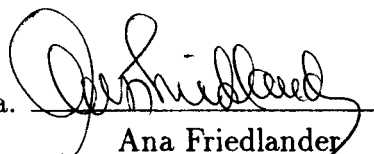


APLICAÇÃO DO PACOTE BOX PARA PROBLEMAS DE SEPARAÇÃO

Este exemplar corresponde a redação final da tese devidamente corrigida e defendida pela Srta. LEONOR, CAMILA QUISPE YAGUA e aprovada pela Comissão Julgadora.

Campinas, 24 de Janeiro de 1995.

Profa. Dra.



Ana Friedlander

Dissertação apresentada ao Instituto de Matemática, Estatística e Ciência da Computação, UNICAMP, como requisito parcial para obtenção do Título de MESTRE em MATEMÁTICA APLICADA.

APLICAÇÃO DO PACOTE BOX PARA PROBLEMAS DE SEPARAÇÃO

Leonor Camila Quispe Yagua

IMECC

Matemática Aplicada

Orientadora: Ana Friedlander

Mestrado em Matemática Aplicada

*” Aos meus queridos pais: Andrés e María
e irmãos: Tila, Rubi e Abel”.*

Agradecimentos

À professora Ana Friedlander pela orientação.

À banca examinadora pelas sugestões.

Ao CNPq pelo apoio financeiro.

Ao FAEP pelo auxílio-ponte.

A *Jesus* pela proteção.

Ao grupo *Charolais* pela acolhida.

Aos meus *pais* pela confiança.

Aos meus irmãos, especialmente a *Tila* pela sua companhia e dedicação e a *Rubi* pelo constante apoio apesar da distância.

Ao professor Walter Torres Montes pelo incentivo.

A todos aqueles que direta ou indiretamente colaboraram com a concretização deste trabalho.

Conteúdo

1	Conceitos Básicos	1
1.1	Introdução	1
1.2	0 problema de Otimização	3
1.3	Algoritmo I (Box)	4
1.4	Resultados de convergência	6
1.5	Identificação das restrições ativas	7
1.6	Minimização Quadrática com restrições em caixa	8
1.6.1	Algoritmo II (Quacan)	14
2	Hiperplanos de separação	17
2.1	Introdução	17
2.2	Problema de Otimização	22
3	Experimentos Práticos	27
3.1	Introdução	27
3.2	Parâmetros de entrada do Box-Quacan	28
3.3	Exemplo	30
3.4	Aplicação à Classificação das Flores	35
3.5	Conclusões	47

Introdução

Os problemas de otimização de grande porte são objeto de estudo de vários pesquisadores devido à sua importância prática.

Conn, Gould e Toint mostraram praticamente que qualquer problema de programação não linear de grande porte pode ser resolvido eficientemente através de técnicas de Lagrangeano aumentado, desde que se tenha à disposição um bom método para resolver problemas com restrições de canalização. Eles concordam com o fato de que restrições de canalização podem ser incorporadas naturalmente em algoritmos de região de confiança. Eles desenvolveram um programa computacional em *Fortran*, para resolver problemas gerais de programação não linear, **Lancelot**, usando Lagrangeano aumentado, no qual as restrições mais complexas vão ser incorporadas à função objetivo.

Para a resolução de problemas irrestritos de grande porte Schlick e Fogelson [13] utilizam o pacote **Tnpack** que usa o método de Newton truncado e o gradiente conjugado pré-condicionado.

Além deles Friedlander, Martínez e Santos [6] também desenvolveram um programa computacional, **Box - Quacan**, para resolver problemas de otimização não linear de grande porte, onde as variáveis estão sujeitas à restrições de canalização, através de técnicas de região de confiança.

Um dos objetivos deste trabalho é demonstrar a potencialidade e eficiência do pacote **Box-Quacan** para a resolução de problemas de classificação de dados e complementar os resultados obtidos na resolução de problemas de classificação de flores com métodos clássicos de estatística.

Os métodos de região de confiança foram objeto de estudo de vários autores.

Fletcher [5] provou propriedades de convergência global para algoritmos de região de confiança. Toint provou propriedades de convergência global para uma classe de métodos de região de confiança para minimização não convexa em espaços de Hilbert e Dennis e Schnabel [4] propuseram um método de região de confiança que consiste numa modificação do método de Newton para obter convergência global.

Este trabalho é desenvolvido em 3 capítulos. No primeiro é feita uma análise dos conceitos e resultados principais da teoria relacionada com a sub-rotina **Box-Quacan**.

No capítulo 2 fazemos um estudo do modelo de otimização proposto por Martínez [10] para problemas de classificação através de hiperplanos de separação.

No capítulo 3, aplicamos o modelo descrito no capítulo 2 para um problema de classificação de flores utilizando o algoritmo **Box - Quacan** na sua resolução. Descrevemos alguns parâmetros de entrada que são necessários para seu uso.

Finalmente, são comparados alguns resultados obtidos em **Box-Quacan** nestes problemas com a resolução dos métodos clássicos de estatística e apresentamos os resultados da implementação computacional.

Capítulo 1

Conceitos Básicos

1.1 Introdução

Devido à sua importância prática, problemas de otimização não linear de grande porte foram objeto de estudo de vários pesquisadores. Conn, Gould e Toint usam em seus experimentos numéricos o pacote **Lancelot**; Murtagh e Saunders o pacote **Minos**; a seguir Conn, Gould e Toint fazem uma comparação numérica entre estes pacotes e Friedlander, Martínez e Santos desenvolvem o pacote **Box-Quacan**.

Neste capítulo apresentamos o pacote de região de confiança chamado **Box-Quacan**. Outros algoritmos gerais de região de confiança são apresentados por Santos [11] para problemas de minimização restrita a um conjunto fechado arbitrário.

O algoritmo desenvolvido resolve problemas de minimização de funções reais de várias variáveis, sujeitas a restrições de canalização (caixa), ou seja:

$$\begin{aligned} \min \quad & f(x) \\ \text{s/a} \quad & x \in \Omega, \end{aligned} \tag{1.1}$$

onde Ω é uma caixa.

Será considerado, em cada iteração um subproblema que consiste em minimizar uma aproximação quadrática da função objetivo dentro de uma região de confiança, que consiste na intersecção da caixa original com a região de confiança. A subrotina **Quacan** é a encarregada de resolver este subproblema.

Estes subproblemas não precisam ser resolvidos com muita precisão para se obter a convergência global do algoritmo principal. Uma relação do valor do modelo quadrático com a solução de um subproblema auxiliar é suficiente para provar que todo ponto limite é estacionário.

A solução de um subproblema quadrático auxiliar (ponto auxiliar) faz o papel do *ponto de Cauchy* aproximado, e será utilizada como aproximação inicial para o subproblema quadrático que calcula a nova aproximação.

O ponto de Cauchy aproximado usado por Conn, Gould & Toint [2] tem boas propriedades de identificação das restrições ativas na solução. É imposto que as restrições que são ativas no ponto de Cauchy aproximado também devem ser ativas no novo ponto calculado, em cada iteração. Paralelamente o ponto auxiliar usado por Friedlander, Martínez & Santos [6], além de jogar um papel fundamental na prova de convergência, também possui a propriedade de identificar o conjunto ativo correto, sob as mesmas condições que o ponto de Cauchy. Porém, não é preciso manter o conjunto ativo do ponto auxiliar, pois não se pode garantir que ele tenha boas propriedades de identificação nas primeiras iterações.

A subrotina **Quacan** tem a capacidade de acrescentar ou descartar restrições ativas de maneira eficiente.

A conservação das mesmas restrições ativas só é justificada em caso da aproximação da solução ser muito boa. Precisa-se um algoritmo potente para minimização de uma função quadrática sobre uma caixa, que tenha a capaci-

dade de acrescentar ou relaxar de maneira eficiente as restrições ativas (O algoritmo **Quacan** satisfaz estas condições).

Quacan é uma generalização do método desenvolvido por Friedlander e Martinez [3] para problemas de otimização de funções quadráticas côncavas com restrições em caixa.

No algoritmo combinam-se técnicas de gradiente projetado com gradientes conjugados, usando a estratégia de restrições ativas.

1.2 O problema de Otimização

Seja o problema

$$\begin{aligned} & \text{minimizar} && f(x) \\ & \text{s/a} && l \leq x \leq u \end{aligned} \tag{1.2}$$

onde $x \in \mathbb{R}^n$ e $f : \mathbb{R}^n \rightarrow \mathbb{R}$ e $f \in C^1$. Sejam $g(x) = \nabla f(x)$, $\{\lambda_1(B), \lambda_2(B), \dots, \lambda_n(B)\}$ o conjunto de autovalores de B tal que $\lambda_1(B) \leq \lambda_2(B) \leq \dots \leq \lambda_n(B)$, onde B é uma matriz simétrica.

Definiremos algumas normas a serem usadas:

$$\begin{aligned} \|x\|_2 &= (|x_1|^2 + \dots + |x_n|^2)^{1/2} \\ &\text{e} \\ \|x\|_\infty &= \max |x_i| \\ &1 \leq i \leq n \end{aligned}$$

Para simplificar usaremos a notação $\| \cdot \|$ em lugar de $\| \cdot \|_2$.

Para utilizar a subrotina *Bor* é preciso fornecer alguns valores, como por exemplo, o ponto factível inicial x_0 , uma matriz simétrica $B_0 \in \mathbb{R}^{n \times n}$ (uma aproximação da hessiana), uma matriz não singular D_0 (matriz de

escalamento) e o raio de confiança inicial Δ^0 que satisfaz $\Delta^0 \geq \Delta_{min}$, onde Δ_{min} é o raio mínimo permitido. Neste caso pode-se usar como matriz de escalamento a matriz identidade.

