REGIÕES DE CONFIANÇA EM PROGRAMAÇÃO MATEMÁTICA

Este exemplar corresponde a redação final da tese devidamente corrigida e defendida pela Sra. Sandra Augusta Santos e aprovada pela Comissão Julgadora.

Campinas, 01 de julho de 1994.

Prof. Dr. José Mario Martínez Perez

Tese apresentada ao Instituto de Matemática, Estatística e Ciência da Computação, UNICAMP, como requisito parcial para obtenção do título de Doutor em Ciências em Matemática Aplicada.

"Este texto que te dou não é para ser visto de perto: ganha sua secreta redondez antes invisível quando é visto de um avião em alto vôo. Então adivinha-se o jogo das ilhas e vêem-se canais e mares."

Clarice Lispector

AGRADECIMENTOS

Ao prof. Martínez, pela orientação exemplar.

Aos professores do IMECC, pelo apoio e incentivo.

À FAPESP, pelo suporte financeiro.

Ao Lúcio, pelo apoio, carinho e "paciência".

À Fátima, pelo cuidadoso trabalho datilográfico.

Ao Quintino, pelo auxílio na impressão das tabelas.

RESUMO

Neste trabalho são propostos três algoritmos de região de confiança para minimização com restrições: RCARB, RCMRI e BOX, desenvolvidos, respectivamente, para problemas com conjuntos arbitrários, restrições de igualdade e variáveis canalizadas. Para RCARB são provados resultados de convergência global (1ª ordem) e é analisada especialmente a minimização em bolas euclidianas, com a apresentação de um conjunto de experimentos numéricos. Para RCMRI são demonstrados resultados de convergência local e global (1ª e 2ª ordens) e é feita uma aplicação para minimização em esferas euclidianas objetivando resolver o Problema do Vetor Inicial em Codificação, sendo apresentados experimentos numéricos. Para o algoritmo BOX são provados resultados de convergência global e identificação das restrições ativas. É feita uma análise detalhada do algoritmo utilizado na resolução do subproblema (QUACAN), destinado a minimizar quadráticas com variáveis canalizadas. É apresentada ainda uma estratégia para minimizar funções convexas com restrições lineares, baseada na utilização de BOX.

ABSTRACT

Three trust region algorithms for constrained minimization are proposed in this work: RCARB, RCMRI and BOX, developed for dealing with arbitrary domains, equality constraints and simple bounds, respectively. Focusing on the algorithm RCARB, global convergence results (1^{st} order) are proved and it is analysed with details the minimization in Euclidean balls, validated by a set of numerical experiments. As regards the algorithm RCMRI, local and global convergence results are proved (1^{st} and 2^{nd} order). It is applied for minimization in Euclidean spheres, particularly intending to solve the Initial Vector Problem in codification theory. Numerical experiments are included. When it comes to the algorithm BOX, both global convergence and identification of the active constraints are proved. It is made a thorough analysis of the algorithm in charge for the resolution of the subproblem (QUACAN), implemented to minimize general quadratics with bound constrained variables and especially developed for large scale problems. Finally, it is presented a strategy for minimizing convex functions with linear constraints, based on using the algorithm BOX.

ÍNDICE

INTRODUÇÃO

Neste trabalho são apresentados e desenvolvidos métodos de região de confiança sob a óptica de programação matemática. São propostos três algoritmos principais, acompanhados dos respectivos resultados de convergência global: RCARB para minimização em conjuntos arbitrários, RCMRI para minimização com restrições de igualdade e BOX para minimização com variáveis canalizadas.

É feita uma análise específica do algoritmo RCARB aplicado a problemas em que o conjunto factível é uma bola euclidiana, para os quais são apresentados experimentos numéricos (problemas-testes clássicos, problemas de regularização e um problema de empacotamento).

Para o algoritmo RCMRI, além dos resultados de convergência global de primeira e segunda ordens, também são apresentados resultados de convergência local. É analisada a aplicação de RCMRI para problemas com restrições do tipo esferas euclidianas. O problema-teste considerado para tal aplicação é o Problema do Vetor Inicial em codificação.

Para o algoritmo BOX é provada a identificação das restrições ativas. Devido ao subproblema específico de BOX, a minimização de uma quadrática geral com restrições de canalização, este algoritmo está fortemente apoiado no algoritmo QUACAN, desenvolvido especialmente para problemas de grande porte. Neste sentido, são analisadas com detalhes as propriedades teóricas de QUACAN bem como são descritas as características importantes relativas à implementação computacional.

Baseada na utilização do algoritmo BOX, é proposta uma estratégia para minimizar funções convexas com restrições lineares, validada com experimentos numéricos para

problemas de estimativa de parâmetros.

Cada capítulo está estruturado, tanto quanto possível, de forma independente dos demais. Maiores detalhes relativos aos conteúdos dos capítulos podem ser obtidos nas respectivas introduções, onde é feito um panorama de cada tópico.

A organização deste trabalho é a seguinte: nos Capítulos 1, 2 e 3 apresentamos os métodos de região de confiança para minimização, respectivamente, em conjuntos arbitrários (RCARB), com restrições de igualdade (RCMRI) e com variáveis canalizadas (BOX). O Capítulo 4 contém a estratégia para minimizar funções convexas com restrições lineares. No Capítulo 5, a resolução do Problema do Vetor Inicial em codificação é tratada como uma aplicação do algoritmo para minimização em esferas. Finalmente, são apresentadas algumas conclusões e sugestões de trabalhos futuros bem como as referências utilizadas ao longo deste trabalho.

CAPÍTULO 1

MÉTODOS DE REGIÃO DE CONFIANÇA PARA MINIMIZAÇÃO EM CONJUNTOS ARBITRÁRIOS

1.1 INTRODUÇÃO

O problema tratado neste capítulo consiste em minimizar uma função diferenciável f em um conjunto fechado arbitrário $D \subset I\!R^n$. Introduzimos um método de região de confiança (Sorensen [1982], Moré e Sorensen [1983], Moré [1978, 1983], Fletcher [1987], Dennis e Schnabel [1983]) para resolver este problema. Contrariamente às abordagens existentes (Vardi [1985], Celis, Dennis e Tapia [1984], Powell e Yuan [1991]), nosso método não utiliza aproximações lineares para D. Desta forma, nossos subproblemas consistem na minimização de uma quadrática na interseção de D com uma bola (região de confiança definida pela norma euclidiana).

Em muitos casos, não existem algoritmos adequados para resolver estes subproblemas. Apesar disso, sabemos tratar com subproblemas em algumas situações relevantes: quando D é uma bola euclidiana, uma esfera, o complemento de uma bola euclidiana ou ainda alguma interseção destes conjuntos.

O caso em que D é uma bola euclidiana foi considerado por Heikenschloss [1990], no contexto de problemas de identificação de parâmetros, e por Vogel [1990] na resolução de equações integrais não lineares. Nesses dois trabalhos, a restrição $x \in D$ tem um papel

regularizador (Tikhonov e Arsenin [1977]), com algumas vantagens práticas. Heikenschloss e Vogel usam a estratégia de Gauss–Newton para definir o modelo quadrático a cada iteração. Neste sentido nossa abordagem é mais geral pois admitimos aproximações não convexas. Além disso, o caso em que uma restrição do tipo bola euclidiana é usada para regularizar problemas mal postos pode ser reduzido a um problema em que a região factível é uma esfera, pois a solução estará obviamente na fronteira. Naturalmente, neste caso o domínio não é convexo, como costuma acontecer na maioria dos problemas em que a região factível é definida por igualdades.

