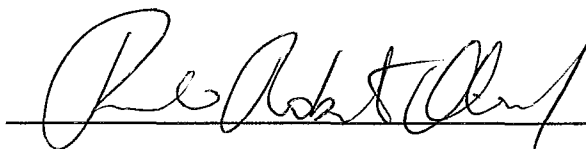


# ESTUDO COMPARATIVO DE MÉTODOS DE MINIMAX

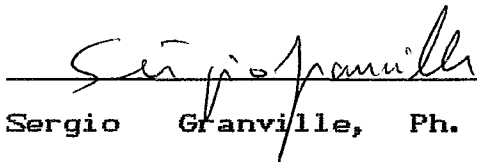
BERTHA ULLOA RUBIO

TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS PROGRAMAS DE  
PÓS-GRADUAÇÃO EM ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE  
JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO  
GRÁU DE MESTRE EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E  
COMPUTAÇÃO

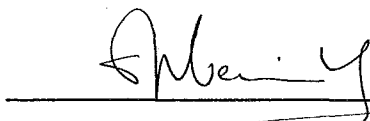
Aprovada por:



Paulo Roberto Oliveira, Dr. Ing



Sergio Granville, Ph. Dr.



Susana Scheimberg de Makler D.Sc.

Rio de Janeiro, RJ - Brasil.

MAIO DE 1993

ULLOA RUBIO BERTHA

ESTUDO COMPARATIVO DE MÉTODOS DE MINIMAX (Rio de Janeiro, 1992)  
XII, 109.p, 29.7 cm (COPPE/UFRJ, M.Sc, Engenharia de Sistemas e  
Computação, 1992).

Tese-Universidade Federal de Rio de Janeiro, COPPE.

1. Otimização em Sistemas.

I. COPPE/UFRJ.

II. TITULO (Série)

DEDICATORIA

A Santiago Esteban e a nuestras hijas:

Sirley Asunción y Cintia Bertha. En

especial mi querida madre.

### AGRADECIMENTOS

Al iniciar este parágrafo de reconocer a personas, instituciones y organizaciones que de una u otra forma hicieron posible que llegara este momento de culminación de un trabajo difícil y arduo, mi pensamiento se vuelca primeramente a mi querida esposa y mis lindas hijitas por el cariño, comprensión y estímulo que siempre me brindaron.

A los profesores de la universidad Federal de Rio de Janeiro; en especial a Paulo Roberto Oliveira por su apoyo y por los incentivos permanentes que permitieron vencer los obstáculos surgidos en el desarrollo de esta tesis.

A los Profesores Sergio Granville y Susana S. de Makler por haber participado como miembro de la banca.

A todo el cuerpo docente del programa de Ingeniería de Sistema e computación en general que me enseñaron y mostraron otros caminos de la matemática aplicada, también a su cuerpo administrativo.

A mis compañeros de cursos con quienes llegué a tener una buena amistad que a crecido con el tiempo, a mis amigos José Arica, Juan Romero y Sandra Isabel de quienes siempre recibí solidaridad y estímulo.

A la CAPES y FAPERJ por su ayuda financiera para cristalizar este trabajo.

Rio de Janeiro, Maio de 1993

Resumo da Tese apresentada à COPPE/UFRJ. como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

## ESTUDO COMPARATIVO DE MÉTODOS DE MINIMAX

Bertha Ulloa Rubio

Maio 1993

Orientador: Paulo Roberto Oliveira.

### RESUMO

É feita uma revisão dos principais métodos para resolução dos problemas de minimax finito por algoritmos da programação suave. Propomos uma classificação em quatro famílias de abordagens metodológicas : as baseadas nos métodos sequenciais quadráticos, nos métodos de projeção (gradiente projetado ), na suavização dos pontos de não-diferenciabilidade, e na regularização do problema dual.

Além de apresentar e analisar os trabalhos referenciados, comparámo-los entre si e realizamos alguns testes computacionais .

Abstract of Thesis presented to COPPE / UFRJ as  
partial  
fulfillment of the requirement for the degree of Master of  
Science (M.Sc.).

Bertha Ulloa Rubio

Maio de 1992

Chairman: Paulo Roberto Oliveira

Department: Sistem Engineering.

## ABSTRACT

A revision is made of the main methods for solution of finite minmax problems by smooth programming algorithms. We propose a classification in four families of methodological approaches; those based on the sequential quadratic methods, on the projection methods (Projected Gradient), on the smoothing of non-differentiability points and on the dual problem regularization.

Besides the presentation and analysis of the principal papers, we also compare them and present some computational tests.

## INDICE

### PÁGINA

### I. INTRODUÇÃO

I.1: Considerações Iniciais	1
I.2: Objetivos do trabalho	1
I.3: Apresentação dos capítulos	2

### II. O PROBLEMA DE MINIMAX FUNDAMENTOS TEÓRICOS

II.1: Introdução	3
II.2: Definições e Propriedades da Função Máximo	3
II.3: Condições Necessárias de Otimalidade	8
II.4: Condições Suficientes para um Mínimo Local	17

### III. PRINCIPAIS FORMULAÇÕES ALGORÍTMICAS

III.1: Introdução	24
III.2: Método de Han	25

III.3: Método de Charalambous e Conn	49
III..4: Método de Zang	63
III.5: Método de Gígola e Gómez	79

#### IV. RESULTADOS NUMÉRICOS

Resultados Numéricos:	100
Implementação do método de Gómez e Gígola	
Bibliografia .	



## CAPÍTULO I

## I. INTRODUÇÃO

## I.1: CONSIDERAÇÕES INICIAIS

Neste trabalho apresentamos alguns métodos para resolver problemas do tipo minimax:

$$\min_{x \in \mathbb{R}^n} \left[ \max_{i=1, \dots, m} f_i(x) \right]; \quad [I.1]$$

onde  $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $i=1, \dots, m$  são funções continuamente diferenciáveis

Os problemas de minimax têm sido objeto de estudo de muitos pesquisadores, que têm desenvolvido diversas técnicas matemáticas para sua solução.

Sua característica não diferenciável é tradicionalmente contornada através de formulações equivalentes e diferenciáveis. Dedicamo-nos apenas aos métodos diferenciáveis.

## I.2: OBJETIVOS

Um primeiro objetivo de nosso trabalho é uma apresentação em quatro categorias de trabalhos correspondentes aos métodos diferenciáveis .

Esta apresentação é feita, em cada caso, a partir de um trabalho que consideramos representativo de cada uma destas famílias.

Desenvolvemos a análise essencial, ora como a feita pelos autores, ora mais detalhada, ora, finalmente, por consecutivos exemplos e demonstrações diversas das dos artigos originais.

Um resumo dos fundamentos teóricos necessários é retirado do clássico trabalho de Dem'yanov [6], em que acrescentamos alguns exemplos.

### I.3: APRESENTAÇÃO DOS CAPÍTULOS

Da teoria básica apresentaremos propriedades de continuidade, diferenciabilidade e condições de otimalidade que serão o tema do capítulo II. A determinação da direção de descida permite classificar os métodos em quatro grandes grupos principais, a serem desenvolvidos no capítulo III.

No capítulo IV apresentamos alguns resultados numéricos

## CAPITULO II

## O PROBLEMA DE MINIMAX : FUNDAMENTOS TEÓRICOS

## II.1: INTRODUÇÃO

Entre os trabalhos relativos a minimax o de V.F. Dem'yanov (1966-1970) é o mais importante no estabelecimento dos fundamentos teóricos.

Neste capítulo vamos focar a teoria básica necessária para a solução de problemas de PNL irrestritos da forma [I.1]. Mesmo sem serem referenciados nos capítulos posteriores, alguns resultados são apresentados por sua importância intrínseca.

## II.2: DEFINIÇÕES E PROPRIEDADES DA FUNÇÃO MÁXIMO

Usaremos a seguinte notação

$$\varphi(x) = \max_i f_i(x), \quad i=1, \dots, m; \quad [II.2]$$

Observe que  $\varphi(x)$  não é necessariamente diferenciável.

O conjunto das funções ativas será representado por

$$I(x) = \left\{ i=1, \dots, m \mid f_i(x) = \varphi(x) \right\},$$

## II.2.1 PROPRIEDADES BASICAS DE $\varphi(x)$

a) Se todas as  $f_i(x)$ ,  $i=1, \dots, m$  são contínuas em  $x_0$  então  $\varphi(x)$  é contínua;

b) Se todas as  $f_i(x)$ ,  $i=1, \dots, m$  são contínuas sobre  $\mathbb{R}^n$  e se para algum  $x_0 \in \mathbb{R}^n$  o conjunto

$$M(x_0) = \left\{ x \in \mathbb{R}^n \mid \varphi(x) \leq \varphi(x_0) \right\} \text{ é limitado,}$$

então existe um ponto  $x^*$  no qual  $\varphi(x)$  atinge o mínimo;

c) Seja  $\Omega \subset \mathbb{R}^n$  conjunto convexo e todas as  $f_i(x)$ ,  $i=1, \dots, m$  convexas em  $\Omega$ . Então  $\varphi(x)$  também o é.

d) Se as funções  $f_i(x)$ ,  $i=1, \dots, m$  são contínuas em  $x_0$ , então para cada  $p \in \mathbb{R}^n$ ,  $\|p\| = 1$  existe  $\alpha_0 > 0$  tal que, se  $\alpha \in [0, \alpha_0]$  temos

$$\varphi(x_0 + \alpha p) = \max_{i=1, \dots, m} f_i(x_0 + \alpha p) = \max_{i \in I(x_0)} f_i(x_0 + \alpha p); \quad [II.3]$$

DEMONSTRAÇÃO (Ver V.F. Dem'yanov e V.N. Malozemov (1970))

Notação: A derivada direcional de  $\varphi$  no ponto  $x_0$ , na direção  $p$ , é dada por

$$\frac{\partial \varphi(x_0)}{\partial p} = \lim_{\alpha \rightarrow 0^+} \frac{\varphi(x_0 + \alpha p) - \varphi(x_0)}{\alpha}$$

se existe o limite.

**TEOREMA 2.1** Suponhamos que as  $f_i(x)$ ,  $i = 1, \dots, m$ , são continuamente diferenciáveis em uma vizinhança  $S_\delta(x_0)$  de  $x_0$  com  $\delta > 0$ . Então  $\varphi(x)$  é diferenciável em qualquer direção  $p$ , ( $|p|=1$ ) e,

$$\frac{\partial \varphi(x_0)}{\partial p} = \max_{i \in I(x_0)} \left( \frac{\partial f_i(x_0)}{\partial x}, p \right); \quad \text{[ II.4 ]}$$

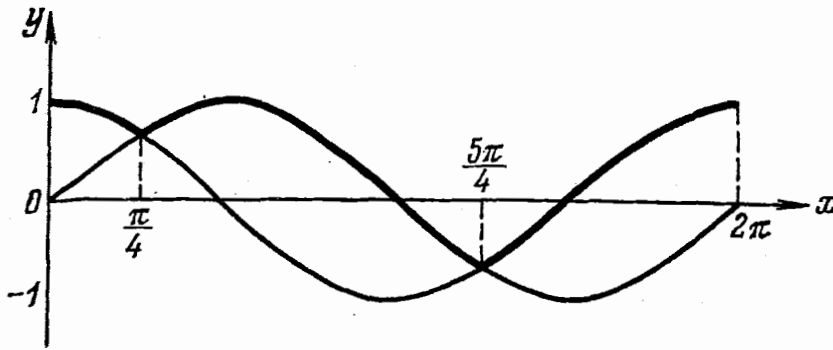
Demonstração (ver, V.F Dem'yanov e V.N Malozemov, 1970)

### II.2.2. - Exemplo (Fig.(2.1))

Sejam

$$f_1(x) = \sin x, \quad f_2(x) = \cos x$$

$$\varphi(x) = \max \{ \sin x, \cos x \}$$



Fig(2.1)

$\varphi(x)$  tem período  $2\pi$ . Considerando  $x \in [0, 2\pi]$

$$I(x) = \begin{cases} \{ 2 \}, & \text{quando } x \in [0, \pi/4) \text{ ou } x \in (5\pi/4, 2\pi], \\ \{ 1, 2 \}, & \text{quando } x = \pi/4 \text{ ou } x = 5\pi/4, \\ \{ 1 \}, & \text{quando } x \in (\pi/4, 5\pi/4), \end{cases}$$

A derivada de  $\varphi(x)$  é descontínua em  $x = \pi/4$  e  $x = 5\pi/4$ .

Em  $x_0 = \pi/4$ , com  $p = -1$

$$\frac{\partial \varphi(\pi/4)}{\partial p} = \max_{i \in I(x_0)} \left[ \frac{\partial f_i(x_0)}{\partial x}, -1 \right]$$

$$= \max \left[ -\cos(\pi/4), \sin(\pi/4) \right]$$

$$= \frac{\sqrt{2}}{2}$$

Similarmente, para  $p = 1$

$$\frac{\partial \varphi(\pi/4)}{\partial p} = \frac{\sqrt{2}}{2}$$

Observe que  $\varphi(x)$  cresce em  $x = \pi/4$  em ambas as direções, isto é trata-se de um ponto de mínimo local com a característica das derivadas direcionais serem positivas.

De fato, tem-se a

## II.2.3 PONTOS ESTACIONÁRIOS E DIREÇÃO DE MÁXIMA DESCIDA

DEFINIÇÃO 1. - Um ponto  $x^*$  no qual

$$\inf_{\|p\|=1} \max_{i \in I(x^*)} \left[ \frac{\partial f_i(x^*)}{\partial x}, p \right] \geq 0$$

Cou

$$\inf_{\|p\|=1} \frac{\partial \varphi(x^*)}{\partial p} \geq 0$$

é um ponto estacionário de  $\varphi(x)$  sobre  $R^n$ .

DEFINIÇÃO 2.- O vetor  $\bar{p}_1$ ,  $\|\bar{p}_1\| = 1$ , é chamado direção de máxima descida da função  $\varphi(x)$  em  $\bar{x}$  se

$$\frac{\partial \varphi(\bar{x})}{\partial \bar{p}_1} = \min_{\|\bar{p}\|=1} \frac{\partial \varphi(\bar{x})}{\partial \bar{p}}$$

Observação:

A minimização da derivada direcional é o critério básico para obtenção de direções de descida na Programação Não Linear. A diferentes normas ( $\|\cdot\|$ ) corresponderão diferentes métodos ("steepest descent" para a norma Euclidiana).

O vetor  $\bar{p}_1$ ,  $\|\bar{p}_1\|=1$ , é chamado direção de máxima subida se

$$\frac{\partial \varphi(\bar{x})}{\partial \bar{p}_1} = \max_{\|\bar{p}\|=1} \frac{\partial \varphi(\bar{x})}{\partial \bar{p}}$$

### II.3: CONDIÇÕES NECESSÁRIAS DE OTIMALIDADE

Previamente apresentamos algumas propriedades adicionais da função  $\varphi(x)$ .

i) Suponhamos que  $f_i(x)$ ,  $i=1, \dots, m$  são continuamente diferenciáveis sobre  $R^n$  e que para algum  $x_0 \in R^n$  o conjunto



$$M(x_0) = \left\{ x \in \mathbb{R}^n \mid \varphi(x) \leq \varphi(x_0) \right\}$$

é limitado; neste caso, a função máximo  $\varphi(x)$  têm derivadas em todas as direções para todo  $x \in \mathbb{R}^n$  e atinge o valor mínimo sobre  $\mathbb{R}^n$ .

ii) Fixemos  $x \in \mathbb{R}^n$ . Consideremos o conjunto

$$H(x) = \left\{ z = \frac{\partial f_i(x)}{\partial x} \mid i \in I(x) \right\}.$$

Denotando  $L(x)$  como a envoltória convexa de  $H(x)$ ,

$$L(x) = \left\{ z = \sum_{i \in I(x)} \alpha_i \frac{\partial f_i(x)}{\partial x} \mid \alpha_i \geq 0, \sum_{i \in I(x)} \alpha_i = 1 \right\},$$

notamos que  $L(x)$  é limitado, fechado e convexo.

**Observação:**

Sob a perspectiva do cálculo subdiferencial,  $L(x)$  é o sub-diferencial  $\partial\varphi(x)$ , conjunto de todos os subgradiantes de  $\varphi$  em  $x$ .

iii) Para  $x \in \mathbb{R}^n$  fixo, introduzimos a função derivada direcional.

$$\phi(p) = \frac{\partial \varphi(x)}{\partial p}; \quad \text{[II.51]}$$

Das propriedades gerais da função máximo obtemos que  $\phi(p)$  é contínua e convexa ( de fato, positivamente homogênea, veja lema 2.2 neste número ) sobre  $\mathbb{R}^n$ .

Estamos interessados no comportamento de  $\phi(p)$  para  $p \in S^0$ , onde  $S^0 = \{ p \in \mathbb{R}^n / \|p\| = 1 \}$ . Como  $S^0$  é limitado e fechado  $\phi(p)$  atinge o mínimo sobre  $S^0$ .

iv) Reescrevamos a definição 2 de direção de máxima descida , utilizando o teorema 2.1:

$$\begin{aligned} \psi(x) &= \min_{\|p\|=1} \phi(p) \\ &= \min_{\|p\|=1} \max_{i \in I(x)} \left[ \frac{\partial f_i(x)}{\partial x}, p \right] \end{aligned}$$

Vamos especificar as condições de otimalidade associadas ao problema de Programação Não Linear [II.1]. A condição natural de que  $\phi$  deve crescer a partir de um ponto de mínimo local em qualquer direção se expressa pelo

**TEOREMA 2.2:** Para que  $x^*$  seja o mínimo de  $\phi(x)$  sobre  $\mathbb{R}^n$  é necessário( se  $\phi(x)$  é convexo, também é suficiente ) que seja estacionário:

$$\inf_{\|p\|=1} \max_{i \in I(x^*)} \left[ \frac{\partial f_i(x^*)}{\partial x}, p \right] \geq 0 ; \quad \text{[II.6]}$$

C ou

$$\frac{\partial \varphi(x^*)}{\partial p} \geq 0, \quad \forall p \in S^0 \quad (III.71)$$

DEMONSTRAÇÃO (  $\Rightarrow$  )

Suponhamos que  $x^*$  é um ponto de mínimo de  $\varphi(x)$  tal que

$$\inf_{\|p\|=1} \frac{\partial \varphi(x^*)}{\partial p} < 0$$

Então existe um vetor  $p_1 \in \mathbb{R}^n$ ,  $\|p_1\|=1$ , tal que

$$\frac{\partial \varphi(x^*)}{\partial p_1} = -a < 0;$$

lembre-se que

$$\frac{\partial \varphi(x^*)}{\partial p_1} = \lim_{\alpha \rightarrow 0^+} \alpha^{-1} \left\{ \varphi(x^* + \alpha p_1) - \varphi(x^*) \right\},$$

daí pode-se afirmar que

$$\varphi(x^* + \alpha p_1) = \varphi(x^*) + \alpha \frac{\partial \varphi(x^*)}{\partial p_1} + o(p_1, \alpha),$$

onde

$$\frac{o(p_1, \alpha)}{\alpha} \xrightarrow{\alpha \rightarrow 0^+} 0$$

Considere  $\alpha_1 > 0$  suficientemente pequeno tal que

$$o(p_1, \alpha_1) \leq \| o(p_1, \alpha_1) \| \leq a/2 \alpha_1.$$

A partir desta desigualdade teremos

$$\begin{aligned} \varphi(x^* + \alpha_1 p_1) &\leq \varphi(x^*) + \alpha_1 \frac{\partial \varphi(x^*)}{\partial p_1} + \| o(p_1; \alpha_1) \| \\ &\leq \varphi(x^*) + \alpha_1 \frac{\partial \varphi(x^*)}{\partial p_1} + \frac{a}{2} \alpha_1, \end{aligned}$$

então

$$\varphi(x^* + \alpha_1 p_1) \leq \varphi(x^*) - a \alpha_1 + \frac{a}{2} \alpha_1$$

$$\varphi(x^* + \alpha_1 p_1) \leq \varphi(x^*) - \frac{a}{2} \alpha_1 < \varphi(x^*),$$

o que contradiz o fato de que  $x^*$  é um mínimo de  $\varphi(x)$ .