Primeiro é resolvido um subproblema fácil :

$$\begin{aligned} \text{minimizar } & Q(z) \equiv 1/2M\|z\|^2 + g^t z \\ \text{s/a } & l \leq x + z \leq u \\ & \|z\|_\infty \leq \Delta \end{aligned}$$

onde M é um limitante para o maior autovalor da matriz simétrica B . A solução deste subproblema é $z^Q(\Delta)$ que é chamada de ponto auxiliar e é usada como aproximação inicial de um novo subproblema que consiste em minimizar uma aproximação quadrática da função f em torno de um ponto x_k na iteração k do algoritmo I.

$$\begin{aligned} \Psi(x) & \approx f(x_k) \\ \Psi(x) & = f(x_k) + \nabla f(x_k)^t(x - x_k) + 1/2(x - x_k)^t B_k(x - x_k) \end{aligned}$$

onde $B_k = \nabla^2 f(x_k)$ ou $B_k \approx \nabla^2 f(x_k)$ para $\|x - x_k\|_\infty \leq \Delta$. É procurado um ponto $\bar{z}(\Delta)$ tal que

$$\begin{aligned} \psi(\bar{z}(\Delta)) & \leq \theta Q(z^Q(\Delta)) \\ l \leq x + \bar{z}(\Delta) & \leq u \\ \|\bar{z}(\Delta)\|_\infty & \leq \Delta \end{aligned}$$

onde $\psi(z) = \frac{1}{2}z^t B z + g^t z$, $\forall z \in \mathbb{R}^n$ e $\theta \in (0, 1]$.

O novo ponto é $\bar{z}(\Delta)$.

1.3 Algoritmo I (Box)

Dados iniciais : x^0 , ponto inicial factível

$$\begin{aligned} \sigma_1, \sigma_2, 0 \leq \sigma_1 \leq \sigma_2 \leq 1 \\ \alpha \in (0, 1) \end{aligned}$$

$\Delta_{min} \geq 0$
 $\theta \in (0, 1]$
 $\Delta_0 \in [10^{-4}, 10]$
 $D_0 = I$
 B_0 matriz simétrica inicial.
 A iteração k , para $k = 0, 1, 2, \dots$

Passo 1 : (Calcular M_k)

$\Delta = \Delta_k$
 calcular $M_k > 0$ tal que $\lambda_n(B_k) \leq M_k$

Passo 2 : (Resolver o subproblema fácil)

$$\begin{aligned}
 & \text{Minimizar} && Q_k(z) \\
 & \text{s/a} && l \leq x_k + z \leq u \\
 & && \|z\|_\infty \leq \Delta
 \end{aligned}$$

onde $Q_k(z) \equiv \frac{1}{2}M_k\|z\|_2^2 + g'_k z$.
 $z_k^Q(\Delta)$ é a solução global do subproblema.
 Se $Q_k(z_k^Q(\Delta)) = 0$ então parar.

Passo 3 : (Calcular o passo)

Calcular $\tilde{z}_k(\Delta)$ tal que

$$\begin{aligned}
 \psi_k(z_k(\Delta)) &\leq \theta Q_k(z_k^Q(\Delta)) \\
 l &\leq x_k + z_k(\Delta) \leq u \\
 \|\tilde{z}_k(\Delta)\|_\infty &\leq \Delta
 \end{aligned}$$

onde $\psi_k(z) = \frac{1}{2}z^t B_k z + g'_k z, \forall z \in \mathbb{R}^n$ e $g_k = \nabla f(x_k)$.

Passo 4 : (Critério de decréscimo suficiente e cálculo do novo ponto)

Se $f(x_k + \tilde{z}_k(\Delta)) \leq f(x_k) + \alpha \psi_k(\tilde{z}_k(\Delta))$
 então

$$\begin{aligned}
 z_k &= \tilde{z}_k(\Delta), \\
 x_{k+1} &= x_k + z_k, \\
 \Delta_k &= \Delta.
 \end{aligned}$$

senão

$$\begin{aligned}\Delta &= \Delta_{new}, \\ \Delta_{new} &\in [\sigma_1 \|z_k(\Delta)\|_\infty, \sigma_2 \Delta]\end{aligned}$$

voltar ao passo 2.

1.4 Resultados de convergência

Dado o problema:

$$\begin{aligned} & \text{minimizar} && f(x) \\ & \text{s/a} && x \in D \end{aligned} \tag{1.3}$$

onde $D = \{x / g_i(x) \leq 0\}, i \in \{1, 2, \dots, m\}$ e $f \in C^1$.

Definição 1 Dizemos que $x_* \in D$ é um ponto estacionário de (1.3) se para toda curva factível α partindo de x_* temos $g(x_*)^T \alpha'(0) \geq 0$.

Definição 2 Uma restrição de desigualdade $g_i(x) \leq 0$, se diz que é ativa em um ponto factível x , se $g_i(x) = 0$ é inactiva se $g_i(x) < 0$.

Lema 1 Se o algoritmo termina no passo 2, $z_k^Q(\Delta)$ é um ponto estacionário do problema original.

Teorema 1 O algoritmo (Box) está bem definido .

Quer dizer que se x_k não é um ponto estacionário, x_{k+1} pode ser calculado a partir de x_k , aplicando finitas vezes os passos (2) - (4).

Demonstração : Ver ref. [6, pag 4]

Teorema 2 (Convergência global do Box) Se $\{x_k\}$ é uma sequência gerada pelo algoritmo I, K_1 é um conjunto infinito de índices tais que :

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

M_k e $|\lambda_1(B_k)|$ são limitados para $k \in K_1$, então x^* é um ponto estacionário do problema:

$$\begin{aligned} & \text{Minimizar} && f(x) \\ & \text{s/a} && l \leq x \leq u \end{aligned}$$

Demonstração : Ver ref. [6]

Observação: Se f é convexa, x^* é um minimizador global

1.5 Identificação das restrições ativas

Para resolver um problema de minimização com restrições é desejável que as restrições ativas na solução sejam identificadas num número finito de iterações.

Seja $A(x) \subset \{1, 2, \dots, 2n\}$ o conjunto de índices das restrições ativas em x , onde $x \in \mathbb{R}^n$ é tal que $l \leq x \leq u$, ou seja,

$$\begin{aligned} i \in A(x) & \text{ se e somente se } [x]_i = l_i \\ n+i \in A(x) & \text{ se e somente se } [x]_i = u_i \end{aligned}$$

Hipótese 1 (Não degeneração dual) Se x^* é um ponto estacionário do problema de:

$$\begin{aligned} & \text{minimizar} && f(x) \\ & \text{s/a} && l < x < u \end{aligned}$$

e $\frac{\partial f}{\partial x_i}(x^*) = 0$, então $l_i < [x^*]_i < u_i$.

Hipótese 2 Para todo $k = 0, 1, 2, \dots$ se $[x_k + z_k^Q(\Delta_k)]_i = l_i$. (respectivamente $[x_k + z_k^Q(\Delta_k)]_i = u_i$) então $[x_{k+1}]_i \equiv [x_k + \tilde{z}_k(\Delta_k)]_i = l_i$ respectivamente $[x_{k+1}]_i \equiv [x_k + \tilde{z}_k(\Delta_k)]_i = u_i$.

Teorema 3 *Suponha que as duas hipóteses são satisfeitas, $\{x_k\}$ é uma sequência infinita gerada pelo algoritmo I e $\{M_k\}$ é limitada. Seja x^* um ponto de acumulação de $\{x_k\}$. Então existe $k_1 \in \mathbb{N}$ tal que $A(x_k) = A(x^*)$ para todo $k \geq k_1$.*

Demonstração : Ver Ref. [6].

1.6 Minimização Quadrática com restrições em caixa

No algoritmo I, precisa-se calcular $z_k^Q(\Delta)$ tal que,

$$\begin{aligned} \psi_k(\bar{z}_k(\Delta)) &\leq \theta Q_k(z_k^Q(\Delta)) \\ l \leq x_k + \bar{z}_k &\leq u \\ \|\bar{z}_k(\Delta)\|_\infty &\leq \Delta \end{aligned}$$

onde z_k^Q é a solução do subproblema fácil.

Se $\bar{z}_k(\Delta) = z_k^Q(\Delta)$, os resultados de convergência são válidos. Para obter uma melhor aproximação considera-se $\bar{z}_k^Q(\Delta)$ uma solução aproximada de :

$$\begin{aligned} &\text{minimizar } \psi_k(z) \\ &\text{s/a } l \leq x_k + z \leq u \\ &\quad \|z\|_\infty \leq \Delta \end{aligned}$$

Para a resolução deste subproblema usa-se a subrotina *Quacan* descrita no algoritmo II .

Se $z = x - x_k$, com algum abuso de notação, o problema se reduz a:

$$\begin{aligned} &\text{minimizar } \psi_k(z) \\ &\text{s/a } z \in \Omega \end{aligned}$$

onde ,

$$\psi(z) = \frac{1}{2} z^t B z + b^t z$$

$$\Omega = \{z \in \mathbb{R}^n / \bar{l} \leq z \leq \bar{u}, \bar{l} \leq \bar{u}\}$$

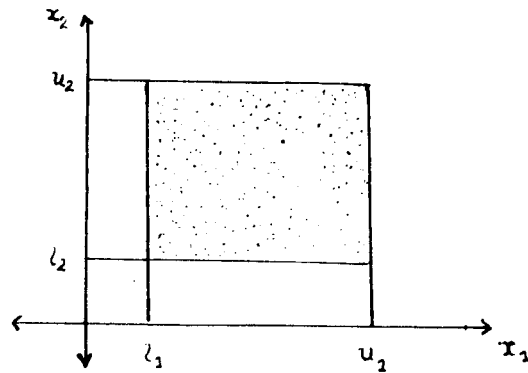
Antes de descrever a subrotina *Quacan*, deve-se lembrar alguns conceitos.

Definição 3 Chama-se *face aberta* de Ω ao conjunto $F_I \subset \Omega$ tal que $I \subset \{1, 2, \dots, 2n\}$ e

$$F_I = \{x \in \Omega / \begin{array}{l} x_i = l_i, \text{ se } i \in I, \\ x_i = u_i, \text{ se } n+i \in I, \\ l_i < x_i < u_i, \text{ outros casos} \end{array}\}.$$

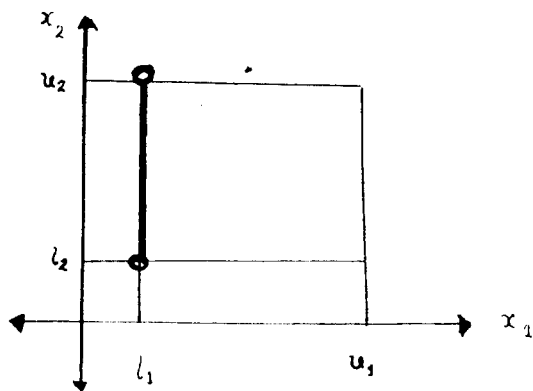
Vejamos um exemplo de face aberta para o caso de $n = 2$.