Uma outra situação em que o domínio não é convexo mas é possível resolver os subproblemas associados ao método de região de confiança ocorre quando o conjunto factível é o complemento de uma bola, ou o complemento da união finita de bolas disjuntas. Estes problemas aparecem quando se quer excluir vizinhanças de pontos indesejáveis de uma lista de possíveis soluções de um problema de otimização, visando encontrar minimizadores globais.

No caso convexo em que o domínio é um politopo também é possível resolver completamente o subproblema de região de confiança usando a norma infinito (Vavasis [1991]).

De acordo com a filosofia atual dos métodos de região de confiança, não é preciso resolver o subproblema exatamente para se obter convergência do algoritmo principal. Em nosso caso, exigimos que o subproblema produza um decréscimo suficiente no modelo, em termos da solução do que chamamos de subproblema *"fácil"*. A solução deste subproblema auxiliar desempenha o papel do clássico ponto de Cauchy, usado em muitos métodos de região de confiança (Conn, Gould e Toint [1988a, 1988b], Toint [1988], Burke, Moré e Toraldo [1990]).

Uma característica adicional do nosso método é a de que o primeiro raio de confiança utilizado a cada iteração é sempre maior que um parâmetro fixo $\Delta_{min} > 0$. Com isso, permitimos passos grandes quando se está longe de solução e eliminamos passos artificialmente pequenos, herdados de iterações anteriores.

Pelo que sabemos, esta é a primeira vez que um algoritmo de região de confiança é analisado no contexto de domínios arbitrários. A abordagem que mais se aproxima da nossa foi feita por Toint [1988], mas ele trabalha com domínios convexos. Conn, Gould and Toint [1988a, 1988b, 1989, 1990] popularizaram a idéia de que restrições de canalização podem ser incorporadas naturalmente em algoritmos de região de confiança usando-se a norma infinito. Eles desenvolveram uma implementação computacional para resolver problemas gerais de programação não linear (LANCELOT) usando Lagrangiano Aumentado, e portanto incorporando as restrições não lineares na função objetivo. Uma

outra maneira de se tratar problemas gerais de programação matemática é usar métodos de região de confiança para definir algoritmos baseados em programação quadrática sequencial e globalmente convergentes (Celis, Dennis e Tapia [1984], Powell e Yuan [1990], Vardi [1985], Williamson [1990], etc), mas neste caso as restrições são linearizadas.

Este capítulo está organizado da seguinte maneira: na Seção 1.2 apresentamos algumas definições, hipóteses e lemas básicos utilizados nos resultados de convergência. Introduzimos uma hipótese de regularidade mais fraca que a hipótese clássica utilizada em programação não linear (Luenberger [1984]). Na Seção 1.3 introduzimos o algoritmo principal (RCARB) e provamos que está bem definido. A Seção 1.4 contém a prova de que todo ponto de acumulação de RCARB é estacionário. Na Seção 1.5 consideramos D como sendo uma bola euclidiana e apresentamos os algoritmos utilizados para resolver os subproblemas neste caso. A implementação computacional e os experimentos numéricos são descritos, respectivamente, nas Seções 1.6 e 1.7. Algumas observações finais são feitas na Seção 1.8.

1.2 DEFINIÇÕES, HIPÓTESES E RESULTADOS BÁSICOS.

Vamos considerar o seguinte problema:

$$\begin{array}{ll} \min & f(x) \\ \text{s/a} & x \in D \end{array} \tag{1.2.1}$$

onde $D \subset I\!R^n$ é fechado, $f \in C^1(A)$ e A é um conjunto aberto que contém D. Usaremos a notação $g = \nabla f$.

DEFINIÇÃO 1.2.1. Dados $x \in D, b > 0$, dizemos que $\alpha : [0, b] \to I\!R^n$ é uma *curva factível partindo de* x se:

a) $\alpha(t) \in D$ para todo $t \in [0, b]$,

b) $\alpha \in C^1([0, b])$,

c) $\alpha(0) = x$, $\alpha'(0) \neq 0$.

TEOREMA 1.2.2. Se x_* é um minimizador local de (1.2.1) então para toda curva factível partindo de x_* temos $g(x_*)^T\alpha'(0) = (f \circ \alpha)'(0) \geq 0$.

Demonstração. Trivial pois 0 é um minimizador local de $f \circ \alpha : [0, b] \to \mathbb{R}$. □

O Teorema 1.2.2 motiva a seguinte definição.

DEFINIÇÃO 1.2.3. Dizemos que $x_* \in D$ é um *ponto estacionário* de (1.2.1) se para toda curva factível α partindo de x_* temos $g(x_*)^T\alpha'(0) \geq 0$.

DEFINIÇÃO 1.2.4. Dado $\alpha : [0, b] \to \mathbb{R}^n, \alpha \in C^1([0, b]), \alpha'(0) \neq 0$, para $\Delta \geq 0$ definimos

$$\tau(\alpha, \Delta) = \min\{t \in [0, b] \mid \|\alpha(t) - \alpha(0)\| = \Delta\} \tag{1.2.2}$$

onde $\|\cdot\|$ é uma norma arbitrária em $\mathbb{R}^n$.

O lema a seguir estabelece algumas propriedades para $\tau(\alpha, \Delta)$.

LEMA 1.2.5. Assumindo-se que $b > 0, \alpha_k : [0, b] \to \mathbb{R}^n, \alpha : [0, b] \to \mathbb{R}^n, \alpha_k, \alpha \in C^1([0, b])$ para todo $k \in \mathbb{N}$, $\alpha'(0) \neq 0$ e

$$\lim_{k\to\infty} \|\alpha_k' - \alpha'\|_\infty = 0 \tag{1.2.3}$$

onde $\|\beta\|_\infty = \max\{\|\beta(t)\| \mid t \in [0, b]\}$, então existem $c_1, c_2, \overline{\Delta} > 0, k_0 \in \mathbb{N}$ tais que $\tau(\alpha_k, \Delta)$ e $\tau(\alpha, \Delta)$ estão bem definidos e

$$\left.\begin{array}{rcccl} c_1\Delta & \leq & \tau(\alpha_k, \Delta) & \leq & c_2\Delta \\ c_1\Delta & \leq & \tau(\alpha, \Delta) & \leq & c_2\Delta \end{array}\right\} \tag{1.2.4}$$

para todo $\Delta \in [0, \overline{\Delta}], k \geq k_0$.

Demonstração. Como $\alpha'(0) \neq 0$, existe $i \in \{1, \ldots, n\}$ tal que $\alpha_i'(0) \neq 0$. Vamos supor, sem perda de generalidade, que $\alpha_i'(0) > 0$. Seja $b_1 \in (0, b]$ tal que $\alpha_i'(t) \geq 2\varepsilon > 0$ para

todo $t \in [0, b_1]$. Pela convergência uniforme de α'_k, existe $k_0 \in \mathbb{N}$ tal que $(\alpha'_k)_i(t) \geq \varepsilon$ para todo $t \in [0, b_1]$, $k \geq k_0$.

