuou a Burmeister numa publicação privada anterior. Desde então, ela foi bastante estudada, principalmente quanto à resolução do problema de complementaridade não linear [6, 8, 15, 21, 31]:

$$\begin{array}{rccc} \phi_1 : & \mathbb{R}^2 & \longrightarrow & \mathbb{R} \\ & (a,b) & \longmapsto & a+b-\sqrt{a^2+b^2} \end{array}$$

A função $u \longmapsto \sqrt{u}$ sendo não diferenciável só na origem, então ϕ_1 não é diferenciável nos pontos tais que $a^2 + b^2 = 0$, ou seja, no único ponto $(0,0)$.

A função de Fischer-Burmeister Penalizada (FBP)

A função de Fischer-Burmeister Penalizada é a uma versão modificada de ϕ_1. Ela foi introduzida por Chen, Chen e Kanzow [3] no estudo do problema de complementaridade não linear:

$$\begin{array}{rccc} \phi_\alpha : & \mathbb{R}^2 & \longrightarrow & \mathbb{R} \\ & (a,b) & \longmapsto & \alpha\left(a+b-\sqrt{a^2+b^2}\right) + (1-\alpha)a^+b^+ \end{array}$$

onde, $\alpha \in]0,1]$ e $u^+ = \max(u,0)$.
Assim temos que ϕ_α é uma combinação convexa das funções ϕ_1 e ϕ_+:

$$\phi_\alpha(a,b) = \alpha\phi_1(a,b) + (1-\alpha)\phi_+(a,b),$$

onde $\phi_+(a,b) = a^+b^+$.

A função $u \longmapsto u^+$ não é diferenciável na origem, mas é C^∞ em qualquer outro ponto e é identicamente nula se $u < 0$. A priori então, terá um problema de diferenciabilidade com ϕ_+, nos pontos onde $a = 0$ ou $b = 0$, ou seja, no conjunto N_0 inteiro.

Pela reformulação (1.17), essas funções NCP geram três funções H_*, H_1 e H_α, cujos zeros são as soluções do sistema KKT original:

$$H_*(x,\lambda,\mu) = \begin{pmatrix} F(x) + h'(x)^T\lambda + g'(x)^T\mu \\ h(x) \\ [\min(g_i(x),\mu_i)]_{1\le i\le q} \end{pmatrix}$$

$$H_1(x,\lambda,\mu) = \begin{pmatrix} F(x) + h'(x)^T\lambda + g'(x)^T\mu \\ h(x) \\ \left[(g_i(x) + \mu_i - \sqrt{g_i^2(x) + \mu_i^2}\right]_{1\le i\le q} \end{pmatrix}$$

$$H_\alpha(x,\lambda,\mu) = \begin{pmatrix} F(x) + h'(x)^T\lambda + g'(x)^T\mu \\ h(x) \\ \left[\alpha\left(g_i(x) + \mu_i - \sqrt{g_i^2(x) + \mu_i^2}\right) + (1-\alpha)g_i(x)^+\mu_i^+\right]_{1\le i\le q} \end{pmatrix}$$

Da não diferenciabilidade de ϕ_*, ϕ_1 e ϕ_α, resulta a não diferenciabilidade das H correspondentes. Assim, embora de tenhamos uma forma mais cômoda para o problema, o tornamos não diferenciável. A Figura 1.3 mostra os conjuntos onde cada uma das ϕ não é diferenciável. No caso em que F é diferenciável, e h e g são duas vezes diferenciáveis, as H apresentam problemas de diferenciabilidade nos mesmos pontos.

O próximo capítulo apresenta os conceitos de cálculo diferenciável generalizado, que são a base do desenvolvimento de métodos capazes de gerir as equações onde a função objetivo é não diferenciável, no sentido clássico.

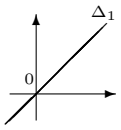
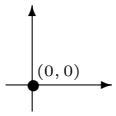
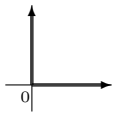
Mínimo	Fischer-Burmeister	FB Penalizada
$\phi_*(a,b) = \min(a,b)$	$\phi_1(a,b) = a + b - \sqrt{a^2+b^2}$	$\phi_\alpha(a,b) = \alpha\phi_1 + (1-\alpha)a_+b_+$
		

Figura 1.3: Pontos de não diferenciabilidade de ϕ_*, ϕ_1 e ϕ_α

Capítulo 2

Preliminares teóricos

Neste capítulo apresentamos os conceitos teóricos nos quais se apóiam os métodos de Newton não diferenciáveis.

Começamos por definições e resultados gerais sobre funções ponto-a-ponto e funções ponto-conjunto. Introduzimos então a derivada e o gradiente generalizados definidos por Clarke [5] para funções reais, assim como diferentes tipos de Jacobianos generalizados para funções vetoriais. Por fim, tratamos da semi-suavidade e de semi-derivada, noções introduzidas e desenvolvidas por Mifflin [25] e Qi [33, 31], e que são os conceitos chaves das provas de convergência dos métodos de Newton propostos no Capítulo 3.

Reconhecemos e avisamos que este capítulo é longo e teoricamente pesado. No entanto, isto vem de uma preocupação nossa em produzir um texto que seja o mais independente, coerente e didático possível. Assim por exemplo, muitas vezes explicitamos os teoremas e proposições de que precisamos, em vez de somente fazer referência aos artigos ou livros em que eles se acham. Também, para entender melhor os diversos conceitos, incluimos as demonstrações de certas proposições, ao mesmo tempo que tentamos ilustrar o discurso com vários exemplos.

2.1 Definições e teoremas gerais

A maioria dos conceitos definidos nesta secção trata de funções da forma padrão:

$$f : D \longrightarrow Y \quad \text{com } D \subset X.$$

Embora esses conceitos possam ser definidos em um quadro bem mais geral do que o da dimensão finita (espaços topológicos, normados, de Banach), restringimo-nos ao caso em que X e Y são espaços de dimensão finita. No entanto, tentamos usar um vocabulário que não seja específico à dimensão finita, de maneira que a transposição ao caso geral seja mais direta.

Freqüentemente, as noções e resultados apresentados aqui são clássicos e já conhecidos, mas valem a pena ser lembrados, pois serão usados em tudo o que segue. Isto servirá também para deixar bem claros as notações e o vocabulário utilizados no resto da dissertação.

Assim, consideramos funções $f : D \longrightarrow Y$ onde $D \subset X = \mathbb{R}^n$ e $Y = \mathbb{R}^m$. Quando não tiver confusão possível, tanto a norma de X quanto a de Y será denotada por $\|\cdot\|$. Por fim, a bola de centro x e de raio ρ do espaço normado X será denotada por $B_X(x, \rho)$.

Definição 2.1 (Funções Lipschitzianas)
Sejam $f : D \longrightarrow Y$, K uma constante positiva e x_0 um ponto de D.

- *f é* localmente K-Lipschitziana em x_0 *(ou no ponto x_0, ou ainda em volta de x_0) se existe uma vizinhança U de x_0 em D tal que:*

$$(\forall x, x' \in U)(\|f(x) - f(x')\| \leq K\|x - x'\|) \tag{2.1}$$

  A constante K é então chamada de coeficiente de Lipschitz de f em x_0.

- *f é K-Lipschitziana (sobre D, ou em D) se (2.1) é válida para $U = D$.*

Notemos que, pela própria definição, uma função localmente Lipschitziana em um ponto x também é localmente Lipschitziana em uma vizinhança de x. Logo, temos os seguintes resultados:

Proposição 2.1
Seja $f : D \longrightarrow Y$ localmente K-Lipschitziana em x_0. Então:

(i) f é contínua em toda vizinhança U de x_0 onde vale (2.1)

(ii) se f for diferenciável em x_0 então $|||f'(x_0)||| \leq K$.

Aqui acima, $|||\cdot|||$ é a norma nos operadores de $L(X, Y)$:

$$|||f||| = \sup_{\|x\|=1} \|f(x)\|$$

Em particular, se f for K-Lipschitziana e diferenciável sobre D então:

$$\sup_{x \in D} |||f'(x)||| \leq K$$

Demonstração

▶ Item (i): Dado a desigualdade (2.1), a propriedade é trivial.

▶ Item (ii): Suponhamos que f é diferenciável em x_0.
Seja U uma vizinhança de x_0 em D onde f é K-Lipschitziana.
Então temos por definição da diferencial em x_0:

$$\text{para } x \in U\text{: } f(x) = f(x_0) + f'(x_0)(x - x_0) + \|x - x_0\|\varepsilon(\|x - x_0\|),$$

onde $\varepsilon : \mathbb{R} \longrightarrow \mathbb{R}$ é tal que $\lim_{u \to 0} \varepsilon(u) = 0$.
Portanto:

$$\begin{aligned} \|f'(x_0)(x - x_0)\| &\leq \|f(x) - f(x_0)\| + \|x - x_0\|\varepsilon(\|x - x_0\|)| \\ &\leq (K + |\varepsilon(\|x - x_0\|)|)\, \|x - x_0\| \end{aligned}$$

isto para todo x em U, vizinhança de x.
Dai, para todo $\delta > 0$, existe $\eta > 0$ tal que para $h \in B(0, \eta)$: :

$$\|f'(x_0)h\| \leq (K + \delta)\|h\|$$

Por isto, por definição da norma de $f'(x_0)$ em $L(X, Y)$, temos para todo $\delta > 0$:

$$\|\|f'(x_0)\|\| \leq K + \delta$$

Enfim, fazendo $\delta \longrightarrow 0^+$, obtemos a desigualdade desejada.

■

Assim quando uma função é localmente Lipschitziana dispomos de um controle da diferencial nos pontos onde essa existe. Reciprocamente, pelo teorema vetorial dos crescimentos finitos, sabemos que se f for diferenciável em um aberto O de D e se a sua diferencial for limitada sobre O, então ela é Lipschitziana sobre O.

Mas este controle fica local e, sem outras hipóteses suplementares, não podemos garantir um controle global das variações de f. Isto quer dizer que localmente Lipschitziana em todos os pontos de D não implica Lipschitziana sobre D. Isto é o que ilustra o seguinte exemplo.

Exemplo 2.1
Consideremos a função:

$$\begin{array}{rccc} f: & \mathbb{R}_+^* & \longrightarrow & \mathbb{R} \\ & x & \longmapsto & \sqrt{x} \end{array}$$

Sabemos que f é derivável (e mesmo C^∞) sobre $\mathbb{R}_+^*$ com:

$$f'(x) = \frac{1}{2\sqrt{x}}$$

Logo, como f' é decrescente sobre $\mathbb{R}_+^*$, temos que ela é limitada sobre qualquer segmento $]a, b[\subset \mathbb{R}_+^*$ com:

$$f'(b) \leq f' \leq f'(a)$$

Portanto, f é localmente Lipschitziana em todos os pontos de $\mathbb{R}_+^*$, mas ela não é Lipschitziana sobre $\mathbb{R}_+^*$, pois a sua derivada não é limitada sobre $\mathbb{R}_+^*$ porque:

$$\exists \lim_{x \to 0^+} f'(x) = +\infty$$

□

Damos agora a definição da semi-continuidade superior para funções a valores reais:

Definição 2.2 (Semi-continuidade superior)
Sejam $f : D \longrightarrow \mathbb{R}$ e x_0 em D.
f é semi-contínua superior *em x_0 se para todo $\varepsilon > 0$ existe uma vizinhança U de x_0 em D tal que:*

$$\text{para todo } x \text{ em } U\text{: } f(x) \leq f(x_0) + \varepsilon$$

ou equivalentemente se:

$$\limsup_{x \to x_0} f(x) \leq f(x_0)$$

f é semi-contínua superior *se ela é* semi-contínua superior *em todos os pontos do seu domínio.*

Claramente, a semi-continuidade superior é uma propriedade mais fraca do que a continuidade: uma função contínua é semi-contínua superior. De fato, ela é um tipo de pre-continuidade no sentido que ela impõe uma certa regularidade local à função no ponto considerado, impedido certas descontinuidades mas autorizando outras.

Por exemplo, no caso $D = X = Y = \mathbb{R}$, a semi-continuidade superior impede descontinuidades do tipo da figura 2.2, mas autoriza as descontinuidades do tipo da figura 2.1.

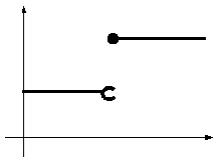


Figura 2.1: Função semi-contínua superior

Aqui vem um outro exemplo de descontinuidade que pode ter uma função semi-contínua superior:

Exemplo 2.2
Consideremos a função $f : [0,1] \longrightarrow \mathbb{R}$ definida como:

$$f(x) = \begin{cases} 0 \text{ se } x \in I_1 := [0, \frac{1}{2}[\cap (\mathbb{R} \backslash \mathbb{Q}) \\ 1 + (\frac{1}{2} - x) \text{ se } x \in I_2 := [0, \frac{1}{2}[\cap \mathbb{Q} \\ 1 \text{ se } x \in [\frac{1}{2}, 1] \end{cases}$$

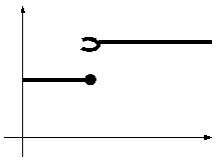


Figura 2.2: Função não semi-contínua superior

Assim temos que:

- Sobre $]\frac{1}{2}, 1]$: f é contínua pois constante. Portanto, também é semi-contínua superior.
- Sobre I_1: nem contínua, nem semi-contínua superior.

  Em efeito, I_2 é denso em $[0, \frac{1}{2}[$, portanto qualquer ponto de I_1 pode ser aproximado por uma seqüência de pontos de I_2. Como $f|_{I_2}$, a restrição de f ao conjunto I_2, é estritamente maior que 1, enquanto f é nula nos pontos de I_1 temos que para qualquer $x \in I_1$:

  $$\nexists \lim_{x' \to x} f(x'),$$

  e por isto f não é contínua em nenhum ponto de I_1.

  Quanto ao limite superior de f em um ponto x de I_1, temos:

  $$\limsup_{x' \to x} f(x') = \max \left(\underbrace{\limsup_{x' \to x, x' \in I_1} f(x')}_{=0}, \underbrace{\limsup_{x' \to x, x' \in I_2} f(x')}_{\geq 1} \right) \geq 1 > 0 = f(x).$$

- Sobre I_2: do mesmo modo, pela densidade de I_1 em $[0, \frac{1}{2}[$, f não é contínua em nenhum ponto de I_2. Mas, ela é semi-contínua superior em todos os pontos de I_2, pois neste caso, temos para qualquer ponto x em I_2:

  $$\limsup_{x' \to x} f(x') = \limsup_{x' \to x, x' \in I_2} f(x'),$$

  e por continuidade de $f|_{I_2}$, tem-se: $\limsup_{x' \to x, x' \in I_2} f(x') = f(x)$.
- Em $x = \frac{1}{2}$: como I_1 é denso em $[0, \frac{1}{2}]$ com $f|_{I_1} = 0$ e $f(\frac{1}{2}) = 1$, então f não pode ser contínua em $\frac{1}{2}$. Mas, de novo, ela é semi-contínua superior pois:

  $$\limsup_{x \to \frac{1}{2}} f(x) = \max \left(\underbrace{\limsup_{x \to \frac{1}{2}, x \in I_1} f(x)}_{=0}, \underbrace{\limsup_{x \to \frac{1}{2}, x \in I_2} f(x)}_{=1}, \underbrace{\limsup_{x \to \frac{1}{2}, x \in [\frac{1}{2}, 1]} f(x)}_{=1} \right) = 1 = f(\frac{1}{2}).$$

□

2.2 Funções ponto-conjunto

Como veremos mais para frente, no caso de uma função localmente Lipschitziana de $\mathbb{R}^n$ em $\mathbb{R}^n$, se definem diferentes tipos de Jacobianos generalizados que em vez de associar a um ponto um só elemento de $\mathcal{M}_n(\mathbb{R})$, lhe associam um conjunto de matrizes.

Nesta secção, damos definições e resultados a respeito desse tipo particular de funções, isto com a intenção de tratar os diferentes Jacobianos com um só ponto de vista global, unificando as diferentes noções. Essa abordagem poderia ser qualificada de topológica. É verdade que deste modo abstraimos um tanto o nosso estudo, mas também enfatizamos as propriedades dos Jacobianos generalizados que são primordiais ao bom funcionamento dos métodos de resolução considerados nesta dissertação. A idéia dessa abordagem acha a sua origem na proposição 2.1.5 de [5], assim como no espírito do conceito de semi-derivada de Qi [31]. No entanto, todas as proposições e demonstrações desta secção são pessoais.

As funções ponto-conjunto podem ser definidas em espaços muito gerais, mas certas definições supõem a existência de uma norma sobre Y. Portanto supomos pelo menos Y é um espaço normado.

Definição 2.3 (Função ponto-conjunto)
Uma função ponto-conjunto*, ou* multifunção *[5], de X em Y é uma função que a um ponto x de X associa um subconjunto de Y:*

$$\begin{array}{rccc} C: & X & \longrightarrow & \mathcal{P}(Y) \\ & x & \longmapsto & C(x) \end{array}$$

No lugar de $C : X \longrightarrow \mathcal{P}(Y)$, usaremos a notação mais compacta

$$C : X \rightrightarrows Y.$$

Definição 2.4 (Fechada)
Uma função ponto-conjunto $C : X \rightrightarrows Y$ é fechada *se o seu gráfico*

$$\mathcal{G}_C := \{(x, y); x \in X, y \in C(x)\}$$

é fechado em $X \times Y$. Ou seja, se verifica-se a seguinte implicação:

$$\left.\begin{array}{l} x_s \longrightarrow x \\ y_s \longrightarrow y \text{ com } y_s \in C(x_s) \end{array}\right\} \Longrightarrow y \in C(x)$$

Definição 2.5 (Localmente limitada)
Uma função ponto-conjunto $C : X \rightrightarrows Y$ é localmente limitada *por $\rho > 0$ em um ponto x de X se existe uma vizinhança U de x em X tal que:*

$$C(U) \subset B_Y(0, \rho)$$

De uma maneira análoga ao caso das funções ponto-a-ponto, define-se uma noção de semi-continuidade superior para as funções ponto-conjunto [5]. Em realidade, esta é o equivalente da continuidade, tanto que essas duas noções coincidem no caso de uma função ponto-conjunto cujos conjuntos imagens são unitários.

Definição 2.6 (Semi-continuidade superior)
Uma função ponto-conjunto $C : X \rightrightarrows Y$ é semi-contínua superior *no ponto x se, para todo $\varepsilon > 0$, existe U vizinhança de x tal que:*

$$\forall y \in U,\ C(y) \subset C(x) + B_Y(0, \varepsilon)$$

C é semi-contínua superior *se é* semi-contínua superior *em todos os pontos de X.*

Em geral, quando os conjuntos imagens são compactos, algumas das propriedades das funções ponto-a-ponto contínuas podem ser estendidas às funções ponto-conjunto semi-contínuas superiores.

Definição 2.7
Diremos que uma função ponto-conjunto $C : X \rightrightarrows Y$ tem uma imagem compacta *(resp.* imagem fechada*) se para todo x em X, o conjunto $C(x)$ é compacto (resp. fechado) em Y.*

Proposição 2.2
Uma função ponto-conjunto semi-contínua superior a imagem fechada é fechada.

Demonstração
Seja $C : X \rightrightarrows Y$ uma função ponto-conjunto semi-contínua superior a imagem fechada.
Queremos provar que o gráfico de C é fechado. Ou seja que:

$$\overline{\mathcal{G}_C} = \mathcal{G}_C.$$

Seja então um ponto (x, y) em $\overline{\mathcal{G}_C}$. Queremos mostrar que $y \in C(x)$.
Como $C(x)$ é fechado por hipótese, isto equivale a provar $y \in \overline{C(x)}$.
De fato vamos provar que para todo $\varepsilon > 0$:

$$y \in \overline{C(x) + B(0, \varepsilon)} := V_\varepsilon.$$

Pois sabemos que:

$$\overline{C(x)} = \bigcap_{\varepsilon > 0} V_\varepsilon.$$

Seja então $\varepsilon > 0$.
Pela semi-continuidade superior de C em x, existe uma vizinhança U de x tal que:

$$C(U) \subset V_\varepsilon. \tag{2.2}$$

Para qualquer vizinhança W de y, o conjunto $U \times W$ é uma vizinhança de (x, y). Logo, como supusemos $(x, y) \in \overline{\mathcal{G}_C}$, temos então:

$$(U \times W) \cap \mathcal{G}_C \neq \emptyset,$$

ou seja,

$$W \cap C(U) \neq \emptyset.$$

Por (2.2), obtemos então que toda vizinhança de y tem uma intersecção não vazia com V_ε.
Ou seja, provamos o que desejávamos:

$$y \in \overline{V_\varepsilon} = V_\varepsilon.$$

■

Definição 2.8 (imagem recíproca)
Sejam uma função ponto-conjunto $C : X \rightrightarrows Y$ e A um subconjunto de Y.
A imagem recíproca *de A por C é o conjunto:*

$$C^{-1}(A) = \{x \in X \mid C(x) \subset A\}.$$

Proposição 2.3 (imagem recíproca de abertos)
Sejam $C : X \rightrightarrows Y$ uma função ponto-conjunto a imagem compacta, e A um aberto de Y. Então, o conjunto $C^{-1}(A)$ é vizinhança dos seus pontos onde C for semi-contínua superior, ou seja:

$$\left\{x \mid x \in C^{-1}(A) \text{ e } C \text{s.c.s em } x\right\} \subseteq \overset{\circ}{C^{-1}}(A)$$

Demonstração
Seja x um ponto em $C^{-1}(A)$ onde C é semi-contínua superior.

Por definição da imagem recíproca temos:

$$C(x) \subset A$$

Como A é aberto, para cada y em $C(x)$, existe $\delta_y > 0$ tal que:

$$B(y, \delta_y) \subset A.$$

Dai, temos a cobertura aberta:

$$C(x) \subset \bigcup_{y \in C(x)} B(y, \frac{\delta_y}{2}).$$

Pela compacidade de $C(x)$, podemos extrair uma subcobertura finita:

$$C(x) \subset \bigcup_{i=1}^{N} B(y_i, \frac{\delta_{y_i}}{2}). \tag{2.3}$$

Logo, definindo $\varepsilon = \frac{1}{2}\min\{\delta_{y_1}, \ldots, \delta_{y_N}\}$, afirmamos que:

$$C(x) + B(0, \varepsilon) \subset A. \tag{2.4}$$

Com efeito, se y pertence a $C(x) + B(0, \varepsilon)$, então existe z em $C(x)$ tal que:

$$\|y - z\| \leq \varepsilon.$$

Porém, pela cobertura (2.3), existe $1 \leq i_0 \leq N$ tal que:

$$\|z - y_{i_0}\| \leq \frac{\delta_{i_0}}{2}.$$

Logo, por desigualdade triangular e pela definição de ε:

$$\|y - y_{i_0}\| \leq \varepsilon + \frac{\delta_{i_0}}{2} \leq \delta_{i_0},$$

ou seja:

$$y \in B(y_{i_0}, \delta_{i_0}) \subset A.$$

Agora, como C é semi-contínua superior em x, sabemos que existe uma vizinhança U de x tal que:

$$C(U) \subset C(x) + B(0, \varepsilon).$$

Por (2.4), obtemos:

$$C(U) \subset A,$$

e a proposição está provada. ■

Proposição 2.4
Seja $C : X \rightrightarrows Y$ uma função ponto-conjunto semi-contínua superior com imagem compacta. Então, a imagem por C de um compacto de X é um compacto de Y.

Demonstração
Seja K um compacto de X, e $(O_i)_{i \in I}$ uma cobertura aberta de $C(K)$:

$$C(K) \subset \bigcup_{i \in I} O_i, \quad O_i \text{ aberto em } Y.$$

Em particular, para todo x em X temos:

$$C(x) \subset \bigcup_{i \in I} O_i.$$

Como $C(x)$ é compacto por hipótese, existe um subconjunto finito $I_x \subset I$ tal que:

$$C(x) \subset \bigcup_{i \in I_x} O_i.$$

Pela Proposição 2.3, para cada x em X, existe então U_x vizinhança de x, que pode ser escolhida aberta, tal que:

$$C(U_x) \subset \bigcup_{i \in I_x} O_i.$$

Agora, a família $(U_x)_{x \in K}$ forma uma cobertura aberta do compacto K. Portanto, existem $x_1, \ldots, x_N$ em K tais que:

$$K \subset \bigcup_{j=1}^{N} U_{x_j}.$$

Dai temos que:

$$C(K) \subset \bigcup_{j=1}^{N} C(U_{x_j}) \subset \bigcup_{j=1}^{N} \bigcup_{i \in I_{x_j}} O_i = \bigcup_{i \in I_{x_1} \cup \cdots \cup I_{x_N}} O_i,$$

onde $I_{x_1} \cup \cdots \cup I_{x_N}$ é um subconjunto finito de I, junto com os I_{x_j}. ■

Um corolário direto desta proposição é apresentado a seguir:

Corolário 2.4.1
Toda função ponto-conjunto $C : X \rightrightarrows Y$ semi-contínua superior com imagem compacta é localmente limitada sobre os compactos de X.

No quadro diferenciável, uma hipótese importante para a validade dos métodos de Newton é a inversibilidade do Jacobiano na solução. No quadro não diferenciável, independentemente do tipo de Jacobiano generalizado considerado, a noção de regularidade equivalente pede a não singularidade de *todos* os elementos do Jacobiano generalizado.

Assim, para uma função ponto-conjunto $C : X \rightrightarrows Y = \mathcal{M}_n(\mathbb{R})$, definimos a *regularidade* da seguinte maneira:

Definição 2.9 (Regularidade)
Seja $C : X \longrightarrow \mathcal{M}_n(\mathbb{R})$ uma função ponto-conjunto. C é regular *no ponto x de X se:*

$$C(x) \subset \mathbb{GL}_n(\mathbb{R})$$

i.e. quando toda V em $C(x)$ é não singular.

Logo, uma propriedade muito interessante é a conservação local da regularidade, para funções ponto-conjunto semi-contínuas superiores. Esta é a proposição:

Teorema 2.1
Sejam uma função ponto-conjunto $C : X \longrightarrow \mathcal{M}_n(\mathbb{R})$ e x em X tais que:

(i) C é semi-contínua superior em x

(ii) $C(x)$ *é compacto*

(iii) C *é regular em* x

Então existe uma vizinhança U *de* x *tal que:*

(a) C *é regular sobre* U*:*

$$C(U) \subset \mathbb{GL}_n(\mathbb{R})$$

(b) existe $c > 0$ *tal que para todo* y *em* U *e* V *em* $C(y)$*:*

$$\|V^{-1}\| \leq c.$$

Demonstração
Para provar esta proposição, provamos antes um lema de ordem geral:
Lema:
Se E é um espaço vetorial normado de dimensão finita, e $A, B \subset E$ são respectivamente compacto e aberto tais que:

$$A \subset B,$$

então existe uma vizinhança compacta de A incluída em B. Ou seja, existem K compacto e O aberto de E tais que:

$$A \subset O \subset K \subset B.$$

Prova do lema:
Como B é um aberto de E e que $A \subset B$, então para cada x em A existe $\rho_x > 0$ tal que:

$$\overline{B(x, \rho_x)} \subset B.$$

Consideremos então a seguinte cobertura aberta de A:

$$A \subset \bigcup_{x \in A} B(x, \rho_x).$$

Pela compacidade de A, podemos extrair uma subcobertura finita e obtemos:

$$A \subset \underbrace{\bigcup_{i=1}^{N} B(x_i, \rho_i)}_{:=O} \subset \underbrace{\bigcup_{i=1}^{N} \overline{B(x_i, \rho_i)}}_{:=K} \subset B,$$

onde O é aberto pela união de abertos, e K é compacto por união finita de bolas, compactas em dimensão finita. □

Voltemos à demonstração da proposição.

Por hipótese, $C(x)$ é um compacto, incluído em $\mathbb{GL}_n(\mathbb{R})$, subconjunto aberto de $\mathcal{M}_n(\mathbb{R})$, espaço normado de dimensão finita.
Portanto, podemos aplicar o lema precedente, e sabemos que o conjunto $C(x)$ possui uma vizinhança compacta em $\mathbb{GL}_n(\mathbb{R})$.
Sejam então K compacto e O aberto tais que:

$$C(x) \subset O \subset K \subset \mathbb{GL}_n(\mathbb{R}).$$

Pela Proposição 2.3, existe uma vizinhança U de x tal que valem as seguintes inclusões, que provam o item (a):

$$C(U) \subset O \subset \mathbb{GL}_n(\mathbb{R}).$$

Enfim, como a inversão matricial é contínua sobre $\mathbb{GL}_n(\mathbb{R})$, a imagem por ela do compacto K é limitada. Em particular, para todo V em $C(U) \subset K$, temos:

$$\|V^{-1}\| \leq \sup_{V \in K} \|V^{-1}\| < +\infty.$$

Isto prova o item (b). ■

2.3 Derivada e gradiente generalizados

Nesta secção apresentamos os conceitos introduzidos por Clarke [5], que servem de base a todo o cálculo diferencial generalizado. Com um propósito didático, começamos apresentando o raciocínio seguido por Clarke, que leva à definição do gradiente generalizado para funções a valores reais localmente Lipschitzianas $f : X \longrightarrow \mathbb{R}$, via noção de derivada direcional generalizada.

Do mesmo modo que na Seção 2.1, embora Clarke defina a derivada e o gradiente generalizados no caso geral de X espaço de Banach, restringimos-nos ao caso em que X é euclidiano, ou seja X é de dimensão finita ($X = \mathbb{R}^n$) e munido da norma euclidiana $\|x\|^2 = \sum_{i=1}^{n} x_i^2$.

2.3.1 Derivada direcional generalizada

Seja $f : X \longrightarrow \mathbb{R}$ localmente Lipschitziana no ponto x. Clarke define a *derivada direcional generalizada* no ponto x na direção v como:

$$f^\circ(x; v) := \limsup_{\substack{y \to x \\ t \to 0^+}} \frac{f(y + tv) - f(y)}{t} \tag{2.5}$$

Podemos notar que ao contrário das definições mais clássicas de derivada, essa definição usa um limite superior em vez de um limite forte. Por isto, a quantidade $f^\circ(x; v)$ sempre

existe e é bem definida em $\overline{\mathbb{R}}$, a reta real completada, seja como for a função f, localmente Lipschitziana ou não.

No entanto, quando f é localmente K-Lipschitziana em volta de x, então para qualquer v em X, y em uma certa vizinhança de x e $t > 0$ próximo suficiente de 0, temos:

$$\left|\frac{f(y+tv)-f(y)}{t}\right| \leq \frac{K\|y+tv-y\|}{|t|} = K\|v\|$$

Portanto, $f^\circ(x;v)$ é seguramente finita e verifica:

$$|f^\circ(x;v)| \leq K\|v\|$$

Neste quadro, a derivada generalizada tem as seguintes propriedades [5, proposição 2.1.2]:

Proposição 2.5
Sejam x em X e $f : X \longrightarrow \mathbb{R}$ localmente K-Lipschitziana no ponto x. Então:

(a) A função $f^\circ(x;\cdot) : v \longmapsto f^\circ(x;v)$ assume os seus valores em $\mathbb{R}$, é positivamente homogênea, subaditiva em X, e verifica:

$$|f^\circ(x;v)| \leq K\,\|v\|$$

(b) a função $f^\circ : (x,v) \longmapsto f^\circ(x;v)$ é semi-contínua superior;

(c) a função $f^\circ(x,\cdot) : v \longmapsto f^\circ(x;v)$ é K-Lipschitziana sobre todo X;

(d) para todo x, v em X: $f^\circ(x;-v) = (-f)^\circ(x;v)$.

Exemplo 2.3
Caso de uma função C^1 numa vizinhança de x

Pelo teorema do valor intermediário, para cada (y,t) existe um ponto $u_{y,t}$ no segmento vetorial $]y, y+tv[$, ou seja $u_{y,t} = y + \alpha_{y,t}tv$ com $\alpha_{y,t} \in]0,1[$, tal que:

$$\frac{f(y+tv)-f(y)}{t} = f'(u_{y,t})v$$

Por construção, quando $y \longrightarrow x$ e $t \longrightarrow 0^+$ temos $u_{y,t} \longrightarrow x$.
Portanto, por continuidade de f' em x temos que:

$$f^\circ(x;v) = \limsup_{\substack{y\to x\\ t\to 0^+}} f'(u_{y,t})v = \lim_{\substack{y\to x\\ t\to 0^+}} f'(u_{y,t})v = f'(x)v$$

Ou seja, a derivada generalizada no ponto x na direção v é a avaliação da derivada (de Fréchet) $f'(x)$ na direção v. Ou de uma outra maneira, essa coincide com a derivada direcional clássica.

Caso da função $f(x) = |x|$

- Para $x \neq 0$, estamos no caso C^1 e portanto temos:

$$f^\circ(x; v) = \begin{cases} v \text{ se } x > 0 \\ -v \text{ se } x < 0 \end{cases}$$

- Para $x = 0$, temos:

$$\frac{f(y + tv) - f(y)}{t} \leq \frac{|y + tv - y|}{t} = |v|$$

enquanto:

$$\exists \lim_{t \to 0^+} \frac{f(0 + tv) - f(0)}{t} = |v|$$

Portanto $f^\circ(0; v) = |v|$.

□

2.3.2 Gradiente generalizado

A partir da derivada generalizada, Clarke define o gradiente generalizado da seguinte maneira:

Definição 2.10 (Gradiente generalizado)
Seja $f : X \longrightarrow \mathbb{R}$, localmente Lipschitziana.

O gradiente generalizado de f no ponto x de X é:

$$\partial f(x) := \{\zeta \in X' \mid \forall v \in X,\ f^\circ(x; v) \geq \langle \zeta, v \rangle\}$$

X' denota o espaço vetorial dos funcionais lineares contínuos de X em $\mathbb{R}$.

Assim, para cada x em X, o gradiente $\partial f(x)$ é um subconjunto de X', isto quer dizer que a função $x \longmapsto \partial f(x)$ é uma função ponto-conjunto de X em X'.

No entanto, no quadro euclidiano em qual nos colocamos, o teorema de Riesz permite de identificar X' ao próprio espaço X, e, na prática, $\partial f(x)$ torna-se então uma função ponto-conjunto de X nele mesmo. Esta multifunção tem as seguintes propriedades [5, proposições 2.1.2 e 2.1.5]:

Proposição 2.6
Se f é K-Lipschitziana em volta de x então:

(a) $\partial f(x)$ é um subconjunto não vazio, convexo, compacto de X, e tem-se:

$$\text{para todo } \xi \in \partial f(x) : \quad \|\xi\| \leq K.$$

(b) Para v em X tem-se:

$$f^\circ(x;v) = \max\{\langle \xi, v\rangle; \xi \in \partial f(x)\}.$$

(c) ∂f é localmente limitada, fechada e semi-contínua superior[1].

Calculemos o gradiente generalizado para as mesmas funções do exemplo 2.3:

Exemplo 2.4
Funções C^1 em uma vizinhança de x

Como vimos no exemplo 2.3, a derivada generalizada de f é:

$$f^\circ(x;v) = f'(x)v = \langle \nabla f(x), v\rangle.$$

Dai vem que os elementos de $\partial f(x)$ são os ξ em X que verificam para todo v em X:

$$\langle \xi, v\rangle \quad \leq \quad \langle \nabla f(x), v\rangle,$$

ou seja, para todo v em X tem-se:

$$\langle \nabla f(x) - \xi, v\rangle \quad \geq 0.$$

Pela linearidade em v, isto impõe que $\xi = \nabla f(x)$ e:

$$\partial f(x) = \{\nabla f(x)\}.$$

A função valor absoluto: $f(x) = |x|$

- se $x \neq 0$, f é C^1 em x, e seguindo o que precede temos:

$$\partial f(x) = f'(x) = \begin{cases} 1 \text{ se } x > 0 \\ -1 \text{ se } x < 0. \end{cases}$$

- se $x = 0$: $\partial f(x)$ é formado pelos $\xi \in \mathbb{R}$ tais que para todo v em $\mathbb{R}$ tenha-se:

$$\xi v \leq f^\circ(0;v) = |v|.$$

Para $v \neq 0$, isso equivale a:

$$\xi \underbrace{\frac{v}{|v|}}_{=\pm 1} \leq 1,$$

ou seja, temos exatamente:

$$|\xi| \leq 1.$$

Assim:

$$\partial f(x) = [-1, 1].$$

[1] A semi-continuidade superior necessita explicitamente que X seja de dimensão finita.

Funções convexas

Seja $f : X \longrightarrow \mathbb{R}$ convexa.

Em análise convexa, a notação $\partial f(x)$ denota o subdiferencial de f no ponto x. Este é o conjunto dos funcionais ξ de X' tais que:

$$\text{para todo } x' \in X, \qquad f(x') \geq \quad \langle \xi, x' - x \rangle + f(x)$$

De fato, se f for também localmente Lipschitziana em x, então o seu gradiente generalizado em x coincide com o subdiferencial em x da análise convexa, fazendo com que as duas notações sejam coerentes [5, proposição 2.2.7].

Também, sabe-se de que f é direcionalmente diferenciável em todos os pontos, e que a sua derivada direcional em um ponto x é a função suporte de $\partial f(x)$. Isto quer dizer que ela verifica para todo v em X:

$$f'_d(x; v) = \sup\{\langle \xi, v \rangle; \xi \in \partial f(x)\}$$

Assim, tem-se também:

$$f^\circ(x; \cdot) = f'_d(x; \cdot)$$

□

Tal e qual, a definição 2.10 do gradiente generalizado não é muito cômoda, pois necessita calcular a derivada generalizada. Por isto, na prática o seu cálculo passa por caracterizações que se apóiam na dimensão finita de X, na regularidade de f ou na compatibilidade do gradiente generalizado com as operações aritméticas.

Para $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, localmente Lipschitziana em uma vizinhança aberta U de x, o *teorema de Rademacher* afirma que o conjunto Ω_f dos pontos de U onde f não é Fréchet-diferenciável é de medida nula. Dai, vem o seguinte teorema, muito útil para o cálculo do gradiente generalizado.

Teorema 2.2 (Caracterização de $\partial f(x)$ em dimensão finita)
Se $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ é localmente Lipschitziana em volta de x, e S um conjunto de medida nula, então o gradiente generalizado de f no ponto x é:

$$\partial f(x) = \text{co}\left\{\lim \nabla f(x_p); x_p \to x, x_p \notin S \cup \Omega_f\right\}.$$

Assim, $\partial f(x)$ é o envelope convexo dos valores de aderência dos gradientes de f quando aproxima-se do ponto x – dentro do conjunto de diferenciabilidade de f.

Exemplo 2.5

- Se f é C^1 em x, por continuidade de ∇f em x, temos imediatamente:

$$\partial f(x) = \text{co}\{\nabla f(x)\} = \{\nabla f(x)\}.$$

- Se $f(x) = |x|$:

$$\partial f(x) = \begin{cases} \{1\} & \text{se } x > 0 \\ \{-1\} & \text{se } x < 0 \\ \text{co}\{-1, 1\} = [-1, 1] & \text{se } x = 0. \end{cases}$$

□

Exemplo 2.6

Consideremos a função definida sobre $\mathbb{R}$ por:

$$f(x) = \begin{cases} x^2 \sin \frac{1}{x} & \text{se } x \neq 0 \\ 0 & \text{se } x = 0. \end{cases}$$

Esta função é muito interessante no sentido que ela é um exemplo de função Fréchet-diferenciável em $x = 0$, mas para a qual acontece a seguinte situação que vai contra a primeira intuição:

$$\partial f(0) \neq \{\nabla f(0)\}.$$

Com efeito, sabemos que f é diferenciável sobre $\mathbb{R}$ com:

$$f'(x) = \begin{cases} 2x \sin \frac{1}{x} - \cos \frac{1}{x} \text{ se } x \neq 0 \\ \lim\limits_{x \to 0} x \sin \dfrac{1}{x} = 0 \text{ se } x = 0. \end{cases}$$

Esta derivada é limitada sobre $\mathbb{R}^*$ pois é contínua e verifica:

$$\exists \lim_{|x| \to +\infty} f'(x) = 1 \qquad \text{e} \qquad \exists \lim_{x \to 0} f'(x) = 0.$$

Logo, como $f'(0) = 0$, f' é limitada sobre todo $\mathbb{R}$ e por isto f é Lipschitziana sobre $\mathbb{R}$.

Calculemos então $\partial f(x)$:

▶ Fora da origem, f é C^∞, portanto C^1, e por isto, para $x \neq 0$:

$$\partial f(x) = \{\nabla f(x)\} = \{2x \sin \frac{1}{x} - \cos \frac{1}{x}\}.$$

▶ Na origem:

- Por um lado temos:

$$\exists f'(0) = 0,$$

- por outro lado, se $x \neq 0$ temos:

$$f'(x) = 2x \sin \frac{1}{x} - \cos \frac{1}{x}.$$

Dai, temos:

$$|f'(x)| \leq 1 + 2|x \sin \frac{1}{x}|$$

Portanto:

$$-1 \leq \liminf_{x \to 0} f'(x) \leq \limsup_{x \to 0} f'(x) \leq 1$$

Pelo Teorema 2.2 então, tem-se:

$$\partial f(0) \subset \text{co}\,([-1,1]) = [-1,1]$$

Porém, para qualquer $a \in [-1,1]$, tomando:

$$x_p = \frac{1}{\arccos a + 2p\pi + \pi}$$

temos $x_p \longrightarrow 0$ e:

$$\exists \lim_{p \to +\infty} f'(x_p) = -\lim_{p \to +\infty} \cos\left(\arccos a + 2p\pi + \pi\right) = \cos(\arccos a) = a$$

Assim, obtemos a inclusão inversa: $[-1,1] \subset \partial f(0)$.
Finalmente tem-se:

$$\partial f(0) = [-1,1] \supsetneq \{\nabla f(0)\}.$$

□

De uma maneira geral, logo que elas existam, as derivadas usuais em x (Gâteaux, Hadamard, Fréchet) pertencem ao conjunto $\partial f(x)$ [5, proposição 2.2.2]. No entanto, o exemplo anterior mostra que a existência da derivada de Fréchet – e portanto a de todas as outras – não garante um gradiente generalizado reduzido a essa derivada só. De fato precisa-se ter uma regularidade mais forte da função.