Exemplo:

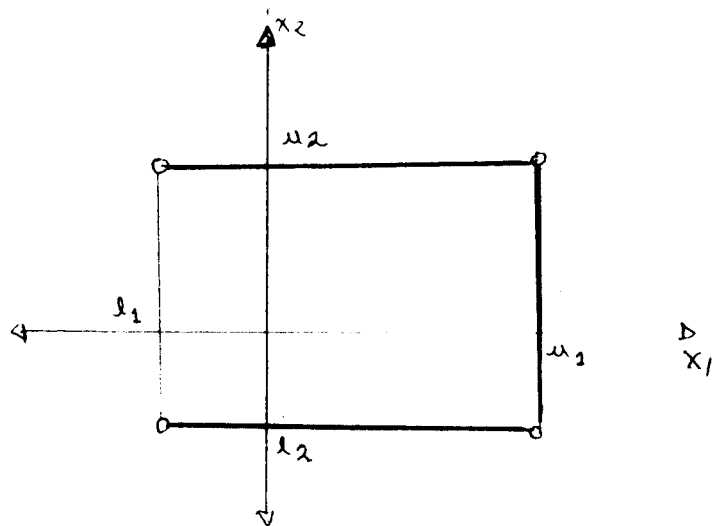


$$F_\phi = \{x \in \Omega / l_i < x_i < u_i, i = 1, 2\}$$

Mais alguns exemplos:



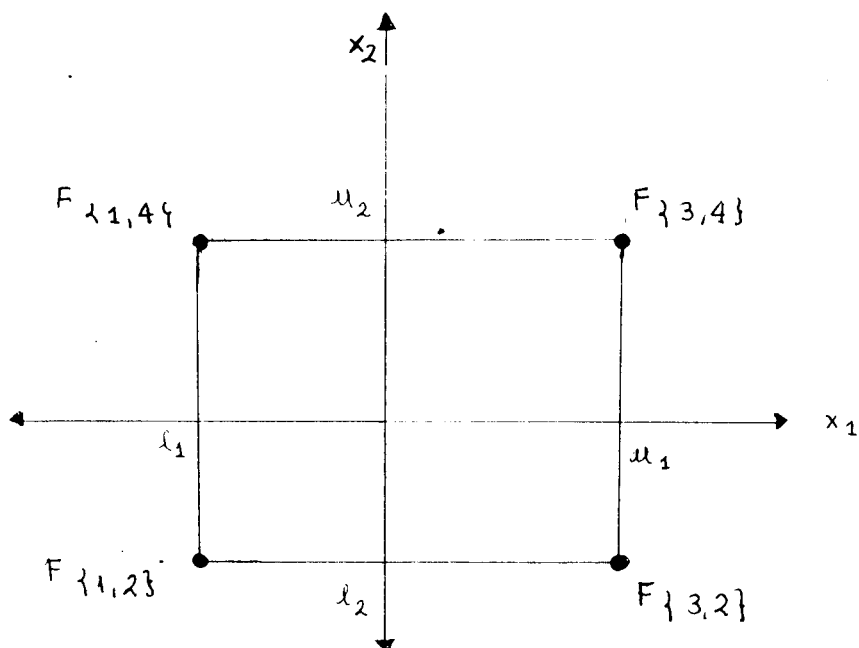
$$F_{\{1\}} = \{x \in \Omega / x_1 = l_1 \text{ e } l_2 < x_2 < u_2\}$$



$$F_{\{2\}} = \{x \in \Omega / x_2 = l_2 \text{ e } l_1 < x_1 < u_1\}$$

$$F_{\{3\}} = \{x \in \Omega / x_1 = u_1 \text{ e } l_2 < x_2 < u_2\}$$

$$F_{\{4\}} = \{x \in \Omega / l_1 < x_1 < u_1 \text{ e } x_2 = u_2\}$$



$$\begin{aligned}
 F_{\{1,4\}} &= \{x \in \Omega / x_1 = l_1, x_2 = u_2\}, \\
 F_{\{3,4\}} &= \{x \in \Omega / x_1 = u_1, x_2 = u_2\}, \\
 F_{\{1,2\}} &= \{x \in \Omega / x_1 = l_1, x_2 = l_2\}, \\
 F_{\{3,2\}} &= \{x \in \Omega / x_1 = u_1, x_2 = l_2\}.
 \end{aligned}$$

Sejam L, K tais que $K \leq L$ e os autovalores de B estejam contidos em $[K, L]$.

Notação:

$$g(z) : -\nabla \psi(z) \equiv -(Bz + b)$$

\bar{F}_I : o fecho de F_I

$V(F_I)$: a menor variedade linear que contém F_I

$S(F_I)$: o subespaço paralelo a $V(F_I)$

$\dim F_I$: a dimensão de $S(F_I)$

Assim, se F_I não for vazio,

$$V(F_I) = \left\{ z \in \mathbb{R}^n / \begin{array}{l} z_i = l_i, \text{ se } i \in I \\ z_{i+n} = u_i, \text{ se } n+i \in I \end{array} \right\}$$

$$S(F_I) = \left\{ z \in \mathbb{R}^n / \begin{array}{l} z_i = 0, \text{ se } i \in I, \\ \text{ou} \\ \text{se } n+i \in I \end{array} \right\},$$

$\bar{F}_I = \Omega \cap V(F_I)$ e $\dim F_I = n - n_I$, onde n_I denota o número de elementos de I .

Para todo $z \in \Omega$ define-se $\bar{g}_p(z) \in \mathbb{R}^n$ o gradiente projetado negativo:

$$\bar{g}_p(z)_i = \begin{array}{l} 0, \text{ se } z_i = l_i \text{ e } \frac{\partial \psi}{\partial z_i}(z) \geq 0 \\ \text{ou} \\ \text{se } z_i = u_i \text{ e } \frac{\partial \psi}{\partial z_i}(z) \leq 0 \\ -\frac{\partial \psi}{\partial z_i}(z), \text{ outros casos.} \end{array}$$

Se z é uma solução de minimizar $\psi(z)$ sujeita a $l \leq z \leq u$, então $\bar{g}_p(z) = 0$.

Para todo $z \in \bar{F}_I$ defini-se $\bar{g}_I(z) \in \mathbb{R}^n$ por:

$$\bar{g}_I(z)_i = \begin{array}{l} 0, \text{ se } i \in I \text{ ou } n+i \in I \\ -\frac{\partial \psi}{\partial z_i}(z), \text{ outros casos.} \end{array}$$

Para todo $z \in \bar{F}_I$ defini-se,

$$\bar{g}_I^c(z)_i = 0, \text{ se } i \notin I \text{ e } n+i \notin I,$$

$$\begin{aligned}
 & 0, \text{ se } i \in I \text{ e } \frac{\partial \psi}{\partial z_i}(z) \geq 0, \\
 & \text{ou} \\
 & \text{se } n+i \in I \text{ e } \frac{\partial \psi}{\partial z_i}(z) \leq 0 \\
 & -\frac{\partial \psi}{\partial z_i}(z), \text{ outros casos.}
 \end{aligned}$$

Logo, para todo $z \in F_I$,

$$\bar{g}_p(z) = \bar{g}_I(z) + \bar{g}_I^c(z).$$

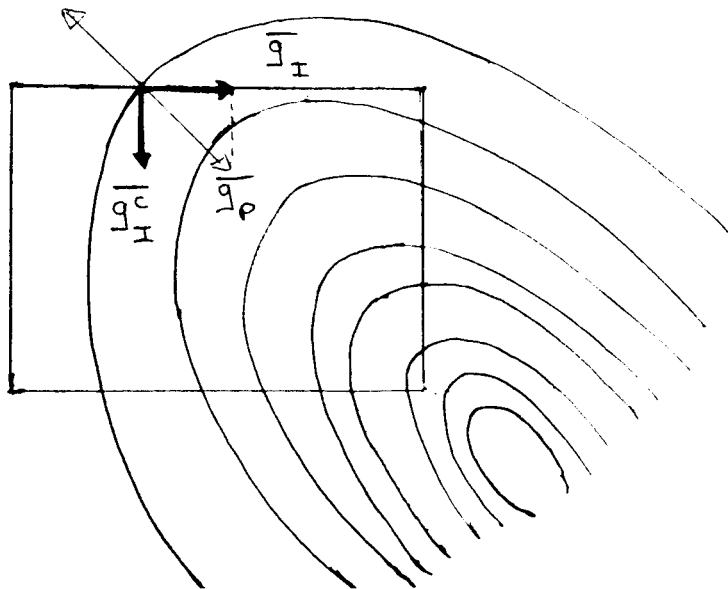


fig1. Direção do Gradiente Projetado

Antes de descrever o algoritmo II, definiremos γ_I e γ .

Se $F_I \neq \phi$,

$$\gamma_I = \min\{u_i - l_i / i \in I \text{ ou } n + i \in I\} \epsilon$$

$$\gamma = \min\{u_i - l_i / i \in \{1, 2, \dots, n\}\}.$$

1.6.1 Algoritmo II (Quacan)

Seja $z^0 \in \Omega$ aproximação inicial arbitrária, $\epsilon \geq 0$ tal que $\epsilon > 0$ se $K < 0$.

Escolher um $\delta > 0$ tal que:

Dados iniciais : $z^0 \in \Omega$ aproximação inicial arbitrária

$\epsilon \geq 0$ tal que $\epsilon > 0$ se $K < 0$

escolher $\delta > 0$ tal que

$$\delta \leq \min\left\{\frac{\epsilon}{\sqrt{L \max\{0, -K\}}}, \sqrt{\frac{L\gamma}{\max\{0, -K\}}}\right\}, \text{ se } L > 0 \text{ ou}$$

$$\delta < \max\left\{\sqrt{\frac{2\gamma\epsilon}{-K}}, \sqrt{\frac{L\gamma}{K}}\right\}, \text{ se } L \leq 0.$$

Dado z_k , então os passos para achar z^{k+1} são:

Passo 1 : Critério de parada

$$\text{Se } \|\bar{g}_p(z_k)\| \leq \epsilon$$

então parar.