Temos agora, para $t \in [0, b_1]$ e para algum $c > 0$ (que depende apenas de $\|\cdot\|$), que

$$\begin{aligned} \|\alpha_k(t) - \alpha_k(0)\| &\geq c|(\alpha_k)_i(t) - (\alpha_k)_i(0)| = \\ c\left|\int_0^t (\alpha'_k)_i(w)dw\right| &= c\int_0^t (\alpha'_k)_i(w)dw \geq c\varepsilon t. \end{aligned} \tag{1.2.5}$$

Analogamente, para todo $t \in [0, b_1]$,

$$\|\alpha(t) - \alpha(0)\| \geq 2c\varepsilon t > c\varepsilon t. \tag{1.2.6}$$

Pela convergência uniforme de α'_k existe $h > 0$ tal que

$$\left.\begin{array}{l} \|\alpha'_k(t)\| \leq h \\ \text{e} \\ \|\alpha'(t)\| \leq h \end{array}\right\}$$

para todo $t \in [0, b_1]$, $k \geq k_0$.

Assim, se $t \in [0, b_1]$, $k \geq k_0$,

$$\|\alpha_k(t) - \alpha_k(0)\| = \|\int_0^t \alpha'_k(w)dw\| \leq \int_0^t \|\alpha'_k(w)\|dw \leq ht. \tag{1.2.7}$$

Analogamente, para todo $t \in [0, b_1]$,

$$\|\alpha(t) - \alpha(0)\| \leq ht. \tag{1.2.8}$$

Definimos $\overline{\Delta} = c\varepsilon b_1$. Por (1.2.5) e (1.2.6), $\|\alpha_k(b_1) - \alpha_k(0)\| \geq \overline{\Delta}$ e $\|\alpha(b_1) - \alpha(0)\| \geq \overline{\Delta}$. Então, para todo $\Delta \in [0, \overline{\Delta}]$ existem $t_k \in [0, b_1]$ tal que $\|\alpha_k(t_k) - \alpha_k(0)\| = \Delta$ e $t \in [0, b_1]$ tal que $\|\alpha(t) - \alpha(0)\| = \Delta$. Por continuidade, concluímos que $\tau(\alpha_k, \Delta) \in [0, b_1]$ e $\tau(\alpha, \Delta) \in [0, b_1]$ estão bem definidos para todo $\Delta \in [0, \overline{\Delta}]$. Portanto, (1.2.4) segue de

(1.2.5), (1.2.6), (1.2.7) e (1.2.8). □

Definimos a seguir a hipótese de regularidade utilizada neste capítulo, tendo em vista o grau de generalidade que estamos assumindo.

DEFINIÇÃO 1.2.6. Dizemos que $x \in D$ é *fracamente regular* se para toda curva factível $\alpha : [0, b] \to D$ partindo de x e para toda seqüência $\{x_k\} \subset D$ que converge para x, existem $b' \in (0, b)$ e $\alpha_k : [0, b'] \to D$ $(k \in I\!N)$ seqüência de curvas factíveis partindo de x_k tal que

$$\lim_{k \to \infty} \|\alpha_k' - \alpha'\|_\infty = 0 \tag{1.2.9}$$

onde $\|\beta\|_\infty = \max\{\|\beta(t)\| \mid t \in [0, b']\}$.

Apresentamos a seguir o resultado que relaciona a Definição 1.2.6 com a definição clássica de regularidade (Luenberger [1984], p. 314).

TEOREMA 1.2.7. Suponhamos que

$$D = \{x \in I\!R^n \mid h(x) = 0, c(x) \leq 0\}$$

onde $h : I\!R^n \to I\!R^m, c : I\!R^n \to I\!R^p, f, h, c \in C^1(I\!R^n)$. Se $\overline{x}$ é um ponto regular então $\overline{x}$ é fracamente regular.

Demonstração. Vamos assumir, sem perda de generalidade, que $p_1 < p$ é tal que

$$c_i(\overline{x}) = 0, \quad i = 1, \ldots, p_1 \tag{1.2.10}$$

e

$$c_i(\overline{x}) < 0, \quad i = p_1 + 1, \ldots, p. \tag{1.2.11}$$

Então, $\nabla h_1(\overline{x}), \ldots, \nabla h_m(\overline{x}), \nabla c_1(\overline{x}), \ldots, \nabla c_{p_1}(\overline{x})$ são linearmente independentes. Definimos

$$A(\overline{x}) = \begin{bmatrix} \nabla h_1(\overline{x})^T \\ \vdots \\ \nabla h_m(\overline{x})^T \\ \nabla c_1(\overline{x})^T \\ \vdots \\ \nabla c_{p_1}(\overline{x})^T \end{bmatrix}. \tag{1.2.12}$$

Sem perda de generalidade, vamos supor que as primeiras $m + p_1$ colunas de $A(\overline{x})$ são linearmente independentes. Assim, a matriz $B(\overline{x}) \in \mathbb{R}^{n\times n}$ é não singular, onde

$$B(\overline{x}) = \begin{bmatrix} & A(\overline{x}) & \\ 0 & & I_{n-(m+p_1)} \end{bmatrix} \tag{1.2.13}$$

Seja a função $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ dada por:

$$\Phi(x) = \begin{bmatrix} h_1(x) \\ \vdots \\ h_m(x) \\ c_1(x) \\ \vdots \\ c_{p_1}(x) \\ x_{m+p_1+1} \\ \vdots \\ x_n \end{bmatrix}. \tag{1.2.14}$$

Então $\Phi'(\overline{x}) = B(\overline{x})$ e

$$\overline{y} = \Phi(\overline{x}) = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \overline{x}_{m+p_1+1} \\ \vdots \\ \overline{x}_n \end{bmatrix}. \tag{1.2.15}$$

Pelo Teorema da Função Inversa existe $\mathcal{N}$ vizinhança aberta de $\overline{x}$ tal que $\Phi : \mathcal{N} \to \Phi(\mathcal{N})$ é uma bijeção e $\Phi^{-1} \in C^1$. Sem perda de generalidade assumimos que

$$c_i(x) < 0 \tag{1.2.16}$$

para todo $x \in \mathcal{N}, i = p_1 + 1, \ldots, p$. Então

$$c_i(\Phi^{-1}(y)) < 0 \tag{1.2.17}$$

para todo $y \in \Phi(\mathcal{N}), i = p_1 + 1, \ldots, p$.

Sejam $\alpha : [0, b] \to D$ uma curva factível partindo de $\overline{x}$ e $b_1 \in [0, b]$ tal que $\alpha(t) \in \mathcal{N}$ para todo $t \in [0, b_1]$.

Definimos $\overline{\alpha} : [0, b_1] \to \Phi(\mathcal{N})$ por

$$\overline{\alpha}(t) = \Phi(\alpha(t)) \tag{1.2.18}$$

para todo $t \in [0, b_1]$. Agora, como $h(\alpha(t)) = 0$ para todo $t \in [0, b_1]$, por (1.2.14) segue que

$$\overline{\alpha}_j(t) = 0 \tag{1.2.19}$$

para $j = 1, \ldots, m, \quad t \in [0, b_1]$. Além disso, como $c_i(\alpha(t)) \leq 0$ para todo $t \in [0, b_1]$, $i = 1, \ldots, p_1$, temos

$$\overline{\alpha}_j(t) \leq 0 \tag{1.2.20}$$

para $j = m + 1, \ldots, m + p_1, \quad t \in [0, b_1]$. Finalmente, como $c_i(\alpha(t)) < 0$ para todo $i = p_1 + 1, \ldots, p, \quad t \in [0, b_1]$, temos

$$c_i(\Phi^{-1}(\overline{\alpha}(t)) < 0 \tag{1.2.21}$$

para $i = p_1 + 1, \ldots, p, \quad t \in [0, b_1]$.