( $\Leftarrow$ )

Suponhamos que  $\varphi(x)$  é convexa e que

$$\inf_{\|p\|=1} \frac{\partial \varphi(x^*)}{\partial p} \geq 0$$

Queremos mostrar que  $x^*$  é mínimo de  $\varphi(x)$  sobre  $\mathbb{R}^n$ .

Suponhamos o contrário. Então existe um  $x_0$  tal que

$$\varphi(x_0) < \varphi(x^*);$$

como  $\varphi(x)$  é convexo, temos para  $\lambda \in [0,1]$ ,

$$\varphi \left( x^* + \lambda(x_0 - x^*) \right) \leq \varphi(x^*) + \left[ \varphi(x_0) - \varphi(x^*) \right] \lambda. \quad \text{III.81}$$

Criamos o vetor unitário

$$p_1 = \frac{x_0 - x^*}{\|x_0 - x^*\|} \quad \text{e calculamos}$$

$$\frac{\partial \varphi(x^*)}{\partial p_1} = \lim_{\alpha \rightarrow 0^+} \frac{1}{\alpha} \left[ \varphi \left( x^* + \frac{\alpha}{\|x_0 - x^*\|} (x_0 - x^*) \right) - \varphi(x^*) \right]. \quad \text{III.91}$$

Em III.81 tomamos  $\lambda = \alpha \left( \|x_0 - x^*\| \right)^{-1}$ , com  $0 < \alpha < \|x_0 - x^*\|$ :

$$\varphi \left( x^* + \frac{\alpha}{\|x_0 - x^*\|} (x_0 - x^*) \right) - \varphi(x^*) \leq \frac{\alpha \left[ \varphi(x_0) - \varphi(x^*) \right]}{\|x_0 - x^*\|};$$

multiplicando esta expressão por  $1/\alpha$  e aplicando o limite

temos

$$\lim_{\alpha \rightarrow 0} \frac{1}{\alpha} \left[ \varphi \left( x^* + \frac{\alpha}{\|x_0 - x^*\|} (x_0 - x^*) \right) - \varphi(x^*) \right] \leq \frac{\varphi(x_0) - \varphi(x^*)}{\|x_0 - x^*\|},$$

isto é,

$$\frac{\partial \varphi(x^*)}{\partial p_1} \leq \frac{\varphi(x_0) - \varphi(x^*)}{\|x_0 - x^*\|} < 0 \quad (\text{contradição})$$

#### OBSERVAÇÕES

1. A condição [III.6] se particulariza em

$$\frac{\partial \varphi(x^*)}{\partial p} = 0,$$

no caso diferenciável (quando  $m = 1$ ,  $\varphi(x) = f_1(x)$ ).

2. A convexidade local em uma vizinhança (convexa) de  $x^*$  é suficiente para a segunda parte do teorema.

TEOREMA 2.3: A desigualdade [III.6] é equivalente a

$$0 \in L(x^*); \quad \text{[III.10]}$$

se  $\varphi$  é (localmente) convexa a condição é também suficiente.

Para a prova ver, V.F. Dem'yanov (1968).

O resultado é consequência imediata das condições de otimalidade de [I.1] reescrito na forma equivalente

$$\begin{aligned} & \min \quad \delta \\ & (x, \delta) \in \mathbb{R}^{n+1} \end{aligned}$$

Sujeito a:

III.111

$$f_i(x) \leq \delta, \quad i=1, \dots, m$$

Asociamos a III.111 a função lagrangeana,

$$L(x, \delta, u) = \delta + \sum_{i=1}^m u_i (f_i(x) - \delta), \quad u \geq 0$$

As condições de K.K.T são

$$1 - \sum_{i=1}^m u_i = 0$$

$$\sum_{i=1}^m u_i \frac{\partial f_i(x^*)}{\partial x} = 0$$

$$u_i (f_i(x^*) - \delta) = 0, \quad u \geq 0$$

o que acarreta III.101.

#### OBSERVAÇÃO

III.101 é também a aplicação direta da condição de otimalidade  $0 \in \partial \varphi(x^*)$  do cálculo sub-diferencial.

A relação da derivada direcional com os gradientes de  $f_i(x)$ , mais precisamente com o sub-diferencial  $L(x)$  é dada pelo,

LEMA 2.2: A seguinte equação é válida:

$$\phi(p) = \max_{z \in L(x)} (z, p)$$

onde  $L(x)$  é o conjunto definido em (ii).

#### OBSERVAÇÃO

No caso diferenciável ( $\phi(x) = f_1(x)$ ), tem-se a linearidade em  $p$   $\phi(p) = (\nabla\phi(x), p)$ .

LEMA 2.3: Se  $\psi(x^*) \geq 0$ ,  $x^*$  é um ponto estacionário de  $\phi(x)$  (Definição 1).

Então  $\psi(x^*) = r(x^*)$ , onde  $r(x^*)$  é o raio da maior esfera com centro na origem que pode ser inscrita em  $L(x^*)$ .

#### GRAFICAMENTE

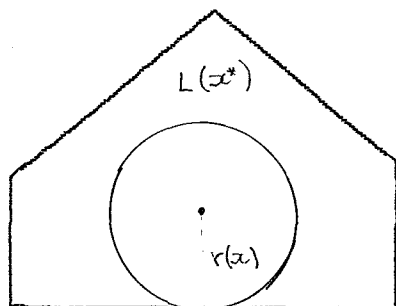


fig. (2.2)

Os resultados a seguir estabelecem condição para obtenção de direção de descida máxima.



Se  $x \in \mathbb{R}^n$  com  $\psi < 0$ ,  $x$  não é ponto estacionário e o teorema 2.3 garante que  $0 \notin L(x)$ .

LEMA 2.4: Se  $\psi(x) < 0$  então

$$\psi(x) = - \|z^*\|,$$

onde  $z^*$  é o ponto mais próximo de  $L(x)$  em relação a origem.

TEOREMA 2.4: Se  $\psi(\bar{x}) < 0$ , então  $\varphi(x)$  tem em  $\bar{x}$  uma única direção de descida máxima  $p(\bar{x})$ ,  $\|p(\bar{x})\| = 1$  dada por

$$p(\bar{x}) = - \frac{z^*}{\|z^*\|}, \quad \frac{\partial \varphi(\bar{x})}{\partial p(\bar{x})} = - \|z^*\|,$$

onde  $z^*$  é o ponto mais próximo de  $L(x)$  em relação à origem.

## II.4: CONDIÇÕES SUFICIENTES PARA UM MÍNIMO LOCAL

TEOREMA 2.5: Se  $x^*$  é ponto estacionário e  $\psi(x^*) = r > 0$  então  $x^*$  é um mínimo local estrito de  $\varphi(x)$  (veja exemplo em [II.2.2]).

TEOREMA 2.6: Seja  $x^*$  um ponto estacionário de  $\varphi(x)$  sobre  $\mathbb{R}^n$  no qual  $\psi(x^*) = 0$ . Suponha também que  $f_i(x)$ ,  $i = 1, \dots, m$  são duas vezes continuamente diferenciáveis em uma vizinhança  $S_\delta(x^*)$ ,  $\delta > 0$  de  $x^*$ .

Se para algum  $\gamma > 0$

$$\min_{p \in G_\gamma} \max_{i \in I_2(x^*, p)} \left[ \frac{\partial^2 f_i(x^*)}{\partial x^2} p, p \right] = a(x^*) > 0, \quad \text{III-121}$$

onde

$$G_\gamma = \left\{ p \mid \|p\| = 1, 0 \leq \frac{\partial \varphi(x^*)}{\partial p} \leq \gamma \right\}$$

$$I_2(x^*, p) = \left\{ i \mid i \in I(x^*), \left[ \frac{\partial f_i(x^*)}{\partial x}, \dots \right] = \frac{\partial \varphi(x^*)}{\partial p} \right\}$$

$$\frac{\partial^2 f_i(x^*)}{\partial x^2} \quad \text{é a hessiana de } f_i,$$

então  $x^*$  é um mínimo local estrito.

Como consequência deste teorema fazemos as seguintes

#### OBSERVAÇÕES

1. III-121 é verdadeira se para todo  $i \in I(x^*)$  as matrizes  $\partial^2 f_i(x^*) / \partial x^2$  são estritamente definidas positivas, isto é, para todo  $i \in I(x^*)$  e  $v \in \mathbb{R}^n$

$$\left( \frac{\partial^2 f_i(x^*)}{\partial x^2} v, v \right) \geq m_i \|v_i\|^2,$$

onde  $m_i > 0$  são independentes de  $v$ .

2. O teorema 2.6 é trivialmente verdadeiro com [II.12] substituído por

$$\min_{\|p\|=1} \max_{i \in I_2(x^*, p)} \left( \frac{\partial^2 f_i(x^*)}{\partial x^2} p, p \right) > 0.$$

3. Se apenas

$$\min_{p \in G_0} \max_{i \in I_2(x^*, p)} \left( \frac{\partial^2 f_i(x^*)}{\partial x^2} p, p \right) > 0, \quad [\text{II.13}]$$

onde

$$G_0 = \left\{ p \mid \|p\| = 1, \frac{\partial \varphi(x^*)}{\partial p} = 0 \right\},$$

não se garante que  $x^*$  seja um mínimo local de  $\varphi(x)$ , conforme se verifica pelo contra-exemplo abaixo:

II.6.1: Contra-Exemplo:

Sejam em  $\mathbb{R}^2$

$$f_1(x) = -x_1 - x_2^2,$$

$$f_2(x) = 3x_1 + x_1^2 + x_2^2$$

Consideremos o ponto  $x^* = (0, 0)$  em que  $f_1$  e  $f_2$  são ativas  
(  $I(x^*) = \{1, 2\}$  ) e

$$\varphi(x^*) = \max_{i=1, 2} f_i(x^*) = 0;$$

temos

$$H(x^*) = \left\{ z = \frac{\partial f_i(x^*)}{\partial x} \quad / \quad i=1, 2 \right\} = \left\{ (-1, 0), (3, 0) \right\},$$

$$L(x^*) = \text{co } H(x^*) = \left\{ (3-4\alpha_1, 0), 0 \leq \alpha_1 \leq 1 \right\}, \quad \alpha_1 \geq 0.$$

Como  $0 \in L(x^*)$ , para  $\alpha_1 = 3/4$ ,  $x^*$  é ponto estacionário de  $\varphi(x)$ , e

$$\psi(x^*) = \min_{\|p\|=1} \max_{z \in L(x^*)} (z, p) = 0,$$

cujo argumento é  $p = (0, \pm 1)$ .

Considere os vetores

$$p_1 = (0, 1) \quad \text{e} \quad p_2 = (0, -1), \quad \text{com}$$

$$I_2(x^*, p_1) = I_2(x^*, p_2) = \{1, 2\}$$

É fácil ver que  $G_0 = \{p_1, p_2\}$ : de fato, para  $p = (a, b)$ ,

$$\left[ \frac{\partial f_1(x^*)}{\partial x}, p \right] = -a, \quad \left[ \frac{\partial f_2(x^*)}{\partial x}, p \right] = 3a$$

logo  $\psi(x^*) = 0$  para  $a = 0$  e  $b = \pm 1$ .

Temos

$$\frac{\partial^2 f_1(x^*)}{\partial x^2} = \begin{pmatrix} 0 & 0 \\ 0 & -2 \end{pmatrix}, \quad \frac{\partial^2 f_2(x^*)}{\partial x^2} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

Logo,

$$\left[ \frac{\partial^2 f_1(x^*)}{\partial x^2}, p_1, p_1 \right] = \left[ \frac{\partial^2 f_1(x^*)}{\partial x^2}, p_2, p_2 \right] = -2,$$

$$\left[ \frac{\partial^2 f_2(x^*)}{\partial x^2}, p_1, p_1 \right] = \left[ \frac{\partial^2 f_2(x^*)}{\partial x^2}, p_2, p_2 \right] = 2,$$

Obtemos:

$$\min_{p \in G_0} \max_{i \in I_2(x^*, p)} \left[ \frac{\partial^2 f_i(x^*)}{\partial x^2}, p, p \right] = 2 > 0,$$

que é III.13).

Mostraremos agora que  $x^*$  não é um mínimo local de  $\varphi(x)$ .

Seja a direção  $V_\varepsilon = (-\varepsilon, 1)$ ,  $1 > \varepsilon > 0$ . Para  $\alpha \geq 0$

$$f_1(x^* + \alpha V_\varepsilon) = \alpha\varepsilon - \alpha^2, \quad f_2(x^* + \alpha V_\varepsilon) = -3\alpha\varepsilon + \alpha^2\varepsilon^2 + \alpha^2$$

e

$$\varphi(x^* + \alpha V_\varepsilon) = \max \left\{ \alpha\varepsilon - \alpha^2, -3\alpha\varepsilon + \alpha^2\varepsilon^2 + \alpha^2 \right\},$$

$$= \begin{cases} f_1(x^* + \alpha V_\varepsilon), & \text{para } \alpha \in [0, \frac{4\varepsilon}{2+\varepsilon^2}], \\ f_2(x^* + \alpha V_\varepsilon), & \text{para } \alpha \geq \frac{4\varepsilon}{2+\varepsilon^2}. \end{cases}$$

GRAFICAMENTE

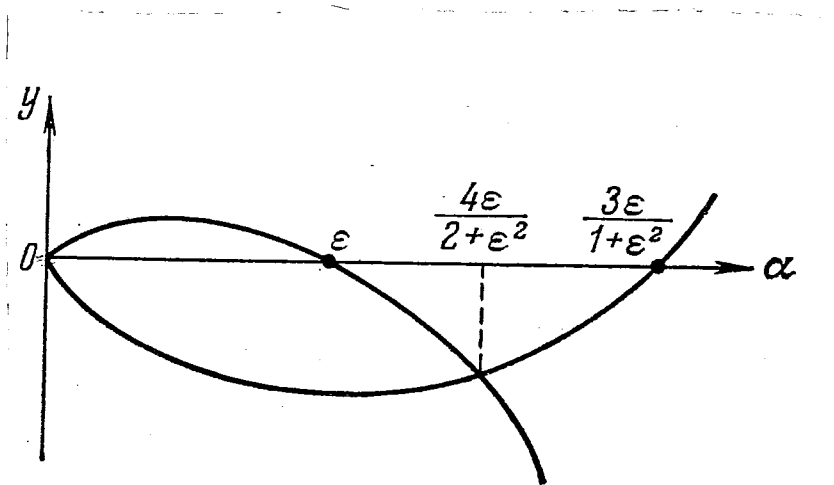


Fig. (2.3)

Denotando

$$\alpha_\varepsilon = \frac{4\varepsilon}{2+\varepsilon^2}, \quad \text{como } 0 < \varepsilon < 1, \text{ tem-se}$$

$$\varphi(x^* + \alpha \frac{V}{\varepsilon}) = f_1\left(\alpha \frac{\varepsilon}{\varepsilon} - \alpha^2 \frac{\varepsilon^2}{\varepsilon^2}\right)$$

$$= \frac{4\varepsilon^2(-2+\varepsilon^2)}{(2+\varepsilon^2)^2} < 0 = \varphi(x^*), \quad \forall \quad 0 < \varepsilon < 1.$$

O que contradiz o fato de que  $x^*$  é um mínimo estrito de  $\varphi(x)$ .

## CAPITULO III

## PRINCIPAIS FORMULAÇÕES ALGORÍTMICAS

## III.1: INTRODUÇÃO.

Neste capítulo, apresentaremos um revisão parcial de alguns métodos para resolver o problema de minimax sem restrições.

Os artigos que escolhemos são os seguintes correspondendo ao que proporíamos como representantes de distintas famílias de abordagens ao problema de minimax.

- HAN, (1977), adapta a metodologia sequencial quadrática da PNL; apresentamos sucintamente, comparando com Han, os de Osborne, Watson (1968) e Murray, Overton (1980).
- CHARALAMBOUS E CONN, (1978), usam uma modificação do método do gradiente projetado; acrescentamos uma comparação deste algoritmo e uma versão "pura" do gradiente projetado.
- ZANG, (1980), utiliza aproximações suaves externas nos pontos não diferenciáveis; a mesma idéia foi utilizado por Bertsekas (1975), porém com funções regularizantes diversas.
- GIGOLA E GOMEZ, (1991), desenvolve um método de regularização baseado numa formulação dual do problema minimax.



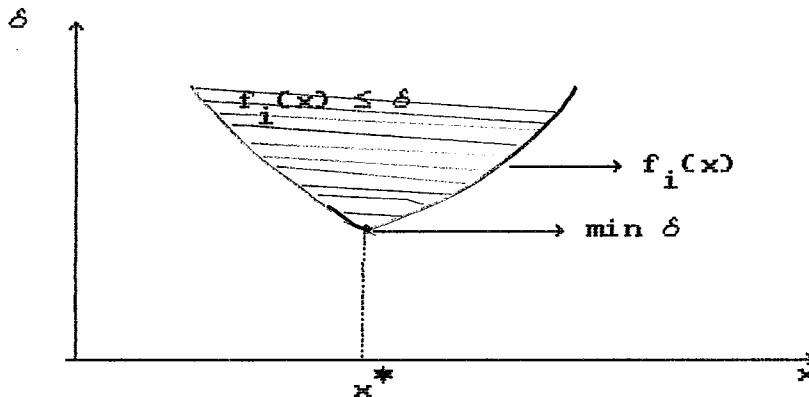
## III.2 . MÉTODO DE HAN

### III.2.1: DESCRIÇÃO E IDÉIAS GERAIS

O método de Han parte da clássica formulação equivalente ao minimax, sob a perspectiva da PNL já vista em [III.11]. De [I.1] e [II.2] formulamos o problema como

$$\begin{aligned} & \text{Min } \delta \\ & (x, \delta) \in \mathbb{R}^{n+1} \\ & \text{sujeito a} \\ & f_i(x) \leq \delta, \quad i=1, \dots, m \end{aligned} \quad \text{[III.2.1]}$$

Para  $m = n = 2$ , temos a seguinte representação:



Conforme já visto a função de Lagrange é definida por

$$L(x, \delta, u) = \delta + \sum_{i=1}^m u_i [f_i(x) - \delta] \quad , \quad u \geq 0$$

Suponha um método iterativo em que se obteve  $x^k$ .  
 Deseja-se uma direção  $p$  de descida para  $f$ , tal que em  $x^k + p$  se  
 tenha também

$$\min \delta$$

sujeito a

[III.2.2]

$$f_i(x^k + p) \leq \delta, \quad i=1, \dots, m.$$

Evidentemente este problema é equivalente ao original  
 [III.2.1], e, por isto, deverá ser de algum modo aproximado. As  
 condições de otimalidade de [III.2.2], a partir da função  
 Lagrangeana,

$$L(p, \delta, u) = \delta + \sum_{i=1}^m u_i [f_i(x^k + p) - \delta], \quad u \geq 0$$

são dadas por

$$\left\{ \begin{array}{l} \sum_{i=1}^m u_i \nabla f_i(x^k + p) = 0, \\ 1 - \sum_{i=1}^m u_i = 0, \\ u_i [f_i(x^k + p) - \delta] = 0, \\ f_i(x^k + p) - \delta \leq 0, \\ u \geq 0. \end{array} \right. \quad \text{[III.2.3]}$$

Uma aproximação de Taylor de 1<sup>a</sup> ordem em torno de  $x^k$  com variável  $p$ , de [III.2.3], fornece

$$\left\{ \begin{array}{l} \text{a)} \quad \sum_{i=1}^m u_i = 1, \\ \text{(b)} \quad \sum_{i=1}^m u_i [ \nabla f_i(x^k) + \nabla^2 f_i(x^k) p ] = 0, \quad \text{[III.2.4]} \\ \text{(c)} \quad u_i [ f_i(x^k) + ( \nabla f_i(x^k) )^T p - \delta ] = 0, \\ \text{(d)} \quad f_i(x^k) + ( \nabla f_i(x^k) )^T p - \delta \leq 0, \\ \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad u \geq 0. \end{array} \right.$$

Verifica-se que estas são as condições de K.K.T do problema quadrático-linear

$$\min \delta + 1/2 p^T B_K p \quad \text{[III.2.5]}$$

sujeito a

$$f_i(x^k) + \nabla f_i(x^k)^T p \leq \delta, \quad i=1, \dots, m$$

com  $B_K$  representando a hessiana da função lagrangeana do problema original:

$$B_K = \sum_{i=1}^m u_i \{ \nabla^2 f_i(x^k) \}. \quad \text{[III.2.6]}$$

Também a partir da equação (b) de [III.2.4] determinamos a direção

$$p^k = -B_k^{-1} \sum_{i=1}^m u_i \nabla f_i(x^k). \quad \text{[III.2.7]}$$

O sub-problema de Programação Quadrática [III.2.5] tem como solução  $(p^k, \delta^k)$ ; o n-vetor  $p^k$  é uma direção de busca a partir de  $x^k$ .