Passo 2 : Verificar se $\|g_I(z_k)\|$ é suficientemente pequena para sair da face.

Se quaisquer uma das condições (a) ou (b) ou (c) ou (d) forem satisfeitas, então ir ao passo (3).

se não

ir ao passo (4).

$$a) L > 0, K \leq 0 \text{ e } \|\bar{g}_I(z_k)\| < \frac{1}{2} \left\{ \frac{\min\{\|\bar{g}_I^c(z_k)\|, L\gamma_I\}^2}{\delta L} + K\delta \right\}$$

$$b) K > 0, \|\bar{g}_I(z_k)\| \geq K\delta \text{ e também satisfaz a desigualdade anterior.}$$

$$c) K > 0, \|\bar{g}_I(z_k)\| < K\delta \text{ e } \|\bar{g}_I(z_k)\| < \sqrt{\frac{K}{L}}, \min\{\|\bar{g}_I^c(z_k)\|, L\gamma_I\}$$

$$d) L \leq 0, \text{ e } \|\bar{g}_I(z_k)\| < \frac{\gamma_I \|\bar{g}_I^c(z_k)\|}{\delta} - \frac{L}{2\delta} \gamma_I^2 + \frac{K}{2} \delta$$

Passo 3 : Achar um novo ponto que não pertença a F_I

Calcular $z_{k+1} \in \Omega - \bar{F}_I$ tal que

$$\psi(z_{k+1}) < \psi(z),$$

$$\forall z \in V(F_I) \text{ tal que } \|z - z_k\| \leq \delta.$$

Passo 4 : Calcular a direção do gradiente conjugado.

Se $k = 0$ ou $z_{k-1} \notin F_I$

então definir $d_k = \bar{g}_I(z_k)$

se não definir $d_k = \bar{g}_I(z_k) + \beta_k d_{k-1}$

onde

$$\beta_k = \frac{\|\bar{g}_I(z_k)\|^2}{\|\bar{g}_I(z_{k-1})\|^2}$$

Passo 5 : Achar o novo ponto sobre \bar{F}_I .

Se $\Psi(z)$ é inferiormente limitada ao longo da semireta $L_k \equiv \{\lambda \geq 0 / z_k + \lambda d_k\}$ ou se o minimizador y_k de $\psi(z)$ sobre L_k não pertence a F_I , obter $z_{k+1} \in \bar{F}_I - F_I$ que satisfaz:

$$\psi(z_{k+1}) < \psi(z_k)$$

se não $z_{k+1} = y_k$.

Em resumo, a idéia do algoritmo é que o fecho da face deve ser abandonada somente se existir um ponto fora de \bar{F}_I que satisfaça uma condição de decréscimo suficiente . Quer dizer que o valor da função no novo ponto deve ser um limitante inferior para os valores de ψ sobre a bola de raio δ , centrada em z_k e restrita a $V(F_I)$.

A existência de um novo ponto é garantida pelas condições (a)-(d) do passo 2. Além disso, se $\|\bar{g}_I(z_k)\| = 0$ e o critério de parada $\|\bar{g}_p(z_k)\| \leq \epsilon$ não é satisfeito então as condições para sair da face se verificam.

Por outro lado , se o critério no passo 2 não é satisfeito, iterações de gradientes conjugados serão executadas dentro de F_I . Neste caso ,se o minimizador ao longo da direção de descida não existe ou não pertence à face F_I , o novo ponto está na fronteira da face atual.

Este algoritmo (*Quacan*) está bem definido e toda sequência $\{z_k\}$ gerada por ele num número finito de iterações satisfaz o critério de parada $\|\bar{g}_p(z_k)\| \leq \epsilon$.

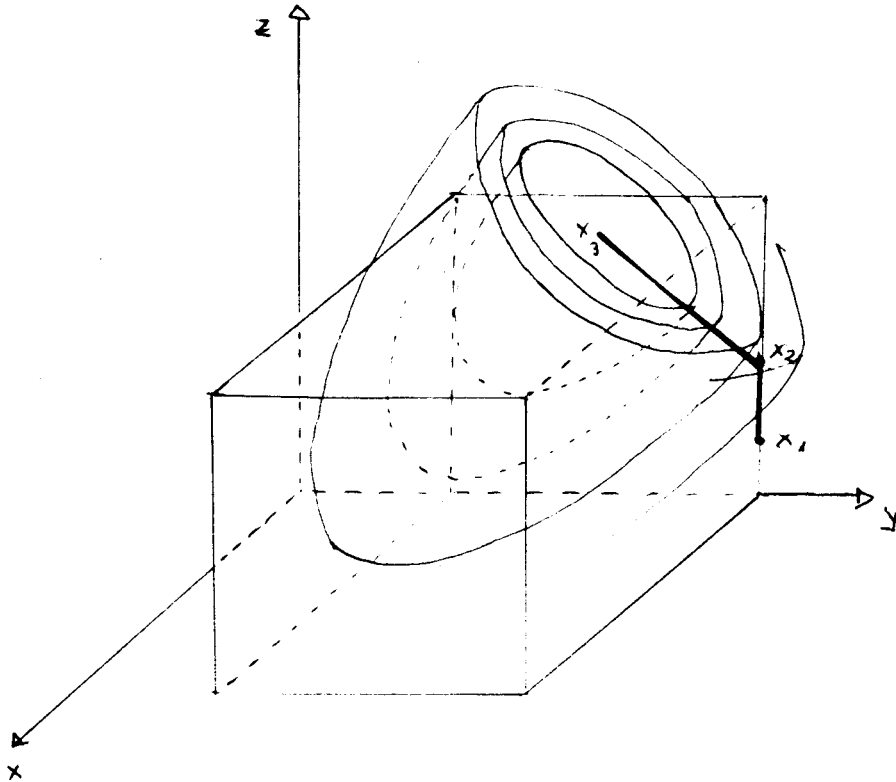


fig2. Direção escolhida em Quacan

Capítulo 2

Hiperplanos de separação

2.1 Introdução

Neste capítulo analisaremos a teoria desenvolvida para o estudo da existência de um hiperplano que faz o papel de separador de dois grupos dados. Assumimos que uma população foi dividida em dois grupos por um critério de mérito desconhecido .

Nosso interesse é verificar a existência do mencionado hiperplano que separe os dois grupos. Este problema foi modelado por Martínez [10] como um problema de otimização convexa.

O problema descrito anteriormente é aplicável em diferentes áreas. Por exemplo, a políticas de contratação em organizações privadas ou governamentais, promoção de indivíduos a um nível superior, determinação de salários, aceitação de artigos em jornais científicos, etc. Muitas vezes a decisão feita nestes casos envolve um julgamento subjetivo .

Assumimos que depois de um processo de seleção os indivíduos de uma certa população foram classificados em duas categorias: superior (S) e inferior (I). A passagem de uma categoria à outra poderia indicar uma promoção

sendo uma mudança positiva para o indivíduo que pertencia a I e agora pertence a S por lhe trazer muitos benefícios, ou pode ser um conjunto de estudantes que cursam uma disciplina e a passagem significa a aprovação nesta.

Em geral a pessoa encarregada de determinar a que categoria o indivíduo pertence, tem que concordar com as qualidades ou habilidades a serem consideradas positivamente com respeito às variáveis.

Há pouca concordância para determinar a melhor maneira de se ponderar as diferentes variáveis .

Se após algum tempo da classificação subjetiva, se desejasse encontrar um critério objetivo que verificasse a separação, uma maneira poderia ser achando um hiperplano que separe os indivíduos que pertencem à categoria I daqueles que pertencem à categoria S . Neste caso, temos que achar os valores dos coeficientes da equação do hiperplano.

A determinação dos coeficientes ou pesos pode também obedecer a restrições adicionais (além da positividade). De fato, poder-se-ia impor alguma relação entre os pesos correspondentes a um determinado ítem, que deveriam estar relacionados aos pesos de outros.

Os indivíduos da população considerada são identificados com pontos de um espaço n -dimensional \mathfrak{R}^n , onde n é o número de ítems considerados .

Os dados do problema são representados pela matriz $Y \in \mathfrak{R}_{m \times n}$ onde m é o número de indivíduos da população e y_{ij} é a nota do indivíduo i no ítem j .

$$y_i = (y_{i1}, y_{i2}, \dots, y_{in})$$

Portanto, a cada indivíduo corresponde uma linha da matriz Y e estas linhas são denotadas por y_1, y_2, \dots, y_m . Além disso, é assumido que para algum $q \in \{1, 2, \dots, m\}$

$$\begin{array}{l} y_1, y_2, \dots, y_q \in S \\ \in \\ y_{q+1}, \dots, y_m \in I. \end{array}$$

O problema é encontrar pesos positivos x_1, \dots, x_n tais que para um $\beta \in \mathbb{R}^+$ dado:

$$\begin{array}{rcl} x_1 y_{11} + \dots + x_n y_{1n} & \geq & \beta \\ \vdots & \ddots & \vdots \\ x_1 y_{q1} + \dots + x_n y_{qn} & \geq & \beta \\ x_1 y_{q+11} + \dots + x_n y_{q+1n} & \leq & \beta \\ \vdots & \ddots & \vdots \\ x_1 y_{m1} + \dots + x_n y_{mn} & \leq & \beta \end{array} \quad (2.1)$$

Usando a notação padrão \langle, \rangle para o produto escalar temos:

$$\begin{array}{l} \langle x, y_i \rangle \geq \beta \quad , \quad i = 1, \dots, q \\ \langle x, y_i \rangle \leq \beta \quad , \quad i = q + 1, \dots, m. \end{array}$$

Além disso x poderá satisfazer restrições adicionais, representadas por :

$$Ax \geq b \quad (2.2)$$

onde $A \in \mathbb{R}_{t \times n}$, e suas linhas são denotadas por a_1, \dots, a_t .

Se a divisão da população em dois grupos de acordo com o critério desconhecido satisfaz o critério matemático, então o hiperplano existe (*fig a*). Nem sempre o hiperplano \mathcal{H} existe (*fig b*). Existem outros casos onde se precisamos separar em mais de dois grupos. Por exemplo si desejamos verificar a divisão feita em três subconjuntos nossa primeira tentativa seria agrupá-los verificando a primeira separação dos subconjuntos agrupados e logo verificar a segunda separação dos subconjuntos restantes (*fig c*).