Definimos

$$\begin{aligned} D_1 = \{y \in \Phi(\mathcal{N}) \mid\ & y_j = 0, j = 1, \ldots, m,\ y_j \leq 0, j = m + 1, \ldots, m + p_1 \\ & \text{e}\ \ c_i(\Phi^{-1}(y)) < 0, i = p_1 + 1, \ldots, p\} \end{aligned} \tag{1.2.22}$$

Por (1.2.18) - (1.2.21), $\overline{\alpha} : [0, b_1] \to D_1$ é uma curva factível partindo de $\overline{y}$.

Tomamos $\{x_k\}$ uma seqüência em D tal que $\lim_{k\to\infty} x_k = \overline{x}$. Seja $k_0 \in I\!N$ tal que $x_k \in \mathcal{N}$ para todo $k \geq k_0$. Definimos, para todo $k \geq k_0$,

$$y_k = \Phi(x_k). \tag{1.2.23}$$

Definimos também, para $k \geq k_0$ e $t \in [0, b_1]$,

$$\overline{\alpha}_k(t) = y_k - \overline{y} + \overline{\alpha}(t). \tag{1.2.24}$$

Sejam $k_1 \in I\!N$ e $b_2 \in [0, b_1]$ tais que

$$c_i(\Phi^{-1}(\overline{\alpha}_k(t))) < 0 \tag{1.2.25}$$

para todo $t \in [0, b_2], i = p_1 + 1, \ldots, p, \quad k > k_1$.

Claramente, $\overline{\alpha}_k$ e $\overline{\alpha}'_k$ convergem uniformemente para $\overline{\alpha}$ e $\overline{\alpha}'$, respectivamente, para $t \in [0, b_2]$. Desta forma, definindo

$$\alpha_k(t) = \Phi^{-1}(\overline{\alpha}_k(t)) \tag{1.2.26}$$

para $t \in [0, b_2], k \geq k_1$, as seqüências α_k e α'_k convergem uniformemente para α e α', respectivamente.

Finalmente, é fácil ver que para $i = 1, \ldots, m$,

$$(\overline{\alpha}_k(t))_i = 0 \tag{1.2.27}$$

e como $\overline{y}_i = 0$ para $i = m + 1, \ldots, m + p_1$, então

$$(\overline{\alpha}_k(t))_i \leq 0 \tag{1.2.28}$$

para $i = m+1, \ldots, m+p_1$. Então, por (1.2.27) e (1.2.28), $\Phi^{-1}(\overline{\alpha}_k(t)) \in D$ para todo $k \geq k_1, t \in [0, b_2]$.

Logo, α_k é uma curva factível para todo $k \geq k_1$. Como a definição de α_k para $k < k_1$ é irrelevante, a prova está completa. □

A hipótese a seguir será usada no resultado de convergência global.

HIPÓTESE 1.2.8. Assumimos que existe uma função contínua $\varphi : [0, \infty) \to [0, \infty)$ tal que $\varphi(0) = 0$ e

$$|f(z) - f(x) - g(x)^T(z-x)| \leq \varphi(\|z-x\|)\|z-x\| \tag{1.2.29}$$

para todo $x, z \in D$.

A condição (1.2.29) garante uma certa uniformidade à função f, mais forte que a existência de plano tangente para todo ponto $x \in D$, assegurada por f ser de classe C^1. É mais exigente que pedir a diferenciabilidade de f segundo Fréchet pois a função φ é a mesma em todo o domínio, porém mais fraca que supor uma condição tipo Lipschitz para $g = \nabla f$, pois então teríamos $\varphi(\|z-x\|) \equiv L\|z-x\|$ onde L é a constante de Lipschitz (Ortega e Rheinboldt [1970], cap. 3). Cabe observar que a Hipótese 1.2.8 é satisfeita quando f é continuamente diferenciável em um conjunto convexo (Ortega e Rheinboldt [1970], p.74). Assim, quando D é uma bola euclidiana ou uma caixa, esta hipótese se verifica automaticamente.

1.3 O ALGORITMO RCARB

Nesta seção introduzimos o algoritmo RCARB, cuja sigla significa Regiões de Confiança em conjuntos Arbitrários. Denotamos por $\|\cdot\|$ uma norma arbitrária em $\mathbb{R}^n$ e a norma matricial correspondente. A cada iteração k do algoritmo a seguir, chamamos de Δ^k o primeiro raio de confiança testado e de Δ_k o raio efetivamente aceito.

Algoritmo 1.3.1. (RCARB)

Sejam $\tau_1, \tau_2, \theta, \Delta_{min}, M, \gamma$ tais que $0 < \tau_1 \leq \tau_2 < 1, \theta \in (0,1], \Delta_{min} > 0, M > 0$ e $\gamma \in (0,1]$. Sejam $x_0 \in D$ um ponto inicial factível, B_0 uma matriz simétrica tal que $\|B_0\| \leq M$ e um raio inicial $\Delta^0 \geq \Delta_{min}$. Dados $x_k \in D, B_k = B_k^T \in \mathbb{R}^{n \times n}$ tal que $\|B_k\| \leq M$ e $\Delta^k \geq \Delta_{min}$, os passos para se obter Δ_k e x_{k+1} são os seguintes:

Passo 0. Faça $\Delta \leftarrow \Delta^k$

Passo 1. Calcule $s_k^Q(\Delta)$ solução global de

$$\begin{array}{ll} \min & Q_k(s) \equiv \frac{1}{2} M \|s\|^2 + g_k^T s \\ \text{s/a} & x_k + s \in D \\ & \|s\| \leq \Delta \end{array} \tag{1.3.1}$$

onde $g_k = g(x_k)$. Se $Q_k(s_k^Q(\Delta)) = 0$, parar.

Passo 2. Calcule $\overline{s}_k(\Delta)$ tal que

$$\left.\begin{array}{c} \psi_k(\overline{s}_k(\Delta)) \leq \gamma Q_k(s_k^Q(\Delta)) \\ x_k + \overline{s}_k(\Delta) \in D \\ \|\overline{s}_k(\Delta)\| \leq \Delta \end{array}\right\} \tag{1.3.2}$$

onde ψ_k é definida por

$$\psi_k(s) \equiv \frac{1}{2} s^T B_k s + g_k^T s \tag{1.3.3}$$

para todo $s \in \mathbb{R}^n$. (Observe que existe $\overline{s}_k(\Delta)$ pois $s_k^Q(\Delta)$ é uma escolha possível).

Passo 3. Se

$$f(x_k + \overline{s}_k(\Delta)) \leq f(x_k) + \theta \psi_k(\overline{s}_k(\Delta)) \tag{1.3.4}$$

então $x_{k+1} = x_k + \overline{s}_k(\Delta)$
$\Delta_k = \Delta$
escolha $\Delta^{k+1} \geq \Delta_{min}$ e $B_{k+1} \in \mathbb{R}^{n \times n}$ simétrica tal que $\|B_{k+1}\| \leq M$

retorne

senão $\Delta \leftarrow \Delta_{novo}$, onde

$$\Delta_{novo} \in [\tau_1 \|\overline{s}_k(\Delta)\|, \tau_2 \Delta] \tag{1.3.5}$$

volte para o Passo 1.

TEOREMA 1.3.2. Se o Algoritmo 1.3.1 pára no Passo 1 ($Q_k(s_k^Q(\Delta)) = 0$), então x_k é um ponto estacionário do problema (1.2.1).

Demonstração. Se $Q_k(s_k^Q(\Delta)) = 0$ então $0 \in \mathbb{R}^n$ é uma solução de (1.3.1). Portanto 0 é um ponto estacionário de (1.3.1), seguindo facilmente que x_k é um ponto estacionário de (1.2.1). □

TEOREMA 1.3.3. Se x_k não é um ponto estacionário de (1.2.1) então x_{k+1} está bem definido pelo Algoritmo 1.3.1.