## FORMULAÇÃO DUAL

[III.2.5] pode ser resolvido através de seu dual, que, conforme veremos, é computacionalmente mais simples.

Mantendo a notação  $u_i$  para os multiplicadores, temos a função dual

$$\tilde{L}(u) = \min_{(p, \delta) \in \mathbb{R}^{n+1}} L(p, \delta, u), u \geq 0$$

onde

$$L(p, \delta, u) = \frac{1}{2} p^T B_k p + \delta + \sum_{i=1}^m u_i [f_i(x^k) + \nabla f_i(x^k)^T p - \delta], \quad \text{[III.2.8]}$$

é a função Lagrangeana associada a [III.2.5].

Por definição, o dual Lagrangeano de [III.2.5] é o problema

$$(D) \max \tilde{L}(u), u \in \mathbb{R}^m, u \geq 0 \quad \text{[III.2.9]}$$

A função dual  $\tilde{L}(u)$  pode ser facilmente explicitada através das condições de otimalidade [III.2.4] :

$$\begin{aligned} \tilde{L}(u) &= \frac{1}{2} \left[ -B_k^{-1} \nabla f(x^k)^T u \right]^T B_k \left[ -B_k^{-1} \nabla f(x^k)^T u \right] + \delta + \\ &\quad u^T f(x^k) + u^T \nabla f(x^k) \left[ -B_k^{-1} \nabla f(x^k)^T u \right] - \delta \sum_{i=1}^m u_i , \\ &= \frac{1}{2} u^T \nabla f(x^k) B_k^{-1} \nabla f(x^k)^T u + \\ &\quad u^T f(x^k) - u^T \nabla f(x^k) B_k^{-1} \nabla f(x^k)^T u , \\ &= - \frac{1}{2} u^T \nabla f(x^k) B_k^{-1} \nabla f(x^k)^T u + u^T f(x^k) . \end{aligned}$$

Com a notação  $\nabla f = J$  (Jacobiana de  $f$ ), chegamos à expressão do dual (na forma de minimização),

$$\min \left[ \frac{1}{2} u^T J_k B_k^{-1} J_k^T u - u^T f(x^k) \right]$$

sujeito a:

III.2.101

$$\sum_{i=1}^m u_i = 1$$

$$u_i \geq 0, \quad \forall i=1, \dots, m$$

Observe que o dual, um funcional quadrático sobre o simplex unitário, é extremamente mais simples do que [III.2.5], que corresponde a uma formulação do tipo geração de restrições (versus geração de colunas de [III.2.10]). Veja também que, conhecido  $u$ ,  $p$  é dado por [III.2.7]

A seguir apresentamos o algoritmo de Han.

### III.2.2: FORMULAÇÃO DO ALGORITMO

Trata-se de um algoritmo de busca linear. Esta será detalhada posteriormente.

PASSO. 0:

São dados  $x^0$  e  $B_0 = I$ , matriz identidade  $n \times n$ .

Para  $K \geq 0$ :

PASSO. 1:

Calcular o vetor  $u^k$ , solução de :

$$\text{Min } \frac{1}{2} u^T J_K H_K J_K^T u - u^T f(x^K)$$

Sujeito a  $m$

$$\sum_{i=1}^m u_i = 1$$

$$u \geq 0,$$

onde

$$J = \nabla f, \quad f(x^k) = (f_1(x^k), \dots, f_m(x^k))^T$$

$$H_k = B_k^{-1}$$

i) Determinar o vetor

$$p^k = - H_k \sum_{i=1}^m u_i^k \nabla f_i(x^k).$$

ii) Se  $p^k = 0$ ,  $x^k$  é ponto estacionário; então

$$\delta_k = \varphi(x^k) = \max \{ f_i(x^k), i=1, \dots, m \} : \text{Parar.}$$

iii) Se  $p^k \neq 0$ , ir para o passo 2.

**PASSO. 2:** Busca Linear

Seja  $x^{k+1}$  o vetor

$$x^{k+1} = x^k + \alpha_k p^k,$$

onde  $\alpha_k$  é o tamanho do passo, tal que " $\varphi$  decresça suficientemente".

**PASSO. 3:**

Atualizar  $B_{k+1}$  por BFGS usando a matriz  $B_k$  que estima a hessiana

$$\nabla_{xx} L(x^k, u^{k-1}).$$

PASSO. 4:

Incrementar  $k \leftarrow k+1$  e retornar ao passo 1.

### III.2.2. a: BUSCA LINEAR

A busca linear é realizada sobre a função original  $\varphi$  a ser minimizada. Isto se justifica duplamente; sob a perspectiva da Programação Não Linear Irrestrita,  $\varphi$  é a função que se quer minimizar, logo é a escolha natural. Do ponto de vista da formulação [III.2.1], uma possível função mérito é a função penalidade exata dada por

$$\delta + \max_i (f_i(x) - \delta) = \max_i f_i(x) = \varphi(x).$$

A proposta a seguir é uma busca aproximada, em que o tamanho de passo  $\alpha_k$  é tomado como o primeiro número da seqüência gerada por:

$$\langle \beta_0^k, \beta_1^k, \dots \rangle; \quad \text{[III-2.11]}$$

satisfazendo

$$\varphi(x^k + \beta p^k) \leq \varphi(x^k) + w \beta \lambda_k; \quad \text{[III-2.12]}$$

onde

$$\lambda_k = - (p^k)^T B_k p^k, \quad 0 < w < 1/2$$



A desigualdade [III.2.12] é consequência da comparação entre a redução real

$$\varphi(x^k + \beta p^k) - \varphi(x^k)$$

e a estimada pelo modelo local em [III.2.5], dado por,

$$\frac{1}{2} (p^k)^T B_k p^k.$$

Vejamos como se gera a seqüência [III-2.11]. Para simplificar a notação, omitiremos o índice  $k$  da iteração corrente.

Inicializamos com  $\beta_0 = 1$ .

Quando  $\beta_j$  não satisfaz [III-2.12], constrói-se uma função quadrática  $\theta_j(\beta)$ , a qual interpola  $\varphi$ :

$$\theta_j(\beta) = \varphi(x^k + \beta p^k) \text{ em } \beta=0 \text{ e } \beta = \beta_j, \quad \text{[III-2.13]}$$

e

$$\theta'_j(0) = \lambda_k$$

(segundo [III-2.12],  $\lambda_k$  pode ser visto como uma aproximação da derivada na direção  $p^k$ ),

Quando  $\beta_{j+1} > 0.1 \beta_j$ , então  $\beta_{j+1}$  é aceito como o valor que minimiza  $\theta_j(\beta)$ ;

$\beta_{j+1}$  é calculada usando-se a seguinte fórmula de Powell [22]

$$\beta_{j+1} = \max \left\{ 0.1 \beta_j, \frac{0.5 \lambda_{\kappa} \beta_j^2}{\lambda_{\kappa} \beta_j + \theta(0) - \theta(\beta_j)} \right\}. \quad \text{[III.2.13]}$$

### III.2.2.b: ATUALIZAÇÃO DA MATRIZ $B_{\kappa}$

Há várias propostas. A mais simples é usar a matriz identidade; uma outra é usar uma matriz  $B_{\kappa}$  que estime a hessiana

$$\nabla_{xx} L(x^{\kappa}, u^{\kappa-1}).$$

Considera-se, para garantia de convergência, as matrizes  $\{B_{\kappa}\}$  que satisfazem a condição de uniforme definição positiva;

$$\eta x^T x \leq x^T B_{\kappa} x \leq \rho x^T x,$$

para  $\rho$  e  $\eta$  positivos,  $\forall x \in \mathbb{R}^n$  e para cada  $\kappa$ .

Existe uma grande variedade de algoritmos para atualização de posto 2 da matriz  $B_{\kappa}$  não cabendo aqui detalhar o estudo dessas fórmulas. Maiores informações podem ser encontradas no livro de Avriel (cap.11, 1976), ou no artigo de

Dennis e Moré (1977), por exemplo. Uma das fórmulas de atualização mais aceitas é a de Broyden, Fletcher, Goldfarb e Shanno (1970), abreviada BFGS, dada por:

$$B_{K+1} = B_K - \frac{B_K S_K S_K^T B_K}{S_K^T B_K S_K} + \frac{y_K y_K^T}{y_K^T S_K},$$

onde:

$$S_K = x_K - x_{K+1}, \quad y_K = \nabla_x L(x^K) - \nabla_x L(x^{K-1}),$$

$$B S = y, \quad S, y \in \mathbb{R}^n \text{ são não nulos. [III.2.14]}$$

Sendo  $B$  simétrica definida positiva, pode-se usar a decomposição de Cholesky, com atualização do fator:

$$B = L L^T, \quad L v = y \quad \text{e} \quad L^T S = v,$$

onde

$L$  é uma matriz triangular não singular e  $v$  um vetor tal que

$$L_{K+1}^T S_K = v,$$

$$v = \left( \frac{y_K^T S_K}{S_K^T B_K S_K} \right)^{1/2} L_K^T S_K,$$

$$L_{K+1} = L_K + \frac{\left( y_K - L_K v \right) v^T}{v^T v},$$

A aproximação inicial  $B_0$  da hessiana é tomada como

$$B_0 = I.$$

Isto significa iniciar o algoritmo com a direção de máxima descida (gradiente)

Também podemos calcular

$$H_{K+1} = B_{K+1}^{-1}$$

com

$$H_{K+1} = \left( I - \frac{S_K y_K^T}{S_K^T y_K} \right) H_K \left( I - \frac{y_K S_K^T}{S_K^T y_K} \right) + \left( \frac{S_K y_K^T}{S_K^T y_K} \right).$$

### III.2.3: CONVERGÊNCIA.

Em resumo, pode-se dizer que o procedimento para escolha do tamanho de passo pode garantir a convergência global.

O método permite determinar um ponto estacionário do problema [III-2.1], onde  $x^*$  satisfaz as condições de KKT para algum vetor  $u^*$  :

$$i) \sum_{i=1} u_i^* \nabla f_i(x^*) = 0$$

$$ii) \sum_{i=1}^m u_i^* = 1 \quad \text{[III.2.15]}$$

$$\text{iii) } u^* \geq 0$$

$$\text{iv) } u_i^* \left( \varphi(x^*) - f_i(x^*) \right) = 0, \quad i=1, \dots, m$$

### Observações sobre a convergência

1.- Se  $p^k = 0$ , então  $\delta_k = \varphi(x^k)$  e  $x^k$  é um ponto estacionário do problema III.2.11, com o vetor  $u^k$  como o multiplicador associado.

De fato as condições de K.K.T aplicadas a [III.2.5] fornecem (derivada nula da função Lagrange em relação a p)

$$B_k p^k + \sum_{i=1}^m u_i^k \nabla f_i(x^k) = 0 \quad \text{[III.2.16]}$$

Quando  $p^k = 0$  obtemos

$$(i) \quad \sum_{i=1}^m u_i^k \nabla f_i(x^k) = 0$$

Além disto, a derivada nula da Lagrangeana associada a [III.2.5] relativamente a  $\delta$  fornece (ii), sendo (iii) verdadeiro por construção.

Por outro lado, as condições de complementaridade se escrevem

$$u_i^k (f_i(x^k) + \nabla f_i(x^k) p^k - \delta_k) = 0. \quad \text{[III.2.17]}$$

Quando  $p = 0$ , obtemos  $u_i^k (f_i(x^k) - \delta_k) = 0$

e, portanto,

$\delta_k = f_i(x^k)$  para  $u_i^k < 1$ , (existe ao menos um tal  $i$ , já que

$$\sum_{i=1}^m u_i^k = 1, u_i^k \geq 0)$$

Portanto,  $x^k$  é ponto estacionário. **|**

$$\geq -(P^k)^T B_K P^k \geq \delta_k - f_i(x^k),$$

(veja [III.2.12])

[III.2.19]

Demonstração

Notemos  $\tilde{I} = \{ i : u_i^k > 0 \}$ . Então, de [III.2.17], obtemos

$$u_i^k \nabla f_i(x^k) p^k = u_i^k (\delta_k - f_i(x^k))$$

que somado em  $i$ , fornece

$$\sum_{i=1}^m u_i^k \nabla f_i(x^k) p^k = \sum_{i \in \tilde{I}} u_i^k (\delta_k - f_i(x^k)); \quad \text{[III.2.20]}$$

usando a definição de  $\varphi$  nesta relação obtemos

$$\sum_{i=1}^m u_i^K \nabla f_i(x^k) p^K \geq \sum_{i \in \tilde{I}} u_i^K \left[ \delta_k - \varphi(x^k) \right],$$

$$= \delta_k - \varphi(x^k),$$

III.2.21

devido a que  $\sum_{i=1}^m u_i^K = 1$ .

Por outro lado, a expressão de  $p^k$  dada por III.2.16] fornece, após multiplicação por  $p^k$ :

$$\sum_{i=1}^m u_i^K \nabla f_i(x^k) p^K = - \left( p^K \right)^T B_K p^K,$$

III.2.22

que, junto com III.2.21] leva à desigualdade

$$- \left( p^K \right)^T B_K p^K \geq \delta_k - \varphi(x^k).$$

|

$$3. - \varphi'(x^k, p^k) \leq - (p^k)^T B_k p^k.$$

( veja Han [13])

Se consideramos  $\beta \in (0, 1)$  então Han mostra que

$$\frac{\varphi(x^k + \beta p^k) - \varphi(x^k)}{\beta} \leq - (p^k)^T B_k p^k + \frac{\beta}{2} M \|p^k\|^2,$$

onde

$$M = \max_{\eta \in (0, \beta)} \left\| f''_i(x^k + \eta p^k) \right\|, \\ i = 1, \dots, m.$$

Esta afirmação é importante porque mostra que o vetor  $p^k$  é qualificado como uma direção de busca no ponto  $x^k$ , desde que  $B_k$  seja uma matriz definida positiva.

4 .- Se o ponto  $x^k$  é solução local do problema [II-1] e a matriz  $B_k$  é definida positiva então  $p^k=0$ , (veja [III.2.7]).

Os seguintes resultados justificam o procedimento de tamanho de passo.

5 .- Se  $p^k \neq 0$  e a matriz  $B_k$  é definida positiva, então o tamanho de passo  $\alpha_k$  tem um valor positivo e  $\alpha_k = \beta_j^k$  para algum  $j$  finito, (veja [III.2.12] e 3, acima ).

6 . Se existem números positivos  $M$  e  $\eta$  tais que

$$(a) \sup_{x \in \mathbb{R}^n} \left\| f''_i(x) \right\| \leq M \text{ para qualquer } i=1, \dots, m;$$



(b)  $x^T B_k x \geq \eta x^T x$  para qualquer  $k=0,1,\dots$ , e qualquer  $x \in \mathbb{R}^n$ , então para todo  $k$ ,  $\alpha_k \geq \min \{0.1 \eta/M, 0.1\}$ .

Este resultado serve para obter a seguinte propriedade de convergência.

7 . Considerando (6) e também a função  $\varphi$  limitada, tem-se que a seqüência  $\{x^k\}$  gerada pelo método termina num ponto estacionário do problema [III-1], ou

$$\lim_{k \rightarrow \infty} \|p^k\| = 0 .$$

Isto permite mostrar que o ponto  $x^k$  cuja direção  $p^k$  tenha norma pequena leva, aproximadamente, às condições necessárias [III-2.15]; de fato isto pode ser tomado mais preciso:

8 . Suponha que existem números não negativos  $M_1$  e  $M_2$  tais que os gradientes  $\nabla f_i(x)$ ,  $i=1,\dots,m$  são limitados por  $M_1$  sobre o conjunto

$$\Omega = \left\{ x : \varphi(x) \leq \varphi(x^0) \right\}$$

e as matrizes  $\{B_k\}$  são limitadas por  $M_2$ . Então:

$$(a) \quad \left\| \sum_{i=1}^m u_i^k \nabla f_i(x^k) \right\| \leq M_2 \|p^k\| ;$$

$$(b) \quad 0 \leq u_i^k \left[ \varphi(x^k) - f_i(x^k) \right] \leq 2 M_1 \|p^k\| ;$$

$$(c) \quad \sum_{i=1}^m u_i^k = 1;$$

$$(d) \quad u^k \geq 0.$$

9.- Suponha que as hipóteses nas propriedades (7) e (8) são verificadas. Então para qualquer tolerância  $\varepsilon > 0$ , o método produz, em um número finito de iterações,  $\bar{x}$  e  $\bar{u}$  satisfazendo:

$$(a) \quad \left\| \sum_{i=1}^m \bar{u}_i \nabla f_i(\bar{x}) \right\| \leq \varepsilon;$$

$$(b) \quad 0 \leq \bar{u}_i \left( \varphi(\bar{x}) - f_i(\bar{x}) \right) \leq \varepsilon;$$

$$(c) \quad \sum_{i=1}^m \bar{u}_i = 1;$$

$$(d) \quad \bar{u} \geq 0.$$

#### OUTRAS OBSERVAÇÕES:

1.- Suponhamos que a função  $\varphi$  é convexa sobre o conjunto  $\Omega_i = \{x : \varphi(x) \leq \varphi(x^i)\}$ , suposto limitado para algum  $i$ ; então

qualquer ponto de acumulação de  $\{x^k\}$  é uma solução de [III.1] em  $\Omega_i$ . Isto permite falar da existência da solução.

2.- Para afirmar a unicidade da solução, utiliza-se o seguinte critério: Se  $\varphi$  é estritamente convexa, então a seqüência  $\{x^k\}$  converge para uma única solução do problema [III.1].

3.- O método tratado por Han permitirá estender a solução a um problema restrito:

$$\min_{x \in \mathbb{R}^n} ( \max_{i=1, \dots, m} f_i(x) ),$$

Sujeito a:

$$g_j(x) \leq 0, \quad j = 1, \dots, q.$$

onde,

$$f_i, g_j : \mathbb{R}^n \longrightarrow \mathbb{R}, \quad i=1, \dots, m, \quad j=1, \dots, q.$$

são funções reais continuamente diferenciáveis, cuja solução é feita pelo processo já explicado anteriormente, com o sub-problema [III.2.5] substituído por:

$$\min_{(p, \delta) \in \mathbb{R}^{n+1}} \frac{1}{2} p^T B p + \delta$$

sujeito a:

$$f_i(x^k) + \nabla f_i(x^k) p \leq \delta, \quad i = 1, \dots, m$$

$$g_j(x^k) + \nabla g_j(x^k) p \leq 0, \quad j = 1, \dots, q$$

onde  $B_k$  é a hessiana da função Lagrangeana do problema original

$$B_K = \sum_{i=1}^m u_i \left[ \nabla^2 f_i(x^K) \right] + \sum_{j=1}^q \lambda_j \left[ \nabla^2 g_j(x^K) \right]$$

e  $p^K$  direção dada por:

$$p^K = -B^{-1} \left[ \sum_{i=1}^m u_i \left[ \nabla f_i(x^K) \right] + \sum_{j=1}^q \lambda_j \left[ \nabla g_j(x^K) \right] \right].$$

A seguir referimos sucintamente a trabalhos com abordagens similares ou próximas da de Han.