O hiperplano $\mathcal{H} = \{y \in \mathbb{R}^n / \langle x, y \rangle = \beta\}$ é chamado hiperplano de separação entre S e I .

O principal interesse é determinar a existência do hiperplano de separação que satisfaz (2.2). Existem casos onde ele não é único (*fig d*).

Propomos modelar este problema como um problema de otimização não linear e para resolvê-lo usamos o método de região de confiança descrito no capítulo anterior.

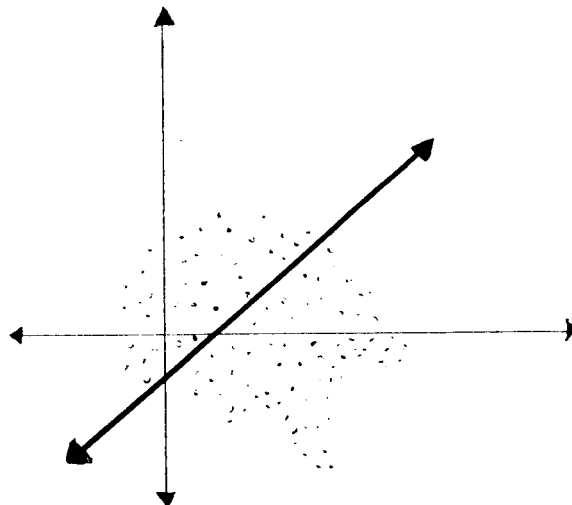


fig (a)

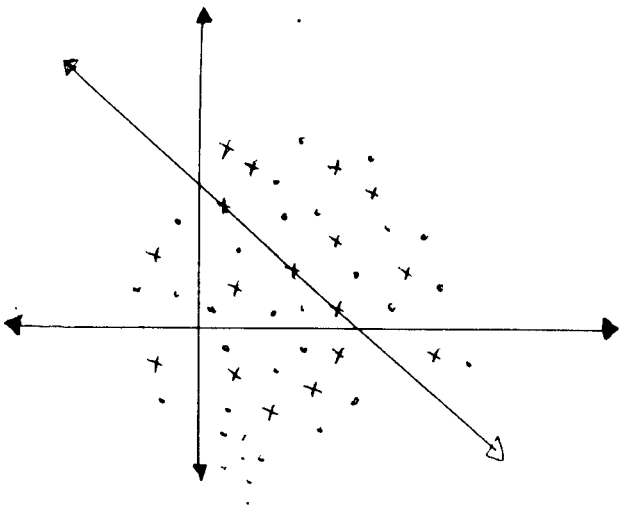


fig (b)

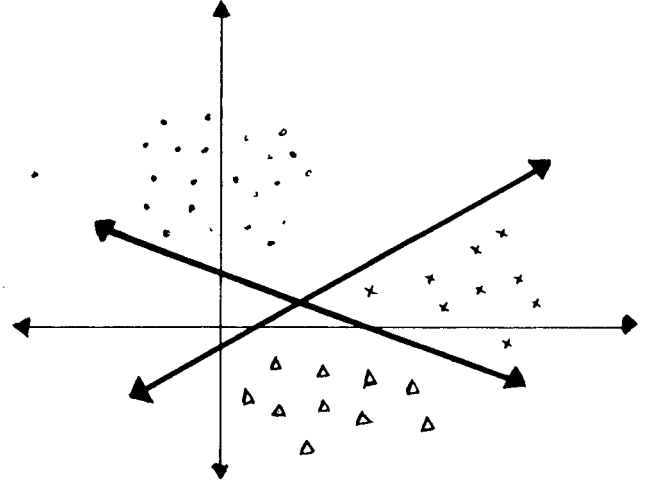


fig (c)

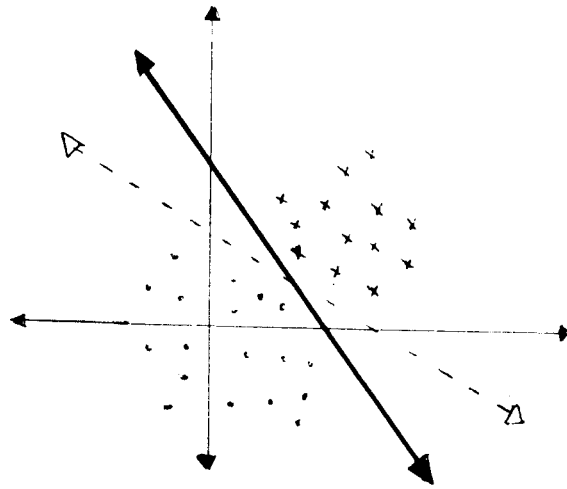


fig (d)

2.2 Problema de Otimização

O problema é achar $x \geq 0$, que satisfaça (2.1) e (2.2). Isto é equivalente ao seguinte problema de otimização :

$$\begin{aligned} & \text{Minimizar } f(x) \\ & \text{s/a } x \geq 0 \end{aligned}$$

onde

$$\begin{aligned} f(x) \equiv & \sum_{i=1}^q \{ \{ \langle x, y_i \rangle - \beta \}_- \}^2 \\ & + \sum_{i=q+1}^m \{ \{ \langle x, y_i \rangle - \beta \}_+ \}^2 \\ & + \sum_{i=1}^t \{ \{ \langle a_i, x \rangle - b_i \}_- \}^2, \end{aligned}$$

$$z_+ = \max\{0, z\},$$

e

$$z_- = \max\{0, -z\}.$$

Se depois de resolver o problema existe um $x^* \in \mathfrak{R}^n$ tal que $f(x^*) = 0$, podemos afirmar que x^* é o minimizador global de f e o hiperplano $\mathcal{H} = \{y \in \mathfrak{R}^n / \langle x^*, y \rangle = \beta\}$ separa S de I . Logo

$$\langle x^*, y_i \rangle \geq \beta, \quad \text{para todo } i \leq q$$

e

$$\langle x^*, y_i \rangle \leq \beta, \quad \text{para todo } i \geq q + 1.$$

Temos que,
se

$$f(x^*) = 0$$

como

$$\begin{aligned} f(x^*) &\equiv \sum_{i=1}^q \{ \{ \langle x^*, y_i \rangle - \beta \}_- \}^2 \\ &\quad + \sum_{i=q+1}^m \{ \{ \langle x^*, y_i \rangle - \beta \}_+ \}^2 \\ &\quad + \sum_{i=1}^t \{ \{ \langle a_i, x^* \rangle - b_i \}_- \}^2 \end{aligned}$$

então

$$\begin{aligned} \{ \langle x^*, y_i \rangle - \beta \}_- &= 0, \text{ para } i = 1, \dots, q \\ \{ \langle x^*, y_i \rangle - \beta \}_+ &= 0, \text{ para } i = q + 1, \dots, m \\ \{ \langle a_i, x^* \rangle - b_i \}_- &= 0, \text{ para } i = 1, \dots, t \end{aligned}$$

pela definição de z_+ e z_- temos :

i)

$$\begin{aligned} \max\{0, -(\langle x^*, y_i \rangle)\} &= 0 \\ -\langle x^*, y_i \rangle + \beta &\leq 0 \\ \langle x^*, y_i \rangle &\geq \beta \end{aligned}$$

ii)

$$\begin{aligned} \max\{0, \langle x^*, y_i \rangle - \beta\} &= 0 \\ \langle x^*, y_i \rangle - \beta &\leq 0 \\ \langle x^*, y_i \rangle &\leq \beta \end{aligned}$$

iii)

$$\begin{aligned} \max\{0, -(\langle a_i, x^* \rangle - b_i)\} &= 0 \\ -(\langle a_i, x^* \rangle - b_i) &\leq 0 \\ \langle a_i, x^* \rangle - b_i &\geq 0 \\ \langle a_i, x^* \rangle &\geq b_i \end{aligned}$$

A função f tem derivadas parciais contínuas e é convexa, portanto, todo ponto de Kuhn-Tucker x^* é um minimizador global. (Luenberger [9]) Isto significa que se encontrarmos um ponto de Kuhn Tucker x^* tal que $f(x^*) > 0$, podemos garantir que não existe um hiperplano de separação que satisfaz $Ax \geq b$ e $x \geq 0$.

De fato,

$$\begin{aligned} \text{se } f(x) > 0 \Rightarrow & \sum_{i=1}^q \{ \{ \langle x, y_i \rangle - \beta \}_- \}^2 \\ & + \sum_{i=q+1}^m \{ \{ \langle x, y_i \rangle - \beta \}_+ \}^2 \\ & + \sum_{i=1}^t \{ \{ \langle a_i, x \rangle - b_i \}_- \}^2 > 0 \end{aligned}$$

portanto

i)

$$\sum_{i=1}^q \{ \{ \langle x, y_i \rangle - \beta \}_- \}^2 > 0$$

ii)

$$\sum_{i=q+1}^m \{ \{ \langle x, y_i \rangle - \beta \}_+ \}^2 > 0$$

iii)

$$\sum_{i=1}^t \{ \{ \langle a_i, x \rangle - b_i \}_- \}^2 > 0.$$

Se for o caso que

$$\sum_{i=1}^q \{ \{ \langle x, y_i \rangle - \beta \}_- \}^2 > 0$$

afirmamos que existe pelo menos uma parcela positiva desde que cada termo da somatória é positivo, ou seja,

existe $i_0 \in \{1, 2, \dots, q\}$ tal que

$$\{\{\langle x, y_{i_0} \rangle - \beta\}_-\}^2 > 0$$

logo

$$\begin{aligned} \{\langle x, y_{i_0} \rangle - \beta\}_- &> 0 \\ \text{ou} \\ -\{\{\langle x, y_{i_0} \rangle - \beta\}_-\} &< 0 \end{aligned}$$

a) Se $\{\langle x, y_{i_0} \rangle - \beta\}_- > 0$, então por definição temos

$$\begin{aligned} \max\{0, -\{\{\langle x, y_{i_0} \rangle - \beta\}_-\}\} &> 0 \\ -\langle x, y_{i_0} \rangle + \beta &> 0 \\ \langle x, y_{i_0} \rangle &< \beta \end{aligned} \tag{2.3}$$

Por outro lado sabemos que $\langle x, y_{i_0} \rangle \geq \beta$ para todo $i \leq q$ logo

$$\begin{aligned} \langle x, y_{i_0} \rangle - \beta &\geq 0 \\ -\langle x, y_{i_0} \rangle + \beta &\leq 0 \\ \beta &\leq \langle x, y_{i_0} \rangle \end{aligned} \tag{2.4}$$

De (2.3) e (2.4) vemos que existe uma contradição. Portanto o hiperplano \mathcal{H} não existe.

b) Para o outro caso, ou seja, quando $-\{\{\langle x, y_{i_0} \rangle - \beta\}_-\} < 0$ afirmamos

que é impossível desde que a função

$$\max\{0, -\{\langle x, y_{i_0} \rangle - \beta\}\} \geq 0.$$

Analogamente acontece nos casos *ii)* e *iii)*.