Demonstração. Como x_k não é um ponto estacionário, existe uma curva factível $\alpha : [0, b] \to D$ partindo de x_k tal que

$$(f \circ \alpha)'(0) = g(x_k)^T \alpha'(0) < 0. \tag{1.3.6}$$

Seja $\overline{\Delta} > 0$ tal que $\tau(\Delta) \equiv \tau(\alpha, \Delta)$ dado por (1.2.2) está bem definido e sejam $c_1, c_2 > 0$ tais que a segunda parte de (1.2.4) valha para todo $\Delta \in [0, \overline{\Delta}]$. Então, para $\Delta \in [0, \overline{\Delta}]$, temos

$$\begin{aligned} Q_k(s_k^Q(\Delta)) &\leq Q_k(\alpha(\tau(\Delta)) - x_k) \\ &= \frac{1}{2} M \|\alpha(\tau(\Delta)) - \alpha(0)\|^2 + g_k^T[\alpha(\tau(\Delta)) - \alpha(0)]. \end{aligned} \tag{1.3.7}$$

Então, por (1.2.4) e (1.3.7),

$$\frac{Q_k(s_k^Q(\Delta))}{\Delta} \leq c_2 \frac{Q_k(s_k^Q(\Delta))}{\tau(\Delta)}$$

$$\leq \quad c_2 \left[\frac{1}{2} M \frac{\|\alpha(\tau(\Delta)) - \alpha(0)\|^2}{\tau(\Delta)} + \frac{g_k^T [\alpha(\tau(\Delta)) - \alpha(0)]}{\tau(\Delta)} \right]. \qquad (1.3.8)$$

Mas, por (1.2.4),

$$\lim_{\Delta \to 0} \frac{\alpha(\tau(\Delta)) - \alpha(0)}{\tau(\Delta)} = \alpha'(0) \qquad (1.3.9)$$

e

$$\lim_{\Delta \to 0} \|\alpha(\tau(\Delta)) - \alpha(0)\| = 0. \qquad (1.3.10)$$

Portanto, por (1.3.8) - (1.3.10) e (1.3.6),

$$\limsup_{\Delta \to 0} \frac{Q_k(s_k^Q(\Delta))}{\Delta} \leq c_2 g_k^T \alpha'(0) < 0.$$

Assim, por (1.3.2),

$$\limsup_{\Delta \to 0} \frac{\psi_k(\bar{s}_k(\Delta))}{\Delta} \leq \gamma c_2 g_k^T \alpha'(0) < 0.$$

Logo, existe $\overline{\overline{\Delta}} > 0$ tal que para todo $\Delta \in (0, \overline{\overline{\Delta}}]$,

$$\frac{\psi_k(\bar{s}_k(\Delta))}{\Delta} \leq \frac{\gamma}{2} c_2 g_k^T \alpha'(0) = c_3 < 0. \qquad (1.3.11)$$

Definimos, para $\Delta > 0$,

$$\rho(\Delta) = \frac{f(x_k + \bar{s}_k(\Delta)) - f(x_k)}{\psi_k(\bar{s}_k(\Delta))}.$$

Então, se $\Delta \in (0, \overline{\overline{\Delta}}]$, por (1.3.11) temos que

$$
\begin{aligned}
|\rho(\Delta)-1| &= \left|\frac{f(x_k+\overline{s}_k(\Delta))-f(x_k)-\psi_k(\overline{s}_k(\Delta))}{\psi_k(\overline{s}_k(\Delta))}\right| \\
&\leq \left|\frac{f(x_k+\overline{s}_k(\Delta))-f(x_k)-g_k^T\overline{s}_k(\Delta)}{c_3\Delta}\right|+\left|\frac{\overline{s}_k(\Delta)^T B_k\overline{s}_k(\Delta)}{2c_3\Delta}\right| \\
&\leq \left|\frac{f(x_k+\overline{s}_k(\Delta))-f(x_k)-g_k^T\overline{s}_k(\Delta)}{c_3\|\overline{s}_k(\Delta)\|}\right|+\frac{M\Delta}{2|c_3|}.
\end{aligned}
$$

Desta forma, pela diferenciabilidade de f,

$$\lim_{\Delta\to 0}\rho(\Delta)=1. \tag{1.3.12}$$

Devido a (1.3.12), após um número finito de reduções (1.3.5), a condição (1.3.4) se verifica. Portanto x_{k+1} está bem definido. □

1.4 CONVERGÊNCIA GLOBAL

Nesta seção apresentamos a prova de que todo ponto limite de uma seqüência gerada pelo Algoritmo 1.3.1 é estacionário. Introduzimos a notação $\lim_{k\in K}$ para denotar $\lim_{\substack{k\in\mathbb{K}\\ k\to\infty}}$, que será usada ao longo deste trabalho.

TEOREMA 1.4.1. Sejam $\{x_k\}$ uma seqüência gerada pelo Algoritmo 1.3.1, $x_*\in D$ fracamente regular e $\lim_{k\in K_1}x_k=x_*$, onde $\mathbb{K}_1$ é um subconjunto infinito de $\mathbb{N}$. Então x_* é um ponto estacionário do problema (1.2.1).

Demonstração. Esta prova está dividida em duas partes, de acordo com o comportamento do raio de confiança aceito Δ_k. Em linhas gerais, as duas possibilidades consideradas são as seguintes: no primeiro caso $\{\Delta_k\}$ tem uma subseqüência que converge para zero, e no segundo caso $\{\Delta_k\}$ está suficientemente longe do zero, isto é,

$$\inf_{k\in\mathbb{K}_1}\Delta_k=0 \tag{1.4.1}$$

ou

$$\inf_{k \in I\!K_1} \Delta_k > 0. \tag{1.4.2}$$

Se x_* não é estacionário, a idéia no primeiro caso é mostrar que existe uma seqüência de raios de confiança rejeitados que vai para zero, o que implica numa contradição. De fato, vamos supor a validade de (1.4.1). Então existe $I\!K_2$, um subconjunto infinito de $I\!K_1$, tal que

$$\lim_{k \in I\!K_2} \Delta_k = 0. \tag{1.4.3}$$

Assim, existe $k_2 \in I\!N$ tal que $\Delta_k < \Delta_{min}$ para todo $k \geq k_2, k \in I\!K_2$. Mas, a cada iteração k tentamos inicialmente o raio $\Delta^k \geq \Delta_{min}$. Portanto, para todo $k \in I\!K_3 \equiv \{k \in I\!K_2 \mid k \geq k_2\}$ existem $\overline{\Delta}_k, s_k^Q(\overline{\Delta}_k), \overline{s}_k(\overline{\Delta}_k)$ tais que $s_k^Q(\overline{\Delta}_k)$ é uma solução de

$$\begin{array}{ll} \min & Q_k(s) \\ \text{s/a} & x_k + s \in D \\ & \|s\| \leq \overline{\Delta}_k \, , \end{array} \tag{1.4.4}$$

$$\psi_k(\overline{s}_k(\overline{\Delta}_k)) \leq \gamma Q_k(s_k^Q(\overline{\Delta}_k)), \tag{1.4.5}$$

$$f(x_k + \overline{s}_k(\overline{\Delta}_k)) > f(x_k) + \theta \psi_k(\overline{s}_k(\overline{\Delta}_k)) \tag{1.4.6}$$