#### 1. - Osborne e Watson (1968)

Descreve um algoritmo que soluciona o problema de minimax irrestrito de módulos de funções.

Trata-se do seguinte problema;

$$\min \max |b_i - f_i(x)|, \quad i=1, \dots, m, \quad x \in \mathbb{R}^n,$$

que equivalente a :

min h

sujeito a:

[III.2.23]

$$|b_i - f_i(x)| \leq h, \quad i=1, \dots, m.$$

Este problema é linearizado segundo

$\min h^k$

sujeito a:

[(III.2.24)]

$$|b_i - f_i(x^k) - \nabla f_i(x^k)d^k| \leq h^k, \quad i=1, \dots, m.$$

Os autores apresentam o seguinte processo para resolver

[(III.2.23)],

i.- Escolher  $x^0$

ii.- Calcular  $d^k$  a partir de [(III.2.24)], este sub-problema é solucionado usando métodos de programação linear, supondo que o posto  $(\nabla f) = n$ .

Denota-se o valor mínimo de  $h^k$  por  $\hat{h}^k$ .

iii.- Busca linear, calcular  $\gamma^k$  minimizando,

$$\max |b_i - f_i(x^k + \gamma^k d^k)|, \quad i=1, \dots, m;$$

iv.- O novo ponto é

$$x^{k+1} = x^k + \gamma^k d^k.$$

Comparação com Han

É interessante comparar o algoritmo que acabamos de descrever com Han, nos casos em que se garanta que  $|b_i - f_i(x)| \leq b_i - f_i(x)$  (ou o posto). Basta, por exemplo, que se tenha, para todo  $i$ ,

$$f_i(x) \leq f_i(x_0) \leq b_i, \quad \text{para algum } x_0.$$

Temos agora o sub-problema

$\min h^k$

S.a

$$b_i - f_i(x^k) - \nabla f_i(x^k)d^k \leq h^k, \quad i=1, \dots, m,$$

cujas condições de K.K.T são

$$(a') \quad \sum_{i=1}^m u_i = 1,$$

$$(b') \quad \sum_{i=1}^m u_i \nabla f_i(x^k) = 0,$$

$$(c') \quad u_i \left[ b_i - f_i(x^k) - \nabla f_i(x^k) d^k - h^k \right] = 0,$$

$$(d') \quad b_i - f_i(x^k) - \nabla f_i(x^k) d^k - h^k \leq 0,$$

$$u \geq 0.$$

Observando [II.2.4] e [III.2.5], somente (b) e (b') são diferentes: temos aqui uma aproximação de  $1^{\text{a}}$  ordem da função lagrangeana em torno de  $x^k$ , dada por

$$\sum_{i=1}^m u_i \left[ f_i(x^k) + \nabla f_i(x^k)^T d \right],$$

enquanto Han com expansão de Taylor até  $2^{\text{a}}$  ordem.

## 2. - Murray e Overton (1980)

O método soluciona o problema [I.1] equivalente a

$$\begin{aligned} \min \quad & x_{n+1} && \text{[III.2.25]} \\ \text{sujeito a:} \end{aligned}$$

$$c_i(x) = x_{n+1} - f_i(\bar{x}) \geq 0, \quad i=1, \dots, m;$$

onde  $x^T = (\bar{x}^T, x_{n+1})$ ,  $(x_1, \dots, x_n)^T = \bar{x} \in \mathbb{R}^n$

Trata-se de um método de restrições ativas, representadas

por  $\hat{c}(x) \in \mathbb{R}^t$  ( $t$  indica número de restrições ativas).

Denotamos a jacobiana das restrições ativas por  $\hat{A}(x) = \nabla \hat{c}(x)$ , isto é :

$$\hat{A}(x) = \nabla \hat{c}(x) = \begin{bmatrix} -\nabla f_1(\bar{x}) & \dots & -\nabla f_t(\bar{x}) \\ 1 & \dots & 1 \end{bmatrix}_{(n+1) \times t}$$

O Lagrangeano de [III.2.25] é dado por:

$$l(x, \lambda) = x_{n+1} - (\lambda)^T \hat{c}(x).$$

Com o objeto de formar o sub-problema quadrático (MSQ), necessitamos de :

$$\nabla_x l(x, \lambda) = e_{n+1} - \hat{A}(x) \lambda, \quad \text{[III.2.26]}$$

onde  $e_{n+1} = (0, \dots, 0, 1)^T \in \mathbb{R}^{n+1}$ ,

e da Hessiana  $(n+1) \times (n+1)$

$$w(x, \lambda) = \nabla_x^2 l(x, \lambda) = - \sum_{i=1}^t \lambda_i \nabla^2 \hat{c}_i(x) = w(x, \lambda),$$

Podemos agora definir o sub-problema

$$\min l(x^k, \lambda^k) + \left( e_{n+1} - \hat{A}(x^k) \lambda^k \right)^T p + \frac{1}{2} p^T w(x^k, \lambda^k) p$$

sujeito a: [III.2.27]

$$\hat{A}(x^k)^T p = - \hat{c}(x^k),$$

ou, mais simplesmente :

$$\min e_{n+1}^T p + \frac{1}{2} p^T w(x^k, \lambda^k) p$$

sujeito a: [III.2.28]

$$\hat{A}(x^k)^T p = - \hat{c}(x^k)$$

Supondo-se  $\hat{A}$  de posto completo,  $\lambda^k$  é solução de

$$\nabla l(x_1^k, \lambda) = e_{n+1} - \hat{A}(x^k)\lambda = 0.$$

Por outro lado, Murray-Overton propõe obter  $p$  utilizando-se da decomposição de  $\mathbb{R}^{n+1}$  nos espaços nulo de  $\hat{A}^T$  e imagem de  $\hat{A}$ , de dimensões respectivas  $n+1-t$  e  $t$ . Assim, com  $w$  definida positivo,  $p$  é calculado como  $p = p_y + p_z$ , onde  $p_y \in \mathbb{R}^t$  e  $p_z \in \mathbb{R}^{n+1-t}$  (para suas expressões ver Murray, Wright, 1981).

A atualização da hessiana é feita usando Quase Newton, garantindo-se que a direção de busca ( $p$ ) seja de descida para a função objetivo do minimax, obtendo-se convergência quadrática local.

Comparação entre Han e Murray-Overton.

Ambos, Han e Murray - Overton, solucionam o problema de minimax usando sub-problemas sob a perspectiva dos métodos sequencial quadráticos (MSQ), o primeiro solucionando o problema dual, e Murray-Overton atacando diretamente o primal, considerando porém apenas as restrições ativas.



### III.3. MÉTODO DE CHARALAMBOUS E CONN

#### III.3.1: INTRODUÇÃO

Considerando o problema [I.1] equivalente a

min  $z$

sujeito a:

[III.3.1]

$$\phi_i(z, x) = z - f_i(x) \geq 0, \quad i=1, \dots, m,$$

os autores apresentam um algoritmo para resolvê-lo, utilizando o critério de  $\epsilon$ -viabilidade, determinando a direção pelo o método do Gradiente Projetado.

Os pontos básicos do algoritmo consistem no cálculo de duas direções. A primeira direção horizontal, objetiva o decréscimo de função  $\varphi$ . A segunda direção é vertical, e procura o decréscimo do erro das funções que estão próximas à função  $\varphi$ .

#### III.3.2 NOTAÇÃO

Necessitamos encontrar a direção de busca no ponto  $x^k$ , a qual é determinada tendo em consideração as funções  $\epsilon$ -ativas do problema minimax em  $x^k$ . Especificamente denotamos

$$1. \rightarrow I(x^k, \varepsilon) = \left\{ i=1, \dots, m \mid \varphi(x^k) - f_i(x^k) < \varepsilon \right\} \quad \text{[III.3.2]}$$

$$= \left\{ i=1, \dots, m \mid z_k - f_i(x^k) < \varepsilon \right\} \quad \text{[III.3.3]}$$

2. ->  $E(x^k, \varepsilon)$ : conjunto das funções e restrições não  $\varepsilon$ -ativas no ponto  $x^k$

$$3. \rightarrow \nabla \phi_j(z, x) = \left[ \frac{\partial \phi_j(z, x)}{\partial z}, \frac{\partial \phi_j(z, x)}{\partial x_1}, \dots, \frac{\partial \phi_j(z, x)}{\partial x_n} \right]^T$$

$$= \left[ 1, -\nabla f_j(x)^T \right]^T$$

4. -> Gradiente do vetor  $z$

$$\nabla z = e = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{n+1}$$

5. -> Label : identifica direção vertical no passo correspondente:

$$\text{Label} = \begin{cases} 6, & \text{se a direção vertical é tomada} \\ 0, & \text{caso contrário.} \end{cases}$$

6.-) VS : indica o tamanho do passo vertical .

$$VS = \begin{cases} 1, & \text{é acertado,} \\ 0, & \text{se o passo não é aceito.} \end{cases}$$

7.-) EPSTOP : valor que define o critério de parada.

8.-)  $\tau_{\max}$  : denota o limite superior do tamanho do passo admitido.

### III.3.3 ALGORITMO

#### PASSO 0

Label =0, k =0 , VS =0,  $x = x^0$

dar o valor de  $\epsilon$  e epstop,

#### PASSO I

Determinar  $I(x^k, \epsilon)$

#### PASSO II

Calcular a matriz de projeção (este passo efetua iterações internas)

(1) Seja  $j = 0$ ,

$P^0 = I$ , matriz identidade de ordem  $(n+1) \times (n+1)$

$A_0 = \emptyset$  (conjunto vazio)

$p^{(j)} = P^{(j)}$  e , (direção de busca horizontal)

[III.3.4]

(2) Quando  $j > 0$ , determinar a matriz  $N^{(j)}$  cujas linhas são os gradientes das restrições  $\epsilon$ -ativas,

$$N^{(j)} = \begin{bmatrix} \nabla \phi_{i_1}^T(z, x^k) \\ \vdots \\ \nabla \phi_{i_j}^T(z, x^k) \end{bmatrix} \quad j \times (n+1)$$

Calcular a matriz de projeção

$$P^{(j)} = I - \left( N^{(j)} \right)^T \left[ N^{(j)} \left( N^{(j)} \right)^T \right]^{-1} N^{(j)},$$

de ordem  $(n+1) \times (n+1)$ ,

Calcular a direção dada por III.3.41;

Determinar

$$i_{j+1} = \left\{ i / i = \arg \max \frac{P^{(j)T} \nabla \phi_i(z, x^k)}{\|P^{(j)}\| \|\nabla \phi_i(z, x^k)\|}, i \in I(x^k, \varepsilon) \setminus A_j \right\}$$

onde  $A_j = \{i_1, i_2, \dots, i_j\}$ , correspondentes ao valor de  $j$  anterior neste passo. Verificamos:

$$\text{se, } P^{(j)T} \nabla \phi_i(z, x^k) \leq 0 \text{ ir para (3), } i \in i_{j+1} ;$$

senão atualizar

$$A_{j+1} = A_j \cup \{i_{j+1}\}$$

$$j \leftarrow j+1$$

Se a cardinalidade de  $A_j = n+1$ , ir para o passo VII; caso contrário calcular uma nova matriz de projeção  $N^{(j)}$  e a nova

direção, continuando o processo.

(3) Denotamos:

$$A(x^k, \varepsilon) = A_j$$

$$P(x^k, \varepsilon) = P^{(j)}$$

$$p = p(x^k, \varepsilon) = p^{(j)}$$

### PASSO III

Se  $\|p\| < \varepsilon_{\text{stop}}$  e  $\text{Label} \neq 6$ , ou  $\varepsilon < \varepsilon_{\text{stop}}$

parar (ótimo é atingido)

Caso contrário ir para o passo IV

### PASSO IV

Busca linear (horizontal):

Determinar  $\tau > 0$  tal que  $\max f_i(x^{k-\tau} \bar{p})$ ,  $i=1, \dots, m$

é minimizado.  $\bar{p}$  é obtido de  $p$ , através de suas  $n$  últimas componentes.

### PASSO V

Condições para determinar o passo vertical: é necessário que o número das restrições  $\varepsilon$ -ativas não mudem em três iterações consecutivas (heurística) e  $\|p\| < 1$ , ou  $\text{Label}=6$  neste caso, ir para o passo VI,

caso contrário, fazer

$$x^{k+1} = x^k$$

$$k \leftarrow k+1$$

e ir para o passo I,

### PASSO VI

Para determinar o passo vertical consideramos

1.-) Determinar  $I(x^k, \varepsilon)$

2. -> Determinar a matriz  $N$  cujas linhas são os gradientes das restrições  $\varepsilon$ -ativas,

$$N = N^{(j)} = \begin{bmatrix} \nabla \phi_{i_1}^T(z, x^k) \\ \vdots \\ \nabla \phi_{i_j}^T(z, x^k) \end{bmatrix} \quad j \times (n+1)$$

$$3. -> x_{temp} = x^k + \bar{v},$$

onde  $\bar{v}$  é o vetor das  $n$  componentes primais de  $v(x^k, \varepsilon)$

$$\varepsilon v(x^k, \varepsilon) = -N^T (N N^T)^{-1} \phi, \quad \phi^T = (\phi_{i_1}, \dots, \phi_{i_j})$$

4. ->

Se  $\max f_i(x_{temp}) < \max f_i(x^k)$ ,  $i=1, \dots, m$ , então

$$VS=1$$

$$x^{k+1} = x_{temp}$$

$$k \leftarrow k+1$$

$$\text{Label} = 0$$

ir para o passo I,

caso contrário

$$VS=0$$

$$x^{k+1} = x^k$$

$$k \leftarrow k+1$$

ir para o passo I

PASSO VII

Quando  $|A_j| = n+1$

fazer

$$\text{Label} = 6$$

$$\varepsilon = \varepsilon/10$$

voltar para o passo VI .

### III. 3. 5: OBSERVAÇÕES

1.- Supõe-se posto  $N^{(j)}$  completo, de modo que

$$N^{(j)} \left[ N^{(j)} \right]^T \text{ é não singular.}$$

Para calcular

$$\left[ N^{(j)} \left[ N^{(j)} \right]^T \right]^{-1}$$

se utiliza a fórmula iterativa de Rosen para problemas não lineares (ver [15]),

2.- A matriz  $N$  representada no passo [III,(2)] é uma matriz cujas linhas são os gradientes de algumas ou todas as restrições  $\varepsilon$ -ativas.

Assim, a matriz  $N$  do passo VI,(2) é a matriz cujas linhas são as gradientes das restrições  $\varepsilon$ -ativas as quais formam uma base do espaço gerado pelos gradientes de todas as restrições ativas.

3.- Quando  $|A_j| = n + 1$  podem acontecer as seguintes possibilidades,

(a) O ótimo é atingido em uma vizinhança.

(b) O ótimo não é atingido e portanto temos que considerar restrições que não estão intervindo na solução.

Esta situação é tratada de duas maneiras:

Primeiro, toma-se o conjunto das direções verticais (Label=6);

segundo, reduz-se o valor de  $\epsilon$ , ( $\epsilon = \epsilon/10$ ).

### III.3.6: RESULTADOS TEÓRICOS

Apresentamos o essencial da teoria desenvolvida no artigo citado.

Previamente necessitamos de algumas hipóteses adicionais.

a) Hipótese 1.- Estrita complementaridade um ponto ótimo  $\hat{x}$  no problema de minimax é caracterizado por satisfazer as condições de Kuhn-Tucker abaixo:

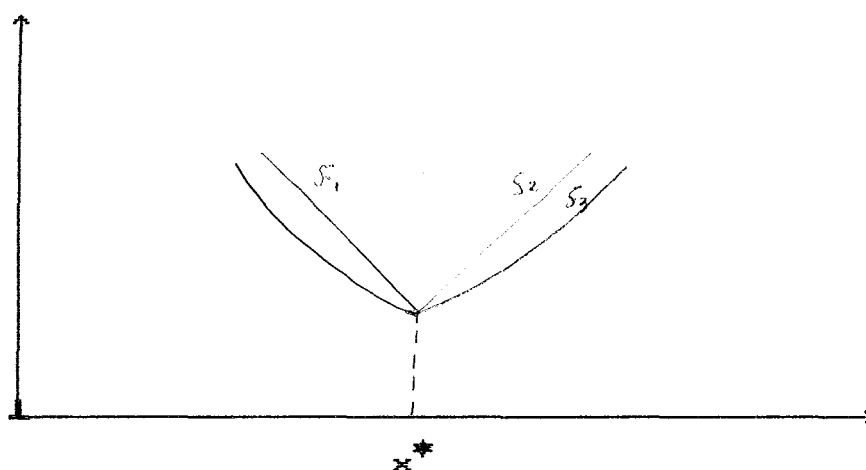
$$\sum_{i \in I(\hat{x}, 0)} u_i \nabla f_i(\hat{x}) = 0$$

$$\sum_{i \in I(\hat{x}, 0)} u_i = 1$$

$$u_i \geq 0, \quad i \in I(\hat{x}, 0)$$

Suporemos  $u_i > 0, \quad i \in I(\hat{x}, 0)$ . Significa não considerar problemas como o do gráfico seguinte:

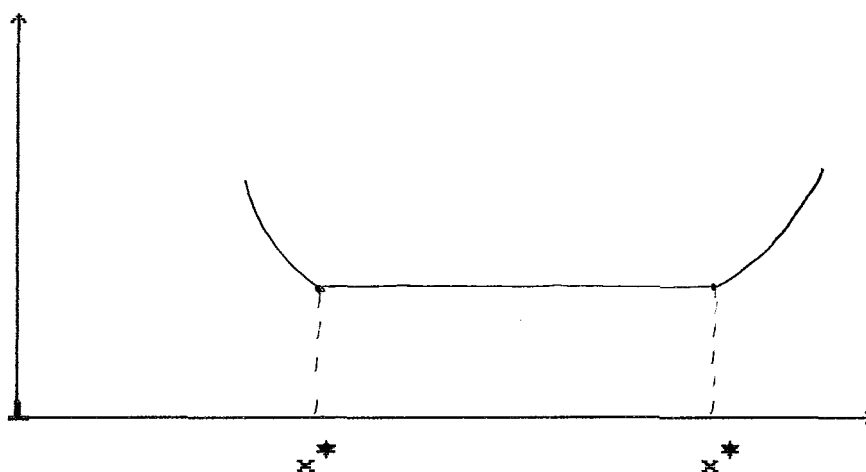




Fig(3.2)

b) Hipótese 2. - Unicidade de solução.

A situação a seguir mostrada em gráfico é impedida por esta hipótese:



Fig(3.3)

TEOREMA 1. -

Sejam as funções  $f_i$   $i=1, \dots, m$ , convexas e continuamente diferenciáveis em uma vizinhança de  $\hat{x}$ . Além disso, as hipóteses 1 e 2 são verificadas. Então, para um compacto  $W \subseteq \mathbb{R}^n$ , uma condição necessária e suficiente para que

$$x \in W, \quad p(x, \varepsilon) = 0, \text{ para } \varepsilon = 0,$$

é que  $x = \hat{x}$ .

Este teorema define o critério de parada do algoritmo.

Observa-se que este teorema garante o passo III do algoritmo.

#### TEOREMA 2. -

Suponhamos que  $f_i, i=1, \dots, m$ , são funções convexas, duas vezes diferenciáveis,  $\hat{x}$  é um ponto de mínimo forte de  $\varphi(x)$ , e a hipótese 1 acontece. Então para qualquer compacto  $W \subseteq \mathbb{R}^n$  e  $\varepsilon$  positivo e suficientemente pequeno, a seqüência  $\{x^i(\varepsilon)\} \subseteq W$  gerada pelo algoritmo converge para  $\hat{x}$ .

#### III.3.7 Charalambous e Conn e o método de gradiente projetado.

As características muito próprias do algoritmo desenvolvido por Charalambous e Conn poderiam mascarar o fato básico de ser ele um método de projeção. Por isto faremos a seguir a tradução do gradiente projetado de Rosen (1961) para o problema [III.3.1], com o cuidado, necessário para a convergência, de se considerar restrições  $\varepsilon$ -ativas, porém sua reconstituição de viabilidade.