Para a resolução de problemas de otimização de funções convexas podemos usar algoritmos com propriedade de convergência global. Neste caso usamos o pacote *Bor-Quacan*.

O gradiente da função é definido por:

$$\nabla f(x) = \sum_{i \in W_1} 2\{\langle x, y_i \rangle - \beta\} y_i + \sum_{i \in W_2} 2\{\langle a_i, x \rangle - b_i\} a_i$$

onde,

$$W_1 = \{ i \in \{1, 2, \dots, q\} / \langle x, y_i \rangle < \beta \}$$

$$\cup \{ i \in \{q+1, \dots, m\} / \langle x, y_i \rangle > \beta \}$$

e

$$W_2 = \{ i \in \{1, \dots, t\} / \langle a_i, x \rangle < b_i \}$$

A matriz hessiana é descontínua nos pontos x tais que $\langle x, y_i \rangle = \beta$ ou $\langle a_i, x \rangle = b_i$. Excluindo esses pontos a hessiana é definida por:

$$\nabla^2 f(x) = 2\left\{ \sum_{i \in W_1} y_i y_i^t + \sum_{i \in W_2} a_i a_i^t \right\}.$$

Capítulo 3

Experimentos Práticos

3.1 Introdução

Neste capítulo testamos o desempenho de **Box-Quacan** descrito no capítulo 1, para problemas de separação.

Definimos os principais parâmetros de entrada da subrotina **Box-Quacan** que o usuário precisará fornecer. Os autores sugerem alguns valores. Castro [3] encontra os parâmetros ótimos, mas para problemas específicos. Em nossos experimentos práticos usamos alguns valores sugeridos pelos autores e outros foram obtidos via testes prévios aplicados numa muestra de 10% da população total de flores.

Aplicaremos o modelo desenvolvido no capítulo anterior para um problema de classificação de flores usando o **Box-Quacan**.

Nosso objetivo é verificar a separação em três grupos dados considerando as medidas de suas diferentes íris com a teoria discutida nos capítulos anteriores.

3.2 Parâmetros de entrada do Box-Quacan

Para o uso da subrotina *Box-Quacan* precisamos definir alguns parâmetros de precisão e controle.

a) De controle:

Nafmax indica o número máximo de avaliações da função f permitidas na subrotina *Box-Quacan*.

Itmax indica o número máximo de iterações aceitas na subrotina *Box-Quacan*.

Kmax indica o número máximo de iterações permitidas na resolução do subproblema quadrático usado em *Quacan*.

b) De precisão:

EpsG fornece a precisão para a convergência ; o método converge quando :

$$\|G_p(x)\| \leq EpsG * \frac{\max\{1, |f(x)|\}}{\max\{1, \|x\|\}} \quad (3.1)$$

onde $G_p(x)$ é o gradiente projetado na região factível.

EpsD fornece a tolerância para o raio da região de confiança .O algoritmo termina quando é incapaz de produzir um decréscimo da função com uma região de confiança de tamanho $EpsD$.

EpsF fornece a tolerância para a função objetivo. O algoritmo termina quando

$$f(x) \leq Fsol + EpsF * \max\{1, |Fsol|\} \quad (3.2)$$

Fsol é uma estimativa para o valor da função objetivo fornecida pelo usuário.

Accuracy determina o grau de precisão da solução do subproblema quadrático

$$\|G_p(x)\| \leq accuracy * \|G_p(x_k)\| \quad (3.3)$$

O valor sugerido é 0.1.

Dont é usado como critério adicional para controlar el progresso de la función objetivo na subrotina *Quacan*. O valor sugerido é 0.01.

Tam0 Dentro do *Quacan* é o tamanho da vizinhança sobre cada face que não vai ser revisitada, quando a correspondente face é abandonada .
Recomenda-se que o parâmetro corresponda a uma fração do diâmetro (1/10 a 1/10000) da caixa do problema original.

Nana Usar $Nana = n$ para uma versão que não aproveita estrutura de dados esparsa e $Nana = 2$ para uma versão esparsa.

Par Auxiliar na definição do tamanho do raio de confiança no início de cada iteração. Deve pertencer a (1,10). A sugestão dos autores é $Par = 4$.

Delta0 É o tamanho do raio de confiança na iteração zero.

DeltaMin É o menor raio possível, para a região de confiança ao início de cada iteração.

Bont Faz o mesmo papel em *Box* que o *Dont* em *Quacan*. O valor sugerido é 0.01.

MepsG Indica que a tolerancia *EpsG* vai ser alterada. O valor sugerido é 1.

frac É uma fração da norma do gradiente projetado no ponto inicial

zlprb,zuprb estes vetores de precisão dupla são os limites do problema original, quer dizer l e u .

Consideramos um problema simples como exemplo.

3.3 Exemplo

Seja a população Ω , formada por 10 alunos que pertencem à mesma turma. Eles cursaram uma determinada disciplina e fizeram três provas obtendo as notas que estão na primeira tabela .

O professor gostaria de aprovar aos alunos y_1, y_3, y_4, y_7 e y_8 baseando-se no conceito de que as provas foram acumulativas e que se eles foram aprovados na terceira prova e as notas das duas primeiras não foram muito baixas, então estão em condições de serem aprovados na matéria.

Para fundamentar melhor esta intuição, ele deseja saber se existem os coeficientes para determinar o plano de separação que resultem nesta decisão.

O professor considera a segunda prova com peso de ao menos dois com respeito à primeira, e a terceira com peso de ao menos três com respeito a primeira, ou seja:

$$\begin{aligned}x_1 &\geq 1 \\x_2 &\geq 2x_1 \\x_3 &\geq 3x_1,\end{aligned}$$

onde x_i é o peso da prova i , para $i = 1, 2, 3$.

Matricialmente temos:

$$A = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix}, b = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

Reordenando os alunos de acordo com a escolha do professor, eles ficam de acordo com a segunda tabela .

Nosso problema consiste em verificar a existência de um hiperplano que

separe fortemente os dois grupos, quer dizer:

$$\begin{array}{rcl}
 x_1 y_{11} + \dots + x_n y_{1n} & > & \beta \\
 \vdots & \ddots & \vdots \\
 x_1 y_{q1} + \dots + x_n y_{qn} & > & \beta \\
 x_1 y_{q+11} + \dots + x_n y_{q+1n} & < & \beta \\
 \vdots & \ddots & \vdots \\
 x_1 y_{m1} + \dots + x_n y_{mn} & < & \beta
 \end{array}$$

onde $y_{i,j}$: é a nota do aluno y_i na prova j .

Substituindo os valores de $y_{i,j}$, $i = 1, \dots, m$ e $j = 1, \dots, n$ temos

$$\begin{array}{rcl}
 4x_1 + 5x_2 + 6x_3 & > & \beta \\
 3x_1 + 7.5x_2 + 5x_3 & > & \beta \\
 3x_1 + 4x_2 + 8x_3 & > & \beta \\
 6x_1 + 8x_2 + 7x_3 & > & \beta \\
 5x_1 + 3x_2 + 7x_3 & > & \beta \\
 3x_1 + 2x_2 + x_3 & < & \beta \\
 7x_1 + 2x_2 + 4x_3 & < & \beta \\
 8x_1 + 5x_2 + 3x_3 & < & \beta \\
 5x_1 + 6x_2 + 4x_3 & < & \beta \\
 3x_1 + 2x_2 + 6x_3 & < & \beta
 \end{array}$$

onde β pode ser escolhido arbitrariamente (neste caso usamos $\beta = 30$).

Notas das provas			
nome do aluno	primeira prova	segunda prova	terceira prova
y_1	4	5	6
y_2	3	2	1
y_3	3	7.5	5
y_4	3	4	8
y_5	7	2	4
y_6	8	5	3
y_7	6	8	7
y_8	5	3	7
y_9	5	6	4
y_{10}	3	2	6

tab I

Notas das provas reordenadas			
nome do aluno	primeira prova	segunda prova	terceira prova
y_1	4	5	6
y_3	3	7.5	5
y_4	3	4	8
y_7	6	8	7
y_8	5	3	7
y_2	3	2	1
y_5	7	2	4
y_6	8	5	3
y_9	5	6	4
y_{10}	3	2	6

tab II

Observação : Apesar de ter uma nota aprovatoria na terceira prova o aluno y_i , ele não vai ser considerado com merito suficiente a aprovar devido a que as duas primeiras foram muito baixas.

Os parâmetros usados neste exemplo são:

$Nafmax$	= 1000
$Itmax$	= 500
$Kmax$	= 500
$EpsG$	= $1.0d - 5$
$EpsD$	= $1.0d - 8$
$EpsF$	= $1.0d - 8$
$Fsol$	= $0.0d0$
$Acuracy$	= $0.01d0$
$Dont$	= $0.01d0$
$Tam0$	= $0.01d0$
$Nana$	= N
Par	= $4.0d0$
$Delta0$	= $10.d0$
$DeltaMin$	= $1.0d0$
$Bont$	= $0.0d0$
$MepsG$	= 1
$Frac$	= $0.0d0$
$zlprb$	= $0.d0$
$zuprb$	= $1000.d0$

Implementamos em linguagem *Fortran 77* na Sun Workstation, o programa principal que chama a subrotina *Bor-Quacan*. O ponto inicial considerado é $x_0 = (1 \ 5 \ 7)$ e a função objetivo avaliada neste ponto é :

$$f(x_0) = 2515.$$

Lembrando a teoria do capítulo anterior sabemos que se $f(x) \neq 0$, então x não é a solução ótima. Precisamos achar um x^* factível tal que $f(x^*) = 0$ para garantir a existência do hiperplano que classifica os alunos em dois grupos, os aprovados e os não aprovados.