e por (1.3.5),

$$\Delta_k \geq \tau_1 \|\overline{s}_k(\overline{\Delta}_k)\|. \tag{1.4.7}$$

Logo, por (1.4.3) e (1.4.7),

$$\lim_{k \in I\!K_3} \|\overline{s}_k(\overline{\Delta}_k)\| = 0. \tag{1.4.8}$$

Suponhamos que x_* não seja estacionário. Então existem $b > 0, \alpha : [0, b] \to D$ uma curva factível partindo de x_* tal que

$$g(x_*)^T \alpha'(0) < 0. \tag{1.4.9}$$

Como x_* é fracamente regular e $\lim_{k\in I\!K_3} x_k = x_*$, existem $b' \in (0,b], \alpha_k : [0,b'] \to D, k \in I\!K_3$, seqüência de curvas factíveis partindo de x_k, tal que

$$\lim_{k\in I\!K_3} \|\alpha_k' - \alpha'\|_\infty = 0. \tag{1.4.10}$$

Por (1.4.10) e pelo Lema 1.2.5 existem $k_0 \in I\!N, \overline{\Delta} > 0$ tais que $\tau(\alpha_k, \Delta)$ e $\tau(\alpha, \Delta)$ estão bem definidos para todo $k \in I\!K_4 \equiv \{k \in I\!K_3 \mid k > k_0\}, \Delta \in [0, \overline{\Delta}]$. Além disso, (1.2.4) vale para todo $k \in I\!K_4, \Delta \in [0, \overline{\Delta}]$. Seja $k_4 \in I\!N$ tal que

$$\|\overline{s}_k(\overline{\Delta}_k)\| \leq \overline{\Delta} \tag{1.4.11}$$

para todo $k \in I\!K_5 \equiv \{k \in I\!K_4 \mid k \geq k_4\}$. Definimos

$$t_k = \tau(\alpha_k, \|\overline{s}_k(\overline{\Delta}_k)\|) \tag{1.4.12}$$

para todo $k \in I\!K_5$. Então, t_k está bem definido e, pelo Lema 1.2.5,

$$c_1\|\overline{s}_k(\overline{\Delta}_k)\| \leq t_k \leq c_2\|\overline{s}_k(\overline{\Delta}_k)\| \tag{1.4.13}$$

para todo $k \in I\!K_5$.

Agora, por (1.3.2), (1.4.12) e (1.3.1),

$$\begin{aligned} \frac{\psi_k(\overline{s}_k(\overline{\Delta}_k))}{t_k} &\leq \frac{\gamma Q_k(s_k^Q(\overline{\Delta}_k))}{t_k} \leq \frac{\gamma Q_k(\alpha_k(t_k) - x_k)}{t_k} \\ &= \gamma\left\{g_k^T\frac{[\alpha_k(t_k) - \alpha_k(0)]}{t_k} + \frac{1}{2}M\frac{\|\alpha_k(t_k) - \alpha_k(0)\|^2}{t_k}\right\} \end{aligned} \tag{1.4.14}$$

para todo $k \in I\!K_5$.

Mas, por (1.4.10),

$$\begin{aligned}\left\|\frac{\alpha_k(t_k)-\alpha_k(0)}{t_k}-\alpha'(0)\right\| &= \left\|\frac{\int_0^{t_k}\alpha_k'(w)dw}{t_k}-\alpha'(0)\right\| \\ =\left\|\frac{\int_0^{t_k}[\alpha_k'(w)-\alpha(0)]dw}{t_k}\right\| &\leq \frac{1}{t_k}\int_0^{t_k}\|\alpha_k'(w)-\alpha'(0)\|dw \\ \leq \frac{1}{t_k}\int_0^{t_k}\|\alpha_k'-\alpha'\|_\infty dw &= \|\alpha_k'-\alpha'\|_\infty \xrightarrow[k\in I\!K_5]{} 0. \end{aligned} \tag{1.4.15}$$

Portanto,

$$\lim_{k\in I\!K_5}\frac{\alpha_k(t_k)-\alpha_k(0)}{t_k}=\alpha'(0), \tag{1.4.16}$$

$$\lim_{k\in I\!K_5}\frac{g_k^T[\alpha_k(t_k)-\alpha_k(0)]}{t_k}=g(x_*)^T\alpha'(0) \tag{1.4.17}$$

e

$$\lim_{k\in I\!K_5}\|\alpha_k(t_k)-\alpha_k(0)\|=0. \tag{1.4.18}$$

Por (1.4.14), (1.4.16) - (1.4.18) e (1.4.9) temos:

$$\limsup_{k\in I\!K_5}\frac{\psi_k(\overline{s}_k(\overline{\Delta}_k))}{t_k}\leq\gamma g(x_*)^T\alpha'(0)<0. \tag{1.4.19}$$

Logo, por (1.4.13) e (1.4.19),

$$\limsup_{k\in I\!K_5}\frac{\psi_k(\overline{s}_k(\overline{\Delta}))}{\|\overline{s}_k(\overline{\Delta}_k)\|}\leq c_2\gamma g(x_*)^T\alpha'(0)<0. \tag{1.4.20}$$

Assim, existe $k_5\in I\!N$ tal que para todo $k\in I\!K_6\equiv\{k\in I\!K_5\mid k\geq k_5\}$ temos

$$\frac{\psi_k(\overline{s}_k(\overline{\Delta}_k))}{\|\overline{s}_k(\overline{\Delta}_k)\|}\leq c_4\equiv\frac{c_2\gamma g(x_*)^T\alpha'(0)}{2}<0. \tag{1.4.21}$$

Definimos, para $k \in I\!K_6$,

$$\overline{\rho}_k = \frac{f(x_k + \overline{s}_k(\overline{\Delta}_k)) - f(x_k)}{\psi_k(\overline{s}_k(\overline{\Delta}_k))}. \tag{1.4.22}$$

Então, por (1.4.21),

$$|\overline{\rho}_k - 1| = \frac{|f(x_k + \overline{s}_k(\overline{\Delta}_k)) - f(x_k) - \psi_k(\overline{s}_k(\overline{\Delta}_k))|}{|\psi_k(\overline{s}_k(\overline{\Delta}_k))|} \tag{1.4.23}$$

$$\leq \frac{|f(x_k + \overline{s}_k(\Delta)) - f(x_k) - g(x_k)^T \overline{s}_k(\overline{\Delta}_k)|}{|c_4| \|\overline{s}_k(\overline{\Delta}_k)\|} + \frac{1}{2} \frac{M \|\overline{s}_k(\overline{\Delta}_k)\|}{|c_4|}. \tag{1.4.24}$$

Desta forma, por (1.2.29), (1.4.24) e (1.4.8),

$$\lim_{k \in I\!K_6} |\overline{\rho}_k - 1| \leq \lim_{k \in I\!K_6} \frac{\varphi(\|\overline{s}_k(\overline{\Delta}_k)\|)}{|c_4|} = 0. \tag{1.4.25}$$

Como (1.4.25) contradiz (1.4.6), concluímos a primeira parte da prova. Em outras palavras, x_k é estacionário quando vale (1.4.1).

Vamos agora analisar o caso em que os raios Δ_k estão suficientemente longe de zero, assumindo que (1.4.2) se verifica.