Definição 2.11 (Quase-diferenciabilidade)
Seja $f : X \longrightarrow \mathbb{R}$ localmente Lipschitziana em x.
Diz-se que f é quase-diferenciável *em x se:*

(i) Todas derivadas direcionais usuais no ponto x existem:

$$\text{para todo } v \text{ em } X, \quad \exists f'_d(x;v) := \lim_{t \to 0^+} \frac{f(x+tv) - f(x)}{t}$$

(ii) essas coincidem com as derivadas direcionais generalizadas correspondentes:

$$\text{para todo } v \text{ em } X, \quad f^\circ(x;v) = f'_d(x;v)$$

As funções quase-diferenciáveis (cf. Mifflin [25] que cita Pshenichnyi [26] a respeito) são simplesmente chamadas de *regulares* por Clarke [5], e têm as seguintes propriedades [5, proposição 2.3.6].

Proposição 2.7
Seja $f : X \longrightarrow \mathbb{R}$ localmente Lipschitziana em x. Então:

(a) Se f é convexa então f é quase-diferenciável em x;

(b) Uma combinação linear positiva de funções quase-diferenciáveis em x é quase-diferenciável em x;

(c) Se f é Gâteaux-diferenciável em x de derivada $f'_G(x)$ e é quase-diferenciável em x, então:

$$\partial f(x) = \{f'_G(x)\}.$$

Assim, no caso da função $f(x) = x^2 \sin \frac{1}{x}$, falta a quase-diferenciabilidade em 0 para que $\partial f(0)$ se reduz ao gradiente de Fréchet. O problema vem do fato que, embora existam todas as derivadas direcionais clássicas, os crescimentos relativos da função *em volta* da origem não convergem a essas derivadas.

De fato a condição de Fréchet-diferenciabilidade em um ponto x não toma em conta as variações entre pontos na vizinhança de x, mas só os crescimentos a partir de x. Por isso, como o observa Clarke [5], a noção diferencial mais ligada à de gradiente generalizado é a de *diferenciabilidade estrita*.

Definição 2.12 (Diferenciabilidade estrita)
Uma função $f : X \longrightarrow Y$ é estritamente diferenciável em x se existe $f'_s(x)$ em $L(X, Y)$ tal que:

$$\exists \lim_{x_1, x_2 \to x} \frac{f(x_2) - f(x_1) - \langle f'_s(x), x_2 - x_1 \rangle}{\|x_2 - x_1\|} = 0$$

Notemos que, escolhendo x_1 e x_2 convenientemente, é fácil ver que a diferenciabilidade estrita implica a diferenciabilidade de Fréchet e que as diferenciais coincidem. Também, observemos que essa definição da diferenciabilidade estrita é diferente da definição dada por Clarke [5]. No entanto, no caso de dimensão finita, as duas definições são equivalentes (cf. a prova da Proposição 2.9 mais para frente).

Proposição 2.8
Se f é C^1(-Fréchet) no ponto x, então f é estritamente diferenciável em x.

Demonstração
Queremos mostrar que:

$$\exists \lim_{x_1,x_2\to x} \frac{f(x_2)-f(x_1)-\langle f'(x), x_2-x_1\rangle}{\|x_2-x_1\|}=0$$

Sejam x_1, x_2 em X.
Como f é Fréchet-diferenciável, pelo teorema do valor intermediário vetorial, existe x_3 em $]x_1, x_2[$ tal que:

$$f(x_2)-f(x_1)=\langle f'(x_3), x_2-x_1\rangle$$

Dai vem que:

$$\begin{aligned}\frac{f(x_2)-f(x_1)-\langle f'(x), x_2-x_1\rangle}{\|x_2-x_1\|} &= \frac{\langle f'(x_3), x_2-x_1\rangle-\langle f'(x), x_2-x_1\rangle}{\|x_2-x_1\|}\\ &\leq \frac{|\|f'(x_3)-f'(x)|\|\,\|x_2-x_1\|}{\|x_2-x_1\|}\\ &= |\|f'(x_3)-f'(x)|\| \longrightarrow 0\end{aligned}$$

Como x_3 tende a x junto com x_1 e x_2, e pela continuidade de f' obtemos a convergência para 0 na desigualdade acima, como desejávamos. ■

O que ganhamos com a diferenciabilidade estrita é exatamente o que faltava à Fréchet-diferencial, isto é um controle dos crescimentos relativos da função *em volta do ponto considerado*. Assim, como anunciado, esta regularidade facilita o cálculo do gradiente generalizado pois temos a seguinte proposição [5, proposição 2.2.4]:

Proposição 2.9
Seja $f : X \longrightarrow \mathbb{R}$ localmente Lipschitziana em x.

(a) Se f é estritamente diferenciável em x então f é quase-diferenciável em x.

(b) f é estritamente diferenciável em $x \Longleftrightarrow \partial f(x)$ reduz-se a um ponto

Demonstração

► (a): Suponhamos f estritamente diferenciável. Seja então $f'_s(x)$ a sua diferencial estrita. Sabemos que f é então direcionalmente diferenciável e que:

$$f'_d(x;\cdot)=\langle f'_s(x), \cdot\rangle.$$

Por outro lado, como f é estritamente diferenciável, temos para todo v fixo em X:

$$\exists \lim_{\substack{y\to x\\ t\to 0^+}} \frac{f(y+tv)-f(y)-\langle f'_s(x), tv\rangle}{t\|v\|}=0.$$

O vetor v é fixo, por isto temos de fato:

$$\exists \lim_{\substack{y\to x\\ t\to 0^+}} \frac{f(y+tv)-f(y)}{t} = \frac{\langle f_s'(x), tv\rangle}{t},$$

e finalmente:

$$\exists \lim_{\substack{y\to x\\ t\to 0^+}} \frac{f(y+tv)-f(y)}{t} = \langle f_s'(x), v\rangle.$$

Logo:

$$f^\circ(x;v) = \limsup_{\substack{y\to x\\ t\to 0^+}} \frac{f(y+tv)-f(y)}{t} = \lim_{\substack{y\to x\\ t\to 0^+}} \frac{f(y+tv)-f(y)}{t} = \langle f_s'(x), v\rangle = f_d'(x;v)$$

Por isto, f é quase-diferenciável.

► (b) $\Longrightarrow$: Se f é estritamente diferenciável em x, então ela também é Gâteaux diferenciável em x e quase-diferenciável pelo o ponto (a) acima. Portanto, a Proposição 2.7 (c) garante que que:

$$\partial f(x) = \{\underbrace{f_G'(x)}_{=f_s'(x)}\}$$

$\Longleftarrow$: Se o gradiente generalizado se reduz a $\{\xi\}$, então pela Proposição 2.6 (b), temos para v em X:

$$f^\circ(x;v) = \max\{\langle\zeta, v\rangle; \quad \zeta\in\partial f(x)\} = \langle\xi, v\rangle$$

Logo, afirmamos então que para todo v em X:

$$\boxed{\exists \lim_{\substack{y\to x\\ t\to 0^+}} \frac{f(y+tv)-f(y)}{t} = \langle\xi, v\rangle}$$

Com efeito, temos:

$$\begin{aligned}\liminf_{\substack{y\to x\\ t\to 0^+}} \frac{f(y+tv)-f(y)}{t} &= -\limsup_{\substack{y\to x\\ t\to 0^+}} \frac{f(y)-f(y+tv)}{t}\\ &= -\limsup_{\substack{y\to x\\ t\to 0^+}} \frac{f(y+tv-tv)-f(y+tv)}{t}\end{aligned}$$

Com a mudança de variável $z = y+tv$ obtemos:

$$\begin{aligned}\liminf_{\substack{y\to x\\ t\to 0^+}} \frac{f(y+tv)-f(y)}{t} &= -\limsup_{\substack{z\to x\\ t\to 0^+}} \frac{f(z-tv)-f(z)}{t}\\ &= -f^\circ(x;-v) = -\langle\xi,-v\rangle = \langle\xi,v\rangle = f^\circ(x;v)\\ &= \limsup_{\substack{y\to x\\ t\to 0^+}} \frac{f(y+tv)-f(z)}{t}\end{aligned}$$

e por isto temos a convergência anunciada.

Também segundo Clarke (demonstração da proposição 2.2.1), como f é localmente Lipschitziana, essa convergência tem que ser *uniformemente compacta* em v.

Sejam x_1, x_2 em X.

Denotando $\bar{v} = \frac{x_2 - x_1}{\|x_2 - x_1\|}$ e $t = \|x_2 - x_1\|$, temos $x_2 = x_1 + t\bar{v}$ e:

$$\frac{f(x_2) - f(x_1) - \langle \xi, x_2 - x_1 \rangle}{\|x_2 - x_1\|} = \left\| \frac{f(x_1 + t\bar{v}) - f(x_1)}{t} - \langle \xi, \bar{v} \rangle \right\|$$

Seja agora $\varepsilon > 0$.

X sendo de dimensão finita, a sua esfera unidade é compacta. Assim, pela convergência uniformemente compacta, existe $\delta = \delta(\varepsilon)$ tal que:

$$0 \leq t, \|x_1 - x\| \leq \delta \Rightarrow \forall \|v\| = 1, \ \left\| \frac{f(x_1 + tv) - f(x_1)}{t} - \langle \xi, v \rangle \right\| \leq \varepsilon$$

Em particular para $v = \bar{v}$ obtemos:

$$\frac{f(x_2) - f(x_1) - \langle \xi, x_2 - x_1 \rangle}{\|x_2 - x_1\|} \leq \varepsilon$$

desde que $\|x_2 - x_1\| \leq 2\delta$.

A diferenciabilidade estrita em x é então provada.

■

Logo, juntando os resultados precedentes temos o corolário:

Corolário 2.9.1

Seja $f : X \longrightarrow \mathbb{R}$, localmente Lipschitziana em x.
Então $\partial f(x)$ é bem definido em uma certa bola $B(x, \varepsilon)$ e as seguintes asserções são equivalentes:

(a) f é C^1 em $B(x, \varepsilon)$

(b) para todo x' em $B(x, \varepsilon)$, $\partial f(x')$ reduz-se a um ponto

Demonstração

- $(a) \Rightarrow (b)$: Se f é C^1 em $B(x, \varepsilon)$, pela Proposição 2.8 ela é estritamente diferenciável em qualquer x' em $B(x, \varepsilon)$, e pela Proposição 2.9(b), $\partial f(x')$ se reduz a $\{\nabla f(x')\}$.
- $(b) \Rightarrow (a)$: pela Proposição 2.9(b), f é (estritamente) diferenciável em $B(x, \varepsilon)$. Também, ∂f é uma função ponto-a-ponto com $\partial f(x') = \{\nabla f(x')\}$ para x' em $B(x, \varepsilon)$. Assim, a semi-continuidade superior de ∂f em $B(x, \varepsilon)$ equivale à continuidade de ∇f em em $B(x, \varepsilon)$.

■

2.3.3 Regras de cálculo do gradiente generalizado

Como no caso das derivadas clássicas, o cálculo do gradiente generalizado pode ser feito usando regras de cálculo ou outros resultados sobre a ação da derivação sobre as somas, produtos, etc. Isto permite calcular o gradiente generalizado de funções compostas por funções mais simples, cujos gradientes já são conhecidos.

No entanto, as fórmulas de cálculo obtidas no caso geral não são exatas, e sem hipóteses de regularidade sobre as funções consideradas, só se consegue calcular um conjunto maior do que o gradiente generalizado.

Proposição 2.10 (Multiplicação por um scalar [5, proposição 2.3.1])
Para todo λ em $\mathbb{R}$ temos: $\partial(\lambda f)(x) = \lambda \partial f(x)$.

Proposição 2.11 (Soma finita [5, proposição 2.3.3])
Para $(f_i)_{i\in I}$ uma família finita de funções localmente Lipschitzianas no ponto x, então $\sum_{i\in I} f_i$ também é localmente Lipschitziana em x e temos:

$$\partial\left(\sum_{i\in I} f_i\right)(x) \subset \sum_{i\in I} \partial f_i(x).$$

Se no máximo uma das f_i não é estritamente diferenciável, então a inclusão acima é uma igualdade.

Proposição 2.12 (Produto [5, proposição 2.3.13])
Sejam f_1, $f_2 : X \longrightarrow \mathbb{R}$ localmente Lipschitzianas em volta de x. Então $f_1 f_2$ também é localmente Lipschitziana em volta de x e:

$$\partial\left(f_1 f_2\right)(x) \subset f_2(x)\partial f_1(x) + f_1(x)\partial f_2(x).$$

Se além disso, $f_1(x) \geq 0$, $f_2(x) \geq 0$, e, f_1 e f_2 são regulares em x, então vale a igualdade e $f_1 f_2$ é regular em x.

Proposição 2.13 (Regra da cadeia [5, teorema 2.3.10])
Sejam x em X, $f : X \longrightarrow Y$ estritamente diferenciável em x e $g : Y \longrightarrow \mathbb{R}$ localmente Lipschitziana em $f(x)$.
Então $g \circ f$ é localmente Lipschitziana em x e:

$$\partial(g \circ f)(x) \subset \partial g(f(x)) \circ f'_s(x)$$

A igualdade vale se g (resp. $-g$) é regular no ponto $f(x)$. Neste case $g \circ f$ (resp. $-g \circ f$) é regular no ponto x.

2.4 Jacobianos generalizados

Vamos considerar agora funções vetoriais $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$, localmente Lipschitzianas em volta do ponto x.

Neste quadro, existem várias definições que generalizam a noção de Jacobiano, cada uma tendo as suas particularidades. Da mesma maneira que no caso das funções reais, o objetivo é achar um objeto que assuma o papel de diferencial, em particular do ponto de vista de primeira aproximação linear da função original.

Foi Clarke [5] quem primeiro definiu um Jacobiano generalizado: o subdiferencial de Clarke ∂f. Este é a generalização direta do gradiente generalizado apresentado na secção precedente. No entanto, por ser mais conveniente para o desenrolar das demonstrações, começamos por definir o B-subdiferencial $\partial_B f$ que foi introduzido por Qi [29] com o fim de melhorar resultados teóricos de convergência dos métodos de Newton não diferenciáveis, originalmente baseados no subdiferencial ∂f. Ele é um subconjunto gerador do subdiferencial de Clarke. Assim esta ordem de apresentação permitirá demonstrar certas propriedades do subdiferencial de Clarke apoiando-se nas do B-subdiferencial, já provadas.

Agora, do ponto de vista prático, há ainda um outro subdiferencial que se revela mais fácil de calcular: o C-subdiferencial introduzido por Han e Sun [17]. Este será o último tipo de subdiferencial de que trataremos.

Daqui para frente, o espaço $\mathcal{M}_{m,n}(\mathbb{R})$ é suposto munido de uma norma qualquer $\|.\|$, por exemplo da norma euclidiana de Frobenius $\|A\| = \mathrm{Tr}(A^T A)$, e, dado que não tem confusão possível, denotaremos também por B a bola unidade associada.

Para $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ diferenciável no ponto x, a matriz Jacobiana de f em x é o seguinte elemento de $\mathcal{M}_{m,n}(\mathbb{R})$:

$$f'(x) = \begin{pmatrix} f_1'(x) \\ \vdots \\ f_m'(x) \end{pmatrix} = \begin{pmatrix} \nabla f_1^T(x) \\ \vdots \\ \nabla f_m^T(x) \end{pmatrix}$$

Lembramos que de novo, vale o teorema de Rademacher, e o conjunto dos pontos onde f não é diferenciável é de medida nula. Este conjunto é denotado Ω_f.

2.4.1 B-subdiferencial: $\partial_B f$

Qi [29] definiu o B-subdiferencial para uma função $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ localmente Lipschitziana em x, da seguinte maneira:

$$\partial_B f(x) = \{\lim f'(x_s); x_s \longrightarrow x, x_s \notin \Omega_f\} \tag{2.6}$$

Ou seja, $\partial_B f(x)$ é o conjunto dos valores de aderência das seqüências das Jacobianas de f quando aproxima-se de x, ficando fora do conjunto Ω_f. Logo, $\partial_B f(x)$ é uma função ponto-conjunto de $\mathbb{R}^n$ em $\mathcal{M}_{m,n}(\mathbb{R})$ que pode ser reescrita da seguinte maneira:

$$\partial_B f(x) = \bigcap_{\varepsilon>0} \overline{\{f'(y); y \in B(x,\varepsilon), y \notin \Omega_f\}}. \tag{2.7}$$

Denotaremos por $\Delta_f(x,\varepsilon)$ o conjunto $\{f'(y); y \in B(x,\varepsilon), y \notin \Omega_f\}$.

Exemplo 2.7
Consideremos de novo a função valor absoluto $f : x \longmapsto |x|$.
Na origem, o seu B-subdiferencial é constituído por todos os valores de aderência das seqüências $(f(x_p))$, onde:

$$x_p \neq 0 \text{ para todo } p, \text{ e } \lim_{p\to+\infty} x_p = 0.$$

Como

$$\exists f'(0^+) = 1 \text{ e } \exists f'(0^-) = -1,$$

então:

$$\partial_B f(0) = \{1, -1\}.$$

□

O B-subdiferencial tem as seguintes propriedades:

Proposição 2.14
Seja $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ localmente K-Lipschitziana em x.

(a) Para todo x em $\mathbb{R}^n$, $\partial_B f(x)$ é um compacto não vazio de $\mathcal{M}_{m,n}(\mathbb{R})$;

(b) $\partial_B f$ é localmente limitada por K em x;

(c) $\partial_B f$ é semi-contínua superior em x;

(d) $\partial_B f$ é fechada;

(e) $\partial_B f(x) \subset \partial_B f_1(x) \times \cdots \times \partial_B f_m(x)$, onde o segundo membro da inclusão é o conjunto das matrizes cuja i-ésima linha pertence ao B-subdiferencial $\partial_B f_i(x)$.

Demonstração

► Item (a):

Pela expressão (2.7), temos diretamente que $\partial_B f(x)$ é fechado como intersecção de fechados.

Como f é localmente K-Lipschitziana em x, pela Proposição 2.1, existe um certo $\varepsilon_0 > 0$ tal que, para todo $y \in B(x, \varepsilon_0)$, as matrizes Jacobianas $f'(y)$ são limitadas por K.

Ou seja, o conjunto $\Delta_f(x, \varepsilon)$ é limitado.

Logo, por inclusão, obtemos que $\partial_B f(x)$ também é limitado.

Sendo em dimensão finita, $\partial_b f(x)$ é afinal compacto.

Agora, pelo teorema de Rademacher existem seqüências de Jacobianas $(f'(x_s))$ com $x_s \longrightarrow x$.

Além disso, uma tal seqüência é limitada pois fica, para s suficientemente grande, dentro do conjunto limitado $\Delta_f(x, \varepsilon_0)$.

Portanto, sendo em dimensão finita, uma tal seqüência possui um valor de aderência que por definição pertence a $\partial_B f(x)$, e portanto $\partial_B f(x)$ é não vazio.

- Item (c):

Seja $\delta > 0$.

Queremos achar $\eta > 0$ tal que:

$$\partial_B f(B(x, \eta)) \subset \partial_B f(x) + B(0, \delta)$$

Como observado no item anterior, para $\varepsilon \leq \varepsilon_0$, o conjunto $\Delta_f(x, \varepsilon)$ é limitado. Sendo fechado por construção, ele é compacto.

Para $p \geq 0$, denotemos então:

$$K_p := \Delta_f(x, 2^{-(p+N)})$$

onde N é tal que $2^{-N} \leq \varepsilon_0$.

Os K_p são compactos, encaixados e temos:

$$\bigcap_{p \geq 0} K_p = \partial_B f(x) \subset \partial_B f(x) + B(0, \delta) := O$$

Afirmamos então que, por O ser aberto, existe $p_0 \geq 0$ tal que:

$$K_{p_0} \subset O$$

Com efeito, raciocinemos pelo absurdo: para todo $p \geq 0$ existe um ponto y_p em $K_p \cap \complement O$.

Como O é aberto, o seu complementar $\complement O$ é fechado, e por isto os conjuntos $K_p \cap \complement O$ são compactos, por ser cada um fechado no K_p respectivo.

Como os K_p são encaixados, a seqüência (y_p) pertence ao compacto K_0, e por isto podemos extrair uma subseqüência $(y_{\varphi(p)})$ convergente a um certo y.

Para cada $p \geq 0$, para todo $q \geq p$ temos $\varphi(q) > \varphi(p) \geq p$ e por isto:

$$y_q \in K_{\varphi(q)} \subset K_{\varphi(p)} \subset K_p$$

Portanto, passando legitimamente ao limite em q, obtemos para todo $p \geq 0$:

$$y \in K_p$$

e então:

$$y \in \bigcap_{p\geq 0} K_p = \partial_B f(x) \subset O.$$

Porém, por outro lado, como $(y_p) \subset \complement O$, o ponto y pertence também ao fechado $\complement O$.

Contradição e resultado.

- (d): $\partial_B f$ é semi-contínua superior com imagem compacta, e por isto com imagem fechada. Portanto, pela Proposição 2.2 ela é fechada.
- (b): $\partial_B f$ é semi-contínua superior com imagem compacta, portanto, pela Proposição 2.4, ela é localmente limitada, pois as bolas são compactas em dimensão finita.
- (e): Esta propriedade vem da própria estrutura dos Jacobianos; pois se uma matriz V pertence a $\partial_B f(x)$, então ela é o limite de uma seqüência do tipo:

$$f'(x_s) = \begin{pmatrix} f_1'(x_s) \\ \vdots \\ f_m'(x_s) \end{pmatrix}.$$

Portanto, cada linha v_i $(1 \leq i \leq m)$ de V é limite da seqüência de gradientes $(f_i'(x_s))$ correspondente. Por isto tem que pertencer a $\partial_B f_i(x)$ e temos

$$V \in \partial_B f_1(x) \times \cdots \times \partial_B f_m(x).$$

■

Nota: Em geral a inclusão em (e) é estrita. Por exemplo, consideremos a função:

$$\begin{array}{rl} f: & \mathbb{R} \longrightarrow \mathbb{R}^2 \\ & x \longmapsto (|x|, |x|) \end{array}$$

Em $x = 0$ temos:

$$\partial_B f(0) = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \end{pmatrix} \right\},$$

enquanto:

$$\partial_B f_1(0) \times \partial_B f_2(0) = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\}.$$

2.4.2 Subdiferencial de Clarke: ∂f

Para $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ localmente Lipschitziana em x, Clarke define um subdiferencial como uma generalização direta da caracterização do gradiente generalizado do teorema 2.2. Esta é a sua definição [5, definição 2.6.1]:

$$\partial f(x) = \text{co}\,\{\lim f'(x_s); x_s \to x, x_s \notin \Omega_f\} \tag{2.8}$$

Assim, temos:

$$\partial f(x) = co[\partial_B f(x)]$$

Logo o jacobiano generalizado de Clarke tem as seguintes propriedades:

Proposição 2.15
Seja $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ localmente K-Lipschitziana em x.

(a) Para todo x em $\mathbb{R}^n$, $\partial f(x)$ é um subconjunto não vazio, convexo e compacto de $\mathcal{M}_{m,n}(\mathbb{R})$;

(b) ∂f é localmente limitada por K em x;

(c) ∂f é fechada;

(d) ∂f é semi-contínua superior;

(e) $\partial f(x) \subset \partial f_1(x) \times \cdots \times \partial f_m(x)$, onde o segundo membro da inclusão é o conjunto das matrizes cuja i-ésima linha pertence ao gradiente generalizado $\partial f_i(x)$.

Demonstração
Tudo vem da identidade $\partial f(x) = co[\partial_B f(x)]$ e das propriedades similares já provadas para o B-subdiferencial:

- (a): O envelope convexo de um compacto não vazio é convexo, compacto não vazio. É um corolário do teorema de Caratheodory que em dimensão finita o envelope convexo conserve a compacidade.
- (d): Como $\partial_B f$ é semi-contínua superior, para todo $\varepsilon > 0$ existe $\delta = \delta_\varepsilon > 0$ tal que se $y \in B(x, \delta)$ então:

  $$\partial_B f(y) \subset \partial_B f(x) + \varepsilon B_{m,n}$$

  Dai vem que:

  $$\partial f(y) = \text{co}\,[\partial_B f(y)] \subset \text{co}\,[\partial_B f(x) + \varepsilon B_{m,n}] \subset \text{co}\,[\partial_B f(x)] + \text{co}\,[\varepsilon B_{m,n}] = \partial f(x) + \varepsilon B_{m,n}$$

- (b): Podemos vê-lo diretamente ou usar a Proposição 2.4 como na demonstração para o B-subdiferencial. Diretamente, como $\partial_B f(x)$ é localmente limitada por K em x, resulta então que para todo y em uma certa vizinhança de x temos:

  $$\partial f(y) = co[\partial_B f(y)] \subset co[B(0, K)] = B(0, K).$$

- (c) de novo resulta da Proposição 2.2 e dos itens (a) e (d).
- (e): Basta se convencer que o envelope convexo de um produto cartesiano está incluído no produto cartesiano dos envelopes convexos.

■

Notas:

1. Rigorosamente, no caso $m = 1$, existe uma pequena incoerência nas notações entre o subdiferencial de Clarke e o gradiente generalizado na sua caracterização do Teorema 2.2.

   Com efeito, neste teorema, o gradiente generalizado é caracterizado como um subconjunto de $\mathbb{R}^n$, ou seja como um conjunto de vetores coluna, enquanto o subdiferencial de Clarke é constituído por vetores linha. Este último corresponde ao gradiente generalizado como na Definição 2.10, como subconjunto de $(\mathbb{R}^n)'$.

   Já que a maioria dos resultados são apresentados para o caso geral (m qualquer), daqui para frente, ∂f denota o *Jacobiano* generalizado, mesmo no caso de funções a valores reais.

2. Para $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, o gradiente generalizado de Clarke tem a seguinte propriedade [5, Proposição 2.3.2]:

   $$\text{se } x \text{ é um extremum local de } f, \text{ então } 0 \in \partial f(x). \tag{2.9}$$

   Logo, no espírito da otimização, não parece ser uma boa idéia considerar conjuntos que não tem essa propriedade. Conjuntos menores podem não possuir o zero mesmo em extremi locais, enquanto com conjuntos maiores, o risco é de adicionar um zero em pontos que não são extremi locais.

   O Exemplo 2.7 ilustra uma situação onde não o B-subdiferencial na origem não possui o zero, embora $x = 0$ seja mínimo local da função valor absoluto.

   De fato, Qi [29] definiu o B-subdiferencial com um propósito completamente contrário à visão da otimização. Com efeito, o objetivo é tirar elementos de $\partial f(x)$ de tal maneira que seja mais fácil cumprir a regularidade (i.e. a inversibilidade dos elementos do subdiferencial). Em particular, no caso de funções reais, isto quer dizer tirar o zero, e perder a condição de otimalidade de primeira ordem (2.9).

3. Ainda no caso $m = 1$, a Proposição 2.6 (b) dava:

   $$f^\circ(x; v) = \max \{\langle \xi, v\rangle ; \xi \in \partial f(x)\}.$$

   De fato, temos o resultado:

   $$f^\circ(x; v) = \max \{\langle \xi, v\rangle ; \xi \in \partial_B f(x)\}.$$

   Com efeito, para v em X, um elemento de $\partial f(x)$ no qual atinge-se o máximo escreve-se:

   $$\xi = \sum_{k=1}^{K_\xi} \lambda_k \xi_k \quad \text{onde } (\xi_k) \subset \partial_B f(x), (\lambda_k) \geq 0 \text{ e } \sum_{k=1}^{K_\xi} \lambda_k = 1.$$

Portanto tem-se, por um lado:

$$\max\{\langle\xi, v\rangle; \xi \in \partial_B f(x)\} \leq \max\{\langle\xi, v\rangle; \xi \in \partial f(x)\},$$

e por outro lado:

$$\max\{\langle\xi, v\rangle; \xi \in \partial f(x)\} = \langle\sum_{k=1}^{K_\xi} \lambda_k \xi_k, v\rangle = \sum_{k=1}^{K_\xi} \lambda_k \langle\xi_k, v\rangle \leq \max\{\langle\xi, v\rangle; \xi \in \partial_B f(x)\}.$$

Por fim, damos duas propriedade interessantes de ∂f: a primeira é um tipo de regra da cadeia, a segunda é a generalização do teorema do valor intermediário. Em particular, esta última permite ter uma fórmula de controle do crescimento da função pelo Jacobiano generalizado.

Proposição 2.16 (Regra da cadeia [5, Teorema 2.6.6])
Sejam $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ localmente Lipschitziana em x, $g : \mathbb{R}^m \longrightarrow \mathbb{R}$ localmente Lipschitziana em $f(x)$. Então, $g \circ f$ é localmente Lipschitziana em x e:

$$\partial(g \circ f)(x) \subset \text{co}[\partial g(F(x))\partial f(x)]$$

Se g for estritamente diferenciável em $f(x)$ então tem-se:

$$\partial(g \circ f)(x) = g'(f(x))\partial f(x)$$

Teorema 2.3 (Valor Médio [5, Proposição 2.6.5])
Se f é Lipschitziana em um subconjunto U aberto convexo de $\mathbb{R}^n$, então para x, y em U, temos:

$$f(y) - f(x) \in co\,(\partial f([x, y])(y - x))$$

onde o segundo membro é o envelope convexo do conjunto: $\{V(y - x); \quad V \in \partial f(v), v \in [x, y]\}$.

2.4.3 C-subdiferencial: $\partial_C f$

O C-subdiferencial foi introduzido por Han e Sun [17]. Para uma função $f : y \longmapsto (f_1(y), \cdots, f_m(y))$ de $\mathbb{R}^n$ em $\mathbb{R}^m$, localmente Lipschitziana em volta de x, ele é definido como:

$$\partial_C f(x) = \partial f_1(x) \times \cdots \times \partial f_m(x) \tag{2.10}$$

Pela Proposição 2.15 (e), temos então as seguintes inclusões:

$$\partial_B f(x) \subseteq \partial f(x) \subseteq \partial_C f(x) \tag{2.11}$$

Assim $\partial_C f(x)$ é um conjunto maior do que os dois primeiros subdiferenciais. No entanto, calcular os seus elementos se revela mais fácil do que para $\partial_B f(x)$ ou $\partial f(x)$, pois o cálculo se faz componente por componente, reduzindo o problema ao cálculo de m *gradientes* generalizados.

O C-subdiferencial tem as seguintes propriedades:

Proposição 2.17
Seja $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$.

(a) Para todo x *em* $\mathbb{R}^n$, $\partial_C f(x)$ *é um subconjunto não vazio, convexo e compacto de* $\mathcal{M}_{m,n}(\mathbb{R})$

(b) $\partial_C f$ *é localmente limitada por* K *em* x

(c) $\partial_C f$ *é fechada*

(d) $\partial_C f$ *é semi-contínua superior*

Demonstração
Todas essas propriedades são propriedades conservadas pelo produto cartesiano finito. Porém pela Proposição 2.6, o gradiente generalizado satisfaz a todas. ■

2.5 Semi-suavidade

A noção de semi-suavidade (semismoothness) foi inicialmente introduzida por Mifflin [25] para funcionais de $\mathbb{R}^n$ em $\mathbb{R}$. Essa foi estendida por Qi e Sun [33] para funções de $\mathbb{R}^n$ em $\mathbb{R}^m$. A semi-suavidade é muito ligada à resolução de equações não lineares não diferenciáveis, pois ela é uma das condições de convergência dos métodos de Newton generalizados para esse tipo de equações.

Definição 2.13 (Semi-suavidade)
Uma função $f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ *é* semi-suave *no ponto* x *de* $\mathbb{R}^n$ *se:*

(a) f *é localmente Lipschitziana em* x

(b) para todo $h \neq 0$ *em* $\mathbb{R}^n$*:*

$$\exists \lim_{\substack{V \in \partial f(x+th') \\ h' \to h, t \to 0^+}} Vh' \tag{2.12}$$

Nota: Denotando $y = x + th'$ no limite acima, temos então:

$$h' \to h, t \to 0^+ \iff y \longrightarrow_h x,$$

onde a convergência na parte direita da equivalência é chamada de convergência a x na direção h. Isto quer dizer que $y \longrightarrow x$ com $y \neq x$ e $\frac{y-x}{\|y-x\|} \longrightarrow h$. Logo, a definição da semi-suavidade assume a seguinte forma mais compacta [28]:

$$\lim_{\substack{V \in \partial f(y) \\ y \to_h x}} Vh.$$

Da existência do limite em (2.12), junto com o Teorema 2.3, se deduz que se f é semi-suave em x, então ela é direcionalmente diferenciável em x e que a derivada direcional coincide com esse limite [33, Proposição 2.1]:

$$f'_d(x; h) = \lim_{\substack{V \in \partial f(x+th') \\ h' \to h, t \to 0+}} Vh' \tag{2.13}$$

Qi e Sun [33] provaram que a semi-suavidade se caracteriza pela convergência uniforme das derivadas direcionais no ponto considerado. Na sua formulação, esse resultado é bem mais cômodo, e é ele que é geralmente usado para provar que uma função é semi-suave. Precisamente, a caracterização da semi-suavidade é a seguinte:

Proposição 2.18
Seja $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ localmente Lipschitziana em x.
Então as seguintes asserções são equivalentes:

(a) f é semi-suave em x;

(b) f é direcionalmente diferenciável em x e para toda $V \in \partial f(x + h), h \to 0$:

$$Vh - f'_d(x; h) = \circ(\|h\|);$$

(c) f é direcionalmente diferenciável em x e:

$$\exists \lim_{\substack{x+h \notin \Omega_f \\ h \to 0}} \frac{f'_d(x + h; h) - f'_d(x; h)}{\|h\|} = 0.$$

Nota: Pela sua simplicidade, a propriedade do item (b) é freqüentemente usada como a própria definição da semi-suavidade [13, 22, 32].

Diz-se que f é *semi-suave de ordem* p se para toda $V \in \partial f(x+h)$ e $h \to 0$ tem-se :

$$Vh - f'_d(x;h) = O(\|h\|^{p+1}).$$

Temos então a seguinte caracterização, similar à da semi-suavidade.

Proposição 2.19
Seja $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ *localmente Lipschitziana em* x.
Então f *é semi-suave de ordem* p *em* x *se e somente se* f *é direcionalmente diferenciável em* x *e existe* $K > 0$ *tal que:*

$$\text{para } x+h \notin \Omega_f, h \to 0, \quad \frac{\|f'_d(x+h;h) - f'_d(x;h)\|}{\|h\|^{p+1}} \leq K.$$

No caso em que $p = 1$, f é chamada também de *fortemente semi-suave* em x [22, 30]. Em regra geral, todas as proposições sobre a semi-suavidade podem ser diretamente generalizadas à semi-suavidade de ordem p.

O item (c) da Proposição 2.18 mostra que a semi-suavidade em x é uma condição de regularidade das derivadas direcionais, na vizinhança de x. Qi [29] chama essa propriedade de *semi-continuidade* da derivada direcional no ponto x.

Exemplo 2.8
Seja a função do Exemplo 2.6:

$$f(x) = \begin{cases} x^2 \sin \frac{1}{x} & \text{se } x \neq 0 \\ 0 & \text{se } x = 0 \end{cases}$$

A função f não é semi-suave em $x = 0$, pois lhe falta a coerência entre as suas derivadas direcionais imposta pela semi-suavidade.
Com efeito, por um lado todas as derivadas direcionais de f em 0 são nulas:

$$f'_d(0;h) \equiv_h f'(0)h = 0$$

Por outro lado, como foi provado no Exemplo 2.6, qualquer ponto de $[-1,1]$ é valor de aderência de $f'_d(h;h)$ quando $h \to 0$.
Logo:

$$\nexists \lim_{\substack{h \neq 0 \\ h \to 0}} \frac{f'_d(h;h)}{\|h\|},$$

e por isto f não satisfaz ao critério do item (c) da Proposição 2.18, e portanto não é semismooth. □

Segundo Mifflin [25], a classe das funções semi-suaves reais é muito larga. Ela engloba as funções C^1, as funções convexas, assim como as funções C^1 por pedaços. Também, ela é estável tanto pelas operações básicas de soma, diferença e produto, bem como pela composição funcional.

Exemplo 2.9
Sejam $f, g : \mathbb{R}^n \longrightarrow \mathbb{R}$ duas funções semi-suaves.
Para $a, b \in \mathbb{R}$ temos:

$$\min(a,b) = \frac{a - b - |a-b|}{2} \quad \text{e} \quad \max(a,b) = \frac{a - b + |a-b|}{2}$$

Portanto, por serem C^1 por pedaços, as funções min e max são semi-suaves.
Logo as funções compostas $\min(f,g)$ e $\max(f,g)$ também são semi-suaves. □

O resultado chave da extensão dessas propriedades de estabilidade à semi-suavidade para funções vetoriais vem da seguinte proposição [33, Corolário 2.4] que faz com que o estudo da semi-suavidade para funções vetoriais se reduz a um estudo separado da semi-suavidade componente por componente.

Proposição 2.20
Seja $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ localmente Lipschitziana em x. Então f é semi-suave em x se e somente se todas as suas componentes são semi-suaves em x.

Por fim, damos um teorema devido a Fischer [13], que estabelece a estabilidade da semi-suavidade forte por composição [13, Teorema 19]. Este resultado simplifica bastante o estudo da semi-suavidade das funções compostas:

Proposição 2.21
Sejam $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ semi-suave de ordem p em x, e $g : \mathbb{R}^m \longrightarrow \mathbb{R}^n$ semi-suave de ordem p em $f(x)$. Então a função composta $h = g \circ f$ é semi-suave de ordem p em x.

2.6 Semi-derivada

Como vimos na Subseção 2.3.3, as fórmulas do cálculo diferencial generalizado não são exatas. Afim de superar essa desvantagem em particular, Qi, Ralph e Zhou [31] definem o conceito de *semi-derivada*. A idéia subjacente é a seguinte: conseguir em um quadro não diferenciável uma função equivalente à derivada, que não seja uma função ponto-conjunto, e que dê uma aproximação linear de primeira ordem da função de origem.

Definição 2.14 (Semi-derivada)
Sejam $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ contínua, e x em $\mathbb{R}^n$.

Uma função $g : \mathbb{R}^n \longrightarrow \mathcal{M}_{n,m}(\mathbb{R})$ *é uma* semi-derivada *de* f *no ponto* x *se:*

(i) g *é localmente limitada em* x*;*

(ii) para $h \longrightarrow 0$*:* $f(x+h) = f(x) + g(x+h)h + \circ(\|h\|)$.

Se diz que g *é uma* semi-derivada forte *de* f *no ponto* x *se para* $h \longrightarrow 0$*, tem-se de fato:*

$$f(x+h) = f(x) + g(x+h)h + O\left(\|h\|^2\right)$$

Diz-se que g é uma uma semi-derivada (resp. semi-derivada forte) de f, se g é uma semi-derivada (resp. semi-derivada forte) de f em todos os pontos de $\mathbb{R}^n$.

É claro que se f possui uma semi-derivada em x, então ela é localmente Lipschitziana em x. Com efeito, para y próximo suficiente de x temos:

$$\| \circ (\|y - x\|)\| \leq \|y - x\|$$

e por isto

$$\|f(y) - f(x)\| \leq (\|g(y)\| + 1)\|y - x\| \leq (M+1)\|y - x\|$$

Qi et al. [31] observam que o quê importa para que uma função seja uma semi-derivada de f no ponto x é a propriedade de convergência (ii) em uma vizinhança de x. Portanto, se mudar o valor de uma semi-derivada em um número finito de pontos, a função obtida também é uma semi-derivada de f no ponto x.

O maior interesse das semi-derivadas é que são funções ponto-a-ponto, e que existem regras de cálculo exato para elas:

Proposição 2.22
Sejam $f_1, f_2 : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ *contínuas,* $\alpha \in \mathbb{R}$*. Então, se* g_1 *e* g_2 *são semi-derivadas (resp. semi-derivadas fortes) em* x *de* f_1 *e* f_2 *respectivamente:*

- $g_1 + \alpha g_2$ *é uma semi-derivada (resp. semi-derivada forte) de* $f_1 + \alpha f_2$ *em* x*;*
- $f_1 g_2 + g_1 f_2$ *é uma semi-derivada (resp. semi-derivada forte) de* $f_1 f_2$ *em* x*;*
- *Se* $f_2(x) \neq 0$*,* $\frac{g_1 f_2 - f_1 g_2}{f_2^2(x)}$ *é uma semi-derivada (resp. semi-derivada forte) de* f_1/f_2 *em* x*;*
- $(g_2 \circ f_1)g_1$ *é uma semi-derivada (resp. semi-derivada forte) de* $f_2 \circ f_1$.

Proposição 2.23
Seja $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ *contínua, de componentes* $f_1, \ldots, f_m$.
Então $g = (g_1, \ldots, g_m)$ *é uma semi-derivada de* f *se e somente se cada para todo* $1 \leq i \leq m$*,* g_i *é uma semi-derivada de* f_i.

Semi-derivadas e semi-suavidade são fortemente ligadas no sentido que, para uma função semi-suave (resp. fortemente semi-suave), qualquer função que associa a x um elemento de $\partial_C f(x)$ é uma semi-derivada (resp. semi-derivada forte) de f. Antes de provar esta propriedade, provamos que tem-se o mesmo resultado se considerar só os elementos de $\partial f(x)$.

Proposição 2.24
Seja $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ localmente Lipschitziana. Então, se f é semi-suave (resp. fortemente semi-suave), qualquer função $g : \mathbb{R}^n \longrightarrow \mathcal{M}_{m,n}(\mathbb{R})$ tal que para todo x tenha-se:

$$g(x) \in \partial f(x)$$

é uma uma semi-derivada (resp. semi-derivada forte) de f.