A função a minimizar  $z$  tem em  $e=(1,0, \dots, 0)^T \in \mathbb{R}^{n+1}$  seu gradiente (constante). Em uma iteração  $k$ , sejam  $i_1, \dots, i_j$  as restrições  $\varepsilon$ -ativas e

$$J_k = \left[ \begin{array}{c} 1, \dots, 1 \\ -\nabla f_{i_1}(x^k), \dots, -\nabla f_{i_j}(x^k) \end{array} \right]^T$$

a jacobiana destas restrições em  $x^k$ , suposta de posto máximo. A

direção a ser considerada pelo método do gradiente projetado é dada por

$$p = -P_k e, \quad P_k = I - J_k^T \left[ J_k \ J_k^T \right]^{-1} J_k.$$

À diferença deste, o algoritmo de Charalambous e Conn não utiliza obrigatoriamente, em cada iteração, todo o conjunto das restrições  $\varepsilon$ -ativas: o teste de direção de descida para  $p$  no passo II(2) permite trabalhar com apenas um sub-conjunto da aquelas restrições.

O passo seguinte poderá ser o IV (a pós passagem por II(3) e III), que é a busca linear, feita sobre o  $\max f_i(x)$ , em  $x^k$ , segundo a direção de descida corrente.

Isto é essencialmente equivalente a minimizar a função custo  $z (z, f_i(x), \forall i)$ , em concordância com o que seria realizado ao gradiente projetado tradicional.

### III.4: MÉTODO DE ZANG

#### III.4.1. MOTIVAÇÃO E DESCRIÇÃO DO MÉTODO

##### III.4.1.1. - MOTIVAÇÃO

Neste item fazemos uma análise do método de Zang (1980) que sugere aproximações suaves nos pontos de não diferenciabilidade.

A aproximação será feita em uma vizinhança arbitrariamente pequena de cada um destes pontos.

O volume da região onde  $\varphi(x)$  é aproximada é controlado por um parâmetro denotado por  $\beta$ . Quando  $\beta \rightarrow 0$  o volume das zonas onde  $\varphi(x)$  é aproximada também se aproxima de zero.

Com a finalidade de suavizar os pontos não diferenciáveis introduzimos duas aproximações: uma delas é continuamente diferenciável e a outra, duas vezes continuamente diferenciável. As propriedades das aproximações permitirão que a função objetivo aproximada seja convexa, caso  $\varphi$  o seja.

##### III.4.1.2. - DESCRIÇÃO DO MÉTODO

Inicialmente observamos que  $\varphi(x)$  pode ser reescrita sob a forma:

$$\varphi(x) = f_1(x) + q \left[ f_2(x) - f_1(x) + q \left[ \dots + q \left[ f_{m-1}(x) - f_{m-2}(x) + q \left[ f_m(x) - f_{m-1}(x) \right] \dots \right] \right] \right], \quad \text{[III.4.1]}$$

onde

$$q(t) = \max(0, t), \quad t \in \mathbb{R}. \quad \text{[III.4.2]}$$

Observa-se que  $\varphi(x)$  é expresso como uma composição de  $q(t)$  e das funções  $f_i(x)$ ,  $i=1, \dots, m$ , continuamente diferenciáveis e convexas ( $\varphi$  também é convexa); somente  $q$  é não diferenciável, sendo  $t = 0$  o ponto de não diferenciabilidade.

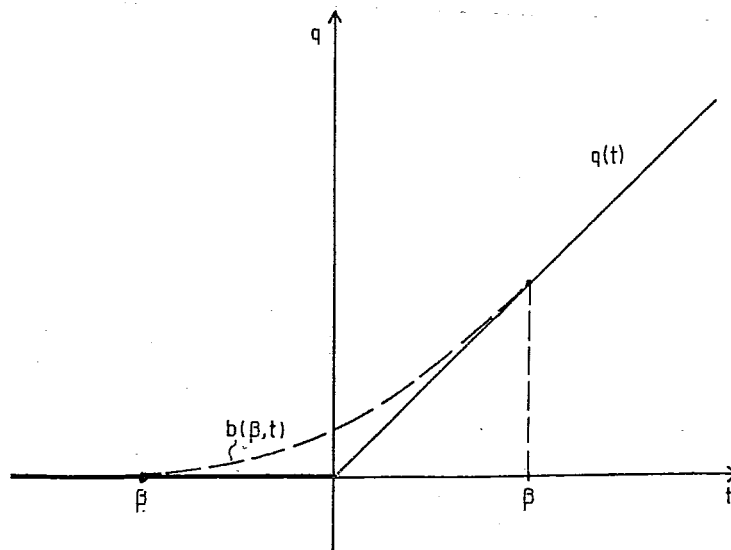


Fig (3.3)

Substituído  $q(t)$  numa vizinhança  $-\beta \leq t \leq \beta$  por alguma função  $b(\beta, t)$  conveniente (ver Fig (3.5)), temos, como resultado  $\bar{q}(\beta, t)$ , uma aproximação para  $q(t)$ .

$$\bar{q}(\beta, t) = \begin{cases} 0 & , \text{ se } t \leq -\beta ; \\ b(\beta, t), & \text{ se } -\beta \leq t \leq \beta; \\ t & , \text{ se } \beta \leq t . \end{cases} \quad \text{[III.4.3]}$$

Assim,  $\varphi$  é aproximada por:

$$\begin{aligned} \bar{\varphi}(\beta, x) = & f_1(x) + \bar{q} \left[ \beta, f_2(x) - f_1(x) + \bar{q} \left[ \beta, \dots + \bar{q} \left[ \beta, \dots, f_{m-1}(x) \right. \right. \right. \\ & \left. \left. \left. - f_{m-2}(x) + \bar{q} \left[ \beta, f_m(x) - f_{m-1}(x) \right] \right] \dots \right] \right], \quad \text{[III.4.4]} \end{aligned}$$

Suporemos que  $\bar{q}(\beta, t)$  tem as seguintes propriedades :

i)  $\bar{q}(\beta, t)$  é convexa e continuamente diferenciável com relação a  $t$ .

Evidentemente  $\bar{\varphi}(\beta, x)$  herda esta propriedade.

Por outro lado, se  $\bar{q}(\beta, t)$  é convexa, então  $b(\beta, t)$  é monotonicamente crescente em  $t$ .

ii) Escolhe-se  $b(\beta, t)$  de modo que

$$\lim_{\beta \rightarrow 0} \bar{q}(\beta, t) = q(t). \quad \text{[III.4.5]}$$

Verifica-se que (i) e (ii) são satisfeitas pelas funções de

aproximação dadas por:

$$b_1(\beta, t) = \frac{t^2}{4\beta} + \frac{t}{2} + \frac{\beta}{4}; \quad \text{[III.4.6]}$$

e

$$b_2(\beta, t) = \frac{-t^4}{16\beta^3} + \frac{3t^2}{8\beta} + \frac{t}{2} + \frac{3\beta}{16}. \quad \text{[III.4.7]}$$

#### VERIFICAÇÃO DE (i)

É imediato que  $\bar{q}_1(\beta, t)$  (correspondente a  $b_1(\beta, t)$ ) é continuamente diferenciável.

Além disto,

$$\frac{d^2 b_1(\beta, t)}{dt^2} = \frac{1}{2\beta} > 0, \quad \beta > 0,$$

portanto,  $b_1(\beta, t)$  é convexa, e assim é  $\bar{q}_1(\beta, t)$ .

$b_2(\beta, t)$  é claramente continuamente diferenciável. Também determinemos a derivada segunda em  $t$

$$\frac{d^2 b_2(\beta, t)}{dt^2} = -\frac{3t^2}{4\beta^3} + \frac{3}{4\beta}.$$

Temos, para

$$a) \quad t=0, \quad \frac{d^2 b_2(\beta, t)}{dt^2} = \frac{3}{4\beta} > 0;$$

$$b) \quad t=\pm\beta, \quad \frac{d^2 b_2(\beta, t)}{dt^2} = 0;$$

$$c) \quad 0 < t < \beta, \quad \frac{d^2 b_2(\beta, t)}{dt^2} = -\frac{3t^2}{4\beta^3} + \frac{3}{4\beta} > 0;$$

$$d) \quad -\beta < t < 0, \quad \frac{d^2 b_2(\beta, t)}{dt^2} > 0.$$

De a, b, c e d temos que

$$\frac{d^2 b_2(\beta, t)}{dt^2} \geq 0,$$

então,  $b_2(\beta, t)$  é convexa, assim como  $\bar{q}_2(\beta, t)$ .

Verificação de (ii):

Trata-se de analisar o intervalo  $-\beta \leq t \leq \beta$ .

Podemos portanto supor que  $t = k\beta^\alpha$ ,  $\alpha \geq 1$ ,  $k > 0$ , e

$$\lim_{\beta \rightarrow 0} b_1(\beta, t) = \lim_{\beta \rightarrow 0} \left[ \frac{t^2}{4\beta} + \frac{t}{2} + \frac{\beta}{4} \right]; \quad \text{[III.4.6.1]}$$

$$= \lim_{\beta \rightarrow 0} \left[ \frac{k^2 \beta^{2\alpha}}{4\beta} + \frac{k\beta^\alpha}{2} + \frac{\beta}{4} \right] = 0,$$

Analogamente se verifica que

$$\lim_{\beta \rightarrow 0} b_2(\beta, t) = 0, \quad \text{para } -\beta \leq t \leq \beta. \quad \blacksquare$$

Outras aproximações para o operador "max" tendo a j-ésima

derivada contínua  $j > 2$  (onde  $b(\beta, t)$  é um polinômio de  $2j$ -ésimo grau, com condições de continuidade em  $t = \pm \beta$ ), são



$$b_3(\beta, t) = \frac{t^5}{32\beta^5} - \frac{5t^4}{32\beta^3} + \frac{15t^2}{32\beta} + \frac{t}{2} + \frac{5\beta}{32},$$

$$b_4(\beta, t) = \frac{5t^6}{256\beta^7} + \frac{7t^5}{64\beta^5} - \frac{35t^4}{128\beta^3} + \frac{35t^2}{64\beta} + \frac{t}{2} + \frac{35\beta}{256}.$$

Com qualquer destas funções aproximantes o problema [I.1] é substituído por :

$$\min_{x \in \mathbb{R}^n} \bar{\varphi}(\beta, x), \quad \text{[III.4.9]}$$

Desta maneira, a minimização de  $\varphi(\beta, x)$  pode ser feita mediante técnicas de minimização que utilizam gradiente.

UMA PROPRIEDADE COMPARATIVA DE  $\bar{q}_1$ ,  $\bar{q}_2$  e  $q$ .

Temos

$$\bar{q}_1(\beta, t) \geq \bar{q}_2(\beta, t) \geq q(t), \text{ isto é} \quad \text{[III.4.10]}$$

$$b_1(\beta, t) \geq b_2(\beta, t), \quad -\beta \leq t \leq \beta$$

Seja

$$\begin{aligned} d &= b_1(\beta, t) - b_2(\beta, t) \\ &= \frac{t^2}{4\beta} + \frac{t}{2} + \frac{\beta}{4} - \left[ \frac{-t^4}{16\beta^3} + \frac{3t^2}{8\beta} + \frac{t}{2} + \frac{3\beta}{16} \right] \\ &= \frac{t^4}{16\beta^3} - \frac{t^2}{8\beta} + \frac{\beta}{16}, \end{aligned}$$

$$= \frac{1}{16\beta^3} \left[ t^4 - 2\beta^2 t^2 + \beta^4 \right],$$

$$= \frac{1}{16\beta^3} \left[ t^2 - \beta^2 \right]^2 \geq 0.$$

Portanto  $b_1(\beta, t) - b_2(\beta, t) \geq 0$ .

Por outro lado vejamos que

$$\bar{q}_2(\beta, t) \geq q(t) \text{ em } [-\beta, \beta]$$

1) como  $q(t) = t$  em  $[0, \beta]$

$$b_2(\beta, t) - t = \begin{cases} \frac{3\beta}{16}, & \text{em } t=0; \\ 0, & \text{em } t=\beta. \end{cases}$$

Se mostramos que  $b_2(\beta, t) - q(t)$  é decrescente em  $\langle 0, \beta \rangle$ , isto é

$\frac{d}{dt} [b_2(\beta, t) - t] < 0$  em  $\langle 0, \beta \rangle$ , teremos o resultado desejado em  $[0, \beta]$ .

De fato

$$\frac{d [b_2(\beta, t) - t]}{dt} = -\frac{t^3}{4\beta^3} + \frac{3t}{4\beta} - \frac{1}{2}; \quad \text{[III.4.11]}$$

substituas  $\eta = \frac{t}{\beta}$  :

$$-\frac{\eta^3}{4} + \frac{3\eta}{4} + \frac{1}{2} = -\frac{1}{4} (\eta+2) (\eta-1)^2 < 0,$$

já que  $0 \leq \eta \leq 1$ .

2) Analogamente  $\bar{q}_2(\beta, t) \geq q(t) = 0$ , em  $-\beta \leq t \leq 0$ .

Conseqüentemente

$$\bar{\varphi}_1(\beta, x) \geq \bar{\varphi}_2(\beta, x) \geq \varphi(x). \quad \text{[III.4.12]}$$

## III.4.2. - EXEMPLO

Considere  $m = 2$ . Por III.4.11

$$\varphi(x) = f_1(x) + q \left[ f_2(x) - f_1(x) \right],$$

onde

$$q(t) = \max(0, t), \quad t \in \mathbb{R}.$$

Segundo a definição temos a seguinte função regularizada .

$$\bar{\varphi}(\beta, x) = f_1(x) + \bar{q} \left[ \beta, f_2(x) - f_1(x) \right],$$

$$\bar{\varphi}(\beta, x) = \begin{cases} f_1(x) & \text{se } f_2(x) - f_1(x) \leq -\beta, \\ f_1(x) + b_j(\beta, t) & \text{se } -\beta \leq f_2(x) - f_1(x) \leq \beta, \\ f_2(x) & \text{se } f_2(x) - f_1(x) \geq \beta. \end{cases}$$

Com  $b_1$ , [III.4.6], temos a aproximação

$$\bar{\varphi}_1(\beta, x) = \begin{cases} f_1(x), & f_2(x) - f_1(x) \leq -\beta, \\ \frac{\left[ f_2(x) - f_1(x) \right]^2}{4\beta} + \frac{\left[ f_2(x) - f_1(x) \right]}{2} + \frac{\beta}{4} + f_1(x), & -\beta \leq f_2(x) - f_1(x) \leq \beta, \\ f_2(x), & \beta \leq f_2(x) - f_1(x). \end{cases}$$

[III.4.13]

Com  $b_2$  temos a segunda aproximação

$$\bar{\varphi}_2(\beta, x) = \begin{cases} f_1(x) & f_2(x) - f_1(x) \leq -\beta, \\ -\frac{[f_2(x) - f_1(x)]^4}{16\beta^3} + \frac{3[f_2(x) - f_1(x)]^2}{8\beta} + \\ \frac{[f_2(x) - f_1(x)]}{2} + \frac{3\beta}{16} + f_1(x), & -\beta \leq f_2(x) - f_1(x) \leq \beta, \\ f_2(x) & \beta \leq f_2(x) - f_1(x), \end{cases}$$

[III.4.14]

Observação.

A função regularizada  $\bar{\varphi}(\beta, x)$ , calculada por [III.4.4], depende dos valores de  $f_i$ ,  $i=1, \dots, m$  e de sua ordem para  $-\beta < x < \beta$ . Esclarecemos esta afirmação mediante o exemplo:

Calculemos  $\bar{\varphi}_1(\beta, x)$  para  $\varphi$  suposto com valores a seguir indicados

Primeiro caso. Sejam  $x$  e  $f_i$  tais que :

Quando  $f_1(x)=1$ ,  $f_2(x)=2$ ,  $f_3(x)=3$ , e  $\beta=4$

$$\bar{\varphi}_1(\beta, x) = f_1(x) + \bar{q}_1 \left\{ \beta, f_2(x) - f_1(x) + \bar{q}_1(\beta, f_3(x) - f_2(x)) \right\}$$

$$\bar{\varphi}_1(\beta, x) = f_1(x) + \bar{q}_1 \left\{ \beta, f_2(x) - f_1(x) + \bar{q}_1(4, 1) \right\}$$

$$\bar{\varphi}_1(\beta, x) = f_1(x) + \bar{q}_1 \left\{ \beta, f_2(x) - f_1(x) + b_1(4, 1) \right\},$$

$$\bar{\varphi}_1(\beta, x) = f_1(x) + \bar{q}_1 \left\{ 4, 1 + \left[ \frac{t^2}{4\beta} + \frac{t}{2} + \frac{\beta}{4} \right] \right\}$$

$$\bar{\varphi}_1(\beta, x) = f_1 + b_1 \left( 4, \frac{41}{16} \right)$$

$$\bar{\varphi}_1(\beta, x) = 3 + \left[ \frac{1681}{4196} + \frac{9}{32} \right] \cong 3.692$$

$$\bar{\varphi}_1(\beta, x) = 3.692 \quad |$$

Segundo caso: Reordenando os valores de  $f_i$ ,  $i=1, 2, 3$ .

$$f_1(x)=2, f_2(x)=1, f_3(x)=3, \beta=4.$$

Analogamente, temos

$$\bar{\varphi}_1(\beta, x) = 2 + b_1 \left( 4, -\frac{7}{16} \right)$$

$$\bar{\varphi}_1(\beta, x) = 3.723 \quad |$$

### III.4.3: ALGORITMO

I.- Dado um ponto  $x_0 \in \mathbb{R}^n$  e um número positivo  $\beta_0$ ,  $\kappa = 0$ ,  $\varepsilon > 0$  dado.

II.-  $\kappa \leftarrow \kappa+1$

Calcule a solução do problema aproximado :

$$\bar{\varphi}(\beta_k, x^k) = \min_{x \in \mathbb{R}^n} \bar{\varphi}(\beta_k, x) \quad \text{[III.4.15]}$$

a) Verifique se  $x^k$  é aceitável pelo problema original:

$$\bar{\varphi}(\beta_k, x^k) = \varphi(x^k)$$

ou

$$\beta_k \leq \varepsilon.$$

PARAR.

b) Caso contrário

$$\beta_{k+1} = \beta_k / \alpha, \quad \alpha > 1$$

Retorne a II.

Observação: Zang [30] utiliza Quase-Newton para resolver

[III.4.15]

#### III.4.4: RESULTADOS TEÓRICOS

A seguir apresentamos algumas propriedades do problema aproximado e do esquema geral do algoritmo sugerido na seção anterior.