Como $\lim_{k \in I\!K_1} x_k = x_*$ e $\{f(x_k)\}$ é estritamente decrescente, temos que

$$\lim_{k \in I\!K_1} f(x_{k+1}) - f(x_k) = 0. \tag{1.4.26}$$

Mas, por (1.3.2) e (1.3.4),

$$f(x_{k+1}) \leq f(x_k) + \theta \psi_k(\overline{s}_k(\Delta_k)) \leq f(x_k) + \theta \gamma Q_k(s_k^Q(\Delta_k)). \tag{1.4.27}$$

Então,

$$\lim_{k\in I\!K_1} g_k^T s_k^Q(\Delta_k) + \frac{1}{2}M\|s_k^Q(\Delta_k)\|^2 = \lim_{k\in I\!K_1} Q_k(s_k^Q(\Delta_k)) = 0. \tag{1.4.28}$$

Definimos $\underline{\Delta} = \inf_{k\in I\!K_1} \Delta_k > 0$. Seja s_* uma solução de

$$\begin{array}{ll} \min & g(x_*)^T s + \frac{1}{2}M\|s\|^2 \\ \text{s/a} & x_* + s \in D \\ & \|s\| \leq \underline{\Delta}/2\ . \end{array} \tag{1.4.29}$$

Seja $k_6 \in I\!K_1$ tal que

$$\|x_k - x_*\| \leq \underline{\Delta}/2 \tag{1.4.30}$$

para todo $k \in I\!K_7 \equiv \{k \in I\!K_1 \mid k \geq k_6\}$.

Definimos, para $k \in I\!K_7$,

$$\widehat{s}_k = x_* + s_* - x_k. \tag{1.4.31}$$

Por (1.4.29) e (1.4.30) temos

$$\|\widehat{s}_k\| \leq \underline{\Delta} \leq \Delta_k \tag{1.4.32}$$

para todo $k \in I\!K_7$. Além disso,

$$x_k + \widehat{s}_k = x_* + s_* \in D. \tag{1.4.33}$$

Por (1.4.32), (1.4.33) e (1.3.1) temos que

$$Q_k(s_k^Q(\Delta_k)) \leq Q_k(\widehat{s}_k) \tag{1.4.34}$$

para todo $k \in I\!K_7$. Então, por (1.4.28), (1.4.31) e (1.4.34),

$$g(x_*)^T s_* + \frac{1}{2} M \|s_*\|^2 = \lim_{k \in I\!K_7} g_k^T \widehat{s}_k(\Delta_k)) + \frac{1}{2} M \|\widehat{s}_k\|^2 \geq \lim_{k \in I\!K_7} Q_k(s_k^Q(\Delta_k)) = 0. \quad (1.4.35)$$

Portanto, 0 é um minimizador de (1.4.29) e pelo Teorema 1.3.2, x_* é estacionário, o que completa a prova. □

1.5 REGIÕES DE CONFIANÇA EM BOLAS EUCLIDIANAS

Nesta seção estudamos o caso em que

$$D = \{x \in I\!R^n \mid \|x\| \leq R\} \quad (1.5.1)$$

onde $R > 0$ e $\|\cdot\|$ é a norma euclidiana em $I\!R^n$.

Assim, o subproblema (1.3.1) fica

$$\begin{array}{ll} \min & Q_k(s) \\ \text{s/a} & \|s + x_k\| \leq R \\ & \|s\| \leq \Delta; . \end{array} \quad (1.5.2)$$

Os três problemas a seguir estão relacionados com (1.5.2):

$$\begin{array}{ll} \min & Q_k(s) \\ \text{s/a} & \|s + x_k\| \leq R\,, \end{array} \quad (1.5.3)$$

$$\begin{array}{ll} \min & Q_k(s) \\ \text{s/a} & \|s\| \leq \Delta \end{array} \quad (1.5.4)$$

e

$$\begin{array}{ll} \min & Q_k(s) \\ \text{s/a} & \|s + x_k\| = R \\ & \|s\| = \Delta\,. \end{array} \quad (1.5.5)$$

Chamemos de s_I o minimizador irrestrito de $Q_k(s)$. Então,

$$s_I = -g_k/M. \tag{1.5.6}$$

Claramente, como $Q_k(s)$ é estritamente convexa e a região factível de (1.5.2) é convexa e compacta, o problema (1.5.2) tem solução única. Além disso, qualquer solução local de (1.5.2) é necessariamente global. Portanto, se a solução de (1.5.3) (ou (1.5.4)) é factível para (1.5.2), então ela é solução de (1.5.2). Se tanto a solução de (1.5.3) quanto a de (1.5.4) não forem factíveis para (1.5.2), então as duas restrições de (1.5.2) são ativas na solução global de (1.5.2), ou seja, esta solução é o minimizador de (1.5.5). Desta forma, o algoritmo a seguir pode ser usado para calcular $s_k^Q(\Delta)$.

ALGORITMO 1.5.1.

Passo 1. Se $\|x_k\| + \Delta \leq R$,
defina $s_k^Q(\Delta) \equiv$ minimizador global de (1.5.4) e pare.
Se $\|x_k\| + R \leq \Delta$,
defina $s_k^Q(\Delta) \equiv$ minimizador global de (1.5.3) e pare.

Passo 2. Calcule $s^1 =$ minimizador global de (1.5.3).
Se

$$\|s^1\| \leq \Delta \tag{1.5.7}$$

faça $s_k^Q(\Delta) = s^1$ e pare.

Passo 3. Calcule $s^2 =$ minimizador global de (1.5.4).
Se

$$\|s^2 + x_k\| \leq R \tag{1.5.8}$$

faça $s_k^Q(\Delta) = s^2$ e pare.

Passo 4. Defina $s_k^Q(\Delta) \equiv$ minimizador global de (1.5.5) e pare.

Os resultados que se seguem mostram que o custo computacional do Algoritmo 1.5.1 é bastante pequeno. No Teorema 1.5.2 provamos que as soluções globais de (1.5.3) e (1.5.4) são simplesmente as projeções do minimizador irrestrito s_I nas respectivas regiões factíveis, e neste caso o cálculo destas projeções é trivial.

TEOREMA 1.5.2. Se $\|x_k + s_I\| \leq R$ então s_I é a solução global de (1.5.3). Caso contrário, a solução global de (1.5.3) é dada por:

$$s^1 = -x_k + R\frac{(s_I + x_k)}{\|s_I + x_k\|}. \tag{1.5.9}$$

Analogamente, o minimizador global de (1.5.4) é s_I se $\|s_I\| \leq \Delta$, ou é dado por

$$s^2 = \frac{\Delta}{\|s_I\|} s_I \tag{1.5.10}$$

se $\|s_I\| > \Delta$.

Demonstração. Temos

$$\begin{aligned} \|s - s_I\|^2 &= \|s\|^2 - 2s_I^T s + \|s_I\|^2 \\ &= \|s\|^2 + 2\frac{g_k^T s}{M} + \|s_I\|^2 = \frac{2}{M} Q_k(s) + \|s_I\|^2. \end{aligned} \tag{1.5.11}$$

Portanto, (1.5.3) é equivalente a

$$\begin{array}{ll} \min & \|s - s_I\|^2 \\ \text{s/a} & \|s + x_k\| \leq R\,. \end{array} \tag{1.5.12}$$

A solução de (1.5.12) é a projeção ortogonal de s_I na bola $\|s + x_k\| \leq R$. Logo, a primeira parte do teorema segue facilmente. A expressão (1.5.10) é obtida de maneira

análoga. □

No lema a seguir começamos a considerar o problema (1.5.5), que consiste na minimização de $Q_k(s)$ na interseção de duas esferas. Provamos que os pontos factíveis deste problema pertencem a um determinado hiperplano.