Demonstração
Suponhamos f semi-suave, o caso fortemente semi-suave é similar.
Seja g tal que:

$$\text{para todo } x \text{ em } \mathbb{R}^n, \quad g(x) \in \partial f(x)$$

Em primeiro lugar, notamos que como ∂f é localmente limitada, segue que g também é localmente limitada.

Seja agora $\varepsilon > 0$, e sejam x, h em $\mathbb{R}^n$.
Queremos provar que, se $\|h\|$ é menor do que um certo $\delta > 0$, então tem-se:

$$\|f(x+h) - f(x) - g(x+h)h\| \leq \varepsilon\|h\|.$$

Para obter isto, mostramos que existe $\delta > 0$ tal que para $\|h\| \leq \delta$ tem-se:

$$\|f(x+h) - f(x) - Mh\| \leq \varepsilon\|h\| \text{ para toda } M \text{ em } \partial f(x+h).$$

Seja então M em $\partial f(x+h)$.
Temos por desigualdade triangular:

$$\|f(x+h) - f(x) - Mh\| \leq \underbrace{\|f(x+h) - f(x) - f'_d(x;h)\|}_{=A(h)} + \underbrace{\|f'_d(x;h) - Mh\|}_{=B(h)}.$$

▶ **controle de $B(h)$:**
Como f é semi-suave em x, pela caracterização 2.18 (b), obtemos diretamente a existência de $\delta > 0$ tal que se $\|h\| \leq \delta$ tenha-se:

$$\text{para toda } M \text{ em } \partial f(x+h), \quad \|f'_d(x;h) - Mh\| \leq \frac{\varepsilon}{2}\|h\|. \tag{2.14}$$

Assim para $\|h\| \leq \delta$ temos:

$$B(h) \leq \frac{\varepsilon}{2}\|h\|.$$

▶ **controle de** $A(h)$:

Pelo teorema do valor intermediário generalizado 2.3, sabemos que existe V_h em $co(\partial f[x, x+h])$ tal que:

$$f(x+h) - f(x) = V_h h = \sum_{i=1}^{K_h} \lambda_i V_i^h h,$$

onde $\lambda_i \geq 0$, $\sum_{i=1}^{K_h} \lambda_i = 1$ e $V_i^h \in \partial f(x + t_i h)$ com $t_i \in [0,1]$.

Logo, se $\|h\| \leq \delta$, a limitação (2.14) é valida para M=V_i^h $(1 \leq i \leq K_h)$, e temos:

$$\|V_i^h t_i h - f'_d(x; t_i h)\| \leq \frac{\varepsilon}{2}\|t_i h\|.$$

Pela homogeneidade positiva de $f'_d(x;.)$, temos então para cada $1 \leq i \leq K_h$:

$$\|V_i^h h - f'_d(x; h)\| \leq \frac{\varepsilon}{2}\|h\|.$$

Portanto:

$$A(h) = \left\| \sum_{i=1}^{K_h} (\lambda_i V_i^h h - f'_d(x;h)) \right\| \leq \sum_{i=1}^{K_h} \lambda_i \|V_i^h h - f'_d(x;h)\| \leq \frac{\varepsilon}{2}\|h\|.$$

■

Disto resulta em particular que, se f for diferenciável em volta de x, como $f'(y) \in \partial f(y)$ para todo y em uma vizinhança de x, então, logo que f for semi-suave em x, a diferencial clássica f' é uma semi-derivada de f em x.

Exemplo 2.10

Seja a função mínimo:

$$\begin{array}{rl} h: & \mathbb{R}^2 \longrightarrow \mathbb{R} \\ & (a,b) \longmapsto \min(a,b) \end{array}$$

A função h é semi-suave, pois é de classe C^1 por pedaços.
Quando $a \neq b$, h é C^1 em (a,b) com:

$$h'(a,b) = \begin{cases} (0,1) & \text{se } a > b \\ (1,0) & \text{se } a < b \end{cases}$$

Quando $a = b$, temos:

$$\partial_B h(a,a) = \{(0,1),(1,0)\}$$

Portanto:

$$\partial h(a,a) = co[\partial_B h(a,a)] = \{(\lambda, 1-\lambda); \lambda \in [0,1]\}$$

Pela semi-suavidade de h, a proposição 2.24 garante que a seguinte função é uma semi-derivada de h em $\mathbb{R}^2$:

$$h_{\frac{1}{2}}(a,b) = \begin{cases} (0,1) & \text{se } a > b \\ (1,0) & \text{se } a < b \\ (\frac{1}{2},\frac{1}{2}) & \text{se } a = b \end{cases}$$

Sejam agora $f, g : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ contínuas, e f_s, g_s semi-derivadas respectivas de f e g. Então, por composição, uma semi-derivada da função $H = \min(f,g)$ é:

$$r(x) = \begin{cases} g_s(x) & \text{se } f(x) > g(x) \\ f_s(x) & \text{se } f(x) < g(x) \\ \frac{1}{2}(f_s + g_s) & \text{se } f(x) = g(x) \end{cases}$$

□

Juntando agora as Proposições 2.24 e 2.23, obtemos facilmente o seguinte resultado:

Proposição 2.25
Seja $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ localmente Lipschitziana. Então se f é semi-suave (resp. fortemente semi-suave), qualquer função $g : \mathbb{R}^n \longrightarrow \mathcal{M}_{m,n}(\mathbb{R})$ tal que para todo x tenha-se:

$$g(x) \in \partial_C f(x)$$

é uma uma semi-derivada (resp. semi-derivada forte) de f.

Dado que o subdiferencial de Clarke $\partial f(x)$ é incluído no C-subdiferencial $\partial_C f(x)$, este resultado é claramente mais forte do que o da Proposição 2.24. Em particular, este resultado é muito importante para a convergência dos métodos de tipo Newton não diferenciáveis. Com efeito, ele valida a idéia de usar os elementos dos Jacobianos generalizados para assumir o papel de derivada no esquema iterativo de Newton.

No próximo capitulo, vamos estudar em detalhe esses métodos e descrever o algoritmo global que usamos para resolver os sistemas KKT.

Capítulo 3

Algoritmos

No capítulo 1 reformulamos o sistema KKT geral (1) como uma equação não linear da forma:

$$H(x) = 0 \tag{3.1}$$

onde $H : \mathbb{R}^N \longrightarrow \mathbb{R}^N$.

No caso de H Fréchet-diferenciável, é conhecido que sob certas condições clássicas, então, o método de Newton converge localmente quadraticamente, assim como os métodos de Newton globalizados (com busca linear, Gauss-Newton) fornecem pontos estacionários do problema (3.1) [24, 34].

No entanto, a nossa intenção é resolver o sistema (3.1) quando H é uma das funções H_*, H_1 e H_α definidas na Seção 1.3.2. Como já observamos, essas possuem pontos de não diferenciabilidade, que impedem portanto o uso do método clássico.

Para tratar esse caso onde H não é Fréchet-diferenciável, várias generalizações do método de Newton foram desenvolvidas, seguindo abordagens diferentes [12, 18, 27, 33]. Os métodos apresentados aqui pertencem à abordagem que consiste em substituir a matriz Jacobiana clássica por um dos Jacobianos generalizados definidos na Seção 2.4. Apoiamos-nos principalmente nos trabalhos de Qi e Sun [33], e de Qi, Ralph e Zhou [31] para dar os resultados de convergência local, enquanto nos servimos do texto de De Luca, Facchinei e Kanzow [6] como referência sobre a globalização do método.

3.1 Método de Newton clássico

Suponhamos H diferenciável. Então o *método de Newton clássico* segue o seguinte esquema iterativo:

$$\begin{cases} x_0 \in \mathbb{R}^N \\ x_{k+1} = x_k - H'(x_k)^{-1} H(x_k). \end{cases} \tag{3.2}$$

A direção:

$$d_k := -H'(x_k)^{-1}H(x_k) \tag{3.3}$$

é conhecida como a *direção de Newton*.

Suponhamos agora que H seja pelo menos de classe C^1. Logo, o seguinte teorema garante a convergência local do esquema iterativo de Newton:

Proposição 3.1 (Convergência local do método de Newton)
Se x_* é uma solução de (3.1) tal que a matriz Jacobiana $H'(x_*)$ seja não singular, então, existe uma vizinhança U de x_* tal que se $x_0 \in U$, a seqüência gerada pelo esquema (3.2) está bem definida e tem as seguintes propriedades:

- *(x_k) converge a x_* superlinearmente;*
- *se existe K constante tal que para todo x em U se cumpre*

$$\|H'(x) - H'(x_*)\| \leq K \|x - x_*\|,$$

então (x_k) converge a x_* quadraticamente.

Em particular, a convergência quadrática ocorre sempre que H é de classe C^2, pois então a derivada segunda de H é localmente limitada, e o teorema dos crescimentos finitos vetorial garante o controle de H' desejado.

Com o fim de obter uma convergência global, o método de Newton puro pode ser modificado pela inclusão de uma busca linear. Em vez da iteração (3.2), usa-se a atualização:

$$x_{k+1} = x_k + t_k d_k, \tag{3.4}$$

onde $t_k \in]0, 1]$ é escolhido tal que uma certa função de mérito θ decresça suficientemente:

$$\theta(x_k + t_k d_k) \leq \theta(x_k) + \beta t_k \theta'(x_k) d_k \qquad (0 < \beta < 1).$$

Um ponto de acumulação de uma seqüência assim gerada é então um ponto crítico da função de mérito. Em geral, usa-se $\theta = \frac{1}{2}\|H\|^2$.

3.2 Métodos de Newton não diferenciáveis

Como para a maioria dos algoritmos, os resultados sobre os métodos de Newton generalizados são de dois tipos.

Os primeiros tratam da convergência local do método, e das condições que garantem que, pelo menos em uma vizinhança de uma solução, a seqüência gerada pelo método seja atraída por essa solução. É o requisito mínimo para um algoritmo ser aceitável.

O segundo tipo de resultado trata da convergência global do método, eventualmente modificado com esse fim. Estas são as maneiras de aumentar o domínio de convergência do método, para que ele convirja mesmo com pontos iniciais bastante longe de qualquer solução.

3.2.1 Método de Newton generalizado local

No caso não diferenciável, a quantidade $H'(x_k)$ não tem mais sentido. Entretanto, conservando como base o esquema do método (3.2), diferentes atualizações da direção de Newton são possíveis:

- Qi e Sun [33] propõem a atualização:

$$x_{k+1} = x_k - V_k^{-1} H(x_k) \qquad \text{com } V_k \in \partial H(x_k). \tag{3.5}$$

- Qi [29] propõe a seguinte modificação de (3.5) a fim de enfraquecer as condições de regularidade do subdiferencial:

$$x_{k+1} = x_k - V_k^{-1} H(x_k) \qquad \text{com } V_k \in \partial_B H(x_k). \tag{3.6}$$

- Chen, Chen e Kanzow [3] usam como base a atualização:

$$x_{k+1} = x_k - V_k^{-1} H(x_k) \qquad \text{com } V_k \in \partial_C H(x_k). \tag{3.7}$$

Usando a noção de semi-derivada introduzida por Qi [31] e definida na Seção 2.6, achamos interessante unificar os pontos de vista, dando um teorema de convergência local geral. Dessa forma, tratamos as três atualizações acima de uma vez só, e enfatizamos as propriedades dos Jacobianos generalizados que têm um papel chave na convergência.

Com esse fim, observamos que (3.5), (3.6) e (3.7) seguem o mesmo esquema comum:

$$x_{k+1} = x_k - g(x_k)^{-1} H(x_k), \qquad \text{com } g(x_k) \in \mathcal{A}(x_k), \tag{3.8}$$

onde $\mathcal{A}$ é uma multifunção de $\mathbb{R}^N$ em $\mathcal{M}_N(\mathbb{R})$:

$$\mathcal{A} : \mathbb{R}^N \rightrightarrows \mathcal{M}_N(\mathbb{R}),$$

e g é uma função escolha da família $(\mathcal{A}(x))_{x\in\mathbb{R}^N}$. Isto quer dizer que g é tal que para todo x em $\mathbb{R}^N$ tenha-se:

$$g(x) \in \mathcal{A}(x)$$

Com essas notações, obtemos o seguinte algoritmo padrão:

Algoritmo 1 (Método de Newton generalizado *(forma padrão)*)
***Passo** 1 **(Preliminar):** Definir $g : \mathbb{R}^N \longrightarrow \mathcal{M}_N(\mathbb{R})$.*

***Passo** 2 **(Inicialização):** Escolher o ponto inicial $x_0 \in \mathbb{R}^N$. Seja $k = 0$.*

Passo** 3 **(Direção de busca):

Resolver na variável d o sistema linear:

$$g(x_k)d + H(x_k) = 0$$

Seja d_k uma solução. Se $d_k = 0$ então terminar, pois achamos uma solução:

$$x^* = x_k.$$

Senão continuar passando ao Passo 4.

***Passo** 4 **(Atualização):** $x_{k+1} = x_k + d_k$, $k \leftarrow k + 1$ e voltar ao Passo 3.*

Temos então o seguinte teorema geral de convergência local para esse algoritmo:

Teorema 3.1 (Convergência local geral)
Sejam $H : \mathbb{R}^N \longrightarrow \mathbb{R}^N$ contínua e x^* uma solução da equação (3.1).
Se g no Passo 1 verifica as condições:

(i) existem uma constante $c > 0$ e uma vizinhança V de x^* tal que para todo x em V:

$$g(x) \text{ é inversível} \quad e \quad \|g(x)^{-1}\| \leq c,$$

(ii) g é uma semi-derivada (resp. semi-derivada forte) de H em x^*,

então, existe uma vizinhança U de x^* tal que para toda inicialização em U, a atualização (3.8) está bem definida em cada passo, e a seqüência gerada por esse esquema converge a x^* superlinearmente (resp. quadraticamente).

Demonstração
Seja $x_0 \in V$.
Por definição de V, sabemos que $g(x_0)$ é inversível e x_1 está bem definido.

Logo temos:

$$\begin{aligned}\|x_1 - x^*\| = \|x_0 - g(x_0)^{-1}H(x_0) - x^*\| &= \|g(x_0)^{-1}[g(x_0)(x_0 - x_*) - H(x_0)]\| \\ &= \|g(x_0)^{-1}[H(x^*) + g(x_0)(x_0 - x_*) - H(x_0)]\| \\ &\leq \|g(x_0)^{-1}\|\|H(x^*) + g(x_0)(x_0 - x_*) - H(x_0)\|\end{aligned}$$

Como g é uma semi-derivada de H em x, temos pela própria definição:

$$\|H(x^*) + g(x_0)(x_0 - x_*) - H(x_0)\| = \circ(\|x_0 - x^*\|).$$

Por fim, com a limitação uniforme de $g(x)^{-1}$ sobre V, obtemos:

$$\|x_1 - x^*\| = \circ(\|x_0 - x^*\|). \tag{3.9}$$

Por isto, existe uma vizinhança $U \subset V$ tal que se x_0 é escolhido em U, então x_1 está bem definido, também pertence a U, e é tal que temos o controle (3.9).

Logo, por recorrência, a seqüência gerada por (3.8) está bem definida e verifica, para todo $k \geq 0$:

$$\|x_{k+1} - x^*\| = \circ(\|x_k - x^*\|).$$

Portanto $x_k \longrightarrow x^*$ superlinearmente.
Do mesmo modo, se g for uma semi-derivada forte de H em x, então teremos, para $k \geq 0$:

$$\|x_{k+1} - x^*\| = O(\|x_k - x^*\|^2)$$

e a convergência seria quadrática. ■

Suponhamos agora que g é definida segundo o esquema (3.8). Por um lado, sabemos pelo Teorema 2.1 que, se a função ponto-conjunto $\mathcal{A}$ é semi-contínua superior, regular em x^* e com $\mathcal{A}(x^*)$ compacto, então qualquer função escolha g cumpre a condição (i).

Por outro lado, provamos na seção 2.4 que $\partial_B H$, ∂H e $\partial_C H$ são semi-contínuos superiores, com imagem compacta e todos incluídos em $\partial_C H$ (inclusão 2.11).

Como, além disso, a Proposição 2.25 garante que se H é semi-suave (resp. fortemente semi-suave) então qualquer função que a x associa um elemento de $\partial_C H(x)$ é uma semi-derivada (resp. semi-derivada forte) de H, obtemos o seguinte teorema, corolário do teorema geral anterior:

Teorema 3.2 (Convergência local)
Sejam $H : \mathbb{R}^N \longrightarrow \mathbb{R}^N$ localmente Lipschitziana e semi-suave (resp. fortemente semi-suave), x^* uma solução da equação (3.1), e g definida tal que para todo x em $\mathbb{R}^N$:

$$g(x) \in \mathcal{A}(x),$$

onde $\mathcal{A} : \mathbb{R}^N \rightrightarrows \mathcal{M}_N(\mathbb{R})$.
Então:

(a) se $\partial_B H$ é regular em x^*, o Algoritmo 1 com $\mathcal{A} = \partial_B H$ converge localmente a x^*;

(b) se ∂H é regular em x^*, o Algoritmo 1 com a escolha $\mathcal{A} = \partial H$ converge localmente a x^*;

(c) se $\partial_C H$ é regular em x^*, o Algoritmo 1 com a escolha $\mathcal{A} = \partial_C H$ converge localmente a x^*;

Ainda mais, essa convergência é superlinear (resp. quadrática).

Nota: Qi [29] chama a condição de regularidade em (a) de *BD-regularidade forte* de H em x^*, enquanto Qi e Pang [28] a chamam simplesmente de *BD-regularidade* de H em x^*. Quanto à condição em (b), Qi [30] usa o termo de *CD-regularidade* de H em x^*.

Pela inclusão (2.11), pedir que todos os elementos do B-subdiferencial sejam não singulares na solução é menos restritivo do que pedi-lo para os outros dois subdiferenciais pois temos as implicações:

$$\partial_C H \text{ regular em } x^* \Rightarrow \partial H \text{ regular em } x^* \Rightarrow \partial_B H \text{ regular em } x^*$$

Disso resulta que se $\partial_C H$ é regular em x^*, então qualquer uma das três escolhas leva a um algoritmo localmente convergente a x^*. Se ∂H é regular em x^*, então tanto a escolha $\mathcal{A} = \partial H$ quanto a escolha $\mathcal{A} = \partial_B H$ levam à convergência local do algoritmo.

Do ponto de vista teórico, um método de Newton que usa a atualização (3.6) (i.e. que faz a escolha $\mathcal{A} = \partial_B H$) é então preferível, pois a sua condição de convergência é a mais fraca das três, e é por isto a mais fácil a conseguir. Qi [29] qualifica então o método baseado no B-subdiferencial de mais resolúvel.

Exemplo 3.1

Seja a função valor absoluto em $\mathbb{R}$:

$$H(x) = |x|.$$

Essa função é semi-suave pois é C^1 por pedaços.

Já vimos no exemplo 2.4 que $\partial H(0) = [-1, 1]$.
Como estamos em dimensão um, também temos trivialmente: $\partial_C H = \partial H(0) = [-1, 1]$.

Assim, neste caso simples, um método de Newton que usa a atualização (3.5) ou (3.7) é inaplicável para resolver a equação $H(x) = 0$, pois temos na solução $x = 0$:

$$0 \in \partial H(0) \subset \partial_C H(0),$$

ou seja, nem $\partial H(0)$ nem $\partial_B H(0)$ é regular.
Mas, por outro lado, temos:

$$0 \notin \partial_B H(0) = \{1, -1\}$$

Portanto, se usarmos a atualização (3.6), vale o Teorema 3.2 (a) que garante a convergência local do método. □

O teorema 3.2 só garante a convergência *local* do Algoritmo 1 a uma solução do problema (3.1). Ou seja, ao iniciar a seqüência próximo o suficiente de uma solução (regular), então aquela converge a essa. Entretanto, se o ponto inicial estiver longe demais de uma solução, não temos garantia de convergência.

Porém, na prática, obviamente, não se sabe a priori quais são as soluções, nem onde deve-se procurá-las. Apresentamos então uma globalização do método, que tem como objetivo aumentar o domínio de convergência do algoritmo.

3.2.2 Método de Newton generalizado global

De uma maneira análoga ao caso diferenciável, as estrégias de globalização dos métodos de Newton generalizados utilizam uma função de mérito θ que fornece uma medida da distância à solução do problema.

A idéia central da globalização é então a seguinte. Longe da solução, nada garante a regularidade de $\partial_B H$. Logo, caso a equação de Newton não for resolúvel, uma direção de busca promissora é uma direção de descida da função de mérito. A direção geralmente usada é então a de descida máxima, ou seja $-\nabla\theta(x_k)$. Junto com isso, uma busca ao longo da direção escolhida é efetuada para garantir um decréscimo suficiente de θ a cada passo.

Por fim, a seqüência gerada pelo algoritmo assim globalizado fornece pontos estacionários da função de mérito. Sob certas condições, esses pontos estacionários também são solução de (3.1).

Tratando-se de equações do tipo (3.1), uma função de mérito natural é:

$$\theta(x) := \frac{1}{2}\|H(x)\|^2.$$

Com efeito, θ é uma medida do erro de estimativa $\|H(x) - H(x^*)\|$, pois é contínua e os seus zeros são exatamente as soluções de (3.1). Portanto, quanto mais próximo x estiver de uma solução, menor fica o valor de $\theta(x)$.

No caso geral, se H for somente suposta semi-suave, θ pode não ser continuamente diferenciável, e não servir para a técnica esboçada acima. Neste caso, métodos de suavização do problema (smoothing methods) podem ser usados [4, 32].

No entanto, em muitas reformulações semi-suave de problemas, a função H, mesmo possuindo pontos de não diferenciabilidade, pode ter a propriedade de que θ é efetivamente de classe C^1 [22]. Neste caso, a teoria que segue pode ser aplicada. Em particular, será provado no capítulo 4 que as funções H_1 e H_α em que estamos interessados têm essa propriedade.

Assim, nesta seção, θ é suposta de classe C^1 em $\mathbb{R}^N$.

O algoritmo global apresentado aqui é devido a De Luca, Facchinei e Kanzow [6]. Em uma abordagem similar à do nosso trabalho, eles desenvolveram o método para resolver equações

do tipo (3.1) que resultam da reformulação semi-suave do problema de complementaridade não linear, usando a função de Fischer-Burmeister.

Apesar da especificidade da função considerada por De Luca et al., o algoritmo proposto em [6] é perfeitamente geral e pode ser aplicado a qualquer função que gera uma função θ continuamente diferenciável.

Como observam De Luca, Facchinei e Kanzow, notemos que quando ela é bem definida, a direção de Newton dada por (3.14) sempre é uma direção de descida para θ.

Com efeito, θ é o resultado da composição entre a função $\gamma : x \longmapsto \frac{1}{2}\|x\|^2$ de classe C^1 e a função localmente Lipschitziana H. A proposição 2.16 dá então que:

$$\partial\theta(x) = H(x)^T \partial H(x)$$

Portanto, já que θ é suposta C^1, o seu subdiferencial é reduzido a $\{\theta'(x)\}$ e temos:

$$\theta'(x) = H(x)^T V \quad \text{com } V \text{ qualquer em } \partial H(x). \tag{3.10}$$

Assim, se d_k é uma direção de Newton do passo k, verifica-se:

$$\theta'(x_k)d_k = H(x_k)^T V_k d_k = -H(x_k)^T H(x_k) = -\|H(x_k)\|^2 < 0.$$

Logo, uma tal direção sempre é uma potencial direção de busca do próximo ponto. Ao mesmo tempo, tratando-se de problemas de complementaridade não linear, é freqüente observar que uma mera busca linear gere passos muito pequenos, resultando em uma convergência lenta da seqüência, ou até na falha do algoritmo.

Assim, De Luca, Facchinei e Kanzow [6] propõem substituir, em cada passo, a busca linear padrão:

$$\theta(x_k + t_k d_k) \leq \theta(x_k) + \beta t_k \theta'(x_k) d_k \tag{3.11}$$

por uma busca linear dita *não monótona* da forma geral

$$\theta(x_k + t_k d_k) \leq \mathcal{W}_k + \beta t_k \theta'(x_k) d_k, \tag{3.12}$$

onde $\mathcal{W}_k$ é tal que:

$$\theta(x_k) \leq \mathcal{W}_k \leq \max_{0 \leq j \leq M_k} \theta(x_{k-j}), \tag{3.13}$$

com $0 < M_k \leq \min(M_k + 1, M)$, onde $M > 0$ é uma certa constante predefinida.

A vantagem deste tipo de busca é que ela não impõe mais o decrescimento monótono da função objetivo a cada iteração, mas sim a cada M iterações. Assim, pode acontecer que θ aumente de x_k para x_{k+1}. Essa liberdade maior na escolha do passo se revela eficiente quanto a diminuir o número de iteração, assim como o número de avaliação de função [16].

Em particular, escolhendo $\mathcal{W}_k = \theta(x_k)$ em cada passo, a busca se reduz à busca linear padrão (3.11). Observamos também que, caso $(\theta(x_k))$ convergir, então, por controle inferior-superior e limitação de (M_k), a desigualdade (3.13) dá que a seqüência $(\mathcal{W}_k)$ também converge ao mesmo limite.

Com esse tipo de busca linear, o algoritmo global geral é o seguinte:

Algoritmo 2 (Método de Newton generalizado globalizado)

***Passo** 1 **(Preliminar):** Definir $g : \mathbb{R}^N \longrightarrow \mathcal{M}_N(\mathbb{R})$.*

***Passo** 2 **(Inicialização):** Escolher os parâmetros: $\rho > 0, p > 2, \beta \in]0, \frac{1}{2}[, M > 0$, e um ponto inicial x_0 em $\mathbb{R}^N$.*

Seja $k = 0$.

***Passo** 3 **(Critério de parada):** Se $\|\nabla\theta(x_k)\| = 0$ ou $H(x_k) = 0$, parar.*

Passo** 4 **(Direção de busca):

Achar d^* uma solução do sistema linear:

$$H(x_k) + g(x_k)d = 0. \tag{3.14}$$

Tomar $d_k = d^*$.

Se (3.14) não for resolúvel ou se d_k não verificar a condição:

$$\nabla\theta(x_k)^T d_k \leq -\rho\|d_k\|^p, \tag{3.15}$$

então tomar $d_k = -\nabla\theta(x_k)$.

Passo** 5 **(Busca linear):

Definir $\mathcal{W}_k$ tal que:

$$\theta(x_k) \leq \mathcal{W}_k \leq \max_{0 \leq j \leq M_k} \theta(x_{k-j}), \tag{3.16}$$

com $0 < M_k \leq \min(M_k + 1, M)$.

Seja m_k o menor inteiro $m \geq 0$ tal que:

$$\theta(x_k + 2^{-m}d_k) \leq \mathcal{W}_k + \beta 2^{-m}\nabla\theta(x_k)^T d_k. \tag{3.17}$$

Seja $t_k = 2^{-m_k}$.

Passo** 6 **(Atualização):

$x_{k+1} = x_k + t_k d_k$ e $k \leftarrow k + 1$. Voltar ao Passo 4.

De Luca, Facchinei e Kanzow [6] provaram um teorema de convergência global para o algoritmo acima, no caso de uma busca linear monótona. Em uma publicação anterior, Grippo, Lampariello e Lucidi [16] trataram do caso de um algoritmo de tipo Newton clássico onde a busca linear não monótona é definida por:

$$\mathcal{W}_k = \max_{0 \le j \le M_k} \theta(x_{k-j}).$$

Se inspirando desses dois artigos, demonstramos o seguinte teorema:

Teorema 3.3 (Convergência global)
Seja H semi-suave (resp. fortemente semi-suave) tal que θ seja de classe C^1 e tal que o conjunto:

$$\Gamma_0 := \left\{ x \in \mathbb{R}^N \mid \theta(x) \le \theta(x_0) \right\}$$

seja compacto.
Se g é uma semi-derivada (resp. semi-derivada forte) de H,
então o Algoritmo 2 está bem definido e gera uma seqüência (x_k) da qual qualquer ponto de acumulação x^* verifica:

(a) x^* é um ponto estacionário de θ, ou seja: $\theta'(x^*) = 0$.

(b) Se existe um elemento não singular em $\partial_B H(x^*)$, então x^* é solução de (3.1)

(c) Se existem um elemento não singular em $\partial_B H(x^*)$, uma vizinhança V de x^* e uma constante $c > 0$ tais que, para todo x em V:

$$g(x) \text{ é inversível} \quad e \quad \|V^{-1}\| \le c,$$

então x^* é solução de (3.1), (x_k) converge a x^*, e a convergência é superlinear (resp. quadrática).

Demonstração

- Item (a):
  Para $k \ge 0$, denotamos $\alpha_k = 2^{-m_k}$.
  Também, para $k \ge 0$, denotemos por $l(k)$ o maior índice entre $k - M_k$ e k tal que:

$$\theta(x_{l(k))}) = \max_{0 \le j \le M_k} \theta(x_{k-j}).$$

  Por construção temos então as seguintes desigualdades:

$$k - M \le k - M_k \le l(k) \le k,$$

e em particular:

$$\lim_{k\to+\infty} l(k) = +\infty.$$

Observamos que, independentemente a escolha feita para d_k, temos para todo $k \geq 0$:

$$\nabla\theta^T(x_k)d_k < 0.$$

Daí, com as novas notações, a busca linear (3.17) junto com (3.16) dá, para todo $k \geq 0$, o seguinte controle:

$$\theta(x_{k+1}) \leq \theta(x_{l(k)}) + \beta\alpha_k\nabla\theta^T(x_k)d_k \leq \theta(x_{l(k)}). \tag{3.18}$$

Logo, observamos que a seqüência $(\theta(x_{l(k)}))$ é decrescente. Com efeito, tem-se:

$$\begin{aligned}\theta(x_{l(k+1)}) &= \max_{0\leq j\leq M_{k+1}} \theta(x_{k+1-j}) \\ &= \max\left\{\max_{0\leq j\leq M_{k+1}-1} \theta(x_{k-j}), \theta(x_{k+1})\right\}.\end{aligned}$$

Porém, por definição de M_k, temos $M_{k+1} - 1 \leq M_k$, por isto:

$$\theta(x_{l(k+1)}) \leq \max\left\{\max_{0\leq j\leq M_k} \theta(x_{k-j}), \theta(x_{k+1})\right\} = \max\left\{\theta(x_{l(k)}), \theta(x_{k+1})\right\},$$

Por fim, com (3.18) obtemos $\theta(x_{k+1}) \leq \theta(x_{l(k)})$, e o decrescimento desejado:

$$\theta(x_{l(k+1)}) \leq \theta(x_{l(k)}).$$

Deste decrescimento deduzimos então que para $k \geq 0$:

$$\theta(x_k) \leq \theta(x_{l(k)}) \leq \theta(x_{l(0)}) = \theta(x_0).$$

Por isto, para todo $k \geq 0$, x_k pertence ao compacto Γ_0. Como, θ é contínua, o conjunto $\theta(\Gamma_0)$ também é compacto, e a seqüência decrescente $(\theta(x_{l(k)})) \subset \theta(\Gamma_0)$ admite um limite quando $k \to +\infty$:

$$\boxed{\exists \lim_{k\to+\infty} \theta(x_{l(k)})} \tag{3.19}$$

Vamos mostrar que de fato, essa convergência vale para a seqüência toda, ou seja:

$$\exists \lim_{k\to+\infty} \theta(x_k))$$

Seja:

$$\hat{l}(k) = l(k + M + 1).$$

Vamos provar primeiro por recorrência que para todo $j \geq 1$ tem-se:

$$\boxed{\text{(a)} \quad \exists \lim_{k\to+\infty} \alpha_{\hat{l}(k)-j}\|d_{\hat{l}(k)-j}\| = 0} \quad \text{e} \quad \boxed{\text{(b)} \quad \exists \lim_{k\to+\infty} \theta(x_{\hat{l}(k)-j}) = \lim_{k\to+\infty} \theta(x_{\hat{l}(k)})} \tag{3.20}$$

► $j = 1$:
De (3.18), obtemos que para todo $k \geq 1$:

$$\theta(x_{\hat{l}(k)}) \leq \theta(x_{l(\hat{l}(k)-1)}) + \alpha_{\hat{l}(k)-1}\nabla\theta(x_{\hat{l}(k)-1})^T d_{\hat{l}(k)-1} \leq \theta(x_{\hat{l}(\hat{l}(k)-1)}),$$

e, pela convergência de $(\theta(x_{\hat{l}(k)}))$ como subseqüência de $(\theta(x_{l(k)}))$ obtemos:

$$\exists \lim_{k\to+\infty} \alpha_{\hat{l}(k)-1}\nabla\theta(x_{\hat{l}(k)-1})^T d_{\hat{l}(k)-1} = 0. \tag{3.21}$$

Porém, para $k \geq 0$, ora d_k é dada pela equação de Newton generalizada (3.14) e verifica a desigualdade (3.15):

$$\nabla\theta(x_k)^T d_k \leq -\rho\|d_k\|^p,$$

ora $d_k = -\nabla\theta(x_k)$ e verifica obviamente a desigualdade:

$$\nabla\theta(x_k)^T d_k \leq -\|d_k\|^2.$$

Assim, para todo $k \geq 0$, d_k verifica:

$$\nabla\theta(x_k)^T d_k \leq -\rho_k\|d_k\|^{p_k}, \tag{3.22}$$

onde $\rho_k \in \{\rho, 1\}$ e $p_k \in \{p, 2\}$.

Daí, temos:

$$\alpha_{\hat{l}(k)-1}\nabla\theta(x_{\hat{l}(k)-1})^T d_{\hat{l}(k)-1} \leq -\underbrace{\alpha_{\hat{l}(k)-1}\rho_{\hat{l}(k)-1}\|d_{\hat{l}(k)-1}\|^{p_{\hat{l}(k)-1}}}_{>0} \leq 0,$$

e por (3.21), obtemos, passando no limite:

$$\lim_{k\to+\infty} \alpha_{\hat{l}(k)-1}\rho_{\hat{l}(k)-1}\|d_{\hat{l}(k)-1}\|^{p_{\hat{l}(k)-1}} = 0.$$

Como $\rho_{\hat{l}(k)-1}$ pega os seus valores em $\{\rho, 1\}$, então necessariamente tem que ter:

$$\lim_{k\to+\infty} \alpha_{\hat{l}(k)-1}\|d_{\hat{l}(k)-1}\|^{p_{\hat{l}(k)-1}} = 0.$$

Agora, como (α_k) é limitada ($(\alpha_k) \subset]0, 1]$) e como $p_k - 1 \geq 0$ para todo $k \geq 0$, então $(\alpha_{\hat{l}(k)-1}^{p_{\hat{l}(k)-1}-1})$ também é limitada e:

$$\lim_{k\to+\infty} \alpha_{\hat{l}(k)-1}^{p_{\hat{l}(k)-1}}\|d_{\hat{l}(k)-1}\|^{p_{\hat{l}(k)-1}} = 0.$$

Logo, dado que $p_k \geq 0$, temos:

$$\lim_{k\to+\infty} \alpha_{\hat{l}(k)-1}\|d_{\hat{l}(k)-1}\| = 0, \tag{3.23}$$

o que é (3.20)(a) para $j = 1$.
Por fim, (3.23) implica que:

$$\|x_{\hat{l}(k)} - x_{\hat{l}(k)-1}\| \longrightarrow 0$$

Portanto, por continuidade uniforme de θ sobre o compacto Γ_0, pela existência do limite (3.19) temos:

$$\exists \lim \theta(x_{\hat{l}(k)-1}) = \lim \theta(x_{\hat{l}(k)}) = \lim \theta(x_{l(k)})$$

- $j \rightsquigarrow j+1$: Suponhamos que (3.20) valha para um certo j. Para k suficientemente grande temos $\hat{l}(k) \geq j+1$ e (3.18) dá:

$$\theta(x_{\hat{l}(k)-j}) \leq \theta(x_{l(\hat{l}(k)-j-1)}) + \alpha_{\hat{l}(k)-j-1}\nabla\theta(x_{\hat{l}(k)-1})^T d_{\hat{l}(k)-j-1} \leq \theta(x_{l(\hat{l}(k)-j-1)}).$$

Pela hipótese de recorrência, temos:

$$\lim \theta(x_{\hat{l}(k)-j}) = \lim \theta(x_{l(k)}),$$

de tal maneira que por controle inferior-superior tem-se:

$$\lim_{k\to+\infty} \alpha_{\hat{l}(k)-(j+1)}\rho_{\hat{l}(k)-(j+1)}\|d_{\hat{l}(k)-(j+1)}\|^{p_{\hat{l}(k)-(j+1)}} = 0.$$

Pelo mesmo raciocínio que usamos no caso $j = 1$, a desigualdade (3.22) e permite de deduzir:

$$\lim_{k\to+\infty} \alpha_{\hat{l}(k)-(j+1)}\|d_{\hat{l}(k)-(j+1)}\| = 0.$$

Também, da mesma maneira, temos que:

$$\|x_{\hat{l}(k)-j} - x_{\hat{l}(k)-(j+1)}\| \longrightarrow 0$$

o que, pela continuidade uniforme de θ sobre Γ_0 e pela hipótese de recorrência, implica:

$$\exists \lim \theta(x_{\hat{l}(k)-(j+1)}) = \lim \theta(x_{\hat{l}(k)-j}) = \lim \theta(x_{l(k)})$$

Seja então $k \geq 0$. Temos $\hat{l}(k) = l(k+M+1) \geq k+1$ e podemos escrever:

$$x_{k+1} = x_{\hat{l}(k)} - \sum_{j=1}^{\hat{l}(k)-k-1} \alpha_{\hat{l}(k)-j} d_{\hat{l}(k)-j}.$$

Como $\hat{l}(k) - k - 1 = l(k+M+1) - k - 1 \leq M$, temos por (3.20)(a):

$$\|x_{k+1} - x_{\hat{l}(k)}\| \longrightarrow 0$$

De novo, pela continuidade uniforme de θ sobre Γ_0 temos:

$$\boxed{\exists \lim_{k\to+\infty} \theta(x_k)}$$

Daí, passando no limite na desigualdade (3.18), temos:

$$\lim_{k\to+\infty} \alpha_k \nabla\theta(x_k)^T d_k = 0.$$

Mais uma vez, pelo mesmo raciocínio usado na recorrência, deduzimos disso que:

$$\boxed{\lim_{k\to+\infty} \alpha_k \|d_k\| = 0} \tag{3.24}$$

Seja então x^* um ponto de acumulação de (x_k). Observamos que isso supõe que o algoritmo não pára, e portanto, que para todo $k \geq 0$, o $\nabla\theta(x_k)$ e $H(x_k)$ são não nulos. Isto implica em particular que d_k seja não nulo para $k \geq 0$.
Vamos provar que:

$$\nabla\theta(x^*) = 0.$$

Dois casos são possíveis:

- **Caso 1:** 0 é valor de aderência de (d_k).
  Então existe uma extração φ de $\mathbb{N}$ tal que

  $$\exists \lim_{k\to+\infty} d_{\varphi(k)} = 0.$$

  Sejam os conjuntos:

  $$K_1 = \left\{k \in \mathbb{N} \mid d_{\varphi(k)} = -\nabla\theta(x_{\varphi(k)})\right\} \quad \text{e} \quad K_2 = \left\{k \in \mathbb{N} \mid g(x_{\varphi(k)})d_{\varphi(k)} = -H(x_{\varphi(k)})\right\}$$

  Como K_1 e K_2 constituem uma partição de $\mathbb{N}$, pelo menos um deles é infinito:
  - Se K_1 é infinito, temos trivialmente:

    $$\nabla\theta(x^*) = -\lim_{k\to+\infty, k\in K_1} d_{\varphi(k)} = 0.$$

  - Se K_2 infinito, então, dispomos de uma subseqüência $(d_{\varphi\circ\psi(k)})$ de $(d_{\varphi(k)})$, convergente a 0, tal que para todo $k \geq 0$ tenha se:

    $$H(x_{\varphi\circ\psi(k)}) = -g(x_{\varphi\circ\psi(k)})d_{\varphi\circ\psi(k)}.$$

    Como semi-derivada, g é localmente limitada, e por tanto limitada sobre os compactos de $\mathbb{R}^N$. Em particular, g é limitada sobre o compacto $\left\{d_{\varphi\circ\psi(k)},\ k \geq 0\right\}$ ($(d_{\varphi\circ\psi(k)})$ é convergente). Portanto:

    $$\exists \lim_{k\to+\infty} H(x_{\varphi\circ\psi(k)}) = 0.$$

Como $\lim_{k\to+\infty} x_{\varphi\circ\psi(k)} = x^*$ e pela continuidade de H temos que ter:

$$H(x^*) = 0.$$

Logo, x^* é minimizador global de θ e por isto $\nabla\theta(x^*) = 0$.