Observamos que o objetivo é mostrar em quanto diferem

o mínimo do problema aproximado e o mínimo da função em estudo. Para isto precisamos dos seguintes lemas que consideramos como hipótese a propriedade, dada por (i)

Lema 4.1. - Para todo  $t \in \mathbb{R}$ , verifica-se

$$0 \leq \bar{q}(\beta, t) - q(t) \leq b(\beta, 0).$$

O máximo valor de  $\bar{q}(\beta, t) - q(t)$  tem relação com o parâmetro  $\beta$ , conforme se tem no

Lema 4.2. - Para todo  $t \in \mathbb{R}$

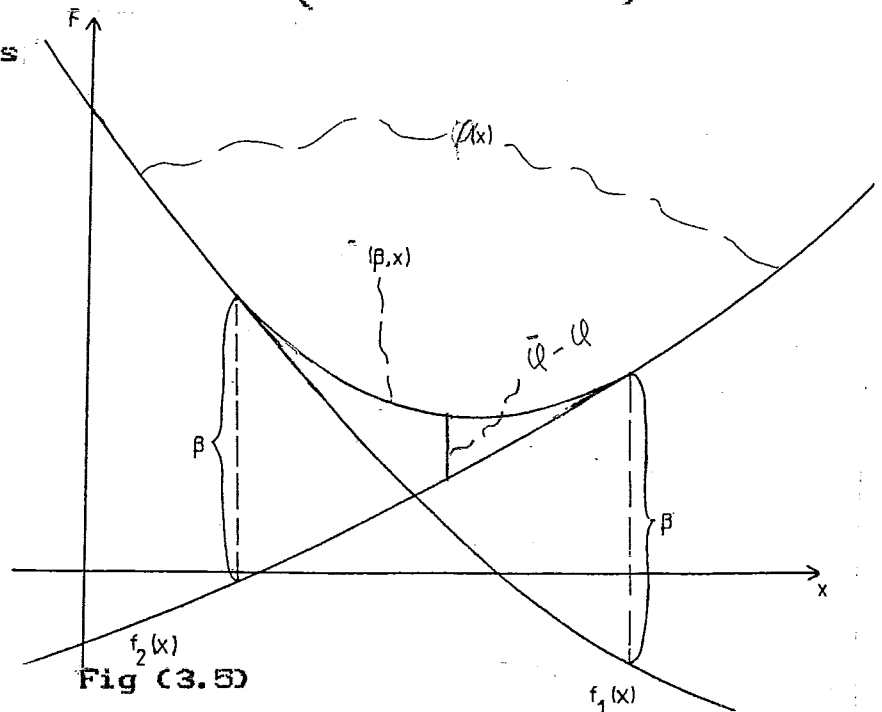
$$\bar{q}(\beta, t+\beta) - q(t) \leq \beta.$$

Dos lemas (4.1) e (4.2) obtém-se o seguinte resultado, considerando (ii) como hipótese para as seguintes proposições (4.3) e (4.4).

PROPOSIÇÃO 4.3. - Para todo  $x \in \mathbb{R}^n$  verifica-se

$$0 \leq \bar{\varphi}(\beta, x) - \varphi(x) \leq \min \left\{ \beta, (m-1) b(\beta, 0) \right\}.$$

Graficamente teremos,



TEOREMA 4.4. - Seja  $x^* \in \mathbb{R}^n$  ponto mínimo de  $\varphi(x)$  e  $\bar{x} \in \mathbb{R}^n$  o ponto mínimo de  $\bar{\varphi}(\beta, x)$ ; então

$$0 \leq \bar{\varphi}(\beta, \bar{x}) - \varphi(x^*) \leq \min \left\{ \beta, (m-1) b(\beta, 0) \right\}.$$

Observação:

O teorema (4.4) relaciona  $\beta$  com o erro residual, sugerindo controle das iterações. Por exemplo,

$$\begin{aligned} \min \left\{ \beta, (m-1) b_1(\beta, 0) \right\} &= \min \left\{ \beta, (m-1) \frac{\beta}{4} \right\} \\ &= \begin{cases} \beta, & \text{se } m \geq 5; \\ (m-1) \frac{\beta}{4}, & \text{senão.} \end{cases} \end{aligned}$$

#### CONVERGÊNCIA

A proposição seguinte fornece um resultado de convergência fraca de  $x^k$  (convergência em  $\varphi(x^k)$ ).

PROPOSIÇÃO 4.5. - Seja  $\{\beta_k\} \rightarrow 0$  uma seqüência de números positivos e suponha que  $x_k$  é solução para

$$\min_{x \in \mathbb{R}^n} \bar{\varphi}(\beta_k, x).$$

Seja  $\bar{x} \in \mathbb{R}^n$  um ponto de acumulação da seqüência  $\{x_k\}$ .

Então  $\varphi(\bar{x}) = \varphi(x^*)$ .

Quando  $I(x^*)$  é um singleton, o problema de minimax é, em alguma vizinhança de  $x^*$ , equivalente ao de minimização de uma



simples função diferenciável. É o que formaliza a,

**PROPOSIÇÃO 4.6.** - Considere  $x^* \in \mathbb{R}^n$  um mínimo local de  $\varphi(x)$ , onde  $\varphi(x^*) = f_1(x^*)$  para um só valor  $1 \leq l < m$ . Então existe um  $\bar{\beta} \in \mathbb{R}^+$  tal que  $x^*$  é um mínimo local de  $\bar{\varphi}(\beta, x)$ , para todo  $\beta$  satisfazendo  $0 < \beta < \bar{\beta}$ .

### III.4.5: COMENTÁRIOS

Zang soluciona o problema de minimax usando aproximações em uma vizinhança arbitrariamente pequena dos pontos não diferenciáveis, trabalhando com

$$\min_{x \in \mathbb{R}^n} \bar{\varphi}(\beta, x);$$

onde  $\bar{\varphi}$  é função aproximada do problema original. Este problema pode ser resolvido por qualquer técnica da Programação Não Linear Irrestrita Diferenciável.

Mencionamos a seguir um trabalho que também realiza aproximação suavizante de  $\varphi$ , Bertsekas (1976).

Embora o artigo possibilite o trato de famílias mais amplas do que a de minimax, fixámo-nos nesta. Também aqui se aproxima cada função  $q(t) = \max(0, t)$ . A motivação é, no entanto, totalmente diversa da do método que acabamos de descrever.

Inicialmente, a função  $q(t)$  é substituída pela equivalente

$$q(t) = \min z \quad \text{[III.4.16]} \\ t \leq z, \quad z \geq 0$$

Em seguida, este problema é aproximado através do Lagrangeano Aumentado (método dos multiplicadores), relativamente à restrição

$t \leq z$ , chegando-se a

$$\min \left[ z + \frac{1}{2c} \left[ \left( \max \left\{ 0, y + c \left[ t - z \right] \right\} \right)^2 - y^2 \right] \right], \quad c > 0, \quad 0 \leq y \leq 1. \quad \text{[III.4.17]}$$

$C$  e  $y$  é um multiplicador de Lagrange associado a [III.4.16].

Visando nossa aplicação, para  $t = f_i(x)$ , a solução de [III.4.17] é dada por:

$$\bar{\gamma}(f_i(x), y^i, c) = \begin{cases} f_i(x) - \frac{(1-y^i)^2}{2c}, & \text{se } \frac{(1-y^i)}{c} \leq f_i(x); \\ y^i f_i(x) + \frac{c(f_i(x))^2}{2}, & \text{se } \frac{y^i}{c} \leq f_i(x) \leq \frac{(1-y^i)}{c}; \\ -\frac{(y^i)^2}{2c}, & \text{se } f_i(x) \leq -\frac{y^i}{c} \end{cases} \quad \text{[III.4.18]}$$

onde  $0 \leq y^i \leq 1$ ,  $c > 0$ ,  $i = 1, \dots, m$ .

O parâmetro  $c$  controla a exatidão da aproximação, e o parâmetro  $y^i$  determina-se a exatidão é maior para valores positivos ou negativos do argumento de  $f_i(x)$  (veja figura 3.7)

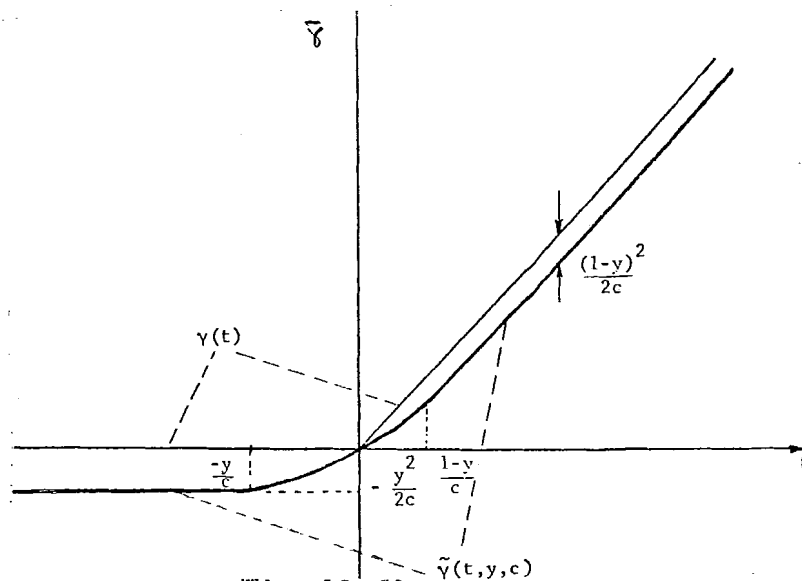


Fig (3.6)

Sendo  $f_i$ ,  $i=1, \dots, m$  funções diferenciáveis, então a função  $\bar{\gamma}(f_i(x), y^i, c)$  é também diferenciável em relação a  $x$ . O gradiente é dado por

$$\nabla \bar{\gamma}(f_i(x), y^i, c) = \begin{cases} \nabla f_i(x) & \text{se } \frac{(1-y^i)}{c} \leq f_i(x); \\ [y^i + c f_i(x)] \nabla f_i(x), & \text{se } \frac{y^i}{c} \leq f_i(x) \leq \frac{(1-y^i)}{c}; \\ 0 & \text{se } f_i(x) \leq -\frac{y^i}{c} \end{cases}$$

Devido à motivação da aproximação, o autor sugere atualizações dos parâmetros  $c$  e  $y^i$  através do método de multiplicadores.

### III.5. MÉTODO DE GÍGOLA E GÓMEZ

#### III.5.1: DEFINIÇÕES E PROPRIEDADES BÁSICAS

Inicialmente se reecreve  $\varphi(x)$  como

$$\varphi(x) = \sup_{u \in U} \left[ \sum_{i=1}^m u_i f_i(x) \right] = \text{Sup}_{u \in U} u^T f(x), \quad \text{[III.5.1]}$$

onde  $U$  é o simplex unitário, conjunto convexo definido por

$$U = \left\{ u \in \mathbb{R}^m \mid \sum_{i=1}^m u_i = 1, u_i \geq 0, i=1, \dots, m \right\}. \quad \text{[III.5.2]}$$

Para demonstração ver Gigola e Gómez (1991) ( Na verdade,  $u^T f(x)$  é a função dual com  $u \in U$ , para  $x$  verificando as condições de otimalidade, correspondente ao problema [III.2.1])

#### OBSERVAÇÃO

Nesta forma, as descontinuidades do gradiente de  $\varphi$  correspondem aos valores de  $x$  para os quais o problema do supremo tem uma multiplicidade de soluções. É, portanto, natural, forçar a unicidade da solução (e a continuidade e unicidade do gradiente) através de uma regularização quadrática. Este é um instrumento conhecido na Programação Matemática, em particular na Programação Linear, como, por exemplo, uma forma de atacar o mau condicionamento dos problemas.

Assim se define a regularização, para  $v \in U$ ,

$$\varphi(x) = \sup_v \sup_{u \in U} \left[ u^T f(x) - \frac{1}{2} \|u - v\|^2 \right] = \sup_{u \in U} h_v(x, \omega), \quad \text{[III.5.3]}$$

Vejamos um exemplo com a finalidade de visualizar melhor o critério de regularização.

### III.5.2: EXEMPLO

Seja  $\varphi(x) = \max(x^2, -x^2 + 1)$ , cuja representação geométrica é

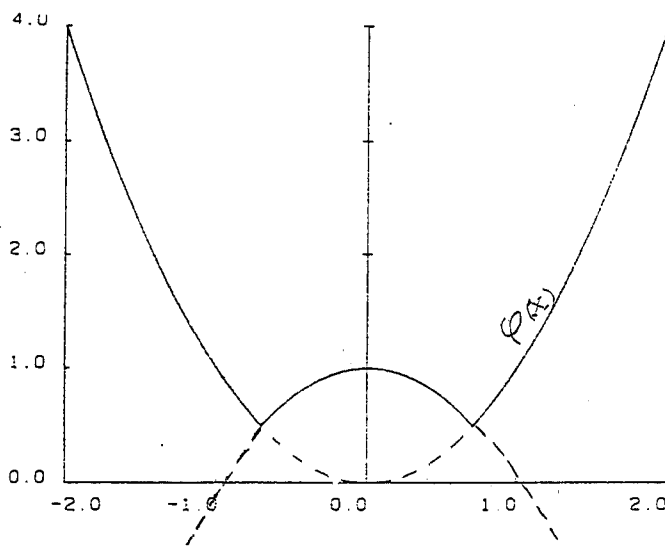


Fig. (3.7)

A função regularizada fica:

$$\varphi_v(x) = \sup_{u_1 \in [0,1]} \left\{ (u_1, 1-u_1) \begin{pmatrix} x^2 \\ -x^2+1 \end{pmatrix} - \frac{1}{2} \|(u_1 - v_1, 1-u_1 - v_2)\|^2 \right\},$$

$$\varphi_v(x) = \sup_{u_1 \in [0,1]} \left\{ (2x^2-1)u_1 - x^2 + 1 - \frac{1}{2}[(u_1 - v_1)^2 + (1-u_1 - v_2)^2] \right\};$$

se, por exemplo,  $v^1 = (0, 1)^T$ , então:

$$\varphi_{\mathbf{v}}(x) = \sup_{u \in [0,1]} \left\{ (2x^2-1) u_1 - u_1^2 + (1-x^2) \right\},$$

que se calcula diretamente, fornecendo (veja fig.(3.19))

$$\varphi_{\mathbf{v}^1}(x) = \begin{cases} -x^2+1, & -\sqrt{1/2} \leq x \leq \sqrt{1/2}, \\ x^2-1, & x \geq \sqrt{3/2} \text{ e } x \leq -\sqrt{3/2}, \\ x^4-2x^2+5/4, & \sqrt{1/2} \leq x \leq \sqrt{3/2} \text{ e } -\sqrt{3/2} \leq x \leq -\sqrt{1/2}, \end{cases}$$

Analogamente temos para  $\mathbf{v}^2 = (1, 0)^T$

$$\varphi_{\mathbf{v}^2}(x) = \begin{cases} x^4 + 1/4, & -\sqrt{1/2} \leq x \leq \sqrt{1/2}, \\ x^2, & x \geq \sqrt{1/2} \text{ e } x \leq -\sqrt{1/2}. \end{cases}$$

Para  $\mathbf{v}^3 = (1/2, 1/2)^T$

$$\varphi_{\mathbf{v}^3}(x) = \begin{cases} x^4 - x^2 + 3/4, & -1 \leq x \leq 1, \\ x^2 - 1/4, & 1 \leq x \text{ e } x \leq -1, \end{cases}$$

Observamos que a solução, quando  $\mathbf{v} = (1/2, 1/2)^T$ , é a solução global  $x = \pm \sqrt{1/2}$ .

$$\varphi_{v^1}(x) = \sup_{u \in \{0,1\}} \left\{ (2x^2-1) u_1 - u_1^2 + (1-x^2) \right\},$$

que se calcula diretamente, fornecendo (veja fig.(3.19))

$$\varphi_{v^1}(x) = \begin{cases} -x^2+1, & -\sqrt{1/2} \leq x \leq \sqrt{1/2}, \\ x^2-1, & x \geq \sqrt{3/2} \text{ e } x \leq -\sqrt{3/2}, \\ x^4-2x^2+5/4, & \sqrt{1/2} \leq x \leq \sqrt{3/2} \text{ e } -\sqrt{3/2} \leq x \leq -\sqrt{1/2}, \end{cases}$$

Analogamente temos para  $v^2 = (1,0)^T$

$$\varphi_{v^2}(x) = \begin{cases} x^4+1/4, & -\sqrt{1/2} \leq x \leq \sqrt{1/2}, \\ x^2, & x \geq \sqrt{1/2} \text{ e } x \leq -\sqrt{1/2}. \end{cases}$$

Para  $v^3 = (1/2, 1/2)^T$

$$\varphi_{v^3}(x) = \begin{cases} x^4 - x^2 + 3/4, & -1 \leq x \leq 1, \\ x^2 - 1/4, & 1 \leq x \text{ e } x \leq -1, \end{cases}$$

Observamos que a solução, quando  $v = (1/2, 1/2)^T$ , é a solução global  $x = \pm \sqrt{1/2}$ .

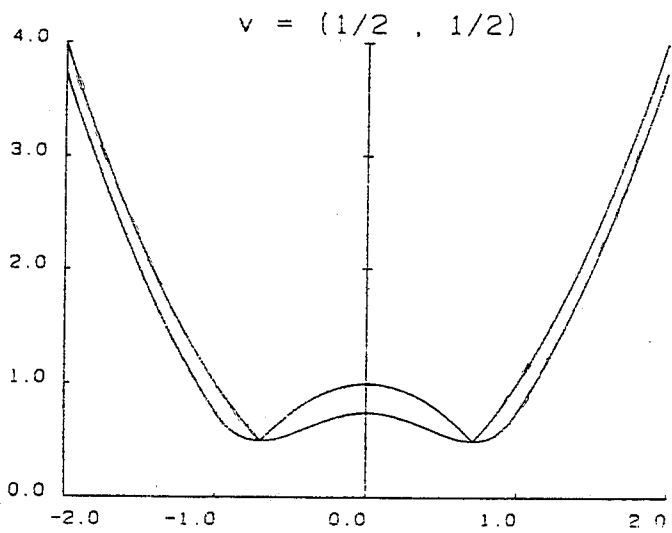
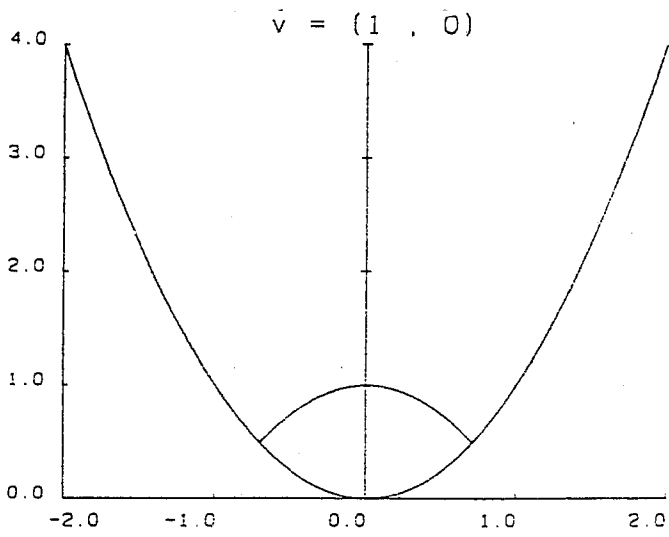
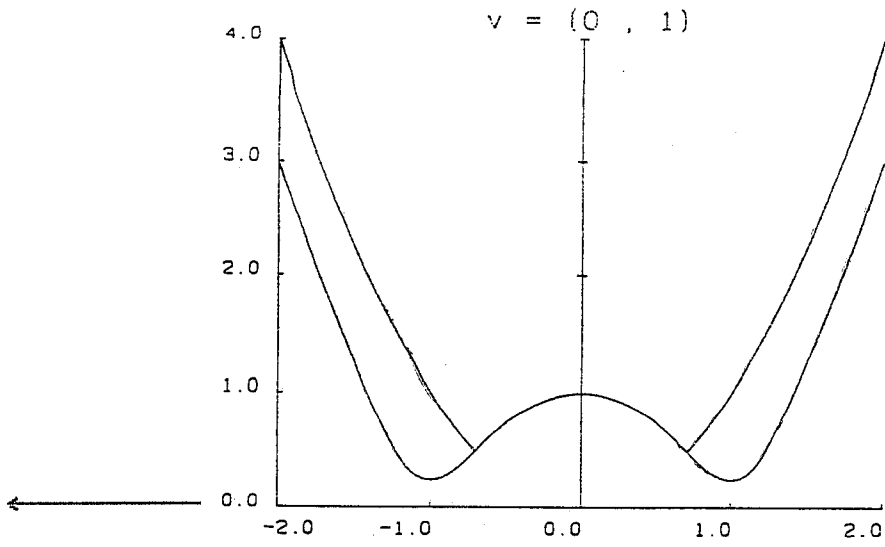


Fig. (3.8)



A função regularizada [III.5.3] tem as seguintes propriedades:

$$i) \varphi(x) = \sup_{v \in U} \varphi_v(x)$$

De fato,

$$\begin{aligned} \sup_{v \in U} \varphi_v(x) &= \sup_{v \in U} \sup_{u \in U} \left[ u^T f(x) - 1/2 \|u-v\|^2 \right], \\ &= \sup_{u \in U} \sup_{v \in U} \left[ u^T f(x) - 1/2 \|u-v\|^2 \right], \\ &= \sup_{u \in U} u^T f(x), \\ &= \varphi(x); \end{aligned}$$

a última igualdade decorrendo de que  $\|u-v\|^2 \geq 0$ , logo  $v = u$  realiza o sup em  $v$ .

$$ii) \varphi(x) - 1/2 (1 + \|v\|^2) \leq \varphi_v(x) \leq \varphi(x).$$

(veja demonstração Gómez e Gígola [10] )

### III.5.3: O ALGORITMO

Inicialização:

Ponto inicial  $(x^0, v^0)$ ,

Para  $k \geq 0$

1) Obtenção de uma aproximação da inversa da Hessiana da Lagrangeana do problema original  $H(x^k)$ , isto é uma, aproximação de

$$\left[ \sum_{i=1}^m v_i^k \nabla^2 f_i(x^k) \right]^{-1}$$

Utiliza-se, por exemplo, BFGS.