Depois de cinco iterações de *Bor* e vinte chamadas de *Quacan* o algoritmo converge para

$$x^* = \begin{bmatrix} 1 \\ 2 \\ 3.25 \end{bmatrix},$$

onde $f(x^*) = 0$. As componentes de x^* são os pesos ótimos a serem usados nas respectivas provas. Agora usando estes pesos o professor poderá separá-los segundo a escolha feita.

Neste caso consideramos a primeira prova integralmente, mas se desejamos relaxar a primeira restrição é suficiente exigir que $x_1 \geq 0$ então a matriz se modificaria a

$$A = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix}, b = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

e o problema seria

$$\begin{array}{ll} \min & f(x) \\ \text{s/a} & x \geq 0. \end{array}$$

Finalmente a matriz A e o vetor b estariam definidos por:

$$A = \begin{bmatrix} -2 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix}, b = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

onde a aproximação inicial é $x_0 = (0 \ 0 \ 0)$ com $f(x_0) = 4500$.

A solução final é obtida depois de três iterações do *Bor* e sete chamadas ao *Quacan* obtendo $f(x^*) = 0$.

$$x^* = \begin{bmatrix} 0.8900191164859 \\ 1.7800382329719 \\ 2.9232887281994 \end{bmatrix}$$

As saídas respectivas dos programas são anexadas ao final.

3.4 Aplicação à Classificação das Flores

O problema de separação pode ser aplicado em diferentes exemplos. Nós estamos interessados em aplicar ao problema de classificação de flores.

Nosso objetivo é verificar a separação com a teoria discutida neste trabalho e talvez tentar melhorar a separação feita em análise multivariada.

A nossa população Ω é constituída por flores. Considerando suas íris, elas foram separadas em três classes: íris setosa, íris versicolor e íris virgínica.

Os dados foram citados por Fisher [8] em 1936, e podem ser observados nas tabelas *IV*, *V* e *VI*.

Para fazer essa classificação foram considerados o comprimento e longitude da pétala e o comprimento e longitude do sépalo.

A mostra para tal problema foi de 150 indivíduos e representará a nossa população. A análise feita para este tipo de problema é encarada do ponto de vista estatístico. Nele se escolhe o método de "cluster" para efetuar a separação. Este método começa inicialmente com seis cluster e, progressivamente, reduz o número de cluster a três. Os dados dos cluster obtidos são

o primeiro é de tamanho 62, constituído por 14 indivíduos do grupo das flores de íris virgínica e 48 elementos do grupo das flores de íris versicolor, o segundo cluster de tamanho 50, constituído integralmente pelas flores de íris setosa e finalmente o último cluster é de tamanho 38 com 36 elementos do grupo das flores de íris virgínica e dois de íris versicolor (*tab III*)

Em nosso caso, para tentar aplicar a teoria do capítulo anterior precisamos ter inicialmente apenas dois grupos para verificar a separação deles. Associamos dos conjuntos de flores que formaram um grupo.

No de cluster	tamanho	Virgínica	Setosa	versicolor
1	62	14	-	48
2	50	-	50	-
3	38	36	-	2

tab III

Em resumo teremos que o nosso primeiro grupo será aquele formado pelas flores de íris setosa e as de íris versicolor, e o segundo grupo será aquele formado pelas flores de íris virgínica.

Como mencionamos anteriormente, este problema será resolvido usando o modelo e os algoritmos apresentados nos capítulos anteriores.

Identificamos os dados do problema com as variáveis descritas no modelo. Agora temos que os x_i não precisam ser obrigatoriamente positivos, logo o limitante inferior pode ser negativo. Consideraremos

$$l \leq u$$

onde, $l = (-1000 \ -1000 \ -1000 \ -1000)$, $u = (1000 \ 1000 \ 1000 \ 1000)$ e o nosso ponto inicial factível é $x_o = (0 \ 0 \ 0 \ 0)$.

Dada a função objetivo:

$$f(x) = \sum_{i=1}^q \{ \{ \langle x, y_i \rangle - \beta \}_- \}^2 + \sum_{i=q+1}^m \{ \{ \langle x, y_i \rangle - \beta \}_+ \}^2$$

avaliamos a função em nosso ponto inicial factível x_o e obtemos $f(x_o) = 50000$. Obviamente estamos afastados da nossa solução ótima.

Para aplicarmos o algoritmo, precisamos fazer alguns cálculos a fim de especificarmos alguns valores dos parâmetros iniciais, tais como:

$$\begin{aligned} \sigma_1 &= 0.1, \\ \sigma_2 &= 0.9, \\ \alpha &= 10^{-4}, \\ \Delta^\circ &= 10 \\ \Delta_{min} &= 1.d0. \\ \beta &= 100 \end{aligned}$$

Depois de algumas iterações de *Bor* conseguimos a solução ótima, verificando, assim, a existência do hiperplano. Especificamente foram seis iterações de *Bor* e nove iterações de *Quacan*.

O vetor solução aproximado é

$$x_1^* = \begin{bmatrix} 9.1454222627725 \\ 5.7928990179917 \\ 4.8435717350834 \\ 0.5762511432536 \end{bmatrix}$$

os valores de alguns deles de maneira que possamos obter uma solução satisfatória.

Foram realizados vários testes com diferentes valores e selecionados aqueles que ofereciam possibilidade de solução

O Bont, por exemplo, foi modificado de valor (de 0.01 a 0.0) ,os valores de parâmetros de controle foram escolhidos suficientemente grandes como para obter uma resposta antes de atingir seus valores e os valores dos raios foram escolhidos levando-se em consideração as dimensões da caixa.

Finalmente, consegue-se uma escolha que na prática mostrou-se eficiente a través dos experimentos efetuados.

tab IV

Flores com iris setosa			
longitude do sépalo	comprimento do sépalo	longitude da pétala	comprimento da pétala
5.1	3.5	1.4	0.2
4.9	3.0	1.4	0.2
4.7	3.2	1.3	0.2
4.6	3.1	1.5	0.2
5.0	3.6	1.4	0.2
5.4	3.9	1.7	0.4
4.6	3.4	1.4	0.3
5.0	3.4	1.5	0.2
4.4	2.9	1.4	0.2
4.9	3.1	1.5	0.1
5.4	3.7	1.5	0.2
4.8	3.4	1.6	0.2
4.8	3.0	1.4	0.1
4.3	3.0	1.1	0.1
5.8	4.0	1.2	0.2
5.7	4.4	1.5	0.4
5.4	3.9	1.3	0.4
5.1	3.5	1.4	0.3
5.7	3.8	1.7	0.3
5.1	3.8	1.5	0.3
5.4	3.4	1.7	0.2
5.1	3.7	1.5	0.4
4.6	3.6	1.0	0.2
5.1	3.3	1.7	0.5
4.6	3.4	1.9	0.2
5.0	3.0	1.6	0.2
5.0	3.4	1.6	0.4
5.2	3.5	1.5	0.2
5.2	3.4	1.4	0.2
4.7	3.2	1.6	0.2
4.8	3.1	1.6	0.2
5.4	3.4	1.5	0.4
5.2	4.1	1.5	0.1
5.5	4.2	1.4	0.2
4.9	3.1	1.5	0.2
5.0	3.2	1.2	0.2
5.5	3.5	1.3	0.2
4.9	3.6	1.4	0.1
4.4	3.0	1.3	0.2
5.1	3.4	1.5	0.2
5.0	3.5	1.3	0.3
4.5	2.3	1.3	0.3
4.4	3.2	1.3	0.2
5.0	3.5	1.6	0.6
5.1	3.8	1.9	0.4
4.8	3.0	1.4	0.3
5.1	3.8	1.6	0.2
4.6	3.2	1.4	0.2
5.3	3.7	1.5	0.2
5.0	3.3	1.4	0.2

tab V

Flores com iris Versicolor			
longitude do sépalo	comprimento do sépalo	longitude da pétala	comprimento da pétala
7.0	3.2	4.7	1.4
6.4	3.2	4.5	1.5
6.9	3.1	4.9	1.5
5.5	2.3	4.0	1.3
6.5	2.8	4.6	1.5
5.7	2.8	4.5	1.3
6.3	3.3	4.7	1.6
4.9	2.4	3.3	1.0
6.6	2.9	4.6	1.3
5.2	2.7	3.9	1.4
5.0	2.0	3.5	1.0
5.9	3.0	4.2	1.5
6.0	2.2	4.0	1.0
6.1	2.9	4.7	1.4
5.6	2.9	3.6	1.3
6.7	3.1	4.4	1.4
5.6	3.0	4.5	1.5
5.8	2.7	4.1	1.0
6.2	2.2	4.5	1.5
5.6	2.5	3.9	1.1
5.9	3.2	4.6	1.8
6.1	2.8	4.0	1.3
6.3	2.5	4.9	1.5
6.1	2.8	4.7	1.2
6.4	2.9	4.3	1.3
6.6	3.0	4.4	1.4
6.8	2.8	4.6	1.4
6.7	3.0	5.0	1.7
6.0	2.9	4.5	1.5
5.7	2.6	3.5	1.0
5.5	2.4	3.6	1.1
5.5	2.4	3.7	1.0
5.6	2.7	3.9	1.2
6.0	2.7	5.1	1.6
5.4	3.0	4.5	1.5
6.0	3.4	4.5	1.6
6.7	3.1	4.7	1.5
6.3	2.3	4.4	1.3
5.6	3.0	4.1	1.3
5.5	2.5	4.0	1.3
5.5	2.6	4.4	1.2
6.1	3.0	4.6	1.4
5.8	2.6	4.0	1.2
5.0	2.3	3.3	1.0
5.6	2.7	4.2	1.3
5.7	3.0	4.2	1.2
5.7	2.9	4.2	1.3
6.2	2.9	4.3	1.3
5.1	2.5	3.0	1.1
5.7	2.8	4.1	1.3