- **Caso 2:** Existe $\delta > 0$ tal que $\|d_k\| > \delta$ para valores de k suficientemente grandes.
Observamos que pela desigualdade (3.22) e pela desigualdade de Cauchy-Schwarz temos para todo $k \geq 0$:

$$\rho_{\varphi(k)}\|d_{\varphi(k)}\|^{p_{\varphi(k)}} \leq \left|\nabla\theta(x_{\varphi(k)})^T d_{\varphi(k)}\right| \leq \left\|\nabla\theta(x_{\varphi(k)})\right\| \left\|d_{\varphi(k)}\right\|$$

Portanto, como $\rho_{\varphi(k)}$ e d_k são não nulos, e como $p_{\varphi(k)} - 1 \geq 0$ temos:

$$\|d_{\varphi(k)}\| \leq \left(\frac{\|\nabla\theta(x_{\varphi(k)})\|}{\rho_{\varphi(k)}}\right)^{\frac{1}{p_{\varphi(k)}-1}}.$$

Pela continuidade de $\nabla\theta$ sobre o compacto $\left\{x_{\varphi(k)}, k \geq 0\right\}$, deduzimos que $d_{\varphi(k)}$ é limitada por um certo $\Delta > 0$. Afinal, temos então:

$$0 < \delta \leq \|d_{\varphi(k)}\| \leq \Delta.$$

Resulta dessa limitação que existe uma extração ψ tal que

$$\exists \lim_{k\to+\infty} d_{\varphi\circ\psi(k)} = d^*,$$

com $\delta \leq \|d^*\|$.
Logo, temos que ter:

$$\alpha_{\varphi\circ\psi(k)} \longrightarrow 0.$$

Porém, $\alpha_k = 2^{-m_k}$, portanto, para valores de k suficientemente grandes, $m_{\varphi\circ\psi(k)}$ é maior que 1 e, pela própria definição de m_k, a desigualdade (3.17) não se verifica para $m = m_{\varphi\circ\psi(k)} - 1$, ou seja:

$$\beta 2\alpha_{\varphi\circ\psi(k)}\nabla\theta(x_{\varphi\circ\psi(k)})^T d_{\varphi\circ\psi(k)} < \theta(x_{\varphi\circ\psi(k)} + 2\alpha_{\varphi\circ\psi(k)}d_{\varphi\circ\psi(k)}) - \mathcal{W}_{\varphi\circ\psi(k)}.$$

Por definição de $\mathcal{W}_k$, temos $\theta(x_{\varphi\circ\psi(k)}) \leq \mathcal{W}_{\varphi\circ\psi(k)}$. Logo:

$$\beta 2\alpha_{\varphi\circ\psi(k)}\nabla\theta(x_{\varphi\circ\psi(k)})^T d_{\varphi\circ\psi(k)} < \theta(x_{\varphi\circ\psi(k)} + 2\alpha_{\varphi\circ\psi(k)}d_{\varphi\circ\psi(k)}) - \theta(x_{\varphi\circ\psi(k)}),$$

e vem:

$$\beta\nabla\theta(x_{\varphi\circ\psi(k)})^T d_{\varphi\circ\psi(k)} < \frac{\theta(x_{\varphi\circ\psi(k)} + 2\alpha_{\varphi\circ\psi(k)}d_{\varphi\circ\psi(k)}) - \theta(x_{\varphi\circ\psi(k)})}{2\alpha_{\varphi\circ\psi(k)}}.$$

Temos $2\alpha_{\varphi\circ\psi(k)} \longrightarrow 0$ e $d_{\varphi\circ\psi(k)} \longrightarrow d^*$. Também, como θ é C^1, ela é estritamente diferenciável. Então, passando corretamente no limite na desigualdade acima obtemos:

$$\beta\nabla\theta(x^*)^T d^* \leq \nabla\theta(x^*)^T d^*.$$

Já que β pertence a $]0, \frac{1}{2}[$, isto quer dizer que:

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

Porém, por outro lado, por (3.22), vale para todo $k \geq 0$:

$$\nabla\theta(x_{\varphi\circ\psi(k)})^T d_{\varphi\circ\psi(k)} \leq -\rho_{\varphi\circ\psi(k)}\|d_{\varphi\circ\psi(k)}\|^{p_{\varphi\circ\psi(k)}} \leq -\rho_{\varphi\circ\psi(k)}\delta^{p_{\varphi\circ\psi(k)}} < \varepsilon,$$

onde $\varepsilon < 0$.
Portanto, no limite tem-se:

$$\nabla\theta(x^*)^T d^* < 0,$$

o que é contraditório.

- Item (b):
  Seja x^* um ponto de acumulação de (x_k).
  Sabemos por (3.10) e pelo Item (a) que:

  $$\theta'(x^*) = H(x^*)^T V = 0$$

  onde V é qualquer em $\partial H(x^*)$.
  Portanto, se existe um elemento não singular em $\partial H(x^*)$ temos que ter $H(x^*) = 0$.
  Ou seja x^* é solução do problema (3.1).

- Item (c):
  Se $\partial_B H$ é regular em x^*, é claro que vale o Item (b), e x^* é solução. Logo, o Teorema 3.2 de convergência local se aplica, e para k suficientemente grande, (x_k) entra na vizinhança de convergência a x^*.

■

Capítulo 4

Aplicação aos sistemas KKT

No capítulo 1 definimos as funções ϕ_*, ϕ_1 e ϕ_α, ao mesmo tempo que as funções H_*, H_1 e H_α associadas que são da seguinte forma padrão:

$$H(z) = H(x, \lambda, \mu) = \begin{pmatrix} F(x) + h'(x)^T\lambda - g'(x)^T\mu \\ h(x) \\ [\phi(g_i(x), \mu_i)]_{1\leq i\leq q} \end{pmatrix}.$$

Lembramos que $z = (x, \lambda, \mu)$ pertence a $\mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^q$ e que as funções F, h, e g são tais que:

$$F : \mathbb{R}^n \longrightarrow \mathbb{R}^n \qquad h : \mathbb{R}^n \longrightarrow \mathbb{R}^p \qquad g : \mathbb{R}^n \longrightarrow \mathbb{R}^q \qquad .$$

Para manipular notações mais compactas, denotamos:

$$\mathcal{F}(z) = \mathcal{F}(x, \lambda, \mu) := F(x) + h'(x)^T\lambda - g'(x)^T\mu$$

assim como:

$$\Phi(x, \mu) := \begin{pmatrix} \phi(g_1(x), \mu_1) \\ \vdots \\ \phi(g_q(x), \mu_q) \end{pmatrix}.$$

O sistema KKT a resolver se reescreve então:

$$H(z) = \begin{pmatrix} \mathcal{F}(z) \\ h(x) \\ \Phi(x, \mu) \end{pmatrix} = 0. \tag{4.1}$$

A nossa intenção é aplicar ao sistema (4.1) o método de Newton generalizado descrito na Subseção 3.2.2. Para isto, temos que provar que cada H verifica as hipóteses necessárias à

aplicação desse método. O requisito mínimo é que as condições de convergência local sejam satisfeitas.

Logo, neste capítulo, vamos interessar-nos em condições suficientes para conseguir, para cada uma das funções NCP estudadas, a semi-suavidade de H, a regularidade de ∂H em uma solução de (4.1) e o caráter C^1 da função de mérito natural $\theta = \frac{1}{2}\|H\|^2$.

4.1 Considerações gerais

Por construção, o sistema (4.1) não é qualquer, e possui uma estrutura especial, que é independente da função NCP envolvida. Podemos então aproveitar dessa estrutura para provar certos resultados gerais sobre o sistema de equações. Uma vez feito esse estudo geral, deixamos aos estudos específicos de cada função apenas as propriedades cujas provas necessitam explicitamente da expressão da função ϕ.

Observamos que por serem gerais, esses resultados podem ser usados no estudo de reformulações semi-suave baseadas em outras funções NCP que não sejam as três estudadas nesta dissertação.

4.1.1 Semi-suavidade

Como Qi [30], definimos diferentes classes funcionais sobre as quais serão baseadas as condições de semi-suavidade das H.

Definição 4.1
Uma função de $\mathbb{R}^n$ em $\mathbb{R}^m$ é:
- LC^1 se ela é C^1 e se a sua derivada é localmente Lipschitziana;
- SC^1 se ela é C^1 e se a sua derivada é semi-suave;
- LC^2 se ela é C^2 e se a sua derivada segunda é localmente Lipschitziana.

Observamos que se f é uma função LC^1, então de fato ela é fortemente semi-suave. Com efeito, para todo x, d em $\mathbb{R}^n$, com $\|d\|$ suficientemente pequeno temos:

$$\frac{\|f'_d(x+d;d) - f'_d(x;d)\|}{\|d\|^2} = \frac{\|f'(x+d)d - f'(x)d\|}{\|d\|^2} \leq \frac{\|f'(x+d) - f'(x)\|\|d\|}{\|d\|^2} \leq K_x$$

onde K_x é o coeficiente de Lipschitz de f' em torno de x. A forte semi-suavidade segue então da caracterização 2.19.

Segundo a proposição 2.20, o estudo da semi-suavidade de funções vetoriais pode ser feito componente por componente.

Logo, observamos que as $n+p$ primeiras componentes de H são simples combinações lineares de F, g, h, e das transpostas das Jacobianas g' e h'.

Portanto, por estabilidade da semi-suavidade (resp. forte semi-suavidade) por combinação linear, obtemos que as $n+p$ primeiras linhas de H são:

(a) semi-suave desde que F é semi-suave, e as funções h e g são SC^1;

(b) fortemente semi-suave desde que F é fortemente semi-suave, e as funções h e g são LC^1.

Resta então estudar a semi-suavidade das q últimas linhas de H, compostas pela função Φ:

$$\Phi(x,\mu) := [\phi(g_i(x),\mu_i)]_{1\leq i\leq q}$$

Suponhamos que g é semi-suave (resp. fortemente semi-suave). Então, graças aos resultados de Mifflin [25] e de Fischer [13] (Proposição 2.21) a respeito da composição de funções semi-suaves, é suficiente provar que ϕ é semi-suave (resp. fortemente semi-suave) para obter o resultado equivalente para Φ.

Como será provado na Seção 4.2, todas as funções NCP estudadas aqui são fortemente semi-suaves. Finalemente, obtemos então a seguinte proposição:

Proposição 4.1 (Semi-suavidade de H)
(a) H é semi-suave se F é semi-suave, e as funções h e g são SC^1.

(b) H é fortemente semi-suave se F é fortemente semi-suave, e as funções h e g são LC^1.

4.1.2 Regularidade na solução

Para cada $z=(x,\lambda,\mu)$ solução de (3.1) definimos os seguintes conjuntos:

- Restrições de desigualdade ativas não degeneradas

$$\mathcal{I}(z) = \{1\leq i\leq q \mid g_i(x)=0, \mu_i>0\}$$

- Restrições de desigualdade ativas degeneradas

$$\mathcal{J}(z) = \{1\leq i\leq q \mid g_i(x)=0, \mu_i=0\}$$

- Restrições de desigualdade não ativas

$$\mathcal{K}(z) = \{1\leq i\leq q \mid g_i(x)>0, \mu_i=0\}$$

Como em uma solução de (3.1) μ e $g(x)$ são complementares, esses três conjuntos formam uma partição do conjunto de índices $\{1,\dots,q\}$:

$$\{1,\dots,q\} = \mathcal{I} \sqcup \mathcal{J} \sqcup \mathcal{K}$$

Qi [30] propõe a seguinte condição de regularidade, que ele chama de extensão da *condição de Robinson* ao caso não diferenciável. Observamos que o segundo requisito dessa condição é uma extensão da condição suficiente de otimalidade de segunda ordem, que garante que uma solução do sistema KKT associado a um problema de minimização também é minimizador local estrito do problema original.

Definição 4.2 (Condição de Robinson generalizada)
Seja um ponto $z = (x, \lambda, \mu)$ solução de (3.1).
Seja $G(z)$ o núcleo das restrições ativas não degeneradas:

$$G(z) = N(h'(x)) \cap \bigcap_{j\in\mathcal{I}(z)} N(g'_j(x))$$

Diz-se que z verifica a condição de Robinson *se:*

(a) os gradientes das restrições ativas são linearmente independentes, ou seja

$$\text{o conjunto } \{\nabla g_j(x)\}_{j\in\mathcal{I}(z)\cup\mathcal{J}(z)} \cup \{\nabla h_i(x)\}_{i=1}^{p} \quad \text{é linearmente independente;}$$

(b) z verifica a condição suficiente de segunda ordem generalizada:

$$\text{para todo } V \in \partial_x\mathcal{F}(z) \text{ e todo } d \in G(z)\backslash\{0\} \text{ tem-se } d^T V d > 0,$$

ou seja: toda $V \in \partial_x\mathcal{F}(z)$ é definida positiva sobre $G(z)$.

Como g é suposta pelo menos C^1, ela é estritamente diferenciável. Então, pelas regras do cálculo diferencial generalizado (em particular a Proposição 2.13), um elemento de $\partial_C H(z)$ é uma matriz em $\mathcal{M}_{n+p+q}(\mathbb{R})$ da forma:

$$\begin{pmatrix} V & h'(x)^T & -g'(x)^T \\ h'(x) & 0 & 0 \\ \Gamma(z)g'(x) & 0 & Z(z) \end{pmatrix} \tag{4.2}$$

onde:

- V é um elemento de $\partial_x\mathcal{F}$
- $\Gamma(z)$ e $Z(z)$ são matrizes diagonais em $\mathcal{M}_q(\mathbb{R})$ cujos elementos diagonais (γ_i) e (ζ_i) são tais que:

$$(\gamma_i, \zeta_i) \in \partial\phi(g_i(x), \mu_i) \qquad (1 \le i \le q).$$

Esta expressão vem do fato de que o C-subdiferencial calcula-se simplesmente componente por componente.

Com hipóteses razoáveis sobre ϕ, podemos provar que o C-subdiferencial de H é regular em qualquer solução que verifica a condição de Robinson. O seguinte teorema é inspirado no estudo geral sobre funções NCP de Qi e Sun [35] e na prova da regularidade de Qi [30]:

Teorema 4.1 (Regularidade)
Seja ϕ uma função NCP tal que para todo a e b complementares em $\mathbb{R}$, e (γ, ζ) em $\partial\phi(a, b)$:

$$\begin{cases} (a) & \gamma \neq 0 \text{ e } \zeta = 0 & \text{se } a = 0 \text{ e } b > 0, \\ (b) & \gamma\zeta \geq 0 \text{ e } \gamma + \zeta \neq 0 & \text{se } a = 0 \text{ e } b = 0, \\ (c) & \gamma = 0 \text{ e } \zeta \neq 0 & \text{se } a > 0 \text{ e } b = 0. \end{cases}$$

Então $\partial_C H$ é regular em toda solução de (3.1) que verifica a condição de Robinson.

Demonstração
Sejam z uma solução de (3.1) que verifica a condição de Robinson, W um elemento de $\partial_C H(z)$, e $d = (d_1, d_2, d_3)$ em $\mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^q$ tal que:

$$W \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix} = 0$$

Queremos mostrar que $d = 0$.

Substituindo W pela sua forma genérica (4.2), obtem-se:

$$\begin{aligned} V d_1 + h'(x)^T d_2 - g'(x)^T d_3 &= 0 && (4.3) \\ h'(x) d_1 &= 0 && (4.4) \\ \gamma_i g_i'(x) d_1 + \zeta_i d_3^i &= 0 \qquad (1 \leq i \leq q) && (4.5) \end{aligned}$$

Como z é solução de (3.1), $g(x)$ e μ são complementares em $\mathbb{R}^q$, e, para cada $1 \leq i \leq q$ estamos em uma das três situações (a), (b) ou (c).
Se $i \in \mathcal{I}(z)$, estamos no caso (a) e temos por (4.5):

$$\gamma_i g_i'(x) d_1 = 0,$$

com $\gamma_i \neq 0$. Portanto:

$$g_i'(x) d_1 = 0,$$

e junto com (4.4), temos então:

$$\boxed{d_1 \in G(z)}$$

Por outro lado, temos:

$$d_3^T g'(x) d_1 = \sum_{i=1}^{q} d_3^i g_i'(x) d_1.$$

Observemos então que por (a), (b) e (c):

- Caso $\gamma_i = 0$ e então $\zeta_i \neq 0$. Por (4.5), Isto implica:

$$d_3^i = 0.$$

- Caso $\gamma_i \neq 0$ e $\zeta_i = 0$, e então (4.5) dá:

$$g_i'(x) d_1 = 0.$$

- Caso $\gamma_i \neq 0$ e $\zeta_i \neq 0$, e (4.5) dá:

$$g_i'(x) d_1 = -\frac{\zeta_i}{\gamma_i} (d_3^i)^2.$$

  Este caso só pode acontecer se $g_i(x) = \mu_i = 0$, e (b) fornece então:

$$\frac{\zeta_i}{\gamma_i} > 0$$

Afinal sobram no somatório somente os termos para os quais γ_i e ζ_i são não nulos. Então:

$$\boxed{d_3^T g'(x) d_1 = - \sum_{i \mid \gamma_i, \zeta_i \neq 0}^{q} \frac{\zeta_i}{\gamma_i} (d_3^i)^2}$$

Porém, multiplicando à esquerda (4.3) por d_1^T obtemos:

$$d_1^T V d_1 + \underbrace{d_1^T h'(x)^T d_2}_{=0} - d_1^T g'(x)^T d_3 = 0.$$

Assim, substituindo $d_1^T g'(x)^T d_3$ pelo somatório, temos:

$$d_1^T V d_1 = - \sum_{i \mid \gamma_i, \zeta_i \neq 0}^{q} \frac{\zeta_i}{\gamma_i} (d_3^i)^2. \tag{4.6}$$

Isto com d_1 em $G(z)$ e cada $\frac{\zeta_i}{\gamma_i} > 0$.
Como z verifica a condição de Robinson, (4.6) implica:

$$\boxed{\begin{array}{l} d_1 = 0 \\ d_3^i = 0 \quad \text{para } i \text{ tal que } \gamma_i, \zeta_i \neq 0 \end{array}}$$

Por fim, substituindo em (4.3) temos:

$$h'(x)^T d_2 - \sum_{i \mid \gamma_i \zeta_i = 0} g_i'(x)^T d_3^i = 0.$$

Já vimos que:

$$\boxed{d_3^i = 0 \qquad \text{para } i \text{ tal que } \gamma_i = 0}.$$

Portanto a igualdade acima se reescreve de fato:

$$h'(x)^T d_2 - \sum_{i \mid \gamma_i \neq 0, \zeta_i = 0} g_i'(x)^T d_3^i = 0.$$

Porém, $\gamma_i \neq 0$ somente acontece se $g_i(x) = 0$. Logo, por independência dos gradientes das restrições ativas, temos:

$$\boxed{\begin{array}{l} d_2 = 0 \\ d_3^i = 0 \quad \text{para } i \text{ tal que } \gamma_i \neq 0, \zeta_i = 0 \end{array}}$$

Afinal obtivemos o desejado: $(d_1, d_2, d_3) = (0, 0, 0)$. ■

4.1.3 Caráter C^1 da função de mérito

Por definição da norma euclidiana, temos para $z = (x, \lambda, \mu)$ em $\mathbb{R}^{n+p+q}$:

$$2\theta(z) = \|H(z)\|^2 = \|\mathcal{F}(z)\|^2 + \|h(x)\|^2 + \|\Phi(x, \mu)\|^2$$

Sabemos que o quadrado da norma euclidiana é de classe C^1. Portanto, supondo F de classe C^1 e as funções h e g de classe C^2, então $\mathcal{F}$ e h são de classe C^1, e por isto θ é de classe C^1 se e somente se a função:

$$\|\Phi\|^2 : x, \mu \longmapsto \|\Phi(x, \mu)\|^2 = \sum_{i=1}^{q} \phi^2(g_i(x), \mu_i)$$

é de classe C^1.

Uma condição suficiente para obter isto é, claramente, que cada um dos termos da soma seja também de classe C^1. Provamos então um lema técnico, que de fato mostra qual propriedade permite a uma função ter um quadrado de classe C^1, sem deixar de possuir pontos de não diferenciabilidade:

Lema 4.1
Seja $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ localmente Lipschitziana, e de classe C^1 exceto no conjunto Ω_f, onde ela não é diferenciável. Seja $v(f)$ o conjunto dos zeros de f. Então:

$$\Omega_f \subset v(f) \Longleftrightarrow \textit{a função } f^2 \textit{ é de classe } C^1$$

Demonstração
A função $\alpha : u \longmapsto u^2$ sendo C^1, ela é estritamente diferenciável.
A proposição 2.16 se aplica então à função $\alpha \circ f = f^2$, dando para todo x em $\mathbb{R}^n$:

$$\partial f^2(x) = 2f(x)\partial f(x)$$

- $\Longrightarrow$: Fora de Ω_f, o gradiente generalizado de f^2 é unitário, reduzido ao gradiente clássico, pois f é C^1 aí.
  Em pontos de Ω_f, a função f é nula por hipótese, e por isto ∂f^2 também é unitário e igual a $\{0\}$.
  Portanto, pelo corolário 2.9.1, f^2 é C^1 e, em cada ponto, o seu gradiente é igual ao único elemento do seu gradiente generalizado.
- $\Longleftarrow$: Seja x em Ω_f.
  Como f não é diferenciável em x, a proposição 2.9 garante que o gradiente generalizado $\partial f(x)$ tem pelo menos dois elementos.
  Portanto, se $f(x)$ não fosse nulo, então o gradiente generalizado $\partial f^2(x)$ não seria unitário, o que contradiria a hipótese que f^2 é de classe C^1.

■

4.1.4 Escolha de um elemento em $\partial_B H(z)$

Suponhamos que F é de classe C^1 e que as funções h e g são de classe C^2. Então, os pontos de não diferenciabilidade de H coincidem com os de Φ:

$$\Omega_H = \Omega_\Phi$$

Logo, fora do conjunto Ω_Φ, H é C^1 e o seu B-subdiferencial se reduz a um elemento: o seu Jacobiano. Por um cálculo clássico obtemos:

$$H'(z) = \begin{pmatrix} \frac{\partial \mathcal{F}}{\partial x} & h'(x)^T & -g'(x)^T \\ h'(x) & 0 & 0 \\ \Gamma(z)g'(x) & 0 & Z(z) \end{pmatrix} \tag{4.7}$$

onde: $\Gamma(z)$ e $Z(z)$ são matrizes diagonais em $\mathcal{M}_q(\mathbb{R})$ cujos elementos diagonais (γ_i) e (ζ_i) são tais que:

$$\gamma_i = \frac{\partial \phi}{\partial a}(g_i(x), \mu_i) \quad \text{e} \quad \zeta_i = \frac{\partial \phi}{\partial b}(g_i(x), \mu_i) \qquad (1 \leq i \leq q)$$

Para um ponto z em Ω_H, H não é diferenciável devido à não diferenciabilidade da função Φ. No entanto, sendo limite de matrizes Jacobianas de H, um elemento do B-subdiferencial $\partial_B H(z)$ conserva a forma matricial (4.7) acima. A diferença é que existe uma multiplicidade

de escolhas possíveis para os elementos diagonais das matrizes $\Gamma(z)$ e $Z(z)$. Eles devem ser escolhidos tais que a seguinte matriz pertença a $\partial_B\Phi(z)$:

$$V = \begin{pmatrix} \Gamma(z)g'(x) & 0 & Z(z) \end{pmatrix} = \begin{pmatrix} \gamma_1 g_1'(x) & & \zeta_1 & \\ \vdots & 0 & & \ddots \\ \gamma_q g_q'(x) & & & \zeta_q \end{pmatrix} \tag{4.8}$$

Para a expressão ficar menos carregada, não explicitamos aqui a dependência à variável z dos coeficientes (γ_i) e (ζ_i).

Obviamente a forma dos coeficientes (γ_i) e (ζ_i) depende da função Φ, ou seja da função ϕ. Logo, para cada função NCP, temos que definir exatamente qual elemento de $\partial_B\Phi(z)$ será escolhido no caso em que z pertence ao conjunto Ω_H.

No entanto, podemos observar que a função Φ também tem uma estrutura especial, que pode ser explorada com o fim de precisar mais um pouco quais são os coeficientes (γ_i) e (ζ_i) no caso geral.

A função Φ é por definição:

$$\Phi(z) = \begin{pmatrix} \phi(g_1(x), \mu_1) \\ \vdots \\ \phi(g_q(x), \mu_q) \end{pmatrix}$$

Portanto, os pontos de não diferenciabilidade de Φ são necessariamente pontos $z = (x, \lambda, \mu)$ tais que o seguinte conjunto de índices seja não vazio:

$$\beta_\phi(z) = \{1 \leq j \leq q \mid (g_j(x), \mu_j) \in \Omega_\phi\}$$

Seja então z um ponto qualquer em $\mathbb{R}^{n+p+q}$, e V um elemento de $\partial_B\Phi(z)$. Então V é da forma (4.8), e por definição do B-subdiferencial, existe um seqüência de pontos $z_s = (x^s, \lambda^s, \mu^s)$ fora de Ω_Φ tal que:

$$z_s \longrightarrow z \qquad \text{e} \qquad V^s := \Phi'(z_s) \longrightarrow V.$$

Para i fora de $\beta_\phi(z)$, a componente Φ_i' do Jacobiano de Φ é bem definida e contínua no ponto z. Por isto a i-ésima linha das matrizes Jacobianas V^s converge a $\Phi_i'(z)$, e

$$V_i = \begin{pmatrix} \frac{\partial\phi}{\partial a}(g_i(x), \mu_i)g_i'(x) & 0 & \frac{\partial\phi}{\partial b}(g_i(x), \mu_i)e_i^T \end{pmatrix},$$

onde o 0 do meio corresponde às p componentes nulas relativas à variável λ e e_i é o i-ésimo vetor da base canônica de $\mathbb{R}^q$.

Ou seja, temos:

$$\boxed{\text{para } i \notin \beta_\phi(z): \quad \gamma_i = \frac{\partial \phi}{\partial a}(g_i(x), \mu_i) \quad \text{e} \quad \zeta_i = \frac{\partial \phi}{\partial b}(g_i(x), \mu_i)} \tag{4.9}$$

Para i em $\beta_\phi(z)$, como Φ é diferenciável no ponto z_s, a i-ésima linha da matriz V^s se escreve:

$$V_i^s = \left(\begin{array}{ccc} \frac{\partial \phi}{\partial a}(g_i(x^s), \mu_i^s) g_i'(x^s) & 0 & \frac{\partial \phi}{\partial b}(g_i(x^s), \mu_i^s) e_i^T \end{array} \right).$$

Logo, como g é de classe C^1, é imediato ver que:

(a) Se $\frac{\partial \phi}{\partial a}$ existe e é contínua no ponto $(g_i(x), \mu_i)$ então necessariamente:

$$\gamma_i = \frac{\partial \phi}{\partial a}(g_i(x), \mu_i).$$

(b) Se $\frac{\partial \phi}{\partial b}$ existe e é contínua no ponto $(g_i(x), \mu_i)$ então necessariamente:

$$\zeta_i = \frac{\partial \phi}{\partial b}(g_i(x), \mu_i).$$

(c) Se $g_i'(x) = 0$ então o valor de γ_i é irrelevante, pois então, independentemente de γ_i, as n primeiras componentes de V_i são nulas.

Observamos que as situações (a) e (b) são mutuamente excludentes uma da outra, pois senão ϕ seria de classe C^1 no ponto $(g_i(x), \mu_i)$. Em compensação, é possível ter ao mesmo tempo as situações (b) e (c). Neste caso, vemos que V_i é totalmente determinada, o que significa que z não é um ponto de não diferenciabilidade de Φ_i. Assim, pode acontecer que um ponto z seja tal que $\beta_\phi(z) \neq \emptyset$, mas fique fora de Ω_Φ.

4.2 Estudo das funções NCP caso a caso

Nesta secção estudamos separadamente as funções do mínimo ϕ_*, de Fischer-Burmeister ϕ_1 e de Fischer-Burmeister penalizada ϕ_α. Enfatizamos novamente o fato de que um ponto importante para a implementação prática do Algoritmo 2 é a definição de um processo de escolha de um elemento dentro do B-subdiferencial. Ou seja, a definição de uma semi-derivada para a função H.

Baseando-se nas considerações gerais anteriores, para cada uma das funções NCP, vamos:
- provar a forte semi-suavidade;
- estabelecer as condições do teorema 4.1 de regularidade de $\partial_B H$ em uma solução;

- provar que os seus pontos de não diferenciabilidade também são zeros (exceto para ϕ_*);
- calcular um elemento em $\partial_B \Phi(z)$.

Lembramos que N_0 é o conjunto dos zeros de qualquer função NCP:

$$N_0 = (\mathbb{R}_+ \times \{0\}) \cup (\{0\} \times \mathbb{R}_+).$$

4.2.1 A função do Mínimo

A função do Mínimo é:

$$\begin{array}{rccc} \phi_* : & \mathbb{R}^2 & \longrightarrow & \mathbb{R} \\ & (a,b) & \longmapsto & \min(a,b) \end{array} \tag{4.10}$$

Essa também tem as outras expressões:

$$\phi_*(a,b) = \frac{1}{2}(a+b-|a-b|) = a - \max\{0, a-b\}.$$

A Figura 4.1 mostra as curvas de nível da função ϕ_*.

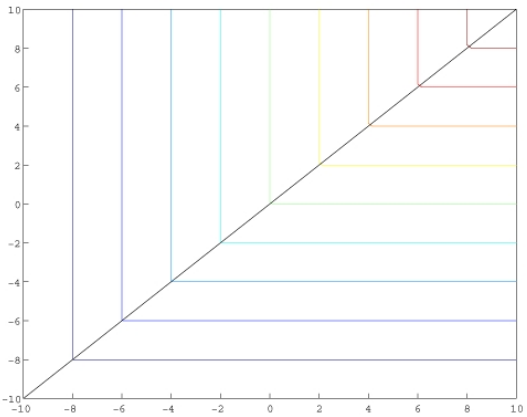


Figura 4.1: Curvas de nível da função Mínimo

Propriedade 1
(a) ϕ_* *é uma função NCP.*

(b) ϕ_* *é 1-Lipschitziana sobre* $\mathbb{R}^2$ *para a norma* $\|\ \|_1$.

(c) ϕ_* *é* C^∞ *em todo ponto onde* $a \neq b$ *e* $\nabla\phi_*(a,b) = \begin{cases} (1,0)^T \text{ se } a < b \\ (0,1)^T \text{ se } b < a. \end{cases}$

Demonstração

- Item (a): imediato.
- Item (c): a definição de ϕ_* que usa a função valor absoluto nós dá imediatamente o caráter C^∞ de ϕ_* fora da primeira bissetriz.

  Logo, quando $a < b$ por exemplo, em uma vizinhança de (a, b) temos $a' < b'$, que implica:

  $$\phi_*(a', b') = a'$$

  Portanto, $\nabla\phi_*(a, b) = (1, 0)^T$.

  O caso $b < a$ é análogo, dando $\nabla\phi_*(a, b) = (0, 1)^T$.
- Item (b): Sejam (a, b) e (a', b') em $\mathbb{R}^2$ temos:

$$\begin{aligned}
|\phi_*(a,b) - \phi_*(a',b')| &= \frac{1}{2}\,|a+b-|a-b| - (a'+b'-|a'-b'|)| \\
&\leq \frac{1}{2}\left[|a-a'| + |b-b'| + \underbrace{||a-b| - |a'-b'||}_{\leq |(a-b)-(a'-b')|}\right] \\
&\leq \frac{1}{2}\,[2|a-a'| + 2|b-b'|] \\
|\phi_*(a,b) - \phi_*(a',b')| &\leq \|(a,b) - (a',b')\|_1 .
\end{aligned}$$

■

Propriedade 2 (Gradiente generalizado)

Para (a, b) em $\mathbb{R}^2$, temos:

$$\partial\phi_*(a,b) = \begin{cases} \{(1,0)\} & \text{se } a < b \\ \{(0,1)\} & \text{se } b < a \\ \{(\lambda, 1-\lambda); \lambda \in [0,1]\} & \text{se } a = b \end{cases}$$

Demonstração

Seja $x = (a, b)$ um ponto qualquer de $\mathbb{R}^2$. Então:

- Caso $a \neq b$: como ϕ_* é C^1 em uma vizinhança de x, ela é também estritamente diferenciável no ponto x (cf. Proposição 2.8). Portanto, a Proposição 2.9 dá que o gradiente generalizado reduz-se ao gradiente clássico $\nabla\phi_*(a, b)$ e a Propriedade 1 (c) fornece o resultado anunciado.
- Caso $a = b$: como ϕ_* é globalmente Lipschitziana, ela é localmente Lipschitziana no ponto (a, a).

  Como é possível convergir a (a, a) tanto por um lado de Δ_1 quanto do outro, temos:

  $$\partial_B\phi_*(a,a) = \{(0,1),(1,0)\}\,.$$

  Logo:

  $$\partial\phi_*(a,a) = co(\partial_B\phi_*(a,a)) = \{(\lambda, 1-\lambda); \lambda \in [0,1]\}\,.$$

■

Condição de regularidade

Sejam a e b complementares em $\mathbb{R}$. Da expressão de $\partial\phi_*$ obtemos:

$$\partial\phi_*(a,b) = \{\phi'_*(a,b)\} = \begin{cases} \{(1,0)\} & \text{se } a = 0 \text{ e } b > 0 \\ \{(\lambda, 1-\lambda); \lambda \in [0,1]\} & \text{se } a = b = 0 \\ \{(0,1)\} & \text{se } a > 0 \text{ e } b = 0. \end{cases}$$

Portanto, é fácil ver que ϕ_* satisfaz as hipóteses do teorema 4.1. Assim, temos o corolário:

Corolário 4.1.1
$\partial_C H_*$ é regular em todo ponto z tal que $H_*(z) = 0$, no qual se verifica a condição de Robinson.

Proposição 4.2 (Forte semi-suavidade)
A função ϕ_* é fortemente semi-suave.

Demonstração
Seja $x = (a, b)$.
De novo, dois casos são possíveis:

- Caso $a \neq b$: então ϕ_* é C^∞ em x e portanto LC^1 em x. Logo, como observado na Subseção 4.1.1, ela é fortemente semi-suave em x.
- Caso $a = b$: Seja h em $\mathbb{R}^2$. Então ϕ_* é direcionalmente diferenciável em $x + h$ e x com:

$$\phi'_*(x+h;h) = \phi'_*(x;h) = \begin{cases} h_2 & \text{se } h_2 < h_1 \\ h_1 & \text{se } h_1 < h_2 \\ h_1 + h_2 & \text{se } h_1 = h_2 \end{cases}$$

  Portanto, para todo h:

$$\frac{\phi'_*(x+h;h) - \phi'_*(x;h)}{\|h\|} = 0,$$

  e é claro que ϕ_* é fortemente semi-suave em x.

■

Escolha de um elemento em $\partial_B \Phi_*(z)$

O conjunto Ω_{ϕ_*} é a primeira bissetriz. Logo o conjunto de índices das componentes potencialmente não diferenciáveis em um ponto $z = (x, \lambda, \mu)$ é:

$$\beta_{\Phi_*}(z) = \{1 \leq i \leq q \mid g_i(x) = \mu_i\}.$$

Segundo a Subseção 4.1.4, para definir um elemento de $\partial_B \Phi_*(z)$, basta definir os coeficientes diagonais das matrizes $\Gamma(z)$ e $Z(z)$. Seja então a definição seguinte:

- Se $i \notin \beta_{\Phi_*}(z)$:

$$\gamma_i = \begin{cases} 0 & \text{se } g_i(x) > \mu_i \\ 1 & \text{se } g_i(x) < \mu_i \end{cases} \quad \text{e} \quad \zeta_i = \begin{cases} 1 & \text{se } g_i(x) > \mu_i \\ 0 & \text{se } g_i(x) < \mu_i \end{cases}$$

- Se $i \in \beta_{\Phi_*}(z)$:

$$\gamma_i = 1 \quad \text{e} \quad \zeta_i = 0$$

Proposição 4.3
Com (γ_i) e (ζ_i) definidos como acima, a matriz V em (4.8) pertence a $\partial_B \Phi_*(z)$.

Demonstração
Seja z em Ω_{Φ_*}, e seja V definida como em (4.8), com (γ_i) e (ζ_i) como acima. Queremos construir uma seqüência (z_s) fora de Ω_{Φ_*} tal que:

$$z_s \longrightarrow z \quad \text{e} \quad \Phi'(z_s) \longrightarrow V.$$

Seja a seqüência (z_s) definida como:

$$z_s = (x_s, \lambda_s, \mu_s) = (x, \lambda, \mu + \varepsilon_s),$$

onde

$$\varepsilon_s^i = \begin{cases} 0 & \text{se } i \notin \beta_{\Phi_*}(z) \\ \frac{1}{s} & \text{se } i \in \beta_{\Phi_*}(z) \end{cases}$$

Para todo $s \geq 0$ e $1 \leq i \leq q$, se i pertence a $\beta_{\Phi_*}(z)$ então:

$$g_i(x_s) = g_i(x) = \mu^i < \mu^i + \varepsilon_s^i = \mu_s^i.$$

Se i não pertence a $\beta_{\Phi_*}(z)$ então:

$$g_i(x_s) = g_i(x) \neq \mu^i = \mu_s^i.$$

Por isto, (z_s) fica fora de Ω_{Φ_*}. Também é imediato ver que $z_s \longrightarrow z$.

Agora, analisamos a convergência de $(\Phi'_*(z_s))$ linha por linha:

- Para $i \notin \beta_{\Phi_*}(z)$: (γ_i) e (ζ_i) são as derivadas parciais de ϕ_* nos pontos $(g_i(x), \mu_i)$ como especificado em (4.9) na Subseção 4.1.4. Logo, a convergência das i-ésima linha de $(\Phi_*(z_s))$ a V_i é automática.
- Para $i \in \beta_{\Phi_*}(z)$:

$$(\Phi'_*(z_s))_i = \left(\frac{\partial \phi_*}{\partial a}(g_i(x_s), \mu_s^i) g_i'(x_s) \quad , \quad 0 \quad , \quad \frac{\partial \phi_*}{\partial b}(g_i(x_s), \mu_s^i) \right).$$

Como $g_i(x_s) = g_i(x) = \mu_i < \mu_s^i$, vem que:

$$\frac{\partial \phi_*}{\partial a}(g_i(x_s), \mu_s^i) = 1 \quad \text{e} \quad \frac{\partial \phi_*}{\partial b}(g_i(x_s), \mu_s^i) = 0$$

Logo:

$$(\Phi'_*(z_s))_i = (g_i'(x_s), 0, 0) \equiv_s (g_i'(x), 0, 0) = V_i$$

Assim $((\Phi'_*(z_s))_i)$ é constante e igual à i-ésima linha de V, e por isto converge a ela.

■

4.2.2 A função de Fisher-Burmeister (FB)

A função NCP dita de Fischer-Burmeister é a função:

$$\begin{array}{rccc} \phi_1: & \mathbb{R}^2 & \longrightarrow & \mathbb{R} \\ & (a,b) & \longmapsto & a+b-\sqrt{a^2+b^2} \end{array} \tag{4.11}$$

A Figura 4.2 mostra as curvas de nível da função ϕ_1. Essas revelam a função de Fischer-Burmeister como uma versão mais suave da função do mínimo. Com efeito, fora da origem, ϕ_1 não apresenta nenhum bico de não diferenciabilidade, apesar de ter curvas de forma parecidas às de ϕ_*.

Introduzida por Fischer [12], esta função NCP foi bastante estudada para resolver problemas de complementaridade não linear (NCP) [6, 8, 15, 21, 31] ou problemas de desigualdade variacional (VIP) [7, 10]. Também, Qi e Jiang [30] fizeram um estudo teórico da reformulação semi-suave aplicada aos sistemas KKT.

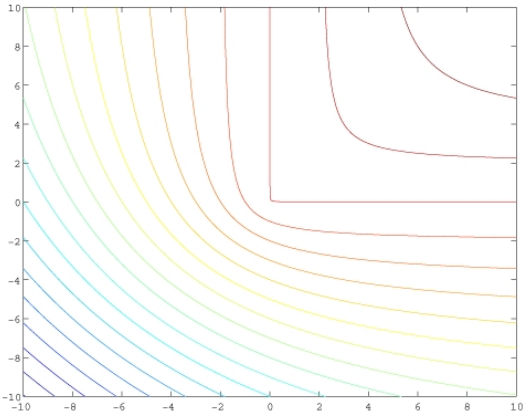


Figura 4.2: Curvas de nível da função de Fischer-Burmeister

Propriedade 3
(a) ϕ_1 *é uma função NCP.*

(b) ϕ_1 *é K-Lipschitziana sobre* $\mathbb{R}^2$ *para a norma* $\|\ \|_2$, *com* $K = 1 + \sqrt{2}$.

(c) ϕ_1 *é* C^∞ *em todo ponto de* $\mathbb{R}^2 \setminus \{(0,0)\}$ *e:*

$$\nabla\phi_1(a,b) = \begin{pmatrix} 1 - \frac{a}{\|(a,b)\|_2} \\ \\ 1 - \frac{b}{\|(a,b)\|_2} \end{pmatrix}.$$

(d) ϕ_1^2 *é de classe* C^1.

Demonstração

- (a) é claro que $N_0 \subset v(\phi_1)$.
  Agora, supondo que $\phi_1(a,b) = 0$ temos:

  $$a + b = \sqrt{a^2 + b^2}.$$

  Ou seja, $a^2 + b^2 + 2ab = a^2 + b^2$ e $a + b \geq 0$.
  Dai vêm: $a, b \geq 0$ com $a = 0$ ou $b = 0$, i.e. $(a,b) \in N_0$.

- (c) vem do fato de que a função raiz é C^∞ sobre $\mathbb{R}_+^*$, e que:

  $$a^2 + b^2 = 0 \Leftrightarrow (a,b) = (0,0).$$

  A expressão da derivada é então obtida aplicando diretamente as regras clássicas de derivação de funções compostas.

- (b) Sejam $a = (a_1, a_2)$ e $b = (b_1, b_2)$ em $\mathbb{R}^2$ temos:

  $$\begin{aligned} |\phi_1(a) - \phi_1(b)| &= |a_1 - b_1 + a_2 - b_2 + \|b\|_2 - \|a\|_2| \\ &\leq \underbrace{|a_1 - b_1| + |a_2 - b_2|}_{=\|a-b\|_1 \leq \sqrt{2}\|a-b\|_2} + \underbrace{|\|b\|_2 - \|a\|_2|}_{\leq \|a-b\|_2} \\ &\leq \left(1 + \sqrt{2}\right) \|a - b\|_2 . \end{aligned}$$

- (d): ϕ_1 é C^1 exceto nos pontos de não diferenciabilidade que também são zeros de ϕ_1. Portanto, o resultado segue do lema 4.1.