2) Cálculo das variáveis duais :

Resolve-se

$$\varphi_{\mathbf{v}}(\mathbf{x}^k) = \text{Sup}_{\mathbf{u} \in U} \left( \mathbf{u}^T \mathbf{f}(\mathbf{x}^k) - 1/2 \|\mathbf{u} - \mathbf{v}\|^2 \right),$$

cuja solução é dada por

$$u_i(\mathbf{x}^k, \mathbf{v}) = \begin{cases} f_i(\mathbf{x}^k) + v_i - \delta, & \text{se } f_i(\mathbf{x}^k) + v_i - \delta > 0, \\ 0, & \text{em outros casos,} \end{cases} \quad \text{[III.5.4]}$$

onde  $\delta$  é tal que

$$\sum_{i=1}^m \max \left( 0, f_i(\mathbf{x}^k) + v_i - \delta \right) = 1, \quad \text{[III.5.5]}$$

3) Cálculo da direção primal

$$\mathbf{p}^k = - \mathbf{H}(\mathbf{x}^k) \nabla_{\mathbf{v}^k} \varphi_{\mathbf{v}^k}(\mathbf{x}^k)$$

$$\varphi_{\mathbf{v}^k}(\mathbf{x}^k) = \sum_{i=1}^m u_i^k \nabla f_i(\mathbf{x}^k)$$

4) Busca linear: tal que a condição de descida

$$\varphi_{\mathbf{v}^k}(\mathbf{x}^k + \alpha \mathbf{p}^k) \leq \varphi_{\mathbf{v}^k}(\mathbf{x}^k) + \alpha w(\mathbf{p}^k)^T \nabla_{\mathbf{v}^k} \varphi_{\mathbf{v}^k}(\mathbf{x}^k),$$

é verificada.

Onde

$\alpha$  é o tamanho do passo,

$w = 0.1$  (valor típico)

Faça

$$x^{k+1} = x^k + \alpha p^k .$$

5. Se

$$\| \nabla_{v^k} \varphi(x^{k+1}) \| < \| \nabla_{v^{k-1}} \varphi(x^k) \| ,$$

ir para o passo 6; caso contrário, fazer

$$x^k \leftarrow x^{k+1} \text{ e voltar ao passo 1.}$$

6. Calcule  $u^{k+1}$  usando [III.5.4] e [III.5.5], com  $v^k \leftarrow u^k$   
e  $x^k \leftarrow x^{k+1}$ .

7. Se

$$\| \nabla_{v^k} \varphi(x^{k+1}) \| < \text{TOL}$$

e

$$\| u^{k+1} - v^k \| < \text{TOL} ,$$

PARE.

Caso contrário, fazer  $v^k \leftarrow u^{k+1}$  e  $k=k+1$  e voltar ao passo 1.

#### CALCULO DE $\delta$

Utilizamos o seguinte algoritmo para calcular  $\delta$ :

a) Geramos o vetor  $\beta$ , ordenando em sentido decrescente os valores

$$\beta_i = f_i(x^k) + v_i ;$$

b) Encontramos um elemento do conjunto  $j_0$ , como o primeiro índice no conjunto  $j_0$ , onde

$$j_0 = \{1, \dots, m-1\}, \quad \text{tal que}$$

$$-j_0 * \beta_{j_0+1} + \sum_{j=1}^{j_0} \beta_j - 1 \geq 0 ;$$

c) Caso contrário,  $j_0 = m$ .

Então

$$\delta = \frac{\sum_{j=1}^{j_0} \beta_j - 1}{j_0}.$$

### OBSERVAÇÕES

1. Para atualizar a Hessiana utiliza-se o método BFGS.
2. O tamanho de passo  $\alpha_k$  pode ser encontrado usando qualquer busca unidimensional, por exemplo a interpolação quadrática.
3. A atualização de  $v^k$  por  $u^{k+1}$  (passo) é justificada pela propriedade i).
4. No algoritmo expressamos, no passo 2, a fórmula para calcular  $u$ ; vamos agora obtê-la.

O problema

$$\text{Sup}_{u \in U} \left\{ u^T f(x) - \frac{1}{2} \|u-v\|^2 \right\}$$

tem a ele associado a função de Lagrange dada por:

$$L(u, \delta, s) = u^T f(x) - \frac{1}{2} \|u-v\|^2 - \delta \left( \sum_{i=1}^m u_i - 1 \right) - s^T u, \quad s \geq 0.$$

Dáí derivamos as condições de otimalidade de 1ª ordem:

$$f_i(x) - u_i + v_i - \delta - s_i = 0, \quad i=1, \dots, m$$

$$\sum_{i=1}^m u_i = 1, \quad u_i \geq 0, \quad i=1, \dots, m$$

$$s_i u_i = 0, \quad i=1, \dots, m$$

Conseqüentemente,

$$u_i = f_i(x) + v_i - \delta - s_i$$

Como  $s_i = 0$  se  $u_i > 0$ , temos

$$u_i = \begin{cases} f_i(x) + v_i - \delta, & \text{se } f_i(x) + v_i - \delta > 0 \\ 0, & \text{senão.} \end{cases}$$

Por outro lado, a fim de verificar as restrições,  $\delta$  é calculado de modo a garantir que

$$\sum_{i=1}^m u_i = 1, \quad \text{isto é:}$$

$$\sum_{i=1}^m u_i = \sum_{i=1}^m \max \left\{ f_i(x) + v_i - \delta, 0 \right\} = 1,$$

o que completa o resultado apresentado no passo 2.

#### III.5.4: RESULTADOS SOBRE CONVERGÊNCIA (CASO CONVEXO)

Consideramos que  $f_i(x)$ ,  $i=1, \dots, m$  são funções convexas.

Para provar a convergência necessitamos previamente do lema abaixo:

Lema:

Dadas as hipóteses abaixo para seqüências e escalares reais:

$$\left\{ \begin{array}{l} 1. \rightarrow a_k \leq c \\ 2. \rightarrow a_k \geq a_{k-1} + b_k - \varepsilon_k \\ 3. \rightarrow a_k \geq c - d b_k - \varepsilon_k \\ 4. \rightarrow b_k \geq 0 \text{ e } \{b_k\} \text{ é limitado } \forall k \\ 5. \rightarrow \langle \varepsilon_k \rangle \longrightarrow 0 \text{ quando } k \longrightarrow \infty \end{array} \right. \quad \text{[III.5.5.1]}$$

então  $\langle b_k \rangle \longrightarrow 0$  e  $\langle a_k \rangle \longrightarrow c$  quando  $k \longrightarrow \infty$

Teorema:

Seja  $(x^k, u^k)$  um par gerado por [III.5.6], satisfazendo

$$\| \nabla \varphi_{u^{k-1}}(x^k) \| \leq \varepsilon^k \quad \text{[III.5.7]}$$

e  $\langle \varepsilon^k \rangle$  uma seqüência que tende a zero. Suponha que a seqüência  $\langle x^k \rangle$  é limitada e a função  $\varphi$  é limitada inferiormente. Então

$$a. - \lim_{k \rightarrow \infty} \| u^k - u^{k-1} \| \longrightarrow 0$$

$$b. - \lim_{k \rightarrow \infty} \varphi_{u^{k-1}}(x^k) = \lim_{k \rightarrow \infty} \varphi(x^k) = \inf_x \varphi(x) = \mu$$

c. - Qualquer limite de sub-seqüência  $\langle x^k \rangle$  convergente faz parte do conjunto solução definido por

$$\left\{ x^* \in \mathbb{R}^n \mid \varphi(x^*) = \mu \right\}$$

Prova:

Para provar o teorema precisamos verificar as propriedades (III.5.5.11).

Previamente identificamos cada uma das seqüências e escalares reais:

$$\text{Seja } g(u) = \inf_{x \in C} (u^T f(x)), \text{ onde } C = \text{co} \left\{ (x / \varphi(x) \leq \varphi(x_0)) \cup [x_k] \right\},$$

(III.5.8)

$x_0$  e  $x^k$  conhecidos.

$$\left\{ \begin{array}{l} a_k = g(u^k) \\ b_k = \frac{\|u^k - u^{k-1}\|}{\sqrt{2}} \\ c = \mu \\ d = 2 \\ \varepsilon_k = M \varepsilon^k. \end{array} \right. \quad \text{(III.5.5.2)}$$

Mostraremos inicialmente a parte (a) do teorema; basta verificar as seguintes propriedades:

$$\left\{ \begin{array}{l} 1. - g(u^k) \leq \mu; \\ 2. - g(u^k) \geq g(u^{k-1}) + \frac{1}{2} \|u^k - u^{k-1}\|^2 - M \varepsilon^k, \quad \text{(III.5.5.3)} \\ 3. - g(u^k) \geq \mu - \frac{\|u^k - u^{k-1}\|}{\sqrt{2}} - M \varepsilon^k; \\ 5. - \lim_{k \rightarrow \infty} M \varepsilon^k = 0, \text{ (temos como hipótese no teorema)} \end{array} \right.$$

Na  $k$ -ésima iteração, pela convexidade de  $f_i$ , obtemos para  $x \in C$

$$\sum_{i=1}^m u_i^k f_i(x) \geq \sum_{i=1}^m u_i^k f_i(x^k) + (x-x^k)^T \sum_{i=1}^m u_i^k \nabla f_i(x^k),$$

ou

$$\sum_{i=1}^m u_i^k f_i(x) \geq \sum_{i=1}^m u_i^k f_i(x^k) + (x-x^k)^T \nabla \varphi_{u^{k-1}}(x^k), \quad \text{[III.5.9]}$$

Como  $\{x^k\}$  é limitada e se tem [III.5.7], existe  $M > 0$  tal que para  $k$  suficientemente grande

$$\sum_{i=1}^m u_i^k f_i(x) \geq \sum_{i=1}^m u_i^k f_i(x^k) - M \varepsilon^k, \quad \text{[III.5.10]}$$

$$\sum_{i=1}^m u_i^k f_i(x) - \frac{1}{2} \|u^k - u^{k-1}\|^2 \geq \sum_{i=1}^m u_i^k f_i(x^k) - \frac{1}{2} \|u^k - u^{k-1}\|^2 - M \varepsilon^k, \quad \text{[III.5.11]}$$

Quando determinamos  $u^{k-1}$  a relação [III.5.3] fica

$$\varphi_{u^{k-1}}(x^k) = \sum_{i=1}^m u_i^k f_i(x^k) - \frac{1}{2} \|u^k - u^{k-1}\|^2, \quad \text{[III.5.12]}$$

com [III.5.12] em [III.5.11] obtemos

$$\varphi_{u^{k-1}}(x^k) \leq \sum_{i=1}^m u_i^k f_i(x) - \frac{1}{2} \|u^k - u^{k-1}\|^2 + M \varepsilon^k, \quad \forall x \in C$$

Então

$$\varphi_{u^{k-1}}(x^k) \leq \inf_{x \in C} \left( \sum_{i=1}^m u_i^k f_i(x) - \frac{1}{2} \|u^k - u^{k-1}\|^2 \right) + M \varepsilon^k;$$



usando a notação [III.5.8] e a propriedade de infimo, obtemos

$$\varphi_{u^{k-1}}(x^k) \leq g(u^k) - \frac{1}{2} \|u^k - u^{k-1}\|^2 + M \varepsilon^k. \quad \text{[III.5.13]}$$

Mais uma vez, retomemos a função regularizadora [III.5.3]

$$\varphi_{u^{k-1}}(x^k) = \sup_{u \in U} \left( \sum_{i=1}^m u_i f_i(x^k) - \frac{1}{2} \|u - u^{k-1}\|^2 \right);$$

é imediato que

$$\varphi_{u^{k-1}}(x^k) \geq \sum_{i=1}^m u_i^{k-1} f_i(x^k); \quad \text{[III.5.14]}$$

e por outro lado

$$\begin{aligned} \sum_{i=1}^m u_i^{k-1} f_i(x^k) &\geq \inf_{x \in C} \left( \sum_{i=1}^m u_i^{k-1} f_i(x) \right); \\ &= g(u^{k-1}), \quad (\text{devido à definição de } g) \end{aligned} \quad \text{[III.5.15]}$$

De [III.5.14] e [III.5.15], obtemos

$$\varphi_{u^{k-1}}(x^k) \geq g(u^{k-1}); \quad \text{[III.5.16]}$$

De [III.5.13] e de [III.5.16], temos

$$g(u^{k-1}) \leq \varphi_{u^{k-1}}(x^k) \leq g(u^k) - \frac{1}{2} \|u^k - u^{k-1}\|^2 + M \varepsilon^k, \quad \text{[III.5.17]}$$

ou

$$g(u^k) \geq g(u^{k-1}) + \frac{1}{2} \|u^k - u^{k-1}\|^2 - M \varepsilon^k, \quad \text{[III.5.18]}$$

que é a propriedade 2.

De [III.5.3] com  $v = u^{k-1}$  (onde supremo é atingido)

$$h_{u^{k-1}}(x, u) = \sum_{i=1}^m u_i f_i(x) - \frac{1}{2} \|u - u^{k-1}\|^2, \quad \text{[III.5.19]}$$

Usando o desenvolvimento de Taylor, reescrevemos  $h_{u^{k-1}}(x, u)$  em  $x=x^k$ , relativamente a  $u$ , em torno de  $u^k$ ,

$$h_{u^{k-1}}(x^k, u) = h_{u^{k-1}}(x^k, u^k) + (u - u^k)^T \nabla_u h_{u^{k-1}}(x^k, u^k) - \frac{1}{2} \|u - u^k\|^2; \quad \text{[III.5.20]}$$

como o supremo de  $h_{u^{k-1}}(x^k, u)$  é obtido em  $u^k$ , então

$$\left( u - u^k \right)^T \left[ \nabla_u h_{u^{k-1}}(x^k, u^k) \right] \leq 0, \quad \forall u \in U; \quad \text{[III.5.21]}$$

aplicando esta desigualdade à expressão anterior, obtemos

$$h_{u^{k-1}}(x^k, u) \leq h_{u^{k-1}}(x^k, u^k) - \frac{1}{2} \|u - u^k\|^2, \quad \forall u \in U \quad \text{[III.5.22]}$$

usando [III.5.19] em [III.5.22]

$$\sum_{i=1}^m u_i f_i(x^k) - \frac{1}{2} \|u - u^{k-1}\|^2 \leq \sum_{i=1}^m u_i^k f_i(x^k) - \frac{1}{2} \|u^k - u^{k-1}\|^2 - \frac{1}{2} \|u - u^k\|^2$$

e

$$\sum_{i=1}^m u_i f_i(x^k) \leq \sum_{i=1}^m u_i^k f_i(x^k) + \frac{1}{2} \left[ \|u - u^{k-1}\|^2 - \|u^k - u^{k-1}\|^2 - \|u - u^k\|^2 \right]$$

$$\begin{aligned}
&= \sum_{i=1}^m u_i^k f_i(x^k) + (u-u^k)^T (u^k - u^{k-1}) \\
&\leq \sum_{i=1}^m u_i^k f_i(x^k) + \|u^k - u^{k-1}\| \sqrt{2}, \quad \text{[III.5.23]}
\end{aligned}$$

O resultado da expressão [III.5.23] devendo-se ao fato de

$$\begin{aligned}
&\|u\| \leq 1, \quad \forall u \in U, \quad \|u\|^2 + \|u^k\|^2 \leq 2, \\
&\|u - u^k\|^2 \leq \|u\|^2 + \|u^k\|^2 \leq 2, \quad \forall u \in U, \quad \text{logo} \\
&\|u - u^k\| \leq \sqrt{2}. \quad \text{[III.5.24]}
\end{aligned}$$

Rescrevamos [III.5.23]:

$$\sum_{i=1}^m u_i^k f_i(x^k) \geq \sum_{i=1}^m u_i f_i(x^k) - \|u^k - u^{k-1}\| \sqrt{2}; \quad \text{[III.5.25]}$$

adicionando  $(-M \varepsilon^k)$  em ambos lados obtemos

$$\sum_{i=1}^m u_i^k f_i(x^k) - M \varepsilon^k \geq \sum_{i=1}^m u_i f_i(x^k) - \|u^k - u^{k-1}\| \sqrt{2} - M \varepsilon^k, \quad \text{[III.5.26]}$$

de [III.5.10] e [III.5.26], temos

$$\sum_{i=1}^m u_i^k f_i(x) \geq \sum_{i=1}^m u_i f_i(x^k) - \|u^k - u^{k-1}\| \sqrt{2} - M \varepsilon^k, \quad \forall x \in C \text{ e } u \in U \quad \text{[III.5.27]}$$

Como por definição de  $\varphi(x)$  se tem

$$\sum_{i=1}^m u_i^k f_i(x^k) \geq \varphi(x^k); \quad \text{[III.5.28]}$$

adicionando  $-\left(\|u^k - u^{k-1}\| \sqrt{2} + M \varepsilon^k\right)$  a esta relação obtemos

$$\sum_{i=1}^m u_i^k f_i(x^k) - \|u^k - u^{k-1}\| \sqrt{2 - M\varepsilon^k} \geq \varphi(x^k) - \|u^k - u^{k-1}\| \sqrt{2 - M\varepsilon^k},$$

[III.5.29]

que com [III.5.29] fornece

$$\sum_{i=1}^m u_i^k f_i(x) \geq \varphi(x^k) - \|u^k - u^{k-1}\| \sqrt{2 - M\varepsilon^k},$$

[III.5.30]

Observa-se, que

$$g(u^k) = \inf_x \sum_{i=1}^m u_i^k f_i(x) \leq \inf_x \sup_{u \in U} \sum_{i=1}^m u_i f_i(x) = \mu$$

$$g(u^k) \leq \mu,$$

[III.5.31]

que é a propriedade 1.

Como  $\inf_x \varphi(x) = \mu$ ,  $\varphi(x) \geq \mu$ , adicionando a esta

desigualdade  $-\left(\|u^k - u^{k-1}\| \sqrt{2 - M\varepsilon^k} + M\varepsilon^k\right)$ , temos

$$\varphi(x^k) - \|u^k - u^{k-1}\| \sqrt{2 - M\varepsilon^k} \geq \mu - \|u^k - u^{k-1}\| \sqrt{2 - M\varepsilon^k},$$

[III.5.32]

de [III.5.32] e [III.5.30] obtemos

$$\sum_{i=1}^m u_i^k f_i(x) \geq \mu - \|u^k - u^{k-1}\| \sqrt{2 - M\varepsilon^k},$$

$$\inf_x \sum_{i=1}^m u_i^k f_i(x) \geq \mu - \|u^k - u^{k-1}\| \sqrt{2 - M\varepsilon^k},$$

$$g(u^k) \geq \mu - \|u^k - u^{k-1}\| \sqrt{2 - M\varepsilon^k},$$

[III.5.33]

que é a propriedade 3. Temos assim  $\|u^k - u^{k-1}\| \rightarrow 0$  ( parte a do teorema )

Alem disto , devido ao lema temos também

$$\lim_{k \rightarrow \infty} g(u^k) = \mu, \quad \text{[III.5.34]}$$

que, com [III.5.17] implica imediatamente em :

$$\lim_{k \rightarrow \infty} \varphi_{u^{k-1}}(x^k) = \mu. \quad \text{[III.5.35]}$$

Falta portanto, para a parte b, mostrar que  $\varphi(x^k) \rightarrow \mu$ .