tab VI

Flores com iris Virginica			
longitude do sépalo	comprimento do sépalo	longitude da pétala	comprimento da pétala
6.3	3.3	6.0	2.5
5.8	2.7	5.1	1.9
7.1	3.0	5.9	2.1
6.3	2.9	5.6	1.8
6.5	3.0	5.8	2.2
7.6	3.0	6.6	2.1
4.9	2.5	4.5	1.7
7.3	2.9	6.3	1.8
6.7	2.5	5.8	1.8
7.2	3.6	6.1	2.5
6.5	3.2	5.1	2.0
6.4	2.7	5.3	1.9
6.8	3.0	5.5	2.1
5.7	2.5	5.0	2.0
5.8	2.8	5.1	2.4
6.4	3.2	5.3	2.3
6.5	3.0	5.5	1.8
7.7	3.8	6.7	2.2
7.7	2.6	6.9	2.3
6.0	2.2	5.0	1.5
6.9	3.2	5.7	2.3
5.6	2.8	4.9	2.0
7.7	2.8	6.7	2.0
6.3	2.7	4.9	1.8
6.7	3.3	5.7	2.1
7.2	3.2	6.0	1.8
6.2	2.8	4.8	1.8
6.1	3.0	4.9	1.8
6.4	2.8	5.6	2.1
7.2	3.0	5.8	1.6
7.4	2.8	6.1	1.9
7.9	3.8	6.4	2.0
6.4	2.8	5.6	2.2
6.3	2.8	5.1	1.5
6.1	2.6	5.6	1.4
7.7	3.0	6.1	2.3
6.3	3.4	5.6	2.4
6.4	3.1	5.5	1.8
6.0	3.0	4.8	1.8
6.9	3.1	5.4	2.1
6.7	3.1	5.6	2.4
6.9	3.1	5.1	2.3
5.8	2.7	5.1	1.9
6.8	3.2	5.9	2.3
6.7	3.3	5.7	2.5
6.7	3.0	5.2	2.3
6.3	2.5	5.0	1.9
6.5	3.0	5.2	2.0
6.2	3.4	5.4	2.3
5.9	3.0	5.1	1.8

PRIMER EJEMPLO DAS NOTAS

```
-----
componentes da matriz A
  1.000000000000000  0.  0.
 -2.000000000000000  1.000000000000000  0.
 -3.000000000000000  0.  1.000000000000000
Componentes do vetor c
  1.000000000000000  0.  0.
```

Box Iteration : 0

First (ten) components of X :

```
  1.000000000000000  5.000000000000000  7.000000000000000
Value of objective function at X : 2515.000000000000
Number of function evaluations up to now : 1
Norm of projected gradient : 1572.4744830998
Quacan converges in 4 iterations
```

Box Iteration : 2

First (ten) components of X :

```
  1.3110241347143  2.4320247859733  2.3093544130534
Value of objective function at X : 2.8205531588466
Number of function evaluations up to now : 4
Norm of projected gradient : 15.274749381150
Quacan converges in 3 iterations
```

Box Iteration : 4

First (ten) components of X :

```
  0.83586818757921  1.6096324461343  4.0440430925222
Value of objective function at X : 3.0973252307954D-02
Number of function evaluations up to now : 6
Norm of projected gradient : 0.12420785804815
Quacan converges in 3 iterations.
```

.....

Convergence of BOX: value of objective function

```
Total number of calls to Quacan : 20
Number of BOX - iterations: 5
Number of function evaluations: 7
First 10 components of solution :
  1.000000000000000  2.000000000000000  3.250000000000000
Value of objective function at solution : 0.
Norm of projected gradient : 0.
```

SEGUNDO EXEMPLO DAS NOTAS

componentes da matriz A
-2.0000000000000000 1.0000000000000000 0.
-3.0000000000000000 0. 1.0000000000000000
Componentes do vetor c
0. 0.
beta= 30.000000000000000

Box Iteration : 0

First (ten) components of X :
0. 0. 0.
Value of objective function at X : 4500.000000000000
Number of function evaluations up to now : 1
Norm of projected gradient : 2868.8848007545
Quacan converges in 2 iterations

Box Iteration : 2

First (ten) components of X :
0.94798748361956 1.7145024113942 2.8919587527615
Value of objective function at X : 0.11346649398422
Number of function evaluations up to now : 3
Norm of projected gradient : 4.9221132013569
Quacan converges in 2 iterations

.....
Convergence of BDX: value of objective function
Total number of calls to Quacan : 7
Number of BDX - iterations: 3
Number of function evaluations: 4
First 10 components of solution :
0.89001911648596 1.7800382329719 2.9232887281994
Value of objective function at solution : 0.
Norm of projected gradient : 0.

SEPARACAO DE FLORES DE IRIS SETOSA E VERSICOLOR DAS VIRGINICAS

Box Iteration : 0

First (ten) components of X :

0. 0. 0. 0.

Value of objective function at X : 500000.00000000

Number of function evaluations up to now : 1

Norm of projected gradient : 102204.862898005

Quacan converges in 1 iterations

Box Iteration : 2

First (ten) components of X :

6.5873879983649 5.9265727789055 2.7705295467481 4.2404580901900

Value of objective function at X : 243.90689783038

Number of function evaluations up to now : 3

Norm of projected gradient : 866.04355212647

Quacan converges in 2 iterations

Box Iteration : 4

First (ten) components of X :

9.0901414436363 5.7383681666001 4.8170785444450 0.53836777468929

Value of objective function at X : 1.7845670041849

Number of function evaluations up to now : 5

Norm of projected gradient : 42.594102983311

Quacan converges in 1 iterations

.....
Convergence of BOX: value of objective function

Total number of calls to Quacan : 9

Number of BOX - iterations: 6

Number of function evaluations: 7

First 10 components of solution :

9.1454222627725 5.7928990179917 4.8435717350834 0.57625114325360

Value of objective function at solution : 0.

Norm of projected gradient : 0.

SEPARACAO DE FLORES DE IRIS SETOSA DAS VERSICOLOR

Box Iteration : 0

First (ten) components of X :

0. 0. 0. 0.

Value of objective function at X : 500000.00000000

Number of function evaluations up to now : 1

Norm of projected gradient : 79255.833854676

Quacan converges in 1 iterations

Box Iteration : 2

First (ten) components of X :

17.190572198828 1.5983659750017 0. 0.

Value of objective function at X : 494.46527028179

Number of function evaluations up to now : 3

Norm of projected gradient : 778.04228064424

Quacan converges in 2 iterations

Box Iteration : 4

First (ten) components of X :

18.795762993735 1.8259188218780 0.24863515829422 0.

Value of objective function at X : 10.5290548583660

Number of function evaluations up to now : 5

Norm of projected gradient : 67.288572522569

Quacan converges in 1 iterations

.....
 Convergence of BOX: value of objective function

Total number of calls to Quacan : 11

Number of BOX - iterations: 6

Number of function evaluations: 7

First 10 components of solution :

19.114184471706 1.9765992309798 0.46423161706460 6.4693597977818D-

Value of objective function at solution : 0.

Norm of projected gradient : 0.

3.5 Conclusões

O pacote *Box-Quacan* foi usado para o problema de separação de dados via hiperplanos de separação.

Os valores dos parâmetros foram colocados à disposição do usuário para que ele possa utilizá-los com facilidade.

Analisando os resultados numéricos obtidos, observamos que o método de regiões de confiança para minimização em caixa é eficiente para esta aplicação.

O desempenho de *Box-Quacan* obtido na classificação das flores complementa o trabalho feito por *Fisher* porque conseguimos o desejado, uma separação forte.

Com esta experiência verificamos a potencialidade do uso deste pacote para os problemas de verificação da classificação de dados.

Em geral, a utilidade da existência do hiperplano de separação é aplicável a diversas áreas, verificando que as escolhas feitas têm fundamento matemático.

Bibliografia

- [1] Conn, A.R., Gould, N.I. and Toint, Ph L.
"Global convergence of a class of Trust region algorithms for optimization with simple bounds"
SIAM Journal on Numerical Analysis, Vol 25, No.2, April 1988
- [2] Conn, Gould e Toint
"Numerical Experiments with the Lancelot package for large - scale nonlinear optimization". Technical Report, Department of Mathematics. FUNDP, Namur, Bélgica, 1990.
- [3] Castro, Suzana Lima de Campos
"Algoritmo Genético aplicado á determinação ótima de parâmetros"
Tese de Mestrado, DMA - IMCECC - UNICAMP, 1994
- [4] Dennis, J.E., Schanabel. R.B.
"Numerical Methods for unconstrained Optimization and nonlinear equations"
New Jersey, Prentice-Hall, Englewood Cliffs, 1983.
- [5] Fletcher, R
" Practical Methods of Optimization "
Second edition. A.Wiley - Interscience Publication 1987.
- [6] Friedlander, A., Matínez, J. M. e Santos, S. A.
"A new trust region algorithm for bound constrained minimization".
Applied Mathematics e Optimization 30 (1994), 325-266.

- [7] Friedlander, A. e Martínez, J.M.
 " *On the maximization of a concave quadratic function with BOX constraints* "
 SIAM Journal on Optimization, 1994.
- [8] Kendall, Maurice
 " *Multivariate Analysis* "
 First Published 1975
- [9] Luenberger, D.G.
 " *Linear and nonlinear programming* "
 2 edition, California, Addison - Wesley publishing Co., 1984
- [10] Martínez, J.M.
 " *Discrimination by means of a trust region method* "
 Intern J. Computer Math, vol 55 ,pp 91-103,1995 OPA.
- [11] Santos, S.A.
 " *Regiões de confiança para programação matemática* "
 Tese de Doutorado, DMA - IMECC - UNICAMP, 1994
- [12] Santos, Sandra Augusta
 " *Métodos de Região de Confiança em conjuntos Arbitrários e Minimização em Bolas* ",
 Tese de Mestrado, DMA - IMECC - UNICAMP, 1991
- [13] Schlick, Tamar, Fogelson, Aaron
 " *Truncated Newton Minimization Package for large scale* "
 Transactions on mathematical Software March 1982, Vol. 18, No. 1.