■

Propriedade 4 (Gradiente generalizado)
O gradiente generalizado de ϕ_1 em $(a,b) \in \mathbb{R}^2$ é:

$$\partial\phi_1(a,b) = \begin{cases} \{\nabla\phi_1(a,b)\} & se\ (a,b) \neq (0,0) \\ (1,1) + B & se\ (a,b) = (0,0) \end{cases}$$

onde B é a bola euclidiana unitária de $\mathbb{R}^2$, ou seja:

$$B = \left\{(a,b) \in \mathbb{R}^2 / \left\|(a,b)\right\|_2 \leq 1\right\}$$

Demonstração

- Se $(a,b) \neq (0,0)$: ϕ_1 é C^1 em (a,b), por isto estritamente diferenciável e o seu gradiente generalizado no ponto (a,b) reduz-se a $\nabla\phi_1(a,b)$.
- Se $(a,b) = (0,0)$: como ϕ_1 é Lipschitziana, ela é localmente Lipschitziana em $(0,0)$. O seu gradiente generalizado na origem é então:

  $$\partial\phi_1(0,0) = \text{co}\left\{\lim \phi_1'(a_s,b_s); (a_s,b_s) \to (0,0), (a_s,b_s) \neq (0,0)\right\}$$

  Seja $((a_s,b_s))$ tendendo a $(0,0)$, tal que $(a_s,b_s) \neq (0,0)$. Temos:

  $$\phi_1'(a_s,b_s) = (1,1) - \left(\frac{a}{\left\|(a,b)\right\|_2}, \frac{b}{\left\|(a,b)\right\|_2}\right) \in (1,1) - B = (1,1) + B$$

  Como (1,1) + B é fechado, obtemos:

  $$\boxed{\partial\phi_1(0,0) \subset (1,1) + B}$$

  Reciprocamente, seja $(\gamma,\zeta) \in B$. Então

  $$(a_s,b_s) = \frac{1}{n}(\gamma,\zeta) \longrightarrow (0,0)$$

  Com:

  $$\phi_1'(a_s,b_s) = (1,1) - \left(\frac{\gamma/n}{\frac{1}{n}}, \frac{\zeta/n}{\frac{1}{n}}\right) = (1,1) - (\gamma,\zeta)$$

  Por isto:

  $$\boxed{(1,1) + B \subset \partial\phi_1(0,0)}$$

■

Condição de Regularidade

Sejam a e b complementares em $\mathbb{R}^2$. Temos:

$$\partial\phi_1(a,b) = \begin{cases} \{(1,0)\} & \text{se } a = 0 \text{ e } b > 0 \\ (1,1) + B & \text{se } a = b = 0 \\ \{(0,1)\} & \text{se } a > 0 \text{ e } b = 0 \end{cases}$$

É claro que ϕ_1 satisfaz as condições do teorema 4.1 no caso $(a,b) \neq (0,0)$. Se a e b forem nulos, temos que provar que:

$$(\gamma,\zeta) \in B \Longrightarrow (1+\gamma)(1+\zeta) \geq 0 \text{ e } 2+\gamma+\zeta \neq 0.$$

Porém, se (γ,ζ) pertence a B então:

$$|\gamma| \leq 1 \text{ e } |\zeta| \leq 1.$$

Portanto:

$$(1+\gamma)(1+\zeta) \geq 0.$$

Além disso, temos por Cauchy-Schwarz:

$$|\gamma+\zeta| \leq \sqrt{2} < 2.$$

Por isto:

$$2+\gamma+\zeta \neq 0.$$

Então, podemos enunciar o seguinte resultado:

Corolário 4.3.1
$\partial_C H_1$ é regular em todo ponto z tal que $H_1(z) = 0$, no qual se verifica a condição de Robinson.

Proposição 4.4 (Forte semi-suavidade)
A função ϕ_1 é fortemente semi-suave.

Demonstração
Seja $x = (a,b)$. Dois casos podem acontecer:

- $x = (a,b) \neq (0,0)$: então ϕ_1 é C^2 em x e portanto LC^1 em x. Logo, como observado na Subseção 4.1.1, ela é fortemente semi-suave em x.

▶ $x = (a, b) = (0, 0)$: Seja $h \neq 0$ em $\mathbb{R}^2$. Então:

$$\exists \phi_1'(0; h) = \lim_{t \to 0^+} \frac{\phi_1(th) - \phi_1(0)}{t} = \phi_1(h).$$

Logo queremos ter:

$$Vh - \phi_1(h) = 0(\|h\|^2), \tag{4.12}$$

onde V pertence a $\partial\phi_1(h)$.

Porém, em $h \neq 0$, ϕ_1 é C^1 e o Jacobiano generalizado $\partial\phi_1(h)$ se reduz a $\{\phi_1'(h)\}$.

Também temos:

$$\begin{aligned} \phi_1'(h)h &= h_1 - \frac{h_1^2}{\|h\|} + h_2 - \frac{h_2^2}{\|h\|} \\ &= h_1 + h_2 - \frac{h_1^2 + h_2^2}{\|h\|} \\ &= h_1 + h_2 - \|h\| = \phi_1(h). \end{aligned}$$

É então claro que temos (4.12).

■

Escolha de um elemento em $\partial_B \Phi_1(z)$

O conjunto Ω_{ϕ_1} é reduzido à origem. Logo o conjunto de índices das componentes que podem ser não diferenciáveis em um ponto $z = (x, \lambda, \mu)$ de Ω_{Φ_1} é:

$$\beta_{\Phi_1}(z) = \{1 \leq i \leq q \mid g_i(x) = \mu_i = 0\}$$

De novo, para definir um elemento de $\partial_B \Phi_1(z)$, basta definirmos coeficientes diagonais das matrizes $\Gamma(z)$ e $Z(z)$. Seja então a seguinte definição:

- Se $i \notin \beta_{\Phi_1}(z)$:

$$\gamma_i = 1 - \frac{g_i(x)}{\sqrt{g_i(x)^2 + \mu_i^2}} \quad \text{e} \quad \zeta_i = 1 - \frac{\mu_i}{\sqrt{g_i(x)^2 + \mu_i^2}}.$$

- Se $i \in \beta_{\Phi_*}(z)$:

$$\gamma_i = 1 \quad \text{e} \quad \zeta_i = 0.$$

Proposição 4.5
Com (γ_i) e (ζ_i) definidos como acima, a matriz V em (4.8) pertence a $\partial_B \Phi_1(z)$.

Demonstração
Seja z em Ω_{Φ_1} e seja V definida como em (4.8), com (γ_i) e (ζ_i) como acima. Queremos construir uma seqüência (z_s) fora de Ω_{Φ_1} tal que:

$$z_s \longrightarrow z \quad \text{e} \quad \Phi'(z_s) \longrightarrow V.$$

Seja de novo a seqüência (z_s) definida como:

$$z_s = (x_s, \lambda_s, \mu_s) = (x, \lambda, \mu + \varepsilon_s),$$

onde

$$\varepsilon_s^i = \begin{cases} 0 & \text{se } i \notin \beta_{\Phi_1}(z) \\ \frac{1}{s} & \text{se } i \in \beta_{\Phi_1}(z) \end{cases}$$

Para todo $s \geq 0$ e $1 \leq i \leq q$, se i pertence a $\beta_{\Phi_1}(z)$ então:

$$g_i(x_s) = g_i(x) = 0 \quad \text{enquanto} \quad \mu_s^i = \mu_i + \varepsilon_s^i = \varepsilon_s^i > 0.$$

Se i não pertence a $\beta_{\Phi_1}(z)$ então:

$$(g_i(x_s), \mu_s^i) = (g_i(x), \mu_i) \neq (0, 0).$$

Por isto, (z_s) fica fora de Ω_{Φ_1}. Também é imediato ver que $z_s \longrightarrow z$.

Agora, analisamos a convergência de $(\Phi_1'(z_s))$ linha por linha:

- Para $i \notin \beta_{\Phi_1}(z)$: (γ_i) e (ζ_i) são as derivadas parciais de ϕ_1 nos pontos $(g_i(x), \mu_i)$ como especificado em (4.9) na Subseção 4.1.4. Logo, a convergência das i-ésima linha de $(\Phi_1(z_s))$ a V_i é automática.
- Para $i \in \beta_{\Phi_1}(z)$:

$$(\Phi_1'(z_s))_i = \left(\frac{\partial \phi_1}{\partial a}(g_i(x_s), \mu_i + \varepsilon_s^i) g_i'(x_s) \quad , \quad 0 \quad , \quad \frac{\partial \phi_1}{\partial b}(g_i(x_s), \mu_i + \varepsilon_s^i) \right),$$

onde:

$$\frac{\partial \phi_1}{\partial a}(g_i(x_s), \mu_i + \varepsilon_s^i) = 1 - \frac{g_i(x)}{\sqrt{g_i(x)^2 + (\mu_i + \varepsilon_s^i)^2}} = 1, \quad \text{pois } g_i(x) = 0;$$

e:

$$\frac{\partial \phi_1}{\partial b}(g_i(x_s), \mu_i + \varepsilon_s^i) = 1 - \frac{\mu_i + \varepsilon_s^i}{\sqrt{g_i(x)^2 + (\mu_i + \varepsilon_s^i)^2}} = 0, \quad \text{pois } g_i(x) = \mu_i = 0 \text{ e } \varepsilon_s^i > 0.$$

Logo, para todo s:

$$(\Phi_1'(z_s))_i = (g_i'(x_s), 0, 0) = (g_i'(x), 0, 0) = V_i.$$

Assim $((\Phi_1'(z_s))_i)$ é constante e igual a i-ésima linha de V, e por isto converge a ela.

■

4.2.3 A função de Fischer-Burmeister Penalizada (FBP)

A função NCP dita de Fischer-Burmeister Penalizada é a uma versão modificada de ϕ_1. Ela é uma combinação convexa entre ϕ_1 e uma função que penaliza a região $\mathbb{R}^2_{++}$ do plano:

$$\begin{array}{rccl} \phi_\alpha & \mathbb{R}^2 & \longrightarrow & \mathbb{R} \\ & (a,b) & \longmapsto & \alpha\left(a+b-\sqrt{a^2+b^2}\right)+(1-\alpha)a_+b_+ \end{array} \tag{4.13}$$

onde $\alpha \in]0,1]$, e para u em $\mathbb{R}$: $u_+ = \max(u,0)$.
Lembrando a notação $\phi_+(a,b) = a_+b_+$ temos:

$$\phi_\alpha = \alpha\phi_1 + (1-\alpha)\phi_+.$$

Esta função NCP foi introduzida por Chen, Chen e Kanzow [3] no estudo da resolução do problema NCP. Eles testaram o seu algoritmo em vários problemas-teste, e obtiveram resultados numéricos muito promissores. Que saibamos, a função de Fischer-Burmeister penalizada ainda não foi usada para resolver sistemas KKT.

A Figura 4.3 mostra as curvas de nível da função ϕ_α para diferentes valores de α. Quanto menor o α, maior é a penalização do quadrante $\mathbb{R}^2_{++}$, e mais achatadas as curvas de nível no resto do plano $\mathbb{R}^2$.

Propriedade 5
(a) ϕ_α é uma função NCP.

(b) ϕ_α é localmente Lipschitziana sobre $\mathbb{R}^2$, mas não é Lipschitziana para $\alpha \neq 1$.

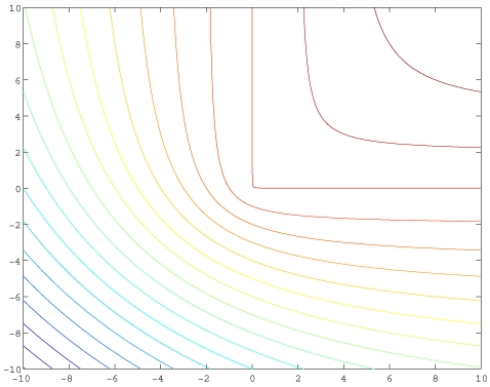
(c) ϕ_α é C^∞ em todo ponto de $\mathbb{R}^2 \backslash N_0$ e:

$$\nabla\phi_\alpha(a,b) = \begin{pmatrix} \alpha(1-\frac{a}{\|(a,b)\|_2})+(1-\alpha)b \\ \\ \alpha(1-\frac{b}{\|(a,b)\|_2})+(1-\alpha)a \end{pmatrix}.$$

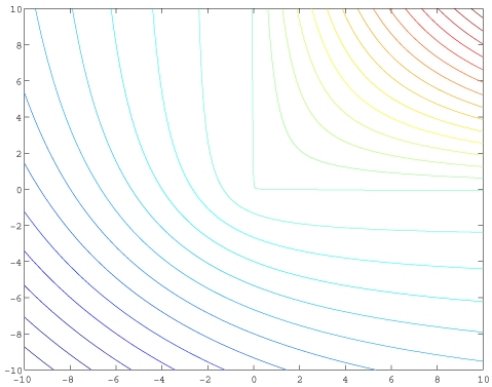
Demonstração

▶ (a): Como $\alpha > 0$, a combinação convexa entre ϕ_1 e ϕ_+ é estrita.
Portanto:

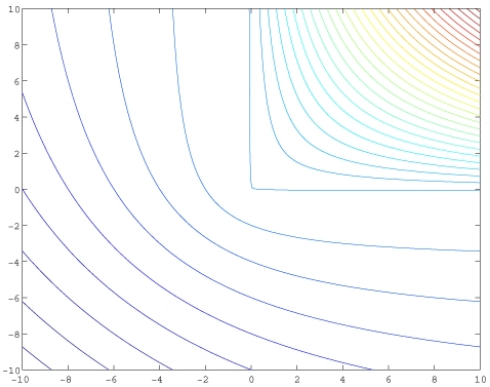
$$\phi_\alpha(a,b) = 0 \Longleftrightarrow \begin{cases} \phi_1(a,b) < 0 \text{ e } \phi_+(a,b) > 0 \\ \phi_1(a,b) = 0 \text{ e } \phi_+(a,b) = 0 \end{cases}$$



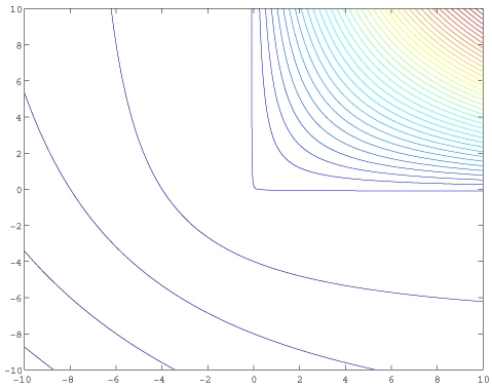
(a) $\alpha = 1 \rightsquigarrow$ Fischer-Burmesiter



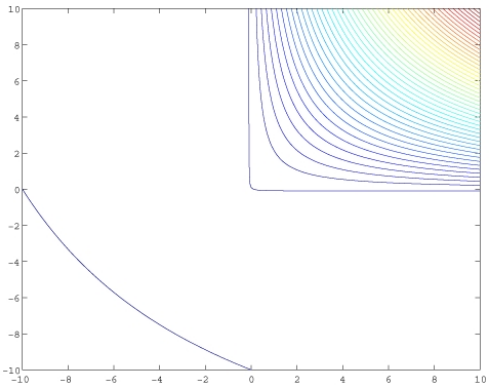
(b) $\alpha = 0.75$



(c) $\alpha = 0.5$



(d) $\alpha = 0.25$



(e) $\alpha = 0.01$

Figura 4.3: Curvas de nível da função de Fischer-Burmeister penalizada

Porém, $\phi_+(a,b) > 0$ se e somente se $a, b > 0$. Neste caso temos $a + b > 0$ e $ab > 0$ o que implica:

$$a + b = \sqrt{(a+b)^2} = \sqrt{a^2 + b^2 + 2ab} > \sqrt{a^2 + b^2},$$

ou seja: $\phi_1(a,b) > 0$.

Logo, como $v(\phi_1) = N_0 \subset v(\phi_+)$ temos:

$$\phi_\alpha(a,b) = 0 \Longleftrightarrow \phi_1(a,b) = 0 \text{ e } \phi_+(a,b) = 0 \Longleftrightarrow (a,b) \in N_0.$$

- (c): Fora de N_0, não há nenhum problema de diferenciabilidade e por teoremas clássicos obtemos facilmente a expressão do gradiente de ϕ_α.
- (b): Já sabemos que ϕ_1 é localmente Lipschitziana. Também, é imediato ver que a função $u \longmapsto u_+$ é 1-Lipschitziana e localmente limitada. Basta então provar que ϕ_+ também é localmente Lipschitziana.

  De fato, para qualquer $(a,b), (a',b')$ em $\mathbb{R}^2$:

$$\begin{aligned} |\phi_+(a,b) - \phi_+(a',b')| &= |a_+b_+ - a'_+b'_+| \\ &= |a_+(b_+ - b'_+) + b'_+(a_+ - a'_+)| \\ &\leq a_+|b_+ - b'_+| + b'_+|a_+ - a'_+| \\ |\phi_+(a,b) - \phi_+(a',b')| &\leq a_+|b - b'| + b'_+|a - a'|. \end{aligned} \tag{4.14}$$

  Portanto, se fixarmos (a,b), então para (a',b') em uma vizinhança de (a,b) temos:

$$|\phi_+(a,b) - \phi_+(a',b')| \leq K\|(a,b) - (a',b')\|_1$$

  e ϕ_+ é localmente Lipschitziana.

Se $\alpha \neq 1$, ϕ_α não é globalmente Lipschitziana pois pela expressão de $\nabla\phi_\alpha$, esse não é limitado. Com efeito, pela presença dos termos $(1-\alpha)b$ e $(1-\alpha)a$ nas duas componentes de $\nabla\phi_\alpha$ temos:

$$\lim_{+\infty} \|\nabla\phi_\alpha\| = +\infty.$$

■

Propriedade 6 (Gradiente generalizado)
O Jacobiano generalizado de ϕ_α para (a,b) em N_0 é:

$$\partial\phi_\alpha(a,b) = \alpha\partial\phi_1(a,b) + (1-\alpha)(b_+\partial a_+, a_+\partial b_+) \tag{4.15}$$

onde $\partial u_+ = \begin{cases} 1 & \text{se } u > 0 \\ 0 & \text{se } u < 0 \\ [0,1] & \text{se } u = 0 \end{cases}$ *é o gradiente generalizado da função $u \longmapsto u_+$ de $\mathbb{R}$ em $\mathbb{R}$.*

Demonstração
Seja (a, b) em N_0.
Separamos o estudo em dois casos:

- $(a, b) \neq (0, 0)$: Então ϕ_1 é C^1 e portanto estritamente diferenciável em (a, b). Logo, pela proposição 2.11 temos:
$$\partial\phi_\alpha(a, b) = \alpha\partial\phi_1(a, b) + (1 - \alpha)\partial\phi_+(a, b).$$
É fácil ver que o gradiente generalizado da função $u \longmapsto u_+$ é dado por ∂u_+.
Logo, o gradiente generalizado da função $\beta : (u, v) \longmapsto u_+$ é:
$$\partial\beta(u, v) = (\partial u_+, 0)$$
Também, é claro que β é regular e positiva ou nula. Portanto, pela proposição 2.12 sobre a diferenciação de um produto, temos:
$$\partial\phi_+(a, b) = b_+(\partial a_+, 0) \ + \ a_+(0, \partial b_+) = (b_+\partial a_+, a_+\partial b_+).$$
Obtemos então (4.15):
$$\partial\phi_\alpha(a, b) = \alpha\partial\phi_1(a, b) + (1 - \alpha)(b_+\partial a_+, a_+\partial b_+).$$
- $(a, b) = (0, 0)$: Esse caso é similar ao anterior, no sentido que ϕ_+ é surpreendentemente estritamente diferenciável em $(0, 0)$, de diferencial nula.
Com efeito, segundo a desigualdade 4.14, para x, y em $\mathbb{R}^2$, temos:
$$|\phi_+(x) - \phi_+(y)| \leq \epsilon(x, y)\|x - y\|,$$
onde $\lim_{x,y\to 0} \epsilon(x, y) = 0$.
Logo, pela proposição 2.9:
$$\partial\phi_+(0, 0) = \{(0, 0)\}.$$
Pela proposição 2.11 temos então:
$$\partial\phi_\alpha(0, 0) = \alpha\partial\phi_1(0, 0),$$
e a expressão (4.15) continua a ser válida neste caso.

■

Propriedade 7 (Forte semi-suavidade)
A função ϕ_α é fortemente semi-suave.

Demonstração
É claro que ϕ_+ é fortemente semi-suave como produto de funções fortemente semi-suaves, pois C^1 por partes. Logo, já que sabemos que ϕ_1 é fortemente semi-suave, ϕ_α é fortemente semi-suave como combinação linear de funções fortemente semi-suaves. ■

Escolha de um elemento em $\partial_B \Phi_\alpha(z)$

No caso da função de Fischer-Burmeister penalizada, o cálculo de um elemento de $\partial_B \Phi_\alpha(z)$ torna-se bem mais delicado do que para as funções NCP precedentes. A dificuldade vem da própria forma do conjunto Ω_{ϕ_α}.

No caso de ϕ_* e ϕ_1, a partir de um ponto z de não diferenciabilidade, era possível sair do conjunto Ω_ϕ somente mexendo com a componente μ de z. Como cada μ_i tem uma ação independente sobre cada componente de Φ, o estudo da convergência podia ser feito componente por componente, e era fácil definir uma seqüência $z_s \longrightarrow z$ com as propriedades desejadas.

O conjunto Ω_{ϕ_α} é constituído pelos dois semi-eixos positivos de $\mathbb{R}^2$, ou seja exatamente N_0. Logo o conjunto de índices das componentes que podem ser não diferenciáveis em um ponto $z = (x, \lambda, \mu)$ de Ω_{Φ_α} é:

$$\beta_{\Phi_\alpha}(z) = \{1 \leq i \leq q \mid g_i(x) = 0 \text{ e } \mu_i \geq 0, \text{ ou } g_i(x) \geq 0 \text{ e } \mu_i = 0\}$$

Assim, para sair do conjunto Ω_{ϕ_α} a partir de um ponto tal que $g_{i_0}(x) = 0$ e $\mu_{i_0} > 0$ (para um certo $1 \leq i_0 \leq q$), é preciso fazer um movimento na componente x. Porém, essa influi sobre todas as q últimas componentes de H_α através das funções g_i. Assim, sem conhecer as g_i, não temos certeza que a seqüência construída fique fora de Ω_{Φ_α}.

Por isto, neste caso, vamos usar o Algoritmo 1 com $\mathcal{A} = \partial_C H$. Com efeito, o cálculo de um elemento de $\partial_C \Phi_\alpha$ é bem mais simples, pois pode ser efetuado componente por componente, sem nos preocuparmos com as correlações entre as g_i.

De fato, tem-se que lembrar que quando se trata de mostrar que V pertence ao B-diferencial de Φ, temos que mostrar que a seqüência $\Phi'(z_s)$ converge a V. Ao contrário, quando se trata do C-subdiferencial, podemos escolher uma seqüência diferente para cada componente, de tal maneira que cada componente pode ser estudada separadamente.

Definimos então os coeficientes (γ_i) e (ζ_i) da seguinte maneira:

- Se $i \notin \beta_{\phi_\alpha}(z)$:

$$\gamma_i = \alpha\left(1 - \frac{g_i(x)}{\sqrt{g_i(x)^2 + \mu_i^2}}\right) + (1-\alpha)\mu_i \quad \text{e} \quad \zeta_i = \alpha\left(1 - \frac{\mu_i}{\sqrt{g_i(x)^2 + \mu_i^2}}\right) + (1-\alpha)g_i(x).$$

- Se $i \in \beta_{\phi_\alpha}(z)$:

  - se $g_i(x) = \mu_i = 0$:

$$\gamma_i = \alpha \quad \text{e} \quad \zeta_i = 0; \tag{4.16}$$

– se $g_i(x) > 0$ e $\mu_i = 0$:

$$\gamma_i = \alpha\left(1 - \frac{g_i(x)}{\sqrt{g_i(x)^2 + \mu_i^2}}\right) \quad \text{e} \quad \zeta_i = \alpha\left(1 - \frac{\mu_i}{\sqrt{g_i(x)^2 + \mu_i^2}}\right) + (1-\alpha)g_i(x); \tag{4.17}$$

– se $g_i(x) = 0$ e $\mu_i > 0$:

$$\gamma_i = \alpha\left(1 - \frac{g_i(x)}{\sqrt{g_i(x)^2 + \mu_i^2}}\right) + (1-\alpha)\mu_i \quad \text{e} \quad \zeta_i = \alpha\left(1 - \frac{\mu_i}{\sqrt{g_i(x)^2 + \mu_i^2}}\right). \tag{4.18}$$

Proposição 4.6
Com (γ_i) e (ζ_i) definidos segundo a regra acima, a matriz V em (4.8) pertence a $\partial_C \Phi_\alpha(z)$.

Demonstração
Seja z em Ω_{Φ_α} e seja V definida como em (4.8), com (γ_i) e (ζ_i) como acima.
Dessa vez queremos mostrar que V pertence ao C-subdiferencial de Φ_α. Ou seja que:

$$V_i \in \partial(\Phi_\alpha)_i(z).$$

- Para $i \notin \beta_{\Phi_\alpha}(z)$: (γ_i) e (ζ_i) são as derivadas parciais de ϕ_α nos pontos $(g_i(x), \mu_i)$ como especificado em (4.9) na Subseção 4.1.4. Logo, para qualquer seqüência (z_s) que seja escolhida fora de Ω_{Φ_α}, a convergência das i-ésima linha de $(\Phi_\alpha(z_s))$ a V_i é automática, e temos:

$$V_i \in \partial_B(\Phi_\alpha)_i(z) \subset \partial(\Phi_\alpha)_i(z).$$

- Para $i \in \beta_{\Phi_\alpha}(z)$:

  - Se $g_i(x) = \mu_i = 0$: Sabemos que ϕ_+ é estritamente diferenciável em $(0,0)$, de diferencial nula. Logo, pelo mesmo raciocínio anterior, temos:

$$\partial(\Phi_\alpha)_i(z) = \alpha\partial(\Phi_1)_i(z),$$

onde:

$$(\Phi_1)_i(z) = \phi_1(g_i(x), \mu_i).$$

Observemos então que $(\Phi_1)_i$ é a composição de ϕ_1 com uma função de classe C^1, que é por isto estritamente diferenciável.

Porém, ϕ_1 é regular (ou quase-diferenciável) na origem, pois para v em $\mathbb{R}^2$:

$$\begin{aligned}
\phi_1^\circ(0;v) &= \max\{\langle \xi, v\rangle \ ; \ \xi \in \partial\phi_1(0)\} \\
&= \max\{\langle (1,1) - \xi, v\rangle \ ; \ \|\xi\| = 1\} \\
&= v_1 + v_2 - \max\{\langle \xi, v\rangle \ ; \ \|\xi\| = 1\} \\
&= v_1 + v_2 - \langle \frac{v}{\|v\|}, v\rangle \\
&= v_1 + v_2 - \|v\| \\
&= \phi_1'(0;v).
\end{aligned}$$

Logo, pela proposição 2.13 temos:

$$\partial(\Phi_1)_i(z) = \partial\phi_1(0) \circ (g_i'(x), 0, e_i).$$

Como:

$$(1,0) \in \partial\phi_1(0),$$

então os coeficientes definidos em (4.16) são tais que:

$$V_i \in \partial(\Phi_\alpha)_i(z).$$

- Se $g_i(x) > 0$ e $\mu_i = 0$:
  Como $g_i(x) > 0$, então a derivada parcial $\frac{\partial\phi_+}{\partial a}$ no ponto $(g_i(x), \mu_i)$ existe e é contínua.
  Já que ϕ_1 é diferenciável nesse ponto, resulta das considerações da Subseção 4.1.4 que o coeficiente γ_i é necessariamente como definido em (4.17).

  Seja a seguinte seqüência $z_s = (x^s, \lambda^s, \mu^s)$:

  $$x^s \equiv_s x, \qquad \lambda^s \equiv_s \lambda, \qquad \mu^s = \mu + \varepsilon_s e_i,$$

  onde $\varepsilon_s \longrightarrow 0^+$.
  Então, para $s \geq 0$, temos:

  $$g_i(x^s) \equiv_s g_i(x) > 0 \qquad \mu_i^s = \varepsilon_s > 0.$$

  Por isto, $(\Phi_\alpha)_i$ é diferenciável em todos pontos de (z_s), z_s converge a z, e:

  $$\frac{\partial\phi_\alpha}{\partial b}(g_i(x^s), \mu_i^s) = \alpha\left(1 - \frac{\mu_i^s}{\sqrt{g_i(x^s)^2 + \mu_i^{s\,2}}}\right) + (1-\alpha)g_i(x^s) \longrightarrow \zeta_i$$

  Portanto, V_i pertence a $\partial(\Phi_\alpha)_i(z)$.
- Se $g_i(x) = 0$ e $\mu_i > 0$:
  Como $\mu_i > 0$, então a derivada parcial $\frac{\partial\phi_+}{\partial b}$ no ponto $(g_i(x), \mu_i)$ existe e é contínua.
  Já que ϕ_1 é diferenciável nesse ponto, resulta das considerações da Subseção 4.1.4 que o coeficiente ζ_i é necessariamente como definido em (4.18).

Ainda segundo a Subseção 4.1.4, se $g_i'(x) = 0$, então o valor de γ_i não importa e para qualquer escolha temos:

$$V_i \in \partial(\Phi_\alpha)_i(z).$$

Por isto, podemos supor que $g_i'(x) \neq 0$.
Seja então a seguinte seqüência $z_s = (x^s, \lambda^s, \mu^s)$:

$$x^s = x + \varepsilon_s \nabla g_i(x), \qquad \lambda^s \equiv_s \lambda, \qquad \mu^s \equiv_s \mu,$$

onde $\varepsilon_s \longrightarrow 0^+$.
Então, como g_i é de classe C^1, temos pelo teorema do valor intermediário vetorial:

$$g_i(x^s) = g_i(x) + \varepsilon_s \nabla g_i(y^s)^T \nabla g_i(x) = \varepsilon_s \nabla g_i(y^s)^T \nabla g_i(x),$$

onde $y^s \in]x, x^s[$.
Pela continuidade de ∇g_i, temos $\nabla g_i(y^s) \longrightarrow \nabla g_i(x)$, e para s grande suficiente:

$$\nabla g_i(y^s)^T \nabla g_i(x) > 0.$$

Por isto, para s grande suficiente:

$$g_i(x^s) > 0.$$

Logo, para s grande suficiente, a seqüência (z_s) fica fora de Ω_{ϕ_α}, converge a z, e temos afinal o desejado:

$$\frac{\partial \phi_\alpha}{\partial a}(g_i(x^s), \mu_i^s) = \alpha \left(1 - \frac{g_i(x^s)}{\sqrt{g_i(x^s)^2 + {\mu_i}^2}}\right) + (1-\alpha)\mu_i \longrightarrow \gamma_i.$$

■

Capítulo 5

Resultados numéricos

Neste capítulo, apresentamos os resultados dos testes efetuados em *MATLAB* . Esses dão uma idéia do comportamento do algoritmo descrito no capítulo 3 na resolução de diferentes sistemas KKT concretos. Para cada um desses problemas, o algoritmo é aplicado aos três sistemas baseados nas funções NCP do mínimo, de Fischer-Burmeister e de Fischer-Burmeister penalizada, para a qual são testados diferentes valores do parâmetro α.

Também, para cada problema são testados vários pontos iniciais. Achamos interessante – e honesto – incluir a integralidade dos resultados numéricos, mesmo para os problemas para os quais o algoritmo convergiu em poucos casos. Por isso, a Seção 5.3 é rica em tabelas e números.

Assim como em vários artigos que serviram de base a esse trabalho, os problemas testes são extraídos da biblioteca *MCPLIB* e da coleção de problemas de minimização de Hock e Schittkowski [20], que chamaremos de *problemas HS*. Globalmente, selecionamos os problemas que aparecem simultaneamente em [3], [6], [7] e [10]. Além desses, testamos dois problemas VIP extraídos de [14] e referenciados em [10].

No caso de um problema NCP, apresentamos os resultados obtidos aplicando o algoritmo a duas reformulações. A primeira é a reformulação do sistema KKT equivalente (1.4), a segunda é a reformulação semi-suave mais usual para esse tipo de problema:

$$\begin{array}{l} x, F(x) \geq 0 \\ \langle x, F(x) \rangle = 0 \end{array} \quad \Leftrightarrow \quad H_\phi(x) = \begin{pmatrix} \phi(F_1(x), x_1) \\ \vdots \\ \phi(F_n(x), x_n) \end{pmatrix} = 0 \tag{5.1}$$

Chamaremos a primeira de *formulação KKT*, a segunda de *formulação NCP*.

O material relativo a esses dois conjuntos de problemas está disponível na internet nos

seguintes endereços[1]:

- M-files, DLL e problemas da *MCPLIB* :
  http://www.gams.com/mpec/matlab.htm e http://www.gams.com/mpec/mcplib.htm
- Problemas HS:
  http://www.uni-bayreuth.de/departments/math/kschittkowski/downloads.htm

Enquanto a biblioteca *MCPLIB* fornece funções de acesso a partir de *MATLAB* , os problemas HS são implementados em *FORTRAN*; Reprogramamos estes em *MATLAB* .

Usamos também o M-arquivo `cpstart.m` escrito por M.C. Ferris, que fornece pontos iniciais para cada problema da *MCPLIB* , diferentes do fornecido pelas rotinas-padrão. Este arquivo está disponível via FTP anônimo, no site da Universidade do Wisconsin-Madison ao endereço ftp://ftp.cs.wisc.edu/math-prog/matlab.

5.1 Detalhes e parâmetros da implementação

Detalhamos aqui alguns pontos técnicos importantes para a implementação. Em particular explicitamos o processo de construção da seqüência $(\mathcal{W}_k)$ na qual se baseia a busca linear não monótona.

5.1.1 Busca linear não monótona

Implementamos a busca linear não monótona descrita em [7], que define a seqüência $(\mathcal{W}_k)$ da seguinte maneira.

Seja um parâmetro $l \geq 1$ fixo. Para $k \geq 0$:

$$\mathcal{W}_k = \max\{\theta(z_j);\ k - l_k + 1 \leq j \leq k\}$$

onde a seqüência (l_k) é construída segundo o seguinte processo:

1. $l_k = 1$ para $1 \leq k \leq 4$.
2. $l_k = 1$ toda vez que a direção de busca é a do gradiente, ou seja quando $d_k = -\nabla\theta(x_k)$.
3. $l_k = \min(l_{k-1} + 1, l)$ em qualquer outro caso.

Observamos que nos quatro primeiros passos e toda vez que a direção do gradiente é escolhida, teremos $\mathcal{W}_k = \theta(x_k)$. Logo, nesses casos, a busca linear tem a forma da busca padrão de Armijo.

[1] Acessados em janeiro 2005

5.1.2 Modificações para a implementação

Seguindo o trabalho de Facchinei, Fischer e Kanzow [7], efetuamos algumas modificações na implementação do Algoritmo 2. Essas permitem melhorar e estabilizar o processo iterativo.

Primeiro introduzimos um teste de descida anterior à busca linear não monótona. Se a seguinte condição se verifica:

$$\theta(z_k + d_k) \leq \gamma\theta(z_k), \qquad \gamma \in]0, 1[\tag{5.2}$$

então escolhemos o passo $\alpha_k = 1$, e não efetuamos nenhuma busca.

Facchinei, Fischer e Kanzow [7] observam que esse teste se justifica pelo fato que já sabemos que o valor otimal é 0, e que a sua introdução não modifica as propriedades de convergência do algoritmo.

Uma segunda modificação consiste em considerar que a equação de Newton (3.14) é resolúvel somente se o número de condição de V_k ficar limitado por uma certa constante τ. De fato, a equação (3.14) pode ser resolúvel mesmo no caso onde V_k é singular. No entanto, por razões de estabilidade numérica, se V_k for tal que:

$$\text{cond}(V_k) > \tau$$

então escolhemos a direção do gradiente como direção de busca.

A terceira modificação vem do fato de que as funções da *MCPLIB* não são todas definidas sobre todo $\mathbb{R}^n$. Logo, um processo de viabilização do ponto corrente por *backtracking* é implementado previamente à busca linear. Mais precisamente, começamos a eventual busca com o primeiro inteiro $j_k \geq 0$ tal que exista

$$F(x_k + 2^{-j_k} d_k).$$

Quando esse processo foi usado durante a resolução de um problema, aparece um "D#"na casa de observação da tabela de resultados (# é corresponde ao número de pontos x_k para os quais foi aplicado o processo).

Por fim, para diminuir o número de avaliação de função quando a direção escolhida é a do gradiente, a busca linear (3.17) começa com $i = \max(j_k, \bar{i})$ onde $\bar{i} = \bar{i}(k)$ é determinado da seguinte maneira:

1. Calcular:

$$\overline{\alpha_k} = \frac{2 \ \max(\theta(z_k), 10^{-7})}{\|\nabla\theta(z_k)\|^2}.$$

2. Usar como valor inicial $\bar{i}$ o maior inteiro $j \in \mathbb{N}$ tal que:

$$2^{-j} \leq \overline{\alpha_k}.$$

Quando $\overline{\alpha_k} < 1$ esse é:

$$\bar{i} = \left\lceil -\frac{\ln \overline{\alpha_k}}{\ln 2} \right\rceil.$$

5.1.3 Critérios de parada

O algoritmo pára caso se cumpra um dos seguintes critérios:

- Precisão desejada alcançada: $\theta(z_k)/n \leq \varepsilon$

- Ponto considerado estacionário: $\|\nabla\theta(z_k)\|/\sqrt{n} \leq \varepsilon_g$

- Número de iterações grande demais: $k \geq k_{max}$

- Passo de busca pequeno demais: m_k é tal que $2^{-m_k} \leq \varepsilon_s$

5.1.4 Pontos iniciais

Os pontos iniciais $z_0 = (x_0, \lambda_0, \mu_0)$ são tais que x_0 é dado pelas bibliotecas ou pelo arquivo `cpstart.m`, enquanto os vetores λ_0 e μ_0 são escolhidos nulos. Quando a dimensão do problema o permite, damos explicitamente o valor do ponto x_0.

Para os problemas da *MCPLIB*, o primeiro ponto inicial é o ponto fornecido pela biblioteca de acesso via *MATLAB* (i.e. o valor retornado pela função `mcpinit`). Este é denotado por x_{mcplib}. Os outros pontos são os fornecidos pelo arquivo `cpstart.m`. Além de x_{mcplib}, nos limitamos a cinco pontos iniciais. Estes são denotados x_0^i, onde i corresponde à posição do ponto no arquivo `cpstart.m`.

Notemos que foram eliminados os pontos que eram muito perto de uma solução ou até iguais a uma solução, assim como os pontos próximos um do outro que davam resultados parecidos. Também excluímos os pontos iniciais que davam um erro de domínio na primeira avaliação de F. Tais pontos foram encontrados para os problemas `hanskoop` e `scarfbsum`. Apesar disso, deixamos os resultados obtidos para `hanskoop` nesses pontos problemáticos, pois mesmo assim o algoritmo convergiu, quando ignorava o processo de viabilização. O algoritmo não convergiu para `scarfbsum` nos pontos suspeitos, por isso não deixamos os resultados.

Para os problemas HS e Fukushima, o ponto inicial é o ponto que vem junto com a definição do problema.

5.1.5 Parâmetros e legenda do resultados

Os diferentes valores para cada parâmetro do algoritmo, assim como a legenda das tabelas de resultados são resumidos nas tabelas 5.1 e 5.2.

Parâmetro	Valor	Descrição
ε	10^{-12}	Valor mínimo para a função de mérito escalada $\theta(z_k)/n$
ε_g	10^{-6}	Valor mínimo para a norma do gradiente escalado $\|\nabla\theta(z_k)\|/n$
ε_s	10^{-15}	Valor mínimo para o passo t_k
k_{max}	300	Número máximo de iteração
ρ	10^{-9}	Coeficiente multiplicativo no teste de ângulo (3.15)
p	2.1	Potência no teste de ângulo (3.15)
γ	0.9	Parâmetro para o teste de descida (5.2)
β	10^{-4}	Parâmetro da busca linear (3.17)

Tabela 5.1: Parâmetros da implementação

Item	Descrição
k	número de iterações antes de cumprir um critério de parada
F-eval	número de avaliação da função F
Newt.	número de passos que usam a direção de Newton
Grad.	número de passos que usam a direção do gradiente
BckT.	número de vezes onde a busca linear diminuiu o passo na direção escolhida
ND	número de pontos da seqüência onde a função NCP é não diferenciável
θ_0	$= \theta(z_0)/n$: valor da função de mérito escalada no ponto inicial
θ_f	$= \theta(z_f)/n$: valor final da função de mérito escalada
$\|\nabla\theta_f\|$	$= \|\nabla\theta(z_f)\|/\sqrt{n}$: norma final escalada do gradiente da função de mérito
Obs.	Observações sobre como se terminou a iteração

Tabela 5.2: Legendas das tabelas de resultados dos problemas testes

5.2 Problemas testes

O problema de Fukushima [14] é um problema VIP cuja função F é monótona, enquanto a única restrição – de desigualdade – é côncava.

Os problemas HS representam as condições KKT dos problemas de minimização da coleção [20]. Todos têm uma função objetivo convexa e restrições de desigualdade côncavas, exceto o

problema HS32 que possui uma restrição de igualdade linear e uma restrição de desigualdade não côncava. A restrição de igualdade do problema HS60 é não linear. Assim, a função F associada aos problemas HS34, HS35, HS43, HS76 e HS100 é monótona.

Quanto aos problemas escolhidos na *MCPLIB*, todos são problemas NCP puros.

5.3 Resultados

O trabalho de Chen, Chen e Kanzow [3] sobre a função de Fischer-Burmeister Penalizada mostra resultados entusiasmantes quanto à sua capacidade de resolver os problemas NCP da *MCPLIB* de maneira eficiente, quando se usa a reformulação NCP (5.1). Os nossos resultados vão no mesmo sentido. Também obtemos resultados similares aos de Facchinei, Fischer e Kanzow [7] e [10] sobre a reformulação KKT do problema VIP com a função de Fischer-Burmeister.

No entanto, com a reformulação KKT dos problemas NCP, os resultados são mitigados quanto a decidir se a função de Fischer-Burmeister Penalizada se destaca realmente da sua versão sem penalização. Além dessas considerações de comparação de funções, aparece claramente que a reformulação NCP dá resultados globalmente melhores. Isso pode ser explicado pelo fato de que a reformulação KKT introduz variáveis, dobrando o tamanho do sistema.