Temos, com

$$\sum_{i=1}^m u_i f_i(x) \geq \sum_{i=1}^m u_i^k f_i(x^k), \quad x \in C$$

e [III.5,30]:

$$\sum_{i=1}^m u_i^k f_i(x) \geq \varphi(x^k) - \|u^k - u^{k-1}\| \sqrt{2} - M \varepsilon^k,$$

$$\inf_x \sum_{i=1}^m u_i^k f_i(x) \geq \varphi(x^k) - \|u^k - u^{k-1}\| \sqrt{2} - M \varepsilon^k, \quad x \in C$$

ou de [III.5.32] obtemos

$$g(u^k) \geq \varphi(x^k) - \|u^k - u^{k-1}\| \sqrt{2} - M \varepsilon^k \geq \mu - \|u^k - u^{k-1}\| \sqrt{2} - M \varepsilon^k \quad \text{[III.5.36]}$$

e

$$\mu \leq \varphi(x^k) \leq g(u^k) + \|u^k - u^{k-1}\| \sqrt{2} + M \varepsilon^k, \quad \text{[III.5.37]}$$

aplicando o limite nestas desigualdades e usando [III.5.34] e

[III.5.35] chegamos ao resultado esperado:

$$\lim_{k \rightarrow \infty} \varphi(x^k) = \mu. \quad \text{[III.5.38]}$$

Resta a parte c.

Suponha que após um número finito de iterações se tem

$$\|\nabla \varphi_{u^{k-1}}(x^k)\|^2 = \|u^k - u^{k-1}\|^2 = 0;$$

então de [III.5.36] e pela definição de  $\varphi_{u^{k-1}}(x)$ , temos

$$\varphi_{u^{k-1}}(x^k) = g(u^k) = \varphi(x^k) = \mu$$

e  $x^k$  é um ponto solução.

Caso contrário, seja  $K \subset \mathbb{N}$  (naturais) tal que a seqüência  $\{x^k\}_{k \in K}$  converge a  $x^*$ . Então, por  $\varphi(x)$  ser contínua e por [III.5.38] tem-se que

$$\mu \leq \varphi(x^*) = \lim_{\substack{k \rightarrow \infty \\ k \in K}} \varphi(x^k) = \mu$$

Então,  $x^*$  é ponto do conjunto solução, e fica mostrado o item (c).

### Observações

1.- Uma interessante interpretação deste método de regularização é feita pelas próprias autoras, através do método de multiplicadores [11]. Para isto, se escreve o problema [I.1]

na forma equivalente, que já conhecemos:

$$\begin{aligned} \min \quad & \delta \\ (x, \delta) \in & \mathbb{R}^{n+1} \\ \text{sujeito a} & \\ f_i(x) \leq & \delta, \quad i = 1, \dots, m. \end{aligned} \quad \text{[III.5.39]}$$

Introduzem variáveis de folga

$$\begin{aligned} \min \quad & \delta \\ (x, \delta, w) \in & \mathbb{R}^{n+1+m} \\ \text{sujeito a} & \\ f_i(x) + w_i^2 - \delta = & 0, \quad i = 1, \dots, m. \end{aligned} \quad \text{[III.5.40]}$$

e definem o lagrangeano aumentado

$$L(x, \delta, w, v, r) = \delta + \sum_{i=1}^m v_i \left( f_i(x) + w_i^2 - \delta \right) + \frac{r}{2} \sum_{i=1}^m \left( f_i(x) + w_i^2 - \delta \right)^2$$

O objetivo agora é expressar a função dual, minimizando  $L$  em  $w, \delta$  e  $x$ .

Temos, inicialmente

$$\inf_w L(x, \delta, w, v, r) = \delta + \sum_{i=1}^m \inf_w \left[ v_i \left( f_i(x) + w_i^2 - \delta \right) + \frac{r}{2} \left( f_i(x) + w_i^2 - \delta \right)^2 \right]$$

que é atingido em

$$w_i^* = \max \left[ 0, - \left( f_i(x) + \frac{v_i}{r} - \delta \right) \right]$$

Consideramos em seguida o  $\inf_{\delta} L(x, \delta, w_i^*, v, r)$  que é atingido em

$\delta = \delta^*$ , quando  $\nabla_{\delta} L(x, \delta^*, w_i^*, v, r) = 0$ , isto é:

$$r \sum_{i=1}^m \left( f_i(x) + \frac{v_i}{r} - \delta^* \right) = 1$$

Denotamos

$$\hat{u}_i = \begin{cases} r \left( f_i(x) + \frac{v_i}{r} - \delta^* \right), & \text{se } r \left( f_i(x) + \frac{v_i}{r} - \delta^* \right) \geq 0 \\ 0, & \text{em outro caso.} \end{cases}$$

com  $\delta^*$  tal que  $\sum u_i = 1$ , então

$$\hat{u} \in U = \left\{ u \in \mathbb{R}^m / \sum_{i=1}^m u_i = 1, u \geq 0 \right\},$$

e

$$\begin{aligned} \inf_{\delta, w} L(x, \delta, w, v, r) &= \sup_{u \in U} \left( u^T f(x) - \frac{1}{2r} \|u-v\|^2 \right) \\ &= \varphi_{v,r}(x), \end{aligned}$$

para mais detalhes ( ver [11] )

Em conclusão comparando a função regularizada e os métodos de multiplicadores temos

$$\varphi_v(x) = \inf_{\delta, w} L(x, \delta, w, v, 1),$$

isto é, corresponde ao uso do Lagrangeano aumentado com  $r=1$ .



### III.5.5 ALGUNS ASPECTOS DAS FORMULAÇÕES DE GÓMEZ-GÍGOLA e HAN

Ambos os algoritmos se basciam na atualização da variável dual  $u$ , obtendo a direção de descida ( de  $2^{\frac{a}{n}}$  ordem )

$$p = - \left( \sum_{i=1}^m u_i \nabla^2 f_i(x) \right)^{-1} \sum_{i=1}^m u_i \nabla f_i(x) ,$$

diretamente ou por BFGS.

A diferença essencial está portanto no processo de obtenção dos multiplicadores  $u_i$ : Han utiliza a natural formulação dual do sub-problema que define a direção primal  $p$ , enquanto Gómez e Gígola os obtém através de uma regularizaçãodo dual do problema original .

Como, de acordo com a parte b do teorema de convergência de Gómez e Gígola  $(x^k, u^{k-1})$  é uma sequência minimizante para o problema original, é razoável supor que, no limite, os processos sejam equivalentes.

## CAPÍTULO IV

## IV. RESULTADOS NUMÉRICOS

Implementamos o algoritmo de Gígola e Gómes em Pascal. Os resultados computacionais são comparados aos de outros algoritmos e à implementação dos autores .

Na implementação numérica usa-se o seguinte:

a) A versão de Han-Powell [19] para o BFGS.

b) A busca linear, realizada sobre a função original  $\varphi_v$  a ser minimizada, utilizando o algoritmo de Brent ( ver [3] ) e interpolação quadrática .

## IV.1. MÉTODO DE MÁXIMA DESCIDA

A fim de verificar o significado da aproximação de 2ª ordem , testamos o método de máxima descida nos mesmos exemplos . Este método corresponde à resolução do problema local

$$\min_{\|p\| \leq 1} \frac{\partial \varphi(x)}{\partial p}$$

isto é,

$$\min_{\|p\| \leq 1} \max_{i \in I(x_0)} \left( \frac{\partial f_i(x)}{\partial x}, p \right),$$

ou ainda ,

min  $\delta$

sujeito a

[IV.11]

$$\left[ \frac{\partial f_i(x)}{\partial x}, p \right] \leq \delta, \quad i \in I(x_0),$$

$$\frac{\|p\|^2}{2} \leq \frac{1}{2}.$$

Este problema tem em,

$$L(p, \delta, u, \lambda) = \delta + \sum_{i \in I(x_0)} u_i \left[ \left[ \frac{\partial f_i(x)}{\partial x}, p \right] - \delta \right] + \frac{\lambda}{2} \left[ (p, p) - 1 \right],$$

$u_i \geq 0$ ,  $\lambda \geq 0$ , sua função de lagrange associada . As condições de otimalidade são:

$$\sum_{i \in I(x_0)} u_i = 1, \quad u_i \geq 0$$

$$\sum_{i \in I(x_0)} u_i \frac{\partial f_i(x)}{\partial x} + \lambda p = 0,$$

além das restrições de [IV.11]. Portanto, a direção de máxima descida é dada, a menos de constante, por

$$p = - \sum_{i \in I(x_0)} u_i \frac{\partial f_i(x)}{\partial x}.$$

Isto naturalmente corresponde a, nas fórmulas de Han e de Gígola-Gómez, aproxima  $\sum u_i \nabla^2 f_i(x)$  pela identidade.

### Aplicação Numérica

As funções testes são dadas por:

#### Problema 1

$$f_1(x) = x_1^4 + x_2^2 ;$$

$$f_2(x) = (2-x_1)^2 + (2-x_2)^2 ;$$

$$f_3(x) = 2 e^{(-x_1 + x_2)} ;$$

cuja solução é :  $x^* = (1, 1)^T$ ,  $\varphi(x^*) = 2$ .

#### Problem 2

$$f_1(x) = x_1^2 + x_2^4 ;$$

$$f_2(x) = (2-x_1)^2 + (2-x_2)^2 ;$$

$$f_3(x) = 2 e^{(-x_1 + x_2)} ;$$

a solução é :  $x^* = (1.13904, 0.89956)^T$ ,  $\varphi(x^*) = 1.95222$  .

Apresentamos abaixo os quadros comparativos dos resultados obtidos com os métodos analisados.

## QUADRO COMPARATIVO CORRESPONDENTE AO PROBLEMA 1

Método	Ponto Inicial	N <sup>o</sup> de Iterações	N <sup>o</sup> Funções calculadas na u.i.	$x^* \cong \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$	Valor alcançado $\varphi(x^*)$
Charalambous e Conn	$\begin{pmatrix} 1 \\ -0.1 \end{pmatrix}$	6	8	$\begin{pmatrix} 1.000008 \\ 1.000019 \end{pmatrix}$	2.000071
Zang	$\begin{pmatrix} 2 \\ 2 \end{pmatrix}$		174		2.000003
G Busca Interpolação	$\begin{pmatrix} 1 \\ -0.1 \end{pmatrix}$	11	18	$\begin{pmatrix} 1.000003 \\ 0.999950 \end{pmatrix}$	2.0000334
G Quadrática	$\begin{pmatrix} 2 \\ 2 \end{pmatrix}$		17		1.952227
I M P L E M E N T A Ç Ã O D E G Ô M E Z E G I G O L A					
(H-P)	$\begin{pmatrix} 1 \\ -0.1 \end{pmatrix}$	17	7	$\begin{pmatrix} 1.000003 \\ 0.999984 \end{pmatrix}$	2.000000
(I.P)	$\begin{pmatrix} 2 \\ 2 \end{pmatrix}$	17	7	$\begin{pmatrix} 1.000002 \\ 0.999995 \end{pmatrix}$	2.000000
(H-P)	$\begin{pmatrix} 1 \\ -0.1 \end{pmatrix}$	17	27	$\begin{pmatrix} 1.000016 \\ 0.999978 \end{pmatrix}$	2.000000
Brent	$\begin{pmatrix} 2 \\ 2 \end{pmatrix}$	20	24	$\begin{pmatrix} 1.000011 \\ 0.999988 \end{pmatrix}$	2.000000
(M.G)	$\begin{pmatrix} 1 \\ -0.1 \end{pmatrix}$	40	7	$\begin{pmatrix} 1.000012 \\ 0.999988 \end{pmatrix}$	2.000000
(I.P)	$\begin{pmatrix} 2 \\ 2 \end{pmatrix}$	28	7	$\begin{pmatrix} 1.000003 \\ 0.999990 \end{pmatrix}$	2.000000
(M.D)	$\begin{pmatrix} 1 \\ -0.1 \end{pmatrix}$	36	22	$\begin{pmatrix} 1.000010 \\ 0.999981 \end{pmatrix}$	2.000000
Brent	$\begin{pmatrix} 2 \\ 2 \end{pmatrix}$	23	28	$\begin{pmatrix} 0.999995 \\ 1.000021 \end{pmatrix}$	2.000000

## QUAADRO COMPARATIVO CORRESPONDENTE AO PROBLEMA 2

Método	Ponto Inicial	N <sup>o</sup> de Iterações	N <sup>o</sup> Funções calculadas na (u. i)	$x^* \approx \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$	Valor alcançado $\varphi(x^*)$	
Charalambous e Conn	$\begin{pmatrix} 1 \\ -0.1 \end{pmatrix}$	8	21	$\begin{pmatrix} 1.138702 \\ 0.899822 \end{pmatrix}$	1.952225	
Zang	$\begin{pmatrix} 2 \\ 2 \end{pmatrix}$		52		1.952227	
G G	Busca Interpolação	$\begin{pmatrix} 1 \\ -0.1 \end{pmatrix}$	11	17	$\begin{pmatrix} 1.139347 \\ 0.899639 \end{pmatrix}$	1.952227
	Quadrática	$\begin{pmatrix} 2 \\ 2 \end{pmatrix}$		17		1.952227
I M P L E M E N T A Ç Ã O D E G Ô M E Z E G I G O L A						
(H-P)	$\begin{pmatrix} 1 \\ -0.1 \end{pmatrix}$	13	7	$\begin{pmatrix} 1.139040 \\ 0.899557 \end{pmatrix}$	1.952224	
(I.P)	$\begin{pmatrix} 2 \\ 2 \end{pmatrix}$	11	7	$\begin{pmatrix} 1.139045 \\ 0.899562 \end{pmatrix}$	1.952224	
(H-P)	$\begin{pmatrix} 1 \\ -0.1 \end{pmatrix}$	11	26	$\begin{pmatrix} 1.139040 \\ 0.899562 \end{pmatrix}$	1.952224	
Brent	$\begin{pmatrix} 2 \\ 2 \end{pmatrix}$	11	24	$\begin{pmatrix} 1.139034 \\ 1.899578 \end{pmatrix}$	1.952224	
(M.G)	$\begin{pmatrix} 1 \\ -0.1 \end{pmatrix}$	11	7	$\begin{pmatrix} 1.139033 \\ 0.899558 \end{pmatrix}$	1.952224	
(I.P)	$\begin{pmatrix} 2 \\ 2 \end{pmatrix}$	11	7	$\begin{pmatrix} 1.139045 \\ 0.899562 \end{pmatrix}$	1.952224	
(M.G)	$\begin{pmatrix} 1 \\ -0.1 \end{pmatrix}$	11	32	$\begin{pmatrix} 1.139034 \\ 0.899569 \end{pmatrix}$	1.952224	
Brent	$\begin{pmatrix} 2 \\ 2 \end{pmatrix}$	11	26	$\begin{pmatrix} 1.139068 \\ 0.899550 \end{pmatrix}$	1.952224	

Notação:

(u. i): última iteração

(H-P): Han-Powell

(M.G): Método de Gradiente

(I.P): Interpolação Quadrática

$\begin{bmatrix} G \\ G \end{bmatrix}$  : Gómez e Gígola.

## B I B L I O G R A F Í A

- [1] Avriel M. (1976), "Non Linear Programing. Analysis and Methods, Prentice-Hall.
- [2] D.H Anderson and M.R Osborne (1977), Discrete, non linear aproximations in polyhedral norms: a Levenberg-like algorithm, Numer. Math, pp.157-170
- [3] D.P Bertsekas, (1975) "Nondifferentiable optimization s derivates", Prentice Hall, Englewood Cliffs, N.J.
- [4] Brent R.P. (1973) "Algorithms for minization without derivates", Prentice Hall, Englewood Cliffs, N.J.
- [5] C. Charalambous and J.W.Bandler (1976), Nolinear minimax optimization as a sequence of least pth optimization with finite values of p, Internat. J. Systems Scie; 7 pp.377-391.
- [6] C.Charalambous and A.R. Conn (1978), An efficient method to solve the minimax problem directly, SIAM J.Numer. Anal. Vol. 15,pp.162-187.
- [7] V.F. Dem'yanov (1968), "Algorithms for some minimax problems", J.Comp. and System Sciences 2: 4, 342-380.
- [8] V.F Dem'yanov and V.N. Malozemov (1974), Introduction to Minimax, John Wiley, New York and Toronto.



- [9] J.E. Denis and R.B Schnabel, "Numerical Methods for unconstrained optimization and nonlinear equations" Prentice -Hall Series in computational mathematics(1983).
- [10] J.E Dennis Jr.and J.J.Moré."Quasi- Newton methods, motivation and theory",SIAM Review 19,(1977),46-49.
- [11] C. Gígola and S.Gómez, (1990) " A Regularization Method for solving the finite convex min-max problem", SIAM J. Numer.Anal. Vol 27, N<sup>o</sup> 6 pp.1621-1634.
- [12] C.Gígola and S.Gómez ,(1990) "Relation Between the regularization and the multipliers method for the min max problem". Inst. de Inves. em Mat. Aplic. y en Sist. Report N<sup>o</sup>569, Univ. Aut. of Mexico .
- [13] C.Gígola and S.Gómez, (1990)" Two second order regularization methods to solve the finite min-max problem". Inst. de Inves. em Mat. Aplic. y en Sist. Report N<sup>o</sup> 568, Univ Aut of Mexico.
- [14] P.E Gill and W.Murray (1974), Newton, Type methods for unconstrained and linearly constrained optimization, Math. Pro. ,7,pp. 311-350.
- [15] S.P.Han, (1981) "Variable methods for minimizing a class of nondifferentiable functions", Mathematical Programming 10, pp

1-13.

[15] S.P Han, (1976) "Superlinearly convergent variable metric algorithms for general nonlinear programming problems", Mathematical Programming, 11 (1,976) pp.263-282.

[16] S. P Han. "Dual variable metric methods for constrained optimization " SIAM Journal an Control and Optimization ", 15 (1977) 546-565.

[17] S.P Han. "A global convergente method for non linear programming",Journal of Optimization Theory and Aplications 22 (1977) 297-309.

[18] C.Lemarechal,"An existension of Davidon methods to non differentiable problens". Mathematical Programming Study 3 (1975) 95--105.

[19] Han Powell, (1984) "Linear and nonlinear programing"

[20] C. Lemarechal, "An existension of Davidon methods to non differentiable problems", Mathematical Programing Study 3(1975) 95-105.

[21] Luenberger, (1984)" Linear and nonlinear programing " Addison-Wesley Publishing Company.

[22] W. Murray and M.L. Overton,"A Projected Lagrangian Algorithm

for nonlinear minimax optimization " SIAM, J. Sci. Stat. Comput. 3 (1980) (345-369).

[23] Murray e Wright (1981) "Practical Optimization" Academic Press, Edition (1981).

[24] M.R Osborne and G.A Watson,(1968) "An algorithm for minimax approximation in the nonlinear case".Computer Journal 12, pp 64-69.

[25] Watson (1979), The minimax solution of an overdetermined system of non linear equations, J.Inst. math. applies., 23, pp. 167-180.

[26] M.J.D Powell,(1977) "A fast algorithm for non linear constrained optimization calculations ", Dundee conference on Numerical Analysis.

[27] P.Wolf, (1975) " A method of conjugate subgradients for minimizing non differentiable functions", Mathematical Programming Study 3 145-173.

[28] J.B.Rosen, (1961), " The gradient projection method for nonlinear programming".Part II: Nonlinear constraints,Journal SIAM, 9 pp 514,532.

[29] Rockafellar.R.T.(1970), "Convex Analysis", Princeton Univ.Press.

[30] I. Zang, (1980), "A Smoothing - Out Technique for min-max optimization", Mathematical Programming 19 .

[31] I. Zang, (1978) "A new arc algorithm for unconstrained optimization ", Mathematical Programming 15, pp, 36-52.