Também observamos que a função do mínimo deu resultados bem melhores do que esperávamos. É verdade que na maioria dos casos o algoritmo não converge ou pára por causa da não diferenciabilidade da função de mérito. No entanto, se nenhuma dessas situações acontecer, o Mínimo geralmente ganha das outras funções NCP, tanto em número de iterações quanto em precisão final. Notemos também o seu excelente comportamento com os problemas Fukushima e HS.

Função NCP	**Iterações**	**F-eval**
Mínimo	1.04	1.05
Fischer-Burmeister	1.34	1.47
Ficher-Burmeister penalizada		
$\alpha = 0.95$	1.27	1.4
$\alpha = 0.75$	1.19	1.25
$\alpha = 0.50$	1.26	1.37
$\alpha = 0.20$	1.35	1.53

(a) Formulação KKT

Função NCP	**Iterações**	**F-eval**
Mínimo	1.55	2.16
Fischer-Burmeister	1.34	1.64
Ficher-Burmeister penalizada		
$\alpha = 0.95$	1.35	1.59
$\alpha = 0.75$	1.3	1.46
$\alpha = 0.50$	1.26	1.39
$\alpha = 0.20$	1.2	1.28

(b) Formulação NCP

Tabela 5.3: Resultados agregados

$x_0 = x_{lib} = (1.000000, 1.000000, 1.000000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	5	6	5	0	0	0	1.7e+000	3.2e-016	1.0e-007	
FB	6	7	6	0	0	0	2.2e+000	1.1e-015	1.6e-007	
FBP0.95	6	7	6	0	0	0	2.1e+000	1.0e-015	1.6e-007	
FB0.75	6	7	6	0	0	0	1.9e+000	9.8e-016	1.7e-007	
FB0.5	6	7	6	0	0	0	1.7e+000	9.2e-016	1.8e-007	
FB0.2	6	7	6	0	0	0	1.6e+000	8.9e-016	1.9e-007	

Tabela 5.4: Problema `Fukushima` : $n = 3, \quad p = 0, \quad q = 1$ (formulação KKT)

$x_0 = x_{lib} = (0.100000, 0.700000, 0.200000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	2	3	2	0	0	0	6.5e+001	0	0	
FB	6	7	6	0	0	6	6.5e+001	5.7e-023	1.1e-011	
FBP0.95	6	7	6	0	0	6	6.5e+001	5.2e-023	9.7e-012	
FB0.75	6	7	6	0	0	6	6.5e+001	3.2e-023	6.0e-012	
FB0.5	6	7	6	0	0	6	6.5e+001	1.4e-023	2.7e-012	
FB0.2	5	6	5	0	0	5	6.5e+001	9.9e-013	2.8e-007	

Tabela 5.5: Problema `HS32` : $n = 3, \quad p = 1, \quad q = 4$ (formulação KKT)

$x_0 = x_{lib} = (0.000000, 1.050000, 2.900000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	0	1	0	0	0	0	1.7e-001	1.7e-001	1.0e+000	$\nexists\nabla\theta(x_f)$
FB	19	68	17	2	15	19	1.7e-001	4.4e-015	5.4e-007	
FBP0.95	23	123	21	2	19	23	1.7e-001	1.7e-014	1.6e-006	
FB0.75	90	1070	88	2	86	90	1.7e-001	4.8e-013	1.8e-006	
FB0.5	82	1016	80	2	78	82	1.7e-001	6.0e-013	1.2e-006	
FB0.2	40	425	38	2	35	40	1.7e-001	4.5e-017	1.9e-008	

Tabela 5.6: Problema `HS34` : $n = 3, \quad p = 0, \quad q = 8$ (formulação KKT)

$x_0 = x_{lib} = (0.500000, 0.500000, 0.500000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	2	3	2	0	0	0	4.8e+000	9.1e-032	2.9e-015	
FB	5	6	5	0	0	5	4.8e+000	7.6e-014	9.6e-007	
FBP0.95	5	6	5	0	0	5	4.8e+000	6.9e-014	8.6e-007	
FB0.75	5	6	5	0	0	5	4.8e+000	4.3e-014	5.4e-007	
FB0.5	5	6	5	0	0	5	4.8e+000	1.9e-014	2.4e-007	
FB0.2	5	6	5	0	0	5	4.8e+000	3.1e-015	3.8e-008	

Tabela 5.7: Problema `HS35` : $n = 3, \quad p = 0, \quad q = 4$ (formulação KKT)

$x_0 = x_{lib} = (0.000000, 0.000000, 0.000000, 0.000000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	8	10	8	0	1	0	6.8e+001	4.5e-024	2.8e-011	
FB	9	16	9	0	3	3	6.8e+001	4.7e-017	7.2e-008	
FBP0.95	9	16	9	0	3	3	6.8e+001	4.5e-017	7.2e-008	
FB0.75	9	16	9	0	3	3	6.8e+001	4.0e-017	7.1e-008	
FB0.5	9	11	9	0	1	1	6.8e+001	5.2e-024	2.7e-011	
FB0.2	9	11	9	0	1	1	6.8e+001	4.7e-024	2.7e-011	

Tabela 5.8: Problema `HS43` : $n = 4, \quad p = 0, \quad q = 3$ (formulação KKT)

$x_0 = x_{lib} = (2.000000, 2.000000, 2.000000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	6	7	6	0	0	0	5.3e+001	1.9e-014	2.1e-006	
FB	6	7	6	0	0	6	5.3e+001	1.9e-014	2.1e-006	
FBP0.95	6	7	6	0	0	6	5.3e+001	1.9e-014	2.1e-006	
FB0.75	6	7	6	0	0	6	5.3e+001	1.9e-014	2.1e-006	
FB0.5	6	7	6	0	0	6	5.3e+001	1.9e-014	2.1e-006	
FB0.2	6	7	6	0	0	6	5.3e+001	1.9e-014	2.1e-006	

Tabela 5.9: Problema `HS60` : $n = 3, \quad p = 1, \quad q = 6$ (formulação KKT)

$x_0 = x_{lib} = (0.500000, 0.500000, 0.500000, 0.500000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	4	8	4	0	2	0	1.3e+000	1.1e-031	1.3e-015	
FB	9	15	9	0	3	9	1.3e+000	7.4e-022	3.8e-011	
FBP0.95	9	15	9	0	3	9	1.3e+000	6.7e-022	3.5e-011	
FB0.75	9	15	9	0	3	9	1.3e+000	4.2e-022	2.2e-011	
FB0.5	7	10	7	0	1	7	1.3e+000	4.4e-014	1.5e-007	
FB0.2	7	9	7	0	1	7	1.3e+000	7.0e-015	2.4e-008	

Tabela 5.10: Problema `HS76` : $n = 4, \quad p = 0, \quad q = 7$ (formulação KKT)

$x_0 = x_{lib}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	9	18	7	2	3	0	8.7e+002	5.4e-014	3.9e-006	
FB	9	21	8	1	4	1	8.7e+002	1.8e-015	1.1e-006	
FBP0.95	8	17	7	1	3	1	8.7e+002	2.3e-017	5.0e-007	
FB0.75	8	17	7	1	3	1	8.7e+002	6.6e-017	1.5e-007	
FB0.5	8	17	7	1	3	1	8.7e+002	4.5e-017	8.2e-008	
FB0.2	10	18	9	1	2	1	8.7e+002	7.0e-022	2.1e-010	

Tabela 5.11: Problema `HS100` : $n = 7, \quad p = 0, \quad q = 4$ (formulação KKT)

$x_0 = x_{mcplib}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	0	1	0	0	0	0	3.2e+002	3.2e+002	2.2e+003	$\nexists\nabla\theta(x_f)$
FB	8	15	7	1	2	1	3.2e+002	3.3e-027	1.7e-012	
FBP0.95	8	15	7	1	2	1	3.2e+002	3.0e-027	1.9e-012	
FB0.75	8	15	7	1	2	1	3.2e+002	1.9e-027	1.9e-012	
FB0.5	8	15	7	1	2	1	3.2e+002	8.7e-028	1.5e-012	
FB0.2	8	15	7	1	2	1	3.2e+002	2.1e-028	1.6e-012	

$x_0 = x^1_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	3879	1	299	299	1	3.4e+002	9.1e-003	2.7e-002	$k \geq k_{max}$
FB	42	324	41	1	35	22	3.4e+002	6.2e-017	1.1e-008	
FBP0.95	44	336	43	1	37	22	3.4e+002	5.6e-017	1.0e-008	
FB0.75	300	3736	20	280	297	18	3.4e+002	7.6e-003	2.6e-001	$k \geq k_{max}$
FB0.5	300	3817	9	291	297	10	3.4e+002	3.8e-003	7.4e-002	$k \geq k_{max}$
FB0.2	300	3861	4	296	296	6	3.4e+002	7.2e-004	6.5e-004	$k \geq k_{max}$

$x_0 = x^2_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	5	6	5	0	0	0	5.0e+003	7.5e-013	1.5e-004	
FB	8	11	8	0	2	8	5.0e+003	8.1e-013	1.4e-004	
FBP0.95	8	11	8	0	2	5	5.0e+003	8.1e-013	1.4e-004	
FB0.75	8	11	8	0	2	8	5.0e+003	8.1e-013	1.4e-004	
FB0.5	8	11	8	0	2	5	5.0e+003	8.1e-013	1.4e-004	
FB0.2	12	25	12	0	6	12	5.0e+003	5.5e-017	1.2e-006	

$x_0 = x^3_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	14	19	11	3	4	0	1.5e+012	8.9e-017	1.7e-006	
FB	18	23	15	3	4	1	1.5e+012	4.0e-022	3.2e-009	
FBP0.95	18	23	15	3	4	1	1.5e+012	4.0e-022	3.2e-009	
FB0.75	18	23	15	3	4	1	1.5e+012	4.0e-022	3.2e-009	
FB0.5	18	23	15	3	4	1	1.5e+012	4.0e-022	3.2e-009	
FB0.2	300	2938	250	50	285	2	1.5e+012	2.7e+002	1.9e+002	$k \geq k_{max}$

$x_0 = x^4_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	3879	1	299	299	1	3.4e+002	9.1e-003	2.7e-002	$k \geq k_{max}$
FB	42	324	41	1	35	22	3.4e+002	6.2e-017	1.1e-008	
FBP0.95	44	336	43	1	37	22	3.4e+002	5.6e-017	1.0e-008	
FB0.75	300	3736	20	280	297	18	3.4e+002	7.6e-003	2.6e-001	$k \geq k_{max}$
FB0.5	300	3817	9	291	297	10	3.4e+002	3.8e-003	7.4e-002	$k \geq k_{max}$
FB0.2	300	3861	4	296	296	6	3.4e+002	7.2e-004	6.5e-004	$k \geq k_{max}$

$x_0 = x^6_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	2988	2	298	298	1	4.3e+003	1.4e+001	1.3e+001	$k \geq k_{max}$
FB	13	18	13	0	3	13	4.3e+003	6.2e-017	1.1e-008	
FBP0.95	13	18	13	0	3	13	4.3e+003	5.6e-017	1.0e-008	
FB0.75	13	18	13	0	3	13	4.3e+003	3.5e-017	6.3e-009	
FB0.5	13	18	13	0	3	13	4.3e+003	1.5e-017	2.8e-009	
FB0.2	12	17	12	0	3	12	4.3e+003	1.6e-012	4.5e-007	$\|\nabla\theta_f\| \leq \varepsilon_g$

Tabela 5.12: Problema `bertsekas` : $n = 15, \quad p = 0, \quad q = 15$ (formulação KKT)

$x_0 = x_{mcplib}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	19	39	0	19	19	0	1.8e-002	6.1e-003	5.2e-007	$\|\nabla\theta_f\| \leq \varepsilon_g$
FB	35	362	35	0	29	0	2.6e-002	7.9e-018	2.5e-007	
FBP0.95	38	378	38	0	32	0	2.4e-002	3.4e-017	6.8e-007	
FB0.75	27	153	26	1	21	0	1.5e-002	3.8e-014	1.5e-005	
FB0.5	34	221	33	1	28	0	6.6e-003	1.1e-018	4.6e-008	
FB0.2	45	344	39	6	39	0	1.0e-003	1.6e-015	7.0e-007	

$x_0 = x^1_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	1	3	0	1	1	0	1.8e-002	9.2e-003	1.4e-001	$\nexists\nabla\theta(x_f)$
FB	36	331	36	0	29	0	7.3e-002	5.2e-013	6.3e-005	
FBP0.95	37	343	37	0	30	0	6.6e-002	2.7e-017	6.1e-007	
FB0.75	38	325	37	1	30	0	4.1e-002	1.3e-018	1.1e-007	
FB0.5	37	322	36	1	32	0	1.8e-002	1.2e-013	2.1e-005	
FB0.2	36	229	35	1	29	0	2.9e-003	5.1e-016	4.0e-007	

$x_0 = x^2_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	0	1	0	0	0	0	1.3e-001	1.3e-001	5.8e-001	$\nexists\nabla\theta(x_f)$
FB	34	295	34	0	27	0	1.0e-001	2.4e-018	1.4e-007	
FBP0.95	34	195	34	0	28	0	4.1e+000	1.1e-014	7.0e-006	
FB0.75	23	79	23	0	15	0	8.1e+001	9.6e-018	2.0e-007	
FB0.5	14	22	14	0	5	0	3.2e+002	6.1e-018	1.1e-007	
FB0.2	12	16	12	0	3	0	8.0e+002	5.0e-013	1.3e-005	

$x_0 = x^3_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	2	4	1	1	1	0	5.0e+003	9.2e-003	1.4e-001	$\nexists\nabla\theta(x_f)$
FB	37	322	37	0	29	0	4.3e+003	5.2e-013	6.3e-005	
FBP0.95	35	125	35	0	13	0	3.7e+013	2.1e-016	1.7e-006	
FB0.75	29	43	29	0	3	0	9.3e+014	1.7e-018	8.6e-008	
FB0.5	27	33	27	0	2	0	3.7e+015	1.7e-014	5.7e-006	
FB0.2	24	25	24	0	0	0	9.5e+015	5.2e-015	1.0e-006	

$x_0 = x^4_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	1	3	0	1	1	0	1.8e-002	9.2e-003	1.4e-001	$\nexists\nabla\theta(x_f)$
FB	36	331	36	0	29	0	7.3e-002	5.2e-013	6.3e-005	
FBP0.95	37	343	37	0	30	0	6.6e-002	2.7e-017	6.1e-007	
FB0.75	38	325	37	1	30	0	4.1e-002	1.3e-018	1.1e-007	
FB0.5	37	322	36	1	32	0	1.8e-002	1.2e-013	2.1e-005	
FB0.2	36	229	35	1	29	0	2.9e-003	5.1e-016	4.0e-007	

$x_0 = x^6_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	2	4	1	1	1	0	1.7e-001	9.2e-003	1.4e-001	$\nexists\nabla\theta(x_f)$
FB	35	295	35	0	27	0	1.7e-001	5.2e-013	6.4e-005	
FBP0.95	32	182	32	0	23	0	1.3e+001	1.9e-017	3.4e-007	
FB0.75	28	99	28	0	19	0	2.6e+002	1.0e-013	2.1e-005	
FB0.5	22	60	22	0	13	0	1.0e+003	2.4e-015	2.2e-006	
FB0.2	17	41	17	0	9	0	2.6e+003	1.6e-014	2.2e-006	

Tabela 5.13: Problema `bertsekas` : $n = 15$ (formulação NCP)

$x_0 = x_{mcplib} = (0.020000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	1529	300	0	298	0	1.2e-003	5.8e-006	3.4e-003	$k \geq k_{max}$
FB	300	5501	300	0	299	1	1.2e-003	5.6e-005	2.5e-002	$k \geq k_{max}$
FBP0.95	300	6260	299	1	299	1	1.2e-003	4.2e-005	1.9e-002	$k \geq k_{max}$
FB0.75	300	5879	300	0	299	1	1.2e-003	3.2e-005	1.5e-002	$k \geq k_{max}$
FB0.5	300	5397	299	1	299	1	1.2e-003	1.2e-005	4.5e-003	$k \geq k_{max}$
FB0.2	300	4008	299	1	299	1	1.2e-003	1.9e-006	3.5e-005	$k \geq k_{max}$

$x_0 = x^1_{cpstart} = (0.000000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	1771	300	0	299	1	5.0e-005	5.7e-006	3.4e-003	$k \geq k_{max}$
FB	300	4744	295	5	300	1	5.0e-005	2.5e-005	1.8e-005	$k \geq k_{max}$
FBP0.95	300	4847	296	4	300	1	5.0e-005	2.4e-005	2.2e-005	$k \geq k_{max}$
FB0.75	300	4043	295	5	299	1	5.0e-005	1.9e-005	4.8e-003	$k \geq k_{max}$
FB0.5	300	4627	296	4	299	1	5.0e-005	1.0e-005	1.5e-004	$k \geq k_{max}$
FB0.2	120	1740	115	5	118	1	5.0e-005	1.9e-006	3.3e-007	$\|\nabla\theta_f\| \leq \varepsilon_g$

$x_0 = x^2_{cpstart} = (0.010000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	1529	300	0	298	0	4.5e-004	5.8e-006	3.4e-003	$k \geq k_{max}$
FB	300	3816	292	8	299	1	4.5e-004	2.5e-005	2.5e-003	$k \geq k_{max}$
FBP0.95	300	3711	293	7	299	1	4.5e-004	2.4e-005	1.6e-003	$k \geq k_{max}$
FB0.75	300	4111	294	6	299	1	4.5e-004	1.8e-005	2.4e-003	$k \geq k_{max}$
FB0.5	300	4713	296	4	299	1	4.5e-004	1.0e-005	9.9e-005	$k \geq k_{max}$
FB0.2	199	3073	194	5	198	1	4.5e-004	1.9e-006	9.4e-007	$\|\nabla\theta_f\| \leq \varepsilon_g$

Tabela 5.14: Problema `billups` : $n = 1, \quad p = 0, \quad q = 1$ (formulação KKT)

$x_0 = x_{mcplib} = (0.020000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	3372	300	0	299	0	1.2e-003	5.5e-006	6.7e-003	$k \geq k_{max}$
FB	98	1659	98	0	97	0	3.5e-003	5.0e-005	4.0e-007	$\|\nabla\theta_f\| \leq \varepsilon_g$
FBP0.95	98	1659	98	0	97	0	3.1e-003	4.5e-005	3.6e-007	$\|\nabla\theta_f\| \leq \varepsilon_g$
FB0.75	90	1458	90	0	89	0	1.9e-003	2.8e-005	6.6e-007	$\|\nabla\theta_f\| \leq \varepsilon_g$
FB0.5	86	1360	86	0	85	0	8.6e-004	1.2e-005	9.9e-007	$\|\nabla\theta_f\| \leq \varepsilon_g$
FB0.2	63	886	63	0	62	0	1.4e-004	2.0e-006	8.0e-007	$\|\nabla\theta_f\| \leq \varepsilon_g$

$x_0 = x^1_{cpstart} = (0.000000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	1529	300	0	298	0	5.0e-005	5.8e-006	6.8e-003	$k \geq k_{max}$
FB	39	850	39	0	38	0	2.0e-004	5.0e-005	7.2e-007	$\|\nabla\theta_f\| \leq \varepsilon_g$
FBP0.95	39	850	39	0	38	0	1.8e-004	4.5e-005	6.5e-007	$\|\nabla\theta_f\| \leq \varepsilon_g$
FB0.75	36	773	36	0	35	0	1.1e-004	2.8e-005	3.1e-007	$\|\nabla\theta_f\| \leq \varepsilon_g$
FB0.5	8	122	8	0	7	0	5.0e-005	1.2e-005	9.3e-007	$\|\nabla\theta_f\| \leq \varepsilon_g$
FB0.2	5	54	5	0	4	0	8.0e-006	2.0e-006	2.1e-007	$\|\nabla\theta_f\| \leq \varepsilon_g$

$x_0 = x^2_{cpstart} = (0.010000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	1959	300	0	299	0	4.5e-004	5.7e-006	6.8e-003	$k \geq k_{max}$
FB	109	1715	109	0	108	0	1.3e-003	5.0e-005	6.5e-007	$\|\nabla\theta_f\| \leq \varepsilon_g$
FBP0.95	109	1715	109	0	108	0	1.2e-003	4.5e-005	5.9e-007	$\|\nabla\theta_f\| \leq \varepsilon_g$
FB0.75	94	1341	94	0	93	0	7.4e-004	2.8e-005	8.2e-007	$\|\nabla\theta_f\| \leq \varepsilon_g$
FB0.5	94	1341	94	0	93	0	3.3e-004	1.2e-005	3.6e-007	$\|\nabla\theta_f\| \leq \varepsilon_g$
FB0.2	10	66	10	0	9	0	5.3e-005	2.0e-006	9.6e-007	$\|\nabla\theta_f\| \leq \varepsilon_g$

Tabela 5.15: Problema `billups` : $n = 1$ (formulação NCP)

$x_0 = x_{mcplib}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	1	2	1	0	0	1	2.4e+002	6.9e+001	1.2e+001	$\nexists\nabla\theta(x_f)$
FB	300	2469	282	18	213	4	2.4e+002	5.5e-006	4.4e-002	$k \geq k_{max}$
FBP0.95	300	2316	288	12	213	255	2.4e+002	5.0e-006	7.7e-003	$k \geq k_{max}$
FB0.75	300	2462	267	33	216	203	2.4e+002	3.1e-006	1.8e-002	$k \geq k_{max}$
FB0.5	300	2211	273	27	203	5	2.4e+002	1.4e-006	2.7e-002	$k \geq k_{max}$
FB0.2	12	33	10	2	4	3	2.4e+002	7.9e-017	1.4e-007	

$x_0 = x^1_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	2794	0	300	300	0	3.7e+002	1.4e+002	1.4e+002	$k \geq k_{max}$
FB	298	2607	26	272	288	2	3.7e+002	2.4e-014	1.9e-005	
FBP0.95	300	3616	288	12	278	2	3.7e+002	8.2e-002	5.1e+001	$k \geq k_{max}$
FB0.75	15	63	14	1	7	2	3.7e+002	5.8e-015	2.4e-005	
FB0.5	300	2262	272	28	209	2	3.7e+002	1.4e-006	1.6e-002	$k \geq k_{max}$
FB0.2	300	1958	179	121	209	3	3.7e+002	2.2e-007	4.9e-003	$k \geq k_{max}$

$x_0 = x^2_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	0	1	0	0	0	0	5.3e+002	5.3e+002	4.4e+003	$\nexists\nabla\theta(x_f)$
FB	300	2445	286	14	213	13	5.3e+002	5.5e-006	5.4e-002	$k \geq k_{max}$
FBP0.95	300	2491	275	25	211	4	5.3e+002	4.9e-006	2.4e-002	$k \geq k_{max}$
FB0.75	300	2392	288	12	211	5	5.3e+002	3.1e-006	4.3e-002	$k \geq k_{max}$
FB0.5	300	2255	274	26	213	4	5.3e+002	1.4e-006	2.5e-002	$k \geq k_{max}$
FB0.2	300	1978	177	123	206	4	5.3e+002	2.2e-007	3.2e-003	$k \geq k_{max}$

$x_0 = x^3_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	0	1	0	0	0	0	2.0e+002	2.0e+002	1.7e+003	$\nexists\nabla\theta(x_f)$
FB	6	11	5	1	1	1	2.0e+002	2.5e-016	2.2e-008	
FBP0.95	6	11	5	1	1	1	2.0e+002	2.2e-016	2.0e-008	
FB0.75	6	11	5	1	1	1	2.0e+002	1.4e-016	1.3e-008	
FB0.5	6	11	5	1	1	1	2.0e+002	6.2e-017	5.6e-009	
FB0.2	6	11	5	1	1	1	2.0e+002	1.0e-017	9.2e-010	

Tabela 5.16: Problema `colvdual` : $n = 20, \quad p = 0, \quad q = 20$ (formulação KKT)

$x_0 = x_{mcplib}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	0	1	0	0	0	0	6.9e+001	6.9e+001	2.8e+002	$\nexists\nabla\theta(x_f)$
FB	300	3542	298	2	277	1	2.7e+002	5.6e-004	8.5e-001	$k \geq k_{max}$
FBP0.95	300	4134	298	2	275	1	2.5e+002	1.2e-003	1.3e+000	$k \geq k_{max}$
FB0.75	300	4376	298	2	278	1	1.5e+002	8.1e-004	4.0e-001	$k \geq k_{max}$
FB0.5	300	3142	299	1	288	1	6.9e+001	4.4e-004	6.6e-001	$k \geq k_{max}$
FB0.2	300	4156	290	10	284	1	1.1e+001	7.7e-005	9.6e-002	$k \geq k_{max}$

$x_0 = x^1_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	3628	0	300	300	0	4.1e+001	3.6e+001	3.6e+001	$k \geq k_{max}$
FB	47	141	47	0	28	0	2.2e+001	9.8e-016	4.0e-007	
FBP0.95	14	24	14	0	5	0	5.4e+002	4.4e-016	4.0e-006	
FB0.75	12	13	12	0	0	0	9.6e+003	3.1e-018	3.0e-008	
FB0.5	11	14	11	0	2	0	3.7e+004	5.9e-016	1.4e-007	
FB0.2	9	11	9	0	1	0	9.3e+004	8.3e-013	3.5e-005	

$x_0 = x^2_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	10177	14	286	298	1	3.3e+002	1.0e+000	2.0e+001	$k \geq k_{max}$
FB	300	3959	299	1	270	1	1.3e+003	5.4e-006	2.2e-002	$k \geq k_{max}$
FBP0.95	13	24	13	0	5	1	1.2e+003	8.3e-020	5.2e-008	
FB0.75	12	18	12	0	3	1	7.4e+002	3.7e-017	5.2e-008	
FB0.5	12	17	12	0	3	1	3.3e+002	8.0e-018	3.2e-007	
FB0.2	300	2855	299	1	269	1	6.5e+001	1.2e-004	3.6e-001	$k \geq k_{max}$

$x_0 = x^3_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	2079	0	300	300	0	1.0e+000	3.1e-002	1.5e+000	$k \geq k_{max}$
FB	16	50	16	0	8	0	3.0e+000	3.6e-019	1.1e-007	
FBP0.95	9	20	9	0	3	0	2.9e+000	3.2e-013	5.9e-005	
FB0.75	7	15	7	0	1	0	2.8e+000	9.0e-022	2.4e-009	
FB0.5	6	12	6	0	1	0	4.1e+000	2.0e-016	1.6e-006	
FB0.2	6	9	6	0	1	0	7.6e+000	1.3e-013	1.1e-005	

Tabela 5.17: Problema `colvdual` : $n = 20$ (formulação NCP)

$x_0 = x_{mcplib}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	3	4	3	0	0	1	2.3e+002	3.1e-020	2.1e-008	
FB	5	6	5	0	0	5	2.3e+002	3.3e-015	2.1e-007	
FBP0.95	5	6	5	0	0	5	2.3e+002	3.0e-015	2.0e-007	
FB0.75	5	6	5	0	0	5	2.3e+002	1.9e-015	1.9e-007	
FB0.5	5	6	5	0	0	5	2.3e+002	8.4e-016	1.9e-007	
FB0.2	5	6	5	0	0	5	2.3e+002	1.4e-016	1.9e-007	

$x_0 = x^1_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	3190	2	298	299	0	2.8e+008	1.4e+003	3.7e+002	$k \geq k_{max}$
FB	300	2841	51	249	287	1	2.8e+008	1.5e+003	2.6e+003	$k \geq k_{max}$
FBP0.95	300	2802	38	262	287	1	2.8e+008	1.5e+003	1.2e+003	$k \geq k_{max}$
FB0.75	23	33	22	1	6	1	2.8e+008	8.5e-015	9.8e-008	
FB0.5	300	3014	34	266	294	1	2.8e+008	2.5e+003	1.8e+003	$k \geq k_{max}$
FB0.2	282	2876	276	6	269	1	2.8e+008	9.5e-017	1.6e-007	

$x_0 = x^2_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	8	11	8	0	2	1	2.7e+004	5.1e-027	7.4e-012	
FB	10	11	10	0	0	10	2.7e+004	3.3e-015	2.1e-007	
FBP0.95	10	11	10	0	0	10	2.7e+004	3.0e-015	2.0e-007	
FB0.75	10	12	10	0	1	10	2.7e+004	1.9e-015	1.9e-007	
FB0.5	10	12	10	0	1	10	2.7e+004	8.4e-016	1.9e-007	
FB0.2	10	12	10	0	1	10	2.7e+004	1.4e-016	1.9e-007	

$x_0 = x^4_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	3338	0	300	300	0	4.5e+002	1.5e+002	6.9e+001	$k \geq k_{max}$
FB	42	331	10	32	35	1	4.5e+002	1.4e-015	1.4e-007	
FBP0.95	13	79	11	2	6	1	4.5e+002	1.3e-015	1.3e-007	
FB0.75	10	39	9	1	3	1	4.5e+002	8.1e-016	1.3e-007	
FB0.5	10	35	9	1	3	1	4.5e+002	3.6e-016	1.2e-007	
FB0.2	11	28	10	1	3	1	4.5e+002	1.3e-016	1.9e-007	

$x_0 = x^5_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	3334	0	300	300	0	4.5e+002	1.5e+002	7.0e+001	$k \geq k_{max}$
FB	42	331	10	32	35	1	4.5e+002	1.4e-015	1.4e-007	
FBP0.95	13	79	11	2	6	1	4.5e+002	1.3e-015	1.3e-007	
FB0.75	10	39	9	1	3	1	4.5e+002	8.1e-016	1.3e-007	
FB0.5	10	35	9	1	3	1	4.5e+002	3.6e-016	1.2e-007	
FB0.2	11	28	10	1	3	1	4.5e+002	1.3e-016	1.9e-007	

$x_0 = x^6_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	0	1	0	0	0	0	2.4e+002	2.4e+002	1.1e+003	$\nexists\nabla\theta(x_f)$
FB	16	109	12	4	8	1	2.4e+002	3.2e-015	2.0e-007	
FBP0.95	15	101	12	3	7	1	2.4e+002	2.9e-015	2.0e-007	
FB0.75	13	89	12	1	5	1	2.4e+002	1.8e-015	1.9e-007	
FB0.5	13	92	12	1	5	1	2.4e+002	8.2e-016	1.9e-007	
FB0.2	11	40	10	1	3	1	2.4e+002	1.4e-016	1.9e-007	

Tabela 5.18: Problema `colvnlp` : $n = 15, \quad p = 0, \quad q = 15$ (formulação KKT)

$x_0 = x_{mcplib}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	3804	0	300	300	0	2.7e-002	1.0e-002	2.0e-001	$k \geq k_{max}$
FB	10	33	10	0	6	0	2.6e-002	1.3e-020	2.9e-009	
FBP0.95	9	20	9	0	5	0	6.9e-002	2.6e-019	4.4e-009	
FB0.75	7	12	7	0	4	0	4.6e-001	7.4e-013	5.8e-006	
FB0.5	6	7	6	0	0	0	1.5e+000	8.4e-016	1.3e-007	
FB0.2	6	7	6	0	0	0	3.4e+000	1.3e-016	3.4e-008	

$x_0 = x^1_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	9438	0	300	300	0	6.9e+005	5.9e-001	3.2e+000	$k \geq k_{max}$
FB	14	18	14	0	2	0	2.8e+006	7.3e-016	5.2e-006	
FBP0.95	14	16	13	1	1	0	2.5e+009	1.9e-013	6.5e-005	
FB0.75	16	18	15	1	1	0	6.2e+010	5.1e-018	3.3e-007	
FB0.5	17	21	14	3	3	0	2.5e+011	1.3e-013	3.5e-005	
FB0.2	16	21	12	4	4	0	6.4e+011	2.4e-013	1.5e-005	

$x_0 = x^2_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	3341	4	296	299	0	3.5e+003	4.7e-002	4.9e-001	$k \geq k_{max}$
FB	16	34	16	0	6	0	1.4e+004	5.8e-018	3.6e-007	
FBP0.95	16	21	16	0	3	0	1.4e+004	7.3e-016	4.6e-006	
FB0.75	12	15	12	0	2	0	4.0e+004	1.9e-015	1.3e-006	
FB0.5	11	13	11	0	1	0	1.3e+005	8.7e-016	6.1e-007	
FB0.2	11	13	11	0	1	0	3.3e+005	1.3e-016	3.2e-008	

$x_0 = x^4_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	2872	0	300	300	0	7.4e+001	9.4e-001	1.6e+001	$k \geq k_{max}$
FB	15	33	15	0	6	0	2.9e+002	2.9e-019	1.2e-007	
FBP0.95	9	14	9	0	3	0	2.6e+002	2.3e-015	4.2e-007	
FB0.75	9	13	9	0	3	0	1.9e+002	3.3e-018	1.2e-008	
FB0.5	8	9	8	0	0	0	1.7e+002	2.6e-017	3.8e-007	
FB0.2	7	8	7	0	0	0	2.6e+002	2.5e-013	2.7e-006	

$x_0 = x^5_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	2859	0	300	300	0	7.4e+001	9.4e-001	1.5e+001	$k \geq k_{max}$
FB	15	33	15	0	6	0	2.9e+002	2.9e-019	1.2e-007	
FBP0.95	9	14	9	0	3	0	2.6e+002	2.3e-015	4.2e-007	
FB0.75	9	13	9	0	3	0	1.9e+002	3.3e-018	1.2e-008	
FB0.5	8	9	8	0	0	0	1.7e+002	2.6e-017	3.8e-007	
FB0.2	7	8	7	0	0	0	2.6e+002	2.5e-013	2.7e-006	

$x_0 = x^6_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	3476	0	300	300	0	5.6e+000	3.0e-002	5.5e-001	$k \geq k_{max}$
FB	13	38	13	0	6	0	2.0e+001	1.6e-015	7.6e-006	
FBP0.95	10	22	10	0	5	0	1.8e+001	1.2e-016	2.0e-006	
FB0.75	7	9	7	0	1	0	1.6e+001	1.1e-015	4.8e-006	
FB0.5	10	14	10	0	2	0	2.1e+001	4.5e-022	2.1e-009	
FB0.2	8	10	8	0	1	0	3.9e+001	2.3e-014	4.5e-006	

Tabela 5.19: Problema `colvnlp` : $n = 15$ (formulação NCP)

$x_0 = x_{mcplib} = (11.500000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	3	5	3	0	1	0	4.8e-001	1.1e-020	2.1e-010	
FB	3	5	3	0	1	3	4.8e-001	1.1e-020	2.1e-010	
FBP0.95	3	5	3	0	1	3	4.8e-001	1.1e-020	2.1e-010	
FB0.75	3	5	3	0	1	3	4.8e-001	1.1e-020	2.1e-010	
FB0.5	3	5	3	0	1	3	4.8e-001	1.1e-020	2.1e-010	
FB0.2	3	5	3	0	1	3	4.8e-001	1.1e-020	2.1e-010	

Tabela 5.20: Problema `cycle` : $n = 1, \quad p = 0, \quad q = 1$ (formulação KKT)

$x_0 = x_{mcplib} = (11.500000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	3	5	3	0	1	0	4.8e-001	1.1e-020	1.5e-010	
FB	3	5	3	0	1	0	4.4e-001	1.8e-015	6.0e-008	
FBP0.95	4	6	4	0	1	0	1.1e+000	2.2e-021	6.3e-011	
FB0.75	5	7	5	0	1	0	6.2e+000	4.5e-026	9.8e-013	
FB0.5	5	7	5	0	1	0	1.9e+001	1.1e-023	2.6e-011	
FB0.2	5	8	5	0	2	0	4.3e+001	1.1e-014	1.2e-006	

Tabela 5.21: Problema `cycle` : $n = 1$ (formulação NCP)

$x_0 = x_{mcplib}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	17	34	17	0	16	1	5.0e-001	0	0	
FB	20	41	20	0	11	1	5.0e-001	4.6e-013	9.6e-007	
FBP0.95	20	41	20	0	11	1	5.0e-001	4.2e-013	8.7e-007	
FB0.75	20	40	20	0	11	1	5.0e-001	2.6e-013	5.4e-007	
FB0.5	20	40	20	0	11	1	5.0e-001	1.2e-013	2.4e-007	
FB0.2	20	39	20	0	10	1	5.0e-001	1.9e-014	3.9e-008	

Tabela 5.22: Problema `explcp` : $n = 16, \quad p = 0, \quad q = 16$ (formulação KKT)

$x_0 = x_{mcplib}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	19	43	19	0	17	1	5.0e-001	0	0	
FB	19	38	19	0	10	0	2.0e+000	4.6e-013	9.6e-007	
FBP0.95	19	38	19	0	10	0	1.8e+000	4.2e-013	8.7e-007	
FB0.75	19	38	19	0	10	0	1.1e+000	2.6e-013	5.4e-007	
FB0.5	19	38	19	0	10	0	5.0e-001	1.2e-013	2.4e-007	
FB0.2	19	38	19	0	10	0	8.0e-002	1.9e-014	3.9e-008	

Tabela 5.23: Problema `explcp` : $n = 16$ (formulação NCP)

$x_0 = x_{mcplib}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	0	1	0	0	0	0	2.2e+002	2.2e+002	9.8e+001	$\nexists\nabla\theta(x_f)$
FB	24	49	22	2	24	2	2.2e+002	5.6e-013	3.7e-006	
FBP0.95	20	45	16	4	16	2	2.2e+002	8.7e-013	1.8e-006	
FB0.75	6	9	4	2	2	2	2.2e+002	2.8e-018	2.2e-009	
FB0.5	5	8	3	2	2	2	2.2e+002	2.7e-013	6.8e-007	
FB0.2	6	9	4	2	2	2	2.2e+002	1.2e-019	2.5e-010	

$x_0 = x^1_{cpstart}$ (error de domínio em `mcpf`)										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	0	1	0	0	0	0	4.1e-001	4.1e-001	4.7e+000	$\nexists\nabla\theta(x_f)$
FB	19	39	18	1	19	1	4.1e-001	2.9e-013	2.7e-006	
FBP0.95	33	121	14	19	33	1	4.1e-001	9.4e-013	3.1e-006	
FB0.75	20	50	16	4	20	1	4.1e-001	8.6e-013	5.0e-006	
FB0.5	98	735	31	67	90	1	4.1e-001	7.0e-016	1.9e-008	
FB0.2	10	20	9	1	9	1	4.1e-001	6.1e-012	9.9e-007	$\|\nabla\theta_f\| \leq \varepsilon_g$

$x_0 = x^2_{cpstart}$ (error de domínio em `mcpf`)										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	0	1	0	0	0	0	1.3e+000	1.3e+000	8.3e+000	$\nexists\nabla\theta(x_f)$
FB	20	41	19	1	20	1	1.3e+000	2.7e-013	2.6e-006	
FBP0.95	35	128	15	20	35	1	1.3e+000	9.9e-013	2.3e-006	
FB0.75	12	24	11	1	11	1	1.3e+000	6.6e-013	1.1e-006	
FB0.5	13	26	12	1	12	1	1.3e+000	6.0e-013	6.5e-007	
FB0.2	12	24	11	1	11	1	1.3e+000	4.2e-012	9.8e-007	$\|\nabla\theta_f\| \leq \varepsilon_g$

$x_0 = x^3_{cpstart}$ (error de domínio em `mcpf`)										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	0	1	0	0	0	0	4.4e-002	4.4e-002	1.5e+000	$\nexists\nabla\theta(x_f)$
FB	97	784	38	59	83	1	4.4e-002	2.5e-020	2.9e-009	
FBP0.95	97	813	33	64	87	1	4.4e-002	1.0e-016	1.4e-008	
FB0.75	103	970	41	62	92	1	4.4e-002	3.9e-017	6.7e-009	
FB0.5	101	774	42	59	83	1	4.4e-002	1.9e-017	4.8e-009	
FB0.2	102	903	40	62	89	2	4.4e-002	9.2e-012	8.7e-007	$\|\nabla\theta_f\| \leq \varepsilon_g$

$x_0 = x^4_{cpstart}$ (error de domínio em `mcpf`)										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	0	1	0	0	0	0	1.1e-001	1.1e-001	2.3e+000	$\nexists\nabla\theta(x_f)$
FB	61	155	12	49	49	1	1.1e-001	7.6e-017	1.3e-008	
FBP0.95	60	154	11	49	49	1	1.1e-001	2.2e-019	6.5e-010	
FB0.75	76	203	19	57	67	1	1.1e-001	8.7e-013	3.1e-006	
FB0.5	75	205	15	60	63	1	1.1e-001	3.0e-020	2.7e-010	
FB0.2	59	155	8	51	51	1	1.1e-001	2.9e-016	9.4e-009	

$x_0 = x^5_{cpstart}$ (error de domínio em `mcpf`)										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	0	1	0	0	0	0	2.3e+002	2.3e+002	1.1e+002	$\nexists\nabla\theta(x_f)$
FB	23	47	22	1	23	1	2.3e+002	5.8e-013	4.0e-006	
FBP0.95	18	36	17	1	17	1	2.3e+002	4.9e-013	1.4e-006	
FB0.75	19	38	18	1	18	1	2.3e+002	4.7e-013	9.0e-007	
FB0.5	97	707	37	60	87	2	2.3e+002	3.2e-017	4.0e-009	
FB0.2	17	33	16	1	15	1	2.3e+002	2.3e-012	9.3e-007	$\|\nabla\theta_f\| \leq \varepsilon_g$

Tabela 5.24: Problema `hanskoop` : $n = 14, \quad p = 0, \quad q = 14$ (formulação KKT)

$x_0 = x_{mcplib}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	26	67	1	25	23	0	2.1e+002	2.8e-035	7.6e-018	
FB	13	18	12	1	3	0	8.5e+002	2.9e-019	2.6e-009	
FBP0.95	20	40	19	1	19	0	7.6e+002	3.5e-013	2.3e-006	
FB0.75	79	668	38	41	64	0	4.8e+002	1.7e-018	1.7e-009	
FB0.5	19	37	18	1	17	0	2.1e+002	3.6e-013	2.1e-006	
FB0.2	19	37	18	1	17	0	3.4e+001	8.7e-013	1.6e-006	

$x_0 = x^1_{cpstart}$ (error de domínio em `mcpf`)										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	207	1688	6	201	201	0	9.7e-002	3.9e-018	1.0e-008	
FB	66	169	14	52	54	0	2.3e-001	7.7e-015	1.5e-006	
FBP0.95	81	199	14	67	69	0	2.1e-001	6.9e-017	3.8e-008	
FB0.75	64	161	17	47	48	0	1.4e-001	1.4e-019	1.3e-009	
FB0.5	81	201	12	69	71	0	6.9e-002	8.3e-013	1.7e-006	
FB0.2	64	163	15	49	50	0	2.4e-002	6.1e-016	8.4e-008	

$x_0 = x^2_{cpstart}$ (error de domínio em `mcpf`)										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	0	1	0	0	0	0	4.2e-001	4.2e-001	4.1e+000	$\nexists\nabla\theta(x_f)$
FB	73	187	12	61	65	0	1.2e+000	1.0e-013	1.2e-006	
FBP0.95	61	158	14	47	48	0	1.1e+000	3.9e-018	8.7e-009	
FB0.75	86	208	18	68	72	0	6.9e-001	1.8e-018	1.8e-009	
FB0.5	78	190	18	60	63	0	3.5e-001	7.5e-020	6.5e-010	
FB0.2	66	167	15	51	52	0	1.5e-001	1.7e-019	2.0e-010	

$x_0 = x^3_{cpstart}$ (error de domínio em `mcpf`)										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	0	1	0	0	0	0	9.6e-003	9.6e-003	2.0e-001	$\nexists\nabla\theta(x_f)$
FB	82	199	16	66	68	1	4.3e-003	7.8e-021	1.3e-010	
FBP0.95	80	196	15	65	67	1	4.0e-003	3.2e-017	8.3e-008	
FB0.75	77	187	18	59	61	1	3.2e-003	1.8e-015	3.5e-007	
FB0.5	65	171	10	55	57	1	2.4e-003	2.7e-018	3.4e-009	
FB0.2	64	168	10	54	55	1	1.5e-003	4.2e-018	1.3e-009	

$x_0 = x^4_{cpstart}$ (error de domínio em `mcpf`)										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	68	181	7	61	64	0	1.0e-002	4.6e-002	9.3e-001	$\nexists\nabla\theta(x_f)$
FB	81	195	22	59	64	0	5.4e-003	7.9e-017	7.5e-008	
FBP0.95	75	187	16	59	62	0	5.2e-003	5.8e-016	2.0e-007	
FB0.75	100	245	23	77	96	0	4.5e-003	4.0e-013	1.2e-006	
FB0.5	69	176	13	56	58	0	3.7e-003	1.7e-013	2.0e-006	
FB0.2	67	170	14	53	54	0	2.9e-003	3.1e-018	1.7e-009	

$x_0 = x^5_{cpstart}$ (error de domínio em `mcpf`)										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	152	639	0	152	152	0	6.8e+001	2.2e-003	8.2e-007	$\|\nabla\theta_f\| \leq \varepsilon_g$
FB	10	11	10	0	0	0	1.9e+002	4.9e-018	3.8e-009	
FBP0.95	71	183	21	50	58	0	2.0e+002	4.1e-015	9.2e-008	
FB0.75	69	174	21	48	54	0	4.9e+002	7.4e-020	3.0e-010	
FB0.5	68	170	17	51	53	0	1.4e+003	8.0e-020	1.4e-009	
FB0.2	65	164	16	49	50	0	3.3e+003	1.7e-013	1.3e-007	

Tabela 5.25: Problema `hanskoop` : $n = 14$ (formulação NCP)

$x_0 = x_{mcplib} = (1.250000, 0.000000, 0.000000, 0.500000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	2	3	2	0	0	1	4.8e+000	1.2e-014	2.2e-006	
FB	2	3	2	0	0	2	4.8e+000	1.2e-014	2.2e-006	
FBP0.95	2	3	2	0	0	2	4.8e+000	1.2e-014	2.2e-006	
FB0.75	2	3	2	0	0	2	4.8e+000	1.2e-014	2.2e-006	
FB0.5	2	3	2	0	0	2	4.8e+000	1.2e-014	2.2e-006	
FB0.2	2	3	2	0	0	2	4.8e+000	1.2e-014	2.2e-006	

$x_0 = x^1_{cpstart} = (0.000000, 0.000000, 0.000000, 0.000000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	1737	297	3	280	1	6.3e+000	1.3e-001	8.3e-001	$k \geq k_{max}$
FB	12	62	12	0	8	1	6.3e+000	9.4e-015	3.2e-007	
FBP0.95	12	62	12	0	8	1	6.3e+000	8.5e-015	3.2e-007	
FB0.75	9	19	9	0	4	1	6.3e+000	2.8e-021	1.0e-009	
FB0.5	9	19	9	0	4	1	6.3e+000	2.8e-021	1.0e-009	
FB0.2	300	4379	287	13	297	1	6.3e+000	4.9e-002	9.2e-001	$k \geq k_{max}$

$x_0 = x^2_{cpstart} = (1.000000, 1.000000, 1.000000, 1.000000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	10594	300	0	300	1	2.6e+001	2.3e+001	1.5e+002	$k \geq k_{max}$
FB	300	4019	294	6	289	39	2.6e+001	6.8e-001	1.8e+001	$k \geq k_{max}$
FBP0.95	300	4052	285	15	294	20	2.6e+001	7.7e-001	5.6e+000	$k \geq k_{max}$
FB0.75	300	4421	292	8	291	19	2.6e+001	1.1e+000	2.4e+001	$k \geq k_{max}$
FB0.5	300	4329	293	7	286	22	2.6e+001	2.8e-001	1.1e+001	$k \geq k_{max}$
FB0.2	300	4392	292	8	290	23	2.6e+001	1.0e+000	2.4e+001	$k \geq k_{max}$

$x_0 = x^3_{cpstart} = (100.000000, 100.000000, 100.000000, 100.000000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	3530	293	7	284	0	1.4e+009	4.7e-001	1.5e+001	$k \geq k_{max}$
FB	300	3536	288	12	279	8	1.4e+009	4.9e-001	5.6e+000	$k \geq k_{max}$
FBP0.95	300	3507	289	11	280	12	1.4e+009	5.4e-001	1.2e+000	$k \geq k_{max}$
FB0.75	300	3172	284	16	278	12	1.4e+009	4.7e-002	1.6e+000	$k \geq k_{max}$
FB0.5	136	1897	132	4	114	12	1.4e+009	1.3e-016	9.6e-008	
FB0.2	79	627	77	2	56	12	1.4e+009	1.2e-023	6.8e-011	

$x_0 = x^5_{cpstart} = (1.000000, 0.000000, 0.000000, 0.000000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	4	5	4	0	0	1	2.3e+000	1.2e-014	2.2e-006	
FB	5	6	5	0	0	5	2.3e+000	3.1e-013	2.7e-006	
FBP0.95	5	6	5	0	0	5	2.3e+000	2.8e-013	2.7e-006	
FB0.75	5	6	5	0	0	5	2.3e+000	1.8e-013	2.6e-006	
FB0.5	5	6	5	0	0	5	2.3e+000	8.9e-014	2.6e-006	
FB0.2	5	6	5	0	0	5	2.3e+000	2.9e-014	2.6e-006	

$x_0 = x^6_{cpstart} = (0.000000, 1.000000, 1.000000, 0.000000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	12565	300	0	300	1	3.2e+000	2.1e+000	1.4e+001	$k \geq k_{max}$
FB	42	555	39	3	38	6	3.2e+000	1.5e-016	4.1e-008	
FBP0.95	42	508	39	3	38	6	3.2e+000	6.1e-015	1.2e-006	
FB0.75	300	3767	282	18	297	6	3.2e+000	8.3e-002	3.8e+000	$k \geq k_{max}$
FB0.5	300	4365	291	9	297	15	3.2e+000	6.7e-001	1.8e+001	$k \geq k_{max}$
FB0.2	300	5727	293	7	299	46	3.2e+000	9.3e-001	4.0e+000	$k \geq k_{max}$

Tabela 5.26: Problema `josephy` : $n = 4, \quad p = 0, \quad q = 4$ (formulação KKT)

$x_0 = x_{mcplib} = (1.250000, 0.000000, 0.000000, 0.500000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	2	3	2	0	0	0	4.9e-003	5.3e-015	9.3e-007	
FB	3	4	3	0	0	0	4.2e-003	2.0e-019	5.2e-009	
FBP0.95	3	4	3	0	0	0	4.3e-003	8.4e-020	3.3e-009	
FB0.75	3	4	3	0	0	0	4.8e-003	1.7e-021	3.9e-010	
FB0.5	2	3	2	0	0	0	5.5e-003	7.8e-013	4.4e-006	
FB0.2	2	3	2	0	0	0	6.3e-003	7.7e-013	1.2e-005	

$x_0 = x^1_{cpstart} = (0.000000, 0.000000, 0.000000, 0.000000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	6	12	5	1	3	0	6.3e+000	4.6e-016	2.8e-007	
FB	7	10	7	0	2	0	2.5e+001	3.3e-024	2.0e-011	
FBP0.95	7	8	7	0	0	0	2.3e+001	6.7e-021	3.8e-010	
FB0.75	6	8	6	0	1	0	1.4e+001	8.0e-014	1.0e-006	
FB0.5	5	8	5	0	1	0	6.3e+000	3.3e-017	2.1e-008	
FB0.2	5	8	5	0	1	0	1.0e+000	5.3e-018	3.4e-009	

$x_0 = x^2_{cpstart} = (1.000000, 1.000000, 1.000000, 1.000000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	7	15	7	0	4	0	5.0e-001	5.6e-022	3.0e-010	
FB	7	12	7	0	2	0	4.3e-001	2.7e-018	1.8e-008	
FBP0.95	5	7	5	0	1	0	7.6e-001	2.6e-013	6.3e-006	
FB0.75	5	6	5	0	0	0	3.1e+000	1.4e-013	3.0e-006	
FB0.5	5	6	5	0	0	0	8.3e+000	2.6e-025	6.3e-012	
FB0.2	6	7	6	0	0	0	1.8e+001	4.5e-024	1.3e-011	

$x_0 = x^3_{cpstart} = (100.000000, 100.000000, 100.000000, 100.000000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	7	13	6	1	3	0	5000	4.6e-016	2.8e-007	
FB	13	20	13	0	5	0	5.0e+003	2.0e-018	1.5e-008	
FBP0.95	14	15	14	0	0	0	3.5e+010	5.7e-013	4.0e-006	
FB0.75	16	17	16	0	0	0	8.7e+011	2.7e-021	1.4e-010	
FB0.5	16	17	16	0	0	0	3.5e+012	3.8e-023	7.8e-011	
FB0.2	16	17	16	0	0	0	9.0e+012	1.3e-016	1.2e-007	

$x_0 = x^5_{cpstart} = (1.000000, 0.000000, 0.000000, 0.000000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	3	4	3	0	0	0	1.6e+000	5.3e-015	9.3e-007	
FB	4	5	4	0	0	0	5.3e+000	2.5e-014	1.0e-006	
FBP0.95	4	5	4	0	0	0	4.8e+000	2.3e-014	9.5e-007	
FB0.75	4	5	4	0	0	0	3.0e+000	1.4e-014	5.9e-007	
FB0.5	4	5	4	0	0	0	1.3e+000	6.3e-015	2.6e-007	
FB0.2	4	5	4	0	0	0	2.1e-001	1.0e-015	4.2e-008	

$x_0 = x^6_{cpstart} = (0.000000, 1.000000, 1.000000, 0.000000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	6	12	6	0	3	0	1.4e+000	2.7e-015	6.7e-007	
FB	7	13	7	0	3	0	4.7e+000	1.1e-015	3.5e-007	
FBP0.95	6	9	6	0	2	0	4.3e+000	1.0e-021	3.5e-010	
FB0.75	6	8	6	0	1	0	2.9e+000	3.4e-017	1.6e-008	
FB0.5	6	9	6	0	1	0	1.8e+000	9.1e-019	7.4e-009	
FB0.2	7	12	7	0	2	0	1.4e+000	1.2e-015	4.4e-007	

Tabela 5.27: Problema `josephy` : $n = 4$ (formulação NCP)

$x_0 = x_{mcplib} = (1.250000, 0.000000, 0.000000, 0.500000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	2	3	2	0	0	1	1.4e+000	1.2e-014	2.6e-006	
FB	2	3	2	0	0	2	1.4e+000	1.2e-014	2.6e-006	
FBP0.95	2	3	2	0	0	2	1.4e+000	1.2e-014	2.6e-006	
FB0.75	2	3	2	0	0	2	1.4e+000	1.2e-014	2.6e-006	
FB0.5	2	3	2	0	0	2	1.4e+000	1.2e-014	2.6e-006	
FB0.2	2	3	2	0	0	2	1.4e+000	1.2e-014	2.6e-006	

$x_0 = x^1_{cpstart} = (0.000000, 0.000000, 0.000000, 0.000000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	2321	299	1	232	1	1.6e+001	7.3e-003	1.2e-001	$k \geq k_{max}$
FB	7	12	7	0	2	1	1.6e+001	2.2e-019	1.1e-008	
FBP0.95	7	12	7	0	2	1	1.6e+001	2.2e-019	1.1e-008	
FB0.75	7	11	7	0	2	1	1.6e+001	6.6e-016	5.8e-007	
FB0.5	7	12	7	0	2	1	1.6e+001	8.7e-013	2.2e-005	
FB0.2	9	15	9	0	2	1	1.6e+001	5.0e-025	1.7e-011	

$x_0 = x^2_{cpstart} = (1.000000, 1.000000, 1.000000, 1.000000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	3410	300	0	273	0	4.0e+001	7.3e-003	1.2e-001	$k \geq k_{max}$
FB	300	4186	281	19	298	6	4.0e+001	1.3e-002	2.7e-001	$k \geq k_{max}$
FBP0.95	300	4317	284	16	298	6	4.0e+001	1.3e-002	3.3e-001	$k \geq k_{max}$
FB0.75	300	4148	278	22	298	7	4.0e+001	9.3e-003	2.7e-001	$k \geq k_{max}$
FB0.5	300	4059	272	28	298	10	4.0e+001	4.8e-003	1.7e-001	$k \geq k_{max}$
FB0.2	300	4045	274	26	298	14	4.0e+001	3.0e-003	2.2e-001	$k \geq k_{max}$

$x_0 = x^3_{cpstart} = (100.000000, 100.000000, 100.000000, 100.000000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	3647	300	0	239	0	1.4e+009	8.7e-003	1.4e-001	$k \geq k_{max}$
FB	300	2674	289	11	278	11	1.4e+009	6.8e-001	8.1e+000	$k \geq k_{max}$
FBP0.95	300	2494	285	15	276	11	1.4e+009	5.2e-001	4.3e+000	$k \geq k_{max}$
FB0.75	300	3220	289	11	277	11	1.4e+009	3.4e-001	2.3e+000	$k \geq k_{max}$
FB0.5	300	2813	290	10	271	11	1.4e+009	3.6e-001	1.2e+001	$k \geq k_{max}$
FB0.2	300	2575	286	14	267	13	1.4e+009	7.5e-002	6.2e+000	$k \geq k_{max}$

$x_0 = x^4_{cpstart} = (1.000000, 0.000000, 1.000000, 0.000000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	1	2	1	0	0	1	1.8e+001	2.3e-030	1.7e-014	
FB	1	2	1	0	0	1	1.8e+001	2.0e-030	1.7e-014	
FBP0.95	1	2	1	0	0	1	1.8e+001	7.0e-030	2.4e-014	
FB0.75	1	2	1	0	0	1	1.8e+001	2.1e-030	2.5e-014	
FB0.5	1	2	1	0	0	1	1.8e+001	9.1e-029	2.1e-013	
FB0.2	1	2	1	0	0	1	1.8e+001	3.9e-029	2.1e-013	

$x_0 = x^5_{cpstart} = (1.000000, 0.000000, 0.000000, 0.000000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	4	5	4	0	0	1	6.3e+000	1.2e-014	2.6e-006	
FB	6	7	6	0	0	6	6.3e+000	2.1e-023	1.0e-011	
FBP0.95	6	7	6	0	0	6	6.3e+000	1.9e-023	9.7e-012	
FB0.75	6	7	6	0	0	6	6.3e+000	1.2e-023	9.0e-012	
FB0.5	6	7	6	0	0	6	6.3e+000	5.3e-024	9.2e-012	
FB0.2	6	7	6	0	0	6	6.3e+000	9.9e-025	9.6e-012	

Tabela 5.28: Problema `kojshin` : $n = 4, \quad p = 0, \quad q = 4$ (formulação KKT)

$x_0 = x_{mcplib} = (1.250000, 0.000000, 0.000000, 0.500000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	2	3	2	0	0	0	4.9e-003	5.3e-015	9.3e-007	
FB	3	4	3	0	0	0	4.2e-003	1.5e-019	4.4e-009	
FBP0.95	3	4	3	0	0	0	4.3e-003	6.9e-020	5.3e-009	
FB0.75	3	4	3	0	0	0	4.8e-003	3.6e-019	1.5e-008	
FB0.5	3	4	3	0	0	0	5.5e-003	4.9e-022	3.2e-011	
FB0.2	2	3	2	0	0	0	6.3e-003	7.7e-013	1.2e-005	

$x_0 = x^1_{cpstart} = (0.000000, 0.000000, 0.000000, 0.000000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	6	17	5	1	4	0	1.6e+001	4.6e-018	2.7e-008	
FB	10	21	10	0	5	0	65	8.2e-019	1.6e-008	
FBP0.95	7	10	7	0	2	0	5.9e+001	5.5e-023	7.7e-011	
FB0.75	8	15	8	0	3	0	3.7e+001	4.9e-019	1.8e-008	
FB0.5	6	10	6	0	2	0	1.6e+001	1.8e-022	8.0e-011	
FB0.2	6	10	6	0	2	0	2.6e+000	6.5e-022	3.6e-010	

$x_0 = x^2_{cpstart} = (1.000000, 1.000000, 1.000000, 1.000000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	7	14	7	0	3	0	5.0e-001	1.7e-023	7.1e-011	
FB	7	15	7	0	4	0	4.3e-001	4.5e-016	3.4e-007	
FBP0.95	6	13	6	0	3	0	8.6e-001	1.6e-013	5.5e-006	
FB0.75	6	10	6	0	1	0	4.2e+000	1.2e-015	7.7e-007	
FB0.5	6	8	6	0	1	0	1.2e+001	9.9e-016	1.0e-007	
FB0.2	6	9	6	0	1	0	2.7e+001	7.5e-014	2.4e-006	

$x_0 = x^3_{cpstart} = (100.000000, 100.000000, 100.000000, 100.000000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	7	18	6	1	4	0	5000	4.6e-018	2.7e-008	
FB	7	10	7	0	2	0	5.0e+003	4.0e-016	2.1e-007	
FBP0.95	15	16	15	0	0	0	3.5e+010	2.1e-021	4.1e-010	
FB0.75	300	3826	299	1	272	0	8.8e+011	1.0e-002	1.6e-001	$k \geq k_{max}$
FB0.5	22	23	22	0	0	0	3.5e+012	5.0e-020	1.5e-009	
FB0.2	17	18	17	0	0	0	9.0e+012	6.6e-014	2.4e-006	

$x_0 = x^4_{cpstart} = (1.000000, 0.000000, 1.000000, 0.000000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	3	4	3	0	0	1	2.5e+000	5.3e-015	9.3e-007	
FB	4	5	4	0	0	1	7.7e+000	4.2e-020	2.3e-009	
FBP0.95	4	5	4	0	0	1	6.9e+000	3.6e-020	2.0e-009	
FB0.75	4	5	4	0	0	1	4.3e+000	1.8e-020	1.0e-009	
FB0.5	4	5	4	0	0	1	1.9e+000	4.9e-021	2.4e-010	
FB0.2	4	5	4	0	0	1	3.1e-001	5.1e-020	1.4e-009	

$x_0 = x^5_{cpstart} = (1.000000, 0.000000, 0.000000, 0.000000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	3	4	3	0	0	0	6.1e+000	5.3e-015	9.3e-007	
FB	5	7	5	0	1	0	2.3e+001	4.2e-015	6.9e-007	
FBP0.95	5	7	5	0	1	0	2.1e+001	4.3e-015	7.1e-007	
FB0.75	5	7	5	0	1	0	1.3e+001	4.5e-015	7.9e-007	
FB0.5	5	7	5	0	1	0	5.8e+000	4.7e-015	8.9e-007	
FB0.2	5	7	5	0	1	0	9.3e-001	5.2e-015	1.0e-006	

Tabela 5.29: Problema `kojshin` : $n = 4$ (formulação NCP)

$x_0 = x_{mcplib} = (4.500000, 11.000000, 0.500000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	7	8	7	0	0	0	1.2e+001	3.6e-016	7.8e-008	
FB	7	8	7	0	0	7	1.2e+001	3.6e-016	7.8e-008	
FBP0.95	7	8	7	0	0	7	1.2e+001	3.6e-016	7.8e-008	
FB0.75	7	8	7	0	0	7	1.2e+001	3.6e-016	7.8e-008	
FB0.5	7	8	7	0	0	7	1.2e+001	3.6e-016	7.8e-008	
FB0.2	7	8	7	0	0	7	1.2e+001	3.6e-016	7.8e-008	

$x_0 = x^1_{cpstart} = (2.500000, 5.500000, 1.500000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	6	10	6	0	1	0	5.0e-001	4.0e-016	8.3e-008	
FB	6	10	6	0	1	6	5.0e-001	4.0e-016	8.3e-008	
FBP0.95	6	10	6	0	1	6	5.0e-001	4.0e-016	8.3e-008	
FB0.75	6	10	6	0	1	6	5.0e-001	4.0e-016	8.3e-008	
FB0.5	6	10	6	0	1	6	5.0e-001	4.0e-016	8.3e-008	
FB0.2	6	10	6	0	1	6	5.0e-001	4.0e-016	8.3e-008	

$x_0 = x^2_{cpstart} = (3.500000, 4.500000, 0.500000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	4	5	4	0	0	0	5.1e-001	5.6e-016	9.8e-008	
FB	4	5	4	0	0	4	5.1e-001	5.6e-016	9.8e-008	
FBP0.95	4	5	4	0	0	4	5.1e-001	5.6e-016	9.8e-008	
FB0.75	4	5	4	0	0	4	5.1e-001	5.6e-016	9.8e-008	
FB0.5	4	5	4	0	0	4	5.1e-001	5.6e-016	9.8e-008	
FB0.2	4	5	4	0	0	4	5.1e-001	5.6e-016	9.8e-008	

$x_0 = x^3_{cpstart} = (2.500000, 1.500000, 1.500000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	5224	90	210	299	0	1.1e+001	1.4e-004	7.7e-002	$k \geq k_{max}$, D89
FB	300	3950	169	131	298	169	1.1e+001	2.7e-005	2.1e-002	$k \geq k_{max}$
FBP0.95	300	4068	167	133	298	167	1.1e+001	2.6e-005	1.5e-002	$k \geq k_{max}$
FB0.75	300	4168	168	132	298	170	1.1e+001	2.6e-005	2.3e-002	$k \geq k_{max}$
FB0.5	300	4093	151	149	298	151	1.1e+001	2.4e-005	1.6e-002	$k \geq k_{max}$
FB0.2	300	4327	134	166	298	134	1.1e+001	3.5e-005	1.7e-002	$k \geq k_{max}$

$x_0 = x^4_{cpstart} = (3.500000, 6.500000, 0.500000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	5	6	5	0	0	0	2.2e+000	4.4e-014	8.6e-007	
FB	5	6	5	0	0	5	2.2e+000	4.4e-014	8.6e-007	
FBP0.95	5	6	5	0	0	5	2.2e+000	4.4e-014	8.6e-007	
FB0.75	5	6	5	0	0	5	2.2e+000	4.4e-014	8.6e-007	
FB0.5	5	6	5	0	0	5	2.2e+000	4.4e-014	8.6e-007	
FB0.2	5	6	5	0	0	5	2.2e+000	4.4e-014	8.6e-007	

$x_0 = x^5_{cpstart} = (3.000000, 0.600000, 0.100000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	10	20	10	0	2	0	3.2e+001	2.7e-017	2.1e-008	
FB	10	20	10	0	2	10	3.2e+001	7.1e-014	1.1e-006	
FBP0.95	10	20	10	0	2	10	3.2e+001	7.1e-014	1.1e-006	
FB0.75	10	20	10	0	2	10	3.2e+001	7.1e-014	1.1e-006	
FB0.5	10	20	10	0	2	10	3.2e+001	7.1e-014	1.1e-006	
FB0.2	10	20	10	0	2	10	3.2e+001	7.1e-014	1.1e-006	

Tabela 5.30: Problema `mathinum` : $n = 3, \quad p = 0, \quad q = 3$ (formulação KKT)

$x_0 = x_{mcplib} = (4.500000, 11.000000, 0.500000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	7	8	7	0	0	0	1.2e+001	3.6e-016	7.3e-008	
FB	8	10	8	0	1	0	2.6e+001	7.3e-023	3.3e-011	
FBP0.95	6	7	6	0	0	0	2.6e+001	1.4e-015	1.4e-007	
FB0.75	5	6	5	0	0	0	3.8e+001	5.4e-015	2.2e-007	
FB0.5	5	6	5	0	0	0	8.0e+001	8.4e-017	2.0e-008	
FB0.2	5	6	5	0	0	0	1.7e+002	3.5e-019	1.4e-009	

$x_0 = x^1_{cpstart} = (2.500000, 5.500000, 1.500000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	21	44	21	0	21	0	5.0e-001	2.7e-013	1.6e-006	D20
FB	4	5	4	0	0	0	3.6e-001	1.4e-017	1.1e-008	
FBP0.95	5	7	5	0	1	0	3.6e-001	4.5e-024	6.8e-012	
FB0.75	6	9	6	0	2	0	3.6e-001	4.5e-014	6.3e-007	
FB0.5	5	6	5	0	0	0	4.2e-001	5.7e-018	5.3e-009	
FB0.2	4	5	4	0	0	0	5.8e-001	1.1e-016	2.5e-008	

$x_0 = x^2_{cpstart} = (3.500000, 4.500000, 0.500000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	4	5	4	0	0	0	5.1e-001	5.6e-016	9.2e-008	
FB	5	6	5	0	0	0	6.6e-001	3.0e-013	2.1e-006	
FBP0.95	5	6	5	0	0	0	6.9e-001	1.1e-015	1.2e-007	
FB0.75	5	6	5	0	0	0	1.0e+000	1.3e-020	3.3e-010	
FB0.5	4	5	4	0	0	0	1.8e+000	8.5e-013	2.1e-006	
FB0.2	4	5	4	0	0	0	3.4e+000	4.7e-014	4.9e-007	

$x_0 = x^3_{cpstart} = (2.500000, 1.500000, 1.500000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	22	44	21	1	21	0	1.0e+001	4.3e-013	1.3e-005	D20
FB	14	55	14	0	9	0	3.1e+001	4.3e-026	2.9e-013	
FBP0.95	9	14	9	0	4	0	2.8e+001	2.4e-013	1.3e-006	
FB0.75	5	6	5	0	0	0	1.9e+001	4.0e-016	4.3e-008	
FB0.5	5	6	5	0	0	0	1.1e+001	1.3e-013	5.1e-007	
FB0.2	5	6	5	0	0	0	8.4e+000	9.0e-014	1.7e-007	

$x_0 = x^4_{cpstart} = (3.500000, 6.500000, 0.500000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	5	6	5	0	0	0	2.2e+000	4.4e-014	8.1e-007	
FB	7	8	7	0	0	0	3.5e+000	3.6e-017	2.3e-008	
FBP0.95	6	7	6	0	0	0	3.6e+000	9.3e-016	1.1e-007	
FB0.75	5	6	5	0	0	0	4.8e+000	3.6e-016	5.6e-008	
FB0.5	5	6	5	0	0	0	9.0e+000	2.5e-019	1.1e-009	
FB0.2	5	6	5	0	0	0	1.8e+001	1.1e-021	7.5e-011	

$x_0 = x^5_{cpstart} = (3.000000, 0.600000, 0.100000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	10	16	10	0	2	0	3.2e+001	3.7e-018	7.4e-009	
FB	10	11	10	0	0	0	1.2e+002	1.5e-016	4.7e-008	
FBP0.95	10	11	10	0	0	0	1.1e+002	9.1e-017	3.5e-008	
FB0.75	10	11	10	0	0	0	6.7e+001	4.3e-017	2.0e-008	
FB0.5	10	11	10	0	0	0	3.3e+001	3.2e-018	4.0e-009	
FB0.2	10	11	10	0	0	0	1.1e+001	4.4e-021	1.5e-010	

Tabela 5.31: Problema `mathinum` : $n = 3$ (formulação NCP)

$x_0 = x_{mcplib} = (1.000000, 0.000000, 1.000000, 1.000000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	3	4	3	0	0	1	6.3e+000	2.4e-031	2.5e-015	
FB	4	5	4	0	0	4	6.3e+000	2.8e-018	5.6e-009	
FBP0.95	4	5	4	0	0	4	6.3e+000	2.7e-018	5.3e-009	
FB0.75	4	5	4	0	0	4	6.3e+000	2.2e-018	4.3e-009	
FB0.5	4	5	4	0	0	4	6.3e+000	1.7e-018	3.4e-009	
FB0.2	4	5	4	0	0	4	6.3e+000	1.4e-018	3.0e-009	

$x_0 = x^1_{cpstart} = (2.500000, 5.500000, 1.500000, 4.500000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	2	3	2	0	0	0	8.5e+000	2.5e-032	2.2e-016	
FB	4	6	3	1	1	4	8.5e+000	1.1e-017	1.0e-008	
FBP0.95	4	6	3	1	1	4	8.5e+000	1.6e-017	1.2e-008	
FB0.75	4	6	3	1	1	4	8.5e+000	1.5e-016	3.6e-008	
FB0.5	4	6	3	1	1	4	8.5e+000	5.6e-015	1.9e-007	
FB0.2	5	7	4	1	1	5	8.5e+000	6.5e-025	1.4e-012	

$x_0 = x^2_{cpstart} = (3.500000, 4.500000, 0.500000, 4.000000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	5401	35	265	300	0	5.4e+001	2.7e-004	2.4e-001	$k \geq k_{max}$, D32
FB	300	4022	23	277	297	23	5.4e+001	3.0e-003	1.5e+000	$k \geq k_{max}$
FBP0.95	300	4003	23	277	297	5	5.4e+001	2.9e-003	1.5e+000	$k \geq k_{max}$
FB0.75	300	6065	22	278	297	4	5.4e+001	2.7e-003	4.1e-001	$k \geq k_{max}$
FB0.5	300	5565	19	281	297	19	5.4e+001	7.0e-004	3.7e-001	$k \geq k_{max}$
FB0.2	300	5611	42	258	298	42	5.4e+001	9.3e-007	6.1e-003	$k \geq k_{max}$

$x_0 = x^3_{cpstart} = (2.500000, 1.500000, 1.500000, 3.500000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	2	3	2	0	0	0	3.2e+000	9.9e-032	4.4e-016	
FB	4	6	3	1	1	3	3.2e+000	9.4e-019	3.0e-009	
FBP0.95	4	6	3	1	1	4	3.2e+000	1.4e-018	3.6e-009	
FB0.75	4	6	3	1	1	4	3.2e+000	1.3e-017	1.0e-008	
FB0.5	4	6	3	1	1	4	3.2e+000	4.9e-016	5.6e-008	
FB0.2	5	7	4	1	1	5	3.2e+000	4.4e-027	1.1e-013	

$x_0 = x^4_{cpstart} = (3.500000, 6.500000, 0.500000, 5.500000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	6084	16	284	300	0	7.7e+001	4.0e-003	4.0e-001	$k \geq k_{max}$, D13
FB	300	5883	26	274	297	4	7.7e+001	3.3e-003	1.5e+000	$k \geq k_{max}$
FBP0.95	300	6005	25	275	297	5	7.7e+001	3.3e-003	3.3e-001	$k \geq k_{max}$
FB0.75	300	5889	24	276	297	25	7.7e+001	3.2e-003	2.2e-001	$k \geq k_{max}$
FB0.5	300	5188	34	266	297	34	7.7e+001	6.0e-004	7.4e-001	$k \geq k_{max}$
FB0.2	300	1940	173	127	299	6	7.7e+001	2.1e-005	1.9e+000	$k \geq k_{max}$

$x_0 = x^7_{cpstart} = (1.000000, 1.000000, 1.000000, 1.000000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	2	3	2	0	0	0	5.8e+000	1.0e-031	8.3e-016	
FB	8	11	7	1	2	6	5.8e+000	6.8e-017	8.8e-008	
FBP0.95	8	13	7	1	3	8	5.8e+000	5.6e-016	2.5e-007	
FB0.75	7	11	6	1	3	7	5.8e+000	2.1e-023	1.3e-011	
FB0.5	8	14	7	1	3	8	5.8e+000	3.8e-015	6.6e-007	
FB0.2	14	55	13	1	8	14	5.8e+000	2.6e-016	1.7e-007	

Tabela 5.32: Problema `mathisum` : $n = 4, \quad p = 0, \quad q = 4$ (formulação KKT)

$x_0 = x_{mcplib} = (1.000000, 0.000000, 1.000000, 1.000000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	8	12	8	0	3	1	5.1e+000	9.7e-018	3.3e-008	D3
FB	8	11	8	0	2	0	1.7e+001	5.6e-017	4.3e-008	
FBP0.95	7	8	7	0	0	0	1.5e+001	1.2e-018	5.9e-009	
FB0.75	5	6	5	0	0	0	9.6e+000	1.5e-013	1.6e-006	
FB0.5	5	6	5	0	0	0	4.7e+000	2.2e-014	4.2e-007	
FB0.2	9	10	9	0	0	0	1.7e+000	1.8e-020	2.8e-010	

$x_0 = x^1_{cpstart} = (2.500000, 5.500000, 1.500000, 4.500000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	25	75	0	25	25	0	3.7e+000	4.1e-001	9.1e-007	$\|\nabla\theta_f\| \leq \varepsilon_g$
FB	6	8	6	0	1	0	3.1e+000	1.1e-018	3.4e-009	
FBP0.95	7	10	7	0	1	0	4.5e+000	2.0e-018	4.3e-009	
FB0.75	7	8	7	0	0	0	1.6e+001	3.2e-024	1.9e-012	
FB0.5	6	14	6	0	1	0	4.5e+001	1.3e-022	1.8e-011	
FB0.2	6	24	6	0	1	0	1.0e+002	6.4e-015	5.2e-008	

$x_0 = x^2_{cpstart} = (3.500000, 4.500000, 0.500000, 4.000000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	35	109	0	35	35	0	5.2e+001	4.1e-001	7.4e-007	$\|\nabla\theta_f\| \leq \varepsilon_g$
FB	8	10	8	0	1	0	1.9e+002	4.6e-017	3.9e-008	
FBP0.95	9	13	9	0	3	0	1.8e+002	4.3e-013	8.8e-007	
FB0.75	9	11	9	0	1	0	1.2e+002	9.9e-023	4.3e-011	
FB0.5	8	9	8	0	0	0	7.0e+001	2.8e-023	1.5e-011	
FB0.2	7	8	7	0	0	0	5.5e+001	1.4e-018	1.4e-009	

$x_0 = x^3_{cpstart} = (2.500000, 1.500000, 1.500000, 3.500000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	201	790	0	201	201	0	3.1e+000	1.3e-001	9.3e-007	$\|\nabla\theta_f\| \leq \varepsilon_g$
FB	9	14	9	0	4	0	4.4e+000	1.6e-025	5.7e-013	
FBP0.95	5	6	5	0	0	0	4.3e+000	1.4e-018	3.6e-009	
FB0.75	7	16	7	0	1	0	5.0e+000	4.3e-020	6.2e-010	
FB0.5	6	9	6	0	1	0	7.9e+000	3.3e-025	9.1e-013	
FB0.2	5	7	5	0	1	0	1.4e+001	8.2e-017	6.8e-009	

$x_0 = x^4_{cpstart} = (3.500000, 6.500000, 0.500000, 5.500000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	16	26	10	6	9	0	7.3e+001	4.5e-021	7.1e-010	D4
FB	8	10	8	0	1	0	2.7e+002	4.9e-018	1.3e-008	
FBP0.95	8	9	8	0	0	0	2.5e+002	2.6e-018	8.8e-009	
FB0.75	8	9	8	0	0	0	1.7e+002	2.4e-018	6.6e-009	
FB0.5	7	8	7	0	0	0	1.3e+002	3.0e-021	1.6e-010	
FB0.2	7	8	7	0	0	0	1.6e+002	2.9e-020	1.9e-010	

$x_0 = x^7_{cpstart} = (1.000000, 1.000000, 1.000000, 1.000000)$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	0	1	0	0	0	0	3.8e+000	3.8e+000	2.8e+001	$\nexists\nabla\theta(x_f)$
FB	6	7	6	0	0	0	1.1e+001	2.5e-024	9.1e-012	
FBP0.95	7	10	7	0	2	0	1.0e+001	6.3e-023	4.3e-011	
FB0.75	8	9	8	0	0	0	6.8e+000	2.3e-017	3.8e-008	
FB0.5	7	8	7	0	0	0	3.8e+000	1.3e-021	1.9e-010	
FB0.2	6	7	6	0	0	0	2.2e+000	3.3e-013	1.2e-006	

Tabela 5.33: Problema `mathisum` : $n = 4$ (formulação NCP)

$x_0 = x_{mcplib}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	3	4	3	0	0	0	2.0e+001	1.7e-017	2.6e-007	
FB	3	4	3	0	0	3	2.0e+001	1.7e-017	2.6e-007	
FBP0.95	3	4	3	0	0	3	2.0e+001	1.7e-017	2.6e-007	
FB0.75	3	4	3	0	0	3	2.0e+001	1.7e-017	2.6e-007	
FB0.5	3	4	3	0	0	3	2.0e+001	1.7e-017	2.6e-007	
FB0.2	3	4	3	0	0	3	2.0e+001	1.7e-017	2.6e-007	

$x_0 = x^1_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	6	7	6	0	0	0	1.0e+004	1.9e-017	1.8e-007	
FB	6	7	6	0	0	6	1.0e+004	1.9e-017	1.8e-007	
FBP0.95	6	7	6	0	0	6	1.0e+004	1.9e-017	1.8e-007	
FB0.75	6	7	6	0	0	6	1.0e+004	1.9e-017	1.8e-007	
FB0.5	6	7	6	0	0	6	1.0e+004	1.9e-017	1.8e-007	
FB0.2	6	7	6	0	0	6	1.0e+004	1.9e-017	1.8e-007	

$x_0 = x^2_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	6	7	6	0	0	0	2.7e+005	4.4e-018	2.3e-007	
FB	6	7	6	0	0	6	2.7e+005	4.4e-018	2.3e-007	
FBP0.95	6	7	6	0	0	6	2.7e+005	4.4e-018	2.3e-007	
FB0.75	6	7	6	0	0	6	2.7e+005	4.4e-018	2.3e-007	
FB0.5	6	7	6	0	0	6	2.7e+005	4.4e-018	2.3e-007	
FB0.2	6	7	6	0	0	6	2.7e+005	4.4e-018	2.3e-007	

$x_0 = x^3_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	5	6	5	0	0	0	1.9e+003	5.6e-016	9.9e-007	
FB	5	6	5	0	0	5	1.9e+003	5.6e-016	9.9e-007	
FBP0.95	5	6	5	0	0	5	1.9e+003	5.6e-016	9.9e-007	
FB0.75	5	6	5	0	0	5	1.9e+003	5.6e-016	9.9e-007	
FB0.5	5	6	5	0	0	5	1.9e+003	5.6e-016	9.9e-007	
FB0.2	5	6	5	0	0	5	1.9e+003	5.6e-016	9.9e-007	

Tabela 5.34: Problema `nash` : $n = 10, \quad p = 0, \quad q = 10$ (formulação KKT)

$x_0 = x_{mcplib}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	5	12	5	0	2	0	9.8e+000	5.4e-020	1.5e-008	
FB	5	8	5	0	1	0	3.1e+001	1.1e-013	9.4e-006	
FBP0.95	5	6	5	0	0	0	2.9e+001	1.2e-014	4.8e-006	
FB0.75	5	6	5	0	0	0	2.4e+001	3.9e-017	5.0e-007	
FB0.5	4	5	4	0	0	0	2.8e+001	2.4e-013	1.3e-005	
FB0.2	4	5	4	0	0	0	4.6e+001	9.4e-020	1.3e-009	

$x_0 = x^1_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	6	7	6	0	0	0	1.0e+004	1.9e-017	1.8e-007	
FB	8	9	8	0	0	0	4.1e+004	9.6e-021	1.0e-008	
FBP0.95	8	9	8	0	0	0	3.7e+004	8.7e-021	9.4e-009	
FB0.75	8	9	8	0	0	0	2.3e+004	5.4e-021	5.9e-009	
FB0.5	8	9	8	0	0	0	1.0e+004	2.4e-021	2.6e-009	
FB0.2	8	9	8	0	0	0	1.6e+003	3.8e-022	4.2e-010	

$x_0 = x^2_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	10	28	10	0	8	0	4.5e+001	6.6e-013	5.8e-005	
FB	11	25	11	0	7	0	4.0e+001	1.8e-023	2.1e-010	
FBP0.95	11	13	11	0	1	0	6.8e+004	1.9e-015	4.4e-006	
FB0.75	11	12	11	0	0	0	1.7e+006	4.1e-017	5.1e-007	
FB0.5	10	11	10	0	0	0	6.6e+006	1.1e-017	1.8e-007	
FB0.2	10	11	10	0	0	0	1.7e+007	1.7e-020	2.8e-009	

$x_0 = x^3_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	6	8	6	0	1	0	1.9e+003	4.1e-024	8.5e-011	D1
FB	7	8	7	0	0	0	7.5e+003	8.6e-019	9.9e-008	
FBP0.95	7	8	7	0	0	0	6.7e+003	1.0e-019	3.3e-008	
FB0.75	7	8	7	0	0	0	4.2e+003	4.8e-020	1.7e-008	
FB0.5	7	8	7	0	0	0	1.9e+003	2.0e-020	7.5e-009	
FB0.2	7	8	7	0	0	0	3.2e+002	3.1e-021	1.2e-009	

Tabela 5.35: Problema `nash` : $n = 10$ (formulação NCP)

$x_0 = x_{mcplib}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	1	6	1	0	1	1	6.9e-003	6.5e-003	1.4e+000	$\nexists\nabla\theta(x_f)$
FB	6	14	6	0	2	3	6.9e-003	4.3e-016	4.5e-008	
FBP0.95	7	14	7	0	2	7	6.9e-003	7.1e-018	3.6e-009	
FB0.75	8	22	7	1	3	7	6.9e-003	4.4e-018	2.2e-009	
FB0.5	7	13	7	0	2	7	6.9e-003	2.0e-018	1.0e-009	
FB0.2	6	10	6	0	2	6	6.9e-003	2.4e-013	3.3e-006	

$x_0 = x^1_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	1	2	1	0	0	1	1.1e+000	7.8e-001	1.2e+000	$\nexists\nabla\theta(x_f)$
FB	9	14	8	1	3	3	1.1e+000	2.0e-016	2.6e-007	
FBP0.95	9	14	8	1	3	8	1.1e+000	2.0e-016	2.6e-007	
FB0.75	10	17	8	2	4	7	1.1e+000	7.3e-018	4.7e-008	
FB0.5	8	10	8	0	1	8	1.1e+000	2.2e-019	8.1e-009	
FB0.2	196	1581	4	192	192	1	1.1e+000	1.6e-011	9.9e-007	$\|\nabla\theta_f\| \leq \varepsilon_g$, D2

$x_0 = x^2_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	0	1	0	0	0	0	2.4e+001	2.4e+001	1.5e+002	$\nexists\nabla\theta(x_f)$
FB	300	2246	0	300	300	1	2.4e+001	1.8e-005	1.1e-002	$k \geq k_{max}$, D2
FBP0.95	300	2409	4	296	296	1	2.4e+001	8.6e-011	5.1e-006	$k \geq k_{max}$, D1
FB0.75	300	2439	4	296	296	1	2.4e+001	2.5e-011	9.8e-006	$k \geq k_{max}$, D1
FB0.5	300	2433	4	296	296	1	2.4e+001	3.0e-011	3.3e-006	$k \geq k_{max}$, D1
FB0.2	300	2430	4	296	296	1	2.4e+001	1.4e-011	1.4e-006	$k \geq k_{max}$, D1

$x_0 = x^3_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	0	1	0	0	0	0	5.9e+003	5.9e+003	8.0e+003	$\nexists\nabla\theta(x_f)$
FB	300	1726	0	300	300	2	5.9e+003	2.7e-005	2.6e-002	$k \geq k_{max}$, D3
FBP0.95	300	2224	4	296	296	2	5.9e+003	8.9e-011	8.6e-006	$k \geq k_{max}$, D4
FB0.75	51	176	5	46	46	2	5.9e+003	7.7e-013	1.5e-005	D3
FB0.5	53	181	5	48	48	2	5.9e+003	8.2e-013	1.6e-005	D3
FB0.2	300	2260	4	296	296	2	5.9e+003	7.2e-011	9.8e-006	$k \geq k_{max}$, D4

$x_0 = x^4_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	0	1	0	0	0	0	3.1e+002	3.1e+002	9.1e+002	$\nexists\nabla\theta(x_f)$
FB	300	2205	0	300	300	1	3.1e+002	5.0e-005	8.2e-003	$k \geq k_{max}$, D3
FBP0.95	300	2235	4	296	296	1	3.1e+002	2.6e-010	7.0e-005	$k \geq k_{max}$, D5
FB0.75	31	110	5	26	26	1	3.1e+002	3.2e-013	1.1e-005	D3
FB0.5	29	98	5	24	24	1	3.1e+002	2.4e-013	9.3e-006	D3
FB0.2	300	2351	4	296	296	1	3.1e+002	4.6e-010	2.7e-005	$k \geq k_{max}$, D5

Tabela 5.36: Problema `powell` : $n = 16, \quad p = 0, \quad q = 16$ (formulação KKT)

$x_0 = x_{mcplib}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	2045	1	299	299	0	3.5e-003	3.8e-007	2.5e-003	$k \geq k_{max}$
FB	5	6	5	0	0	0	4.3e-003	5.0e-016	2.7e-007	
FBP0.95	5	6	5	0	0	0	3.8e-003	4.5e-016	2.4e-007	
FB0.75	5	6	5	0	0	0	2.4e-003	2.8e-016	1.5e-007	
FB0.5	5	6	5	0	0	0	1.1e-003	1.3e-016	6.6e-008	
FB0.2	5	6	5	0	0	0	1.7e-004	2.0e-017	1.1e-008	

$x_0 = x^1_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	2186	4	296	298	0	5.6e-001	4.1e-007	1.2e-003	$k \geq k_{max}$
FB	8	11	8	0	2	0	2.3e+000	2.7e-016	2.1e-007	
FBP0.95	8	11	8	0	2	0	2.0e+000	2.4e-016	1.9e-007	
FB0.75	8	11	8	0	2	0	1.3e+000	1.6e-016	1.2e-007	
FB0.5	9	11	9	0	1	0	5.6e-001	5.8e-013	4.9e-006	
FB0.2	10	12	10	0	1	0	9.0e-002	8.0e-016	7.9e-008	

$x_0 = x^2_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	2319	0	300	300	0	1.2e+001	6.1e-006	1.1e-002	$k \geq k_{max}$, D2
FB	9	19	8	1	2	0	4.7e+001	8.9e-015	2.1e-006	D1
FBP0.95	9	18	8	1	1	0	4.3e+001	6.5e-019	9.7e-009	D1
FB0.75	11	21	9	2	2	0	2.7e+001	2.8e-013	2.4e-006	D1
FB0.5	9	18	7	2	2	0	1.2e+001	2.1e-013	1.5e-006	D1
FB0.2	9	16	7	2	2	0	1.9e+000	1.2e-016	5.0e-008	D1

$x_0 = x^3_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	0	1	0	0	0	0	2.9e+003	2.9e+003	4.0e+003	$\nexists\nabla\theta(x_f)$
FB	13	28	12	1	2	1	1.2e+004	3.7e-017	8.1e-008	D1
FBP0.95	10	24	8	2	2	1	1.1e+004	3.6e-013	2.1e-006	D1
FB0.75	70	549	13	57	57	1	6.6e+003	9.9e-013	1.3e-005	D1
FB0.5	300	3014	10	290	290	1	2.9e+003	7.5e-012	9.6e-006	$k \geq k_{max}$, D1
FB0.2	300	2336	6	294	294	1	4.7e+002	1.7e-008	6.9e-004	$k \geq k_{max}$, D1

$x_0 = x^4_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	2145	0	300	300	0	1.5e+002	4.3e-006	8.7e-003	$k \geq k_{max}$, D3
FB	11	21	10	1	1	0	6.1e+002	1.1e-017	4.4e-008	D1
FBP0.95	11	21	10	1	1	0	5.5e+002	2.1e-019	6.4e-009	D1
FB0.75	11	27	9	2	3	0	3.4e+002	1.2e-018	1.1e-008	D1
FB0.5	27	175	8	19	19	0	1.5e+002	8.7e-013	1.1e-005	D1
FB0.2	300	2756	8	292	292	0	2.4e+001	5.8e-012	7.4e-006	$k \geq k_{max}$, D1

Tabela 5.37: Problema `powell` : $n = 16$ (formulação NCP)

$x_0 = x_{mcplib}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	2	3	2	0	0	1	6.9e-002	7.8e-024	6.2e-011	
FB	2	3	2	0	0	2	6.9e-002	7.8e-024	6.2e-011	
FBP0.95	2	3	2	0	0	2	6.9e-002	7.8e-024	6.2e-011	
FB0.75	2	3	2	0	0	2	6.9e-002	7.8e-024	6.2e-011	
FB0.5	2	3	2	0	0	2	6.9e-002	7.8e-024	6.2e-011	
FB0.2	2	3	2	0	0	2	6.9e-002	7.8e-024	6.2e-011	

$x_0 = x^1_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	0	1	0	0	0	0	1.3e+001	1.3e+001	1.5e+010	$\nexists\nabla\theta(x_f)$
FB	49	115	11	38	39	1	1.3e+001	5.6e-018	3.5e-009	
FBP0.95	49	115	11	38	39	1	1.3e+001	5.0e-018	3.1e-009	
FB0.75	49	115	11	38	39	1	1.3e+001	3.1e-018	2.1e-009	
FB0.5	49	114	11	38	39	1	1.3e+001	1.4e-018	1.2e-009	
FB0.2	123	785	11	112	114	1	1.3e+001	2.2e-013	8.4e-007	

$x_0 = x^2_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	0	1	0	0	0	0	1.4e+001	1.4e+001	4.9e+009	$\nexists\nabla\theta(x_f)$
FB	51	116	14	37	39	1	1.4e+001	1.3e-017	5.2e-009	
FBP0.95	51	116	14	37	39	1	1.4e+001	1.1e-017	4.7e-009	
FB0.75	51	116	14	37	39	1	1.4e+001	7.1e-018	3.1e-009	
FB0.5	51	114	14	37	38	3	1.4e+001	1.4e-018	1.2e-009	
FB0.2	50	111	13	37	38	2	1.4e+001	2.2e-013	8.4e-007	

$x_0 = x^3_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	0	1	0	0	0	0	5.3e+001	5.3e+001	1.3e+011	$\nexists\nabla\theta(x_f)$
FB	52	104	15	37	39	1	5.3e+001	5.6e-018	3.5e-009	
FBP0.95	52	103	15	37	39	1	5.3e+001	5.0e-018	3.1e-009	
FB0.75	52	100	15	37	39	1	5.3e+001	3.1e-018	2.1e-009	
FB0.5	54	100	17	37	39	2	5.3e+001	3.1e-018	1.8e-009	
FB0.2	50	94	13	37	39	1	5.3e+001	2.3e-013	8.4e-007	

$x_0 = x^4_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	0	1	0	0	0	0	3.3e+001	3.3e+001	1.2e+011	$\nexists\nabla\theta(x_f)$
FB	54	108	17	37	40	1	3.3e+001	5.6e-018	3.5e-009	
FBP0.95	54	108	17	37	40	1	3.3e+001	5.0e-018	3.1e-009	
FB0.75	54	102	17	37	39	1	3.3e+001	3.1e-018	2.1e-009	
FB0.5	52	100	15	37	40	1	3.3e+001	3.1e-018	1.8e-009	
FB0.2	51	97	14	37	40	1	3.3e+001	3.4e-013	1.0e-006	

Tabela 5.38: Problema `scarfanum` : $n = 13, \quad p = 0, \quad q = 13$ (formulação KKT)

$x_0 = x_{mcplib}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	2	3	2	0	0	0	2.4e-005	7.8e-024	6.2e-011	
FB	4	5	4	0	0	0	2.4e-005	1.8e-016	1.8e-007	
FBP0.95	4	5	4	0	0	0	2.2e-005	1.6e-016	1.6e-007	
FB0.75	4	5	4	0	0	0	1.5e-005	9.6e-017	9.7e-008	
FB0.5	4	5	4	0	0	0	8.8e-006	3.7e-017	4.1e-008	
FB0.2	3	4	3	0	0	0	5.8e-006	8.5e-013	2.6e-006	

$x_0 = x^1_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	2082	0	300	300	0	8.1e+000	6.8e-002	4.9e-001	$k \geq k_{max}$
FB	87	201	25	62	64	0	2.9e+001	5.6e-018	3.1e-008	
FBP0.95	97	808	38	59	77	0	2.6e+001	1.2e-017	4.3e-008	
FB0.75	15	21	14	1	4	0	1.6e+001	9.1e-013	3.7e-005	
FB0.5	14	22	13	1	5	0	7.3e+000	5.7e-013	3.8e-006	
FB0.2	100	723	36	64	80	0	1.2e+000	3.4e-013	1.6e-006	

$x_0 = x^2_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	2162	0	300	300	0	8.7e+000	7.0e-002	4.5e-001	$k \geq k_{max}$
FB	80	488	28	52	62	0	3.2e+001	5.6e-018	3.1e-008	
FBP0.95	93	665	34	59	74	0	2.9e+001	1.2e-017	4.3e-008	
FB0.75	98	726	39	59	77	0	1.8e+001	3.1e-018	1.8e-008	
FB0.5	96	703	37	59	76	0	7.9e+000	3.2e-018	1.2e-008	
FB0.2	93	731	36	57	74	0	1.3e+000	3.4e-013	1.6e-006	

$x_0 = x^3_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	0	1	0	0	0	0	4.0e+001	4.0e+001	1.3e+010	$\nexists\nabla\theta(x_f)$
FB	11	15	10	1	3	0	1.5e+002	1.0e-022	5.3e-010	
FBP0.95	16	26	15	1	8	0	1.4e+002	1.9e-016	6.8e-007	
FB0.75	9	12	8	1	2	0	8.7e+001	6.7e-013	3.2e-005	
FB0.5	300	4045	266	34	299	0	4.0e+001	2.7e-003	3.4e+005	$k \geq k_{max}$
FB0.2	7	11	6	1	2	0	9.2e+000	2.7e-012	4.3e-007	$\|\nabla\theta_f\| \leq \varepsilon_g$

$x_0 = x^4_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	1601	0	300	300	0	2.3e+001	7.8e-002	1.2e+000	$k \geq k_{max}$
FB	13	18	12	1	4	0	8.5e+001	4.2e-023	3.8e-010	
FBP0.95	104	853	45	59	87	0	7.7e+001	9.2e-015	4.8e-006	
FB0.75	10	13	9	1	2	0	4.8e+001	9.0e-013	3.7e-005	
FB0.5	10	13	9	1	2	0	2.3e+001	4.0e-013	1.7e-005	
FB0.2	10	13	9	1	2	0	8.1e+000	6.4e-014	2.6e-006	

Tabela 5.39: Problema `scarfanum` : $n = 13$ (formulação NCP)

$x_0 = x_{mcplib}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	2	3	2	0	0	1	1.0e-001	7.8e-024	5.9e-011	
FB	2	3	2	0	0	2	1.0e-001	7.8e-024	5.9e-011	
FBP0.95	2	3	2	0	0	2	1.0e-001	7.8e-024	5.9e-011	
FB0.75	2	3	2	0	0	2	1.0e-001	7.8e-024	5.9e-011	
FB0.5	2	3	2	0	0	2	1.0e-001	7.8e-024	5.9e-011	
FB0.2	2	3	2	0	0	2	1.0e-001	7.8e-024	5.9e-011	

$x_0 = x^1_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	0	1	0	0	0	0	7.3e+001	7.3e+001	2.2e+011	$\nexists\nabla\theta(x_f)$
FB	53	99	13	40	41	1	7.3e+001	1.2e-017	5.0e-009	
FBP0.95	53	99	13	40	41	1	7.3e+001	1.1e-017	4.5e-009	
FB0.75	53	99	13	40	41	2	7.3e+001	6.6e-018	2.9e-009	
FB0.5	53	98	13	40	41	1	7.3e+001	2.9e-018	1.6e-009	
FB0.2	52	96	12	40	41	2	7.3e+001	3.1e-013	8.7e-007	

$x_0 = x^2_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	0	1	0	0	0	0	8.0e+001	8.0e+001	1.9e+011	$\nexists\nabla\theta(x_f)$
FB	56	105	14	42	43	1	8.0e+001	1.2e-017	5.0e-009	
FBP0.95	56	105	14	42	43	1	8.0e+001	1.1e-017	4.5e-009	
FB0.75	56	105	14	42	43	1	8.0e+001	6.6e-018	2.9e-009	
FB0.5	56	104	14	42	43	1	8.0e+001	2.9e-018	1.6e-009	
FB0.2	54	101	12	42	43	1	8.0e+001	3.1e-013	8.7e-007	

$x_0 = x^3_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	0	1	0	0	0	0	5.6e+001	5.6e+001	2.0e+012	$\nexists\nabla\theta(x_f)$
FB	69	138	13	56	57	1	5.6e+001	1.2e-017	5.0e-009	
FBP0.95	69	137	13	56	57	1	5.6e+001	1.1e-017	4.6e-009	
FB0.75	69	137	13	56	57	1	5.6e+001	6.8e-018	3.0e-009	
FB0.5	69	137	13	56	57	1	5.6e+001	3.0e-018	1.7e-009	
FB0.2	66	127	9	57	58	1	5.6e+001	3.2e-013	8.8e-007	

$x_0 = x^4_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	0	1	0	0	0	0	8.9e+001	8.9e+001	1.6e+011	$\nexists\nabla\theta(x_f)$
FB	46	91	11	35	36	1	8.9e+001	1.2e-017	5.0e-009	
FBP0.95	46	91	11	35	36	1	8.9e+001	1.1e-017	4.6e-009	
FB0.75	46	90	11	35	36	1	8.9e+001	6.8e-018	3.0e-009	
FB0.5	46	90	11	35	36	1	8.9e+001	3.0e-018	1.7e-009	
FB0.2	45	87	10	35	36	1	8.9e+001	3.2e-013	8.8e-007	

Tabela 5.40: Problema `scarfasum` : $n = 14, \quad p = 0, \quad q = 14$ (formulação KKT)

$x_0 = x_{mcplib}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	2	3	2	0	0	0	2.2e-005	7.8e-024	5.9e-011	
FB	4	5	4	0	0	0	2.3e-005	1.7e-016	1.8e-007	
FBP0.95	4	5	4	0	0	0	2.1e-005	1.5e-016	1.6e-007	
FB0.75	4	5	4	0	0	0	1.4e-005	9.0e-017	9.9e-008	
FB0.5	4	5	4	0	0	0	8.2e-006	3.5e-017	4.1e-008	
FB0.2	3	4	3	0	0	0	5.4e-006	7.9e-013	2.5e-006	

$x_0 = x^1_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	125	562	10	115	117	0	6.9e+001	2.5e-020	3.3e-009	
FB	14	19	13	1	3	0	2.6e+002	5.2e-018	3.2e-008	
FBP0.95	15	19	14	1	2	0	2.4e+002	1.1e-017	4.3e-008	
FB0.75	16	21	15	1	3	0	1.5e+002	6.8e-018	2.7e-008	
FB0.5	16	21	15	1	3	0	6.5e+001	3.0e-018	1.2e-008	
FB0.2	15	20	14	1	3	0	1.0e+001	3.2e-013	1.6e-006	

$x_0 = x^2_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	2421	0	300	300	0	7.5e+001	7.3e-002	4.7e-001	$k \geq k_{max}$
FB	14	19	13	1	3	0	2.9e+002	5.2e-018	3.2e-008	
FBP0.95	15	19	14	1	2	0	2.6e+002	1.1e-017	4.3e-008	
FB0.75	16	21	15	1	3	0	1.6e+002	6.8e-018	2.7e-008	
FB0.5	16	21	15	1	3	0	7.2e+001	3.0e-018	1.2e-008	
FB0.2	15	20	14	1	3	0	1.1e+001	3.2e-013	1.6e-006	

$x_0 = x^3_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	100	200	2	98	99	0	5.1e+001	8.3e-001	1.8e+011	$\nexists\nabla\theta(x_f)$
FB	98	178	20	78	79	0	1.9e+002	1.2e-017	4.7e-008	
FBP0.95	98	178	21	77	79	0	1.7e+002	1.1e-017	4.3e-008	
FB0.75	101	184	21	80	82	0	1.1e+002	2.9e-018	1.8e-008	
FB0.5	101	206	23	78	83	0	4.8e+001	1.3e-018	7.9e-009	
FB0.2	99	179	20	79	79	0	7.8e+000	3.2e-013	1.6e-006	

$x_0 = x^4_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	85	178	10	75	80	0	8.1e+001	1.9e-023	9.1e-011	
FB	15	20	14	1	4	0	3.1e+002	2.8e-013	1.0e-005	
FBP0.95	16	23	15	1	6	0	2.8e+002	2.9e-023	7.3e-011	
FB0.75	15	20	14	1	4	0	1.7e+002	1.6e-013	5.7e-006	
FB0.5	93	603	34	59	73	0	7.8e+001	3.0e-018	1.2e-008	
FB0.2	18	25	17	1	6	0	1.2e+001	4.0e-017	2.5e-008	

Tabela 5.41: Problema `scarfasum` : $n = 14$ (formulação NCP)

$x_0 = x_{mcplib}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	0	1	0	0	0	0	9.2e+001	9.2e+001	3.2e+004	$\nexists\nabla\theta(x_f)$
FB	300	4616	0	300	300	2	9.2e+001	2.4e+001	3.4e+001	$k \geq k_{max}$
FBP0.95	300	4616	0	300	300	2	9.2e+001	2.4e+001	3.4e+001	$k \geq k_{max}$
FB0.75	300	4616	0	300	300	2	9.2e+001	2.4e+001	3.4e+001	$k \geq k_{max}$
FB0.5	300	4616	0	300	300	2	9.2e+001	2.4e+001	3.4e+001	$k \geq k_{max}$
FB0.2	300	4616	0	300	300	2	9.2e+001	2.4e+001	3.4e+001	$k \geq k_{max}$

Tabela 5.42: Problema `scarfbsum` : $n = 40, \quad p = 0, \quad q = 40$ (formulação KKT)

$x_0 = x_{mcplib}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	3716	0	300	300	0	6.8e+001	2.6e-003	5.2e-001	$k \geq k_{max}$
FB	300	3040	67	233	298	0	2.7e+002	3.7e-004	1.1e+000	$k \geq k_{max}$
FBP0.95	117	693	115	2	110	0	2.4e+002	1.5e-013	8.8e-004	
FB0.75	300	2999	86	214	298	0	1.5e+002	1.0e-004	3.5e-001	$k \geq k_{max}$
FB0.5	300	2810	78	222	298	0	6.8e+001	8.4e-005	3.8e-001	$k \geq k_{max}$
FB0.2	300	2657	96	204	298	0	1.1e+001	5.7e-006	1.5e-002	$k \geq k_{max}$

Tabela 5.43: Problema `scarfbsum` : $n = 40$ (formulação NCP)

$x_0 = x_{mcplib}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	0	1	0	0	0	0	3.9e+000	3.9e+000	5.9e+000	$\nexists\nabla\theta(x_f)$
FB	4	6	3	1	1	4	3.9e+000	2.9e-014	6.0e-007	
FBP0.95	4	6	3	1	1	4	3.9e+000	2.5e-014	5.6e-007	
FB0.75	4	6	3	1	1	4	3.9e+000	1.9e-014	4.9e-007	
FB0.5	4	6	3	1	1	4	3.9e+000	1.6e-014	4.5e-007	
FB0.2	4	6	3	1	1	4	3.9e+000	1.5e-014	4.4e-007	

$x_0 = x^1_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	3	11	3	0	1	2	1.1e+002	3.5e+001	8.4e+000	$\nexists\nabla\theta(x_f)$
FB	9	16	9	0	3	9	1.1e+002	1.4e-013	5.9e-007	
FBP0.95	9	16	9	0	3	9	1.1e+002	1.3e-013	5.4e-007	
FB0.75	9	13	9	0	2	9	1.1e+002	7.5e-014	3.7e-007	
FB0.5	9	11	9	0	1	9	1.1e+002	3.5e-014	2.2e-007	
FB0.2	10	13	10	0	2	10	1.1e+002	9.1e-015	1.7e-007	

$x_0 = x^2_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	1139	2	298	298	0	1.4e+002	6.2e-005	4.8e-003	$k \geq k_{max}$
FB	300	1133	6	294	296	8	1.4e+002	1.0e-010	7.0e-006	$k \geq k_{max}$
FBP0.95	300	1132	6	294	295	8	1.4e+002	3.2e-011	3.6e-006	$k \geq k_{max}$
FB0.75	28	91	6	22	22	8	1.4e+002	1.3e-012	9.0e-007	$\|\nabla\theta_f\| \leq \varepsilon_g$
FB0.5	23	83	6	17	17	8	1.4e+002	2.0e-012	8.8e-007	$\|\nabla\theta_f\| \leq \varepsilon_g$
FB0.2	26	116	6	20	20	8	1.4e+002	8.6e-013	1.1e-006	

$x_0 = x^3_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	1140	4	296	298	1	2.7e+003	1.4e-004	1.1e-002	$k \geq k_{max}$
FB	37	125	8	29	30	10	2.7e+003	9.7e-013	1.2e-006	
FBP0.95	37	125	8	29	30	10	2.7e+003	9.9e-013	1.4e-006	
FB0.75	29	91	8	21	22	10	2.7e+003	1.5e-012	9.2e-007	$\|\nabla\theta_f\| \leq \varepsilon_g$
FB0.5	48	180	8	40	42	10	2.7e+003	3.1e-012	9.8e-007	$\|\nabla\theta_f\| \leq \varepsilon_g$
FB0.2	27	63	9	18	20	9	2.7e+003	9.9e-013	2.4e-006	

Tabela 5.44: Problema `sppe` : $n = 27, \quad p = 0, \quad q = 27$ (formulação KKT)

$x_0 = x_{mcplib}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	300	1117	0	300	300	0	1.3e+000	5.7e-006	1.5e-003	$k \geq k_{max}$
FB	6	9	4	2	2	0	3.0e+000	2.2e-013	5.5e-007	
FBP0.95	8	13	4	4	4	0	2.8e+000	6.7e-017	1.5e-008	
FB0.75	10	34	4	6	6	0	2.1e+000	2.4e-016	2.3e-008	
FB0.5	27	156	4	23	23	0	2.3e+000	8.3e-022	3.4e-011	
FB0.2	36	257	3	33	33	0	3.7e+000	2.6e-012	9.7e-007	$\|\nabla\theta_f\| \leq \varepsilon_g$

$x_0 = x^1_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	2	4	2	0	1	2	1.1e+002	1.8e+001	6.7e+000	$\nexists\nabla\theta(x_f)$
FB	8	9	8	0	0	3	4.4e+002	1.9e-015	8.7e-008	
FBP0.95	10	13	8	2	2	4	4.0e+002	1.8e-016	1.6e-008	
FB0.75	10	20	8	2	3	7	2.5e+002	2.2e-015	6.8e-008	
FB0.5	10	19	8	2	3	5	1.1e+002	1.5e-015	4.0e-008	
FB0.2	12	25	10	2	5	5	1.8e+001	6.9e-016	7.4e-008	

$x_0 = x^2_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	8	12	8	0	2	7	5.2e+001	3.4e-015	1.2e-007	
FB	6	7	6	0	0	0	1.1e+002	2.1e-022	4.2e-011	
FBP0.95	6	7	6	0	0	0	1.5e+002	6.4e-024	5.5e-012	
FB0.75	6	7	6	0	0	0	7.9e+002	4.5e-014	6.0e-007	
FB0.5	7	8	7	0	0	0	2.7e+003	1.2e-013	6.9e-007	
FB0.2	8	9	8	0	0	0	6.6e+003	2.8e-016	3.1e-008	

$x_0 = x^3_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	5	12	5	0	2	3	2.4e+002	1.2e+001	6.1e+000	$\nexists\nabla\theta(x_f)$
FB	8	9	8	0	0	0	5.9e+002	1.3e-015	7.4e-008	
FBP0.95	8	9	8	0	0	0	5.1e+003	2.3e-024	2.7e-012	
FB0.75	9	10	9	0	0	0	9.9e+004	1.0e-021	1.3e-010	
FB0.5	9	10	9	0	0	0	3.9e+005	3.6e-013	1.2e-006	
FB0.2	10	11	10	0	0	0	9.8e+005	5.2e-016	4.2e-008	

Tabela 5.45: Problema `sppe` : $n = 27$ (formulação NCP)

$x_0 = x_{mcplib}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	2	3	2	0	0	1	8.3e+000	5.7e-021	3.2e-010	
FB	2	3	2	0	0	2	8.3e+000	5.7e-021	3.2e-010	
FBP0.95	2	3	2	0	0	2	8.3e+000	5.7e-021	3.2e-010	
FB0.75	2	3	2	0	0	2	8.3e+000	5.7e-021	3.2e-010	
FB0.5	2	3	2	0	0	2	8.3e+000	5.7e-021	3.2e-010	
FB0.2	2	3	2	0	0	2	8.3e+000	5.7e-021	3.2e-010	

$x_0 = x^1_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	16	38	12	4	10	1	1.8e+002	4.2e-015	2.0e-007	
FB	9	14	9	0	2	9	1.8e+002	9.0e-015	2.2e-007	
FBP0.95	9	14	9	0	2	9	1.8e+002	8.5e-015	2.2e-007	
FB0.75	9	13	9	0	2	9	1.8e+002	9.1e-016	5.5e-008	
FB0.5	9	12	9	0	1	9	1.8e+002	2.5e-015	1.6e-007	
FB0.2	9	14	9	0	2	9	1.8e+002	5.6e-015	4.5e-008	

$x_0 = x^2_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	34	61	34	0	10	0	1.8e+002	5.5e-019	2.5e-009	
FB	21	42	21	0	10	2	1.8e+002	1.5e-016	2.7e-008	
FBP0.95	21	42	21	0	10	21	1.8e+002	1.4e-016	2.6e-008	
FB0.75	15	31	15	0	7	15	1.8e+002	2.6e-013	1.0e-006	
FB0.5	15	31	15	0	7	15	1.8e+002	1.3e-013	8.9e-007	
FB0.2	50	550	49	1	43	38	1.8e+002	2.7e-017	2.2e-008	

$x_0 = x^3_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	11	33	11	0	5	0	1.8e+002	5.1e-017	2.4e-008	
FB	15	44	15	0	8	4	1.8e+002	1.7e-014	2.3e-007	
FBP0.95	15	44	15	0	8	15	1.8e+002	1.6e-014	2.2e-007	
FB0.75	15	45	15	0	9	15	1.8e+002	3.5e-017	7.5e-009	
FB0.5	15	45	15	0	9	15	1.8e+002	1.6e-017	4.4e-009	
FB0.2	33	250	32	1	28	26	1.8e+002	5.1e-014	9.5e-007	

$x_0 = x^4_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	25	78	22	3	18	1	2.2e+003	3.8e-017	1.9e-008	
FB	113	1005	108	5	95	13	2.2e+003	6.9e-016	4.0e-008	
FBP0.95	188	1888	178	10	171	48	2.2e+003	3.7e-013	9.0e-007	
FB0.75	300	3530	285	15	288	65	2.2e+003	2.0e+000	6.1e+000	$k \geq k_{max}$
FB0.5	51	462	50	1	36	42	2.2e+003	1.1e-015	6.0e-008	
FB0.2	232	1999	230	2	219	34	2.2e+003	4.1e-015	8.2e-008	

Tabela 5.46: Problema `tobin` : $n = 42, \quad p = 0, \quad q = 42$ (formulação KKT)

$x_0 = x_{mcplib}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	2	3	2	0	0	0	1.6e-003	5.7e-021	3.1e-010	
FB	2	3	2	0	0	0	1.6e-003	6.2e-018	1.7e-008	
FBP0.95	2	3	2	0	0	0	2.1e-003	6.8e-018	1.8e-008	
FB0.75	2	3	2	0	0	0	8.3e-003	2.6e-017	1.6e-007	
FB0.5	2	3	2	0	0	0	2.6e-002	9.9e-017	6.4e-007	
FB0.2	2	3	2	0	0	0	6.2e-002	2.7e-016	1.6e-006	

$x_0 = x^1_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	40	111	36	4	22	0	1.8e+002	6.9e-018	7.2e-009	
FB	10	13	10	0	2	0	7.3e+002	3.9e-020	2.6e-010	
FBP0.95	7	10	7	0	1	0	6.6e+002	3.2e-014	1.0e-006	
FB0.75	8	12	8	0	2	0	4.1e+002	3.5e-014	6.8e-007	
FB0.5	8	13	8	0	3	0	1.8e+002	1.4e-016	1.5e-007	
FB0.2	9	16	9	0	5	0	2.9e+001	4.7e-016	9.8e-007	

$x_0 = x^2_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	36	117	34	2	19	0	1.8e+002	4.1e-015	1.7e-007	
FB	8	10	8	0	1	0	7.0e+002	9.9e-016	4.5e-008	
FBP0.95	8	12	8	0	2	0	6.3e+002	2.1e-013	8.9e-007	
FB0.75	8	13	8	0	2	0	4.0e+002	2.1e-013	3.6e-006	
FB0.5	8	12	8	0	2	0	1.8e+002	1.6e-015	5.1e-007	
FB0.2	7	12	7	0	3	0	3.0e+001	2.4e-016	9.1e-007	

$x_0 = x^3_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	20	39	20	0	9	0	1.8e+002	3.0e-015	1.7e-007	
FB	9	11	9	0	1	0	6.9e+002	2.6e-014	4.3e-007	
FBP0.95	10	16	10	0	4	0	6.3e+002	1.0e-024	6.1e-012	
FB0.75	10	19	10	0	4	0	3.9e+002	4.5e-014	6.6e-007	
FB0.5	9	15	9	0	3	0	1.8e+002	2.0e-018	1.8e-008	
FB0.2	8	15	8	0	4	0	3.3e+001	8.5e-016	2.6e-006	

$x_0 = x^4_{cpstart}$										
NCP	k	F-eval	Newt.	Grad	BckT.	ND	θ_0	θ_f	$\|\nabla\theta_f\|$	Obs.
Min	20	52	18	2	9	0	1.7e+002	3.9e-013	1.9e-006	
FB	10	13	10	0	2	0	5.6e+002	2.2e-016	2.1e-008	
FBP0.95	11	16	11	0	3	0	2.0e+003	7.4e-013	1.7e-006	
FB0.75	13	20	13	0	3	0	3.3e+004	3.7e-013	1.0e-005	
FB0.5	13	18	13	0	3	0	1.3e+005	1.1e-014	3.0e-006	
FB0.2	13	17	13	0	3	0	3.3e+005	1.7e-018	3.1e-008	

Tabela 5.47: Problema `tobin` : $n = 42$ (formulação NCP)

Conclusão

Nesta dissertação apresentamos um método de tipo Newton não diferenciável, para resolver sistemas de Karush-Kuhn-Tucker gerais. A abordagem é similar à da reformulação do problema de complementaridade não linear em um sistema de equações semi-suave, por meio de funções NCP.

Neste caso, estudamos em detalhes três funções NCP, e em particular a função de Fischer-Burmeister penalizada, que até agora só tinha sido usada na reformulação do problema de complementaridade não linear.

Foi desenvolvida uma teoria geral para resolução de sistemas não diferenciáveis com métodos tipo Newton e justificada sua aplicação para a resolução dos sistemas KKT com reformulações NCP.

Estas estratégias não diferenciáveis são justificadas dado que sua implementação é viável, são eficientes computacionalmente, e é possível aplicar métodos tipo Newton sem a restritiva hipótese de complementaridade estrita.

A estratégia introduzida nesta dissertação, resolução de sistemas KKT com a função de Fischer-Burmeister penalizada se mostrou competitiva e para alguns parâmetros de penalização foi capaz de resolver de forma eficiente alguns problemas que as outras estratégias não resolveram. De qualquer forma, os resultados numéricos não indicam uma estratégia das três testadas que seja claramente dominante. Embora em problemas sem dificuldade em relação à diferenciação exista uma clara preferência pela função mínimo, esta é a que apresenta maior dificuldade em casos complicados. Uma possibilidade a ser usada é tomar a função de Fischer-Burmeister penalizada apelando a diferentes parâmetros quando a função mínimo falhe.

Este trabalho deixa um material muito interessante para a resolução de sistemas não diferenciáveis, com as condições teóricas de sua aplicação.

No futuro, resta confirmar estas conclusões sugeridas com experimentos mais completos, e ver a possibilidade de funcionamento de uma estratégia com atualização do parâmetro de penalização em cada iteração. Outra possibilidade é desenvolver métodos de Newton inexatos não diferenciáveis. Também é possível pensar em algum tipo de pertubação nas funções NCP para aplicar métodos tipo quase-Newton.

Bibliografia

[1] R Andreani, J.M. Martinez, and M.L. Schuverdt. The cpld condition implies the quasenormality constraint qualification. to appear in Journal of Optimization Theory and Applications.

[2] D Bertsekas. *Non Linear Programming*. Athena Scientific, 2nd edition, 1999.

[3] Bintong Chen, Xiaojun Chen, and Christian Kanzow. A penalized fischer-burmeister ncp-function: theorical investigation and numerical results, 1997.

[4] X. Chen, L. Qi, and D. Sun. Global and superlinear convergence of the smoothing newton method and its application to general box constrained variational inequalities. *Mathematics of Computations*, 67(222):519 – 540, April 1998.

[5] F.H. Clarke. *Optimization and Nonsmooth Analysis*. John Wiley & Sons, New York, 1983.

[6] T. De Luca, F. Facchinei, and C. Kanzow. A semismooth equation approach to the solution of nonlinear complimentarity. *Mathematical Programming*, 75:407 – 439, 1996.

[7] Francisco Facchinei, Andreas Fischer, and Christian Kanzow. A semismooth newton method for variational inequalities: the case of box constraints. Technical report, Institute of Applied Mathematics, University of Hamburg, Hamburg, December 1995. Preprint.

[8] Francisco Facchinei and João Soares. A new merit function for nonlinear complementarity problems and a related algorithm. *SIAM Journal of Optimization*, 7(1):225 – 247, February 1997.

[9] Franscico Facchinei and Jong-Shi Pang. *Finite-dimensional variational inequality and complementary problem*, volume I and II. Springer-Verlag, 2003.

[10] Fransisco Facchinei, Andreas Fischer, and Christian Kanzow. A semismooth newton method for variational inequalities: Theorical results and preliminaries numerical experience. Technical report, Institute of Applied Mathematics, University of Hamburg, December 1995.

[11] M.C. Ferris and J-S Pang. Engineering and economic applications of complementary problem. *SIAM Review*, 39:669 – 713, 1997.

[12] Andreas Fischer. A special newton-type optimization method. *Optimization*, 24:269–284, 1992.

[13] Andreas Fischer. Solution of monotone complementarity problems with locally lipschitzian functions. *Mathematical Programming*, 76:513 – 532, 1997.

[14] M. Fukushima. A relaxed projection method for variational inequalities. *Mathematical Programming*, 35:58–70, 1986.

[15] Carl Geiger and Christian Kanzow. On the resolution of monotone complementarity problems. Technical report, Institute of Applied Mathematics, University of Hamburg, April 1994.

[16] L. Grippo, F Lampariello, and S. Lucidi. A nonmonotone line search technique for newton's method. *SIAM Journal of Numerical Analysis*, 23(4):707 – 716, August 1986.

[17] J Han and D Sun. Newton and quase-newton methods for normal maps with polyhedral sets. *Journal of Optimization Theory and Applications*, 94:659 – 676, 1997.

[18] Shih-Ping Han, Jong-Shi Pang, and Narayan Rangaraj. Globally convergent newton methods for nonsmooth equations. *Mathematics of Operations Research*, 17:586 – 607, August 1992.

[19] P.T. Harker and J-S Pang. Finite-dimensional variational inequality and nonlinear complementary problem. *Mathematics of Operations Research*, 17:586 – 607, August 1992.

[20] W Hock and K Schittkowski. Test examples for nonlinear programming codes. *Lecture Notes in Economics and Mathematical Systems*, 187, 1981.

[21] Houyuan Jiang. Global convergence analysis of the generalized newton and gauss-newton methods of the fischer-burmeister equation for the complementary problem. *Mathematics of Operations Research*, 24(3):529 – 543, August 1999.

[22] Houyuan Jiang and Daniel Ralph. Global and local superlinear convergence analysis of newton-type methods for semismooth equations with smooth least squares. Technical report, The University of Melbourne, December 1997.

[23] D.G. Luemberger. *Linear and Nonlinear Programming*. Addison-Wesley, 1984.

[24] J.M. Martínez and S.A. Santos. Métodos computacionais de otimizaçao, 1995.

[25] robert Mifflin. Semismooth and semiconvex functions in constrained optimization. *SIAM Journal on Control and Optimization*, 15(6):959 – 972, November 1977.

[26] Pshenichnyi B. N. *Necessary Conditions for an Extremum*. Marcel Dekker, New York, 1971.

[27] Jong-Shi Pang. Newton's methods for b-differentiable equations. *Mathematics of Operations Research*, 15(2):311 – 341, May 1990.

[28] Jong-Shi Pang and Liqun Qi. Nonsmooth equations: motivation and algorithms. *SIAM, J. Optimization*, 3(3):443 – 465, August 1993.

[29] Liqun Qi. Convergence analysis of some algorithms for solving nonsmooth equations. *Mathematics of Operations Research*, 18(1):227–244, February 1993.

[30] Liqun Qi and Houyuan Jiang. Semismooth karush-kuhn-tucker equations and convergence analysis of newton and quasi-newton methods for solving these equations. *Mathematics of Operations Research*, 22(2), May 1997.

[31] Liqun Qi, Danny Ralph, and Guanglu Zhou. Semiderivative functions and reformulation methods for solving complimentarity and variational problems. *Nonlinear Optimization and Applications*, 2, 1998.

[32] Liqun Qi and Defeng Sun. Nonsmooth equations and smoothing newton methods. Technical report, The University of New South Wales, April 1998.

[33] Liqun Qi and Jie Sun. A nonsmooth version of newton's method. *Mathematical Programming*, 58:353–367, 1993.

[34] Werner C. Rheinboldt. *Methods for solving systems of nonlinear equations*. Regional conference series in applied mathematics. SIAM, 2nd edition, 1974.

[35] Defeng Sun and Liqun Qi. On ncp-functions. *Computational Optimization and Applications*, 00:1–21, 1998.