LEMA 1.5.3. Sejam s um ponto factível de (1.5.5) e $x_k \neq 0$. Então

$$x_k^T(x_k + s - y_c) = 0 \tag{1.5.13}$$

onde

$$y_c = \left(\frac{R^2 - \Delta^2}{2\|x_k\|^2} + \frac{1}{2}\right) x_k. \tag{1.5.14}$$

Demonstração. Como $\|s + x_k\|^2 = R^2$, temos que

$$\|x_k\|^2 + 2s^T x_k + \|s\|^2 = R^2. \tag{1.5.15}$$

Mas $\|s\|^2 = \Delta^2$, e então de (1.5.15) segue que

$$s^T x_k = \frac{R^2 - \Delta^2 - \|x_k\|^2}{2}. \tag{1.5.16}$$

Acrescentando $\|x_k\|^2$ nos dois membros de (1.5.16),

$$\begin{aligned} x_k^T(s + x_k) &= \frac{R^2 - \Delta^2 + \|x_k\|^2}{2} \\ &= \left(\frac{R^2 - \Delta^2}{2\|x_k\|^2} + \frac{1}{2}\right)\|x_k\|^2 = y_c^T x_k. \end{aligned} \tag{1.5.17}$$

Logo, (1.5.13) segue de (1.5.17). □

Completamos a caracterização do conjunto factível de (1.5.5) no lema a seguir, onde provamos que este conjunto é uma esfera contida no hiperplano definido no Lema 1.5.3.

LEMA 1.5.4. Seja $x_k \neq 0$. Definimos

$$\mathcal{H} = \{s \in I\!R^n \mid x_k^T(x_k + s - y_c) = 0\} \tag{1.5.18}$$

onde y_c é dado por (1.5.14) e

$$\mathcal{F} = \{s \in I\!R^n \mid \|s + x_k\| = R, \|s\| = \Delta\}. \tag{1.5.19}$$

Então

$$\mathcal{F} = \{s \in \mathcal{H} \mid \|x_k + s - y_c\| = \beta\} \tag{1.5.20}$$

onde

$$\beta = \sqrt{R^2 - \left(\frac{R^2 - \Delta^2}{2\|x_k\|} + \frac{\|x_k\|}{2}\right)^2}. \tag{1.5.21}$$

Demonstração. No Lema 1.5.3 provamos que $\mathcal{F}$, o conjunto dos pontos factíveis do problema (1.5.5), está contido no hiperplano $\mathcal{H}$. Tomemos $s \in \mathcal{F}$. Por (1.5.18), x_k é ortogonal a $x_k + s - y_c$. Então, pelo teorema de Pitágoras e por (1.5.14), temos

$$\begin{aligned} \|x_k + s - y_c\|^2 &= \|x_k + s\|^2 - \|y_c\|^2 = R^2 - \|y_c\|^2 \\ &= R^2 - \left(\frac{R^2 - \Delta^2}{2\|x_k\|} + \frac{\|x_k\|}{2}\right)^2 = \beta^2. \end{aligned}$$

Portanto,

$$\mathcal{F} \subset \{s \in \mathcal{H} \mid \|x_k + s - y_c\| = \beta\}. \tag{1.5.22}$$

Reciprocamente, vamos supor que $s \in \mathcal{H}$ e $\|x_k + s - y_c\| = \beta$. Então, por (1.5.13), (1.5.14) e (1.5.21),

$$\|x_k + s\|^2 = \|x_k + s - y_c\|^2 + \|y_c\|^2 = \beta^2 + \|y_c\|^2 = R^2. \tag{1.5.23}$$

Além disso,

$$\begin{aligned}
\|s\|^2 &= \|x_k + s - y_c\|^2 + \|y_c - x_k\|^2 = \beta^2 + \left\| \left(\frac{R^2 - \Delta^2}{2\|x_k\|^2} - \frac{1}{2} \right) x_k \right\|^2 = \\
&= R^2 - \left(\frac{R^2 - \Delta^2}{2\|x_k\|} + \frac{\|x_k\|}{2} \right)^2 + \left(\frac{R^2 - \Delta^2}{2\|x_k\|} - \frac{\|x_k\|}{2} \right)^2 = \\
&= R^2 - 4\frac{(R^2 - \Delta^2)}{2\|x_k\|} \cdot \frac{\|x_k\|}{2} = \Delta^2.
\end{aligned} \tag{1.5.24}$$

Logo, (1.5.20) segue de (1.5.22) – (1.5.24). □

O Lema 1.5.5 a seguir estabelece que o interior da bola de dimensão $n-1$ cuja fronteira é o conjunto factível de (1.5.5) contém apenas pontos interiores da região factível de (1.5.2).

LEMA 1.5.5. Sejam $x_k \neq 0, s \in \mathcal{H}$ e $\|x_k + s - y_c\| < \beta$. Então $\|x_k + s\| < R$ e $\|s\| < \Delta$.

Demonstração. Trivial, usando os mesmos cálculos utilizados para obter (1.5.23) e (1.5.24). □

No próximo teorema completamos a caracterização do minimizador de (1.5.5).

TEOREMA 1.5.6. Se a direção $s_k^Q(\Delta)$ é calculada no Passo 4 do Algoritmo 1.5.1, então

$$s_k^Q(\Delta) = y_c - x_k + \beta \frac{(x_k + s_P - y_c)}{\|x_k + s_P - y_c\|} \tag{1.5.25}$$

onde

$$s_P = s_I + \frac{x_k^T y_c - \|x_k\|^2 - x_k^T s_I}{\|x_k\|^2} x_k \tag{1.5.26}$$

e s_I, y_c, β são dados, respectivamente, por (1.5.6), (1.5.14) e (1.5.21).

Demonstração. Claramente, se o Algoritmo 1.5.1 não pára nos três primeiros passos, a solução global de (1.5.12) pertence a $\mathcal{F}$. Vamos considerar o seguinte problema:

$$\begin{array}{ll} \min & Q_k(s) \\ \text{s/a} & s \in \mathcal{H} \\ & \|x_k + s - y_c\| \leq \beta \,, \end{array} \tag{1.5.27}$$

onde $\mathcal{H}, y_c$ e β são definidos por (1.5.18), (1.5.14) e (1.5.21), respectivamente.

Pela análise anterior e pelo Lema 1.5.4, o minimizador global de (1.5.2) também é minimizador global de:

$$\begin{array}{ll} \min & Q_k(s) \\ \text{s/a} & s \in \mathcal{H} \\ & \|x_k + s - y_c\| = \beta \,. \end{array} \tag{1.5.28}$$

Mas, pelo Lema 1.5.5, os pontos factíveis de (1.5.27) também são factíveis para (1.5.2). Portanto, um mininizador global de (1.5.28) também tem que ser um minimizador global de (1.5.27). Logo, o minimizador global de (1.5.2) é o (único) minimizador global de (1.5.27) e sabemos que a restrição de desigualdade é ativa neste minimizador. Agora, por (1.5.11), (1.5.27) é equivalente a

$$\begin{array}{ll} \min & \|s - s_I\|^2 \\ \text{s/a} & s \in \mathcal{H} \\ & \|x_k + s - y_c\| \leq \beta \,. \end{array} \tag{1.5.29}$$

Além disso, pelo teorema de Pitágoras, (1.5.29) é equivalente a

$$\begin{array}{ll} \min & \|s - s_P\|^2 \\ \text{s/a} & \|x_k + s - y_c\| \leq \beta \,. \end{array} \tag{1.5.30}$$

onde s_P é a projeção de s_I em $\mathcal{H}$, dada por (1.5.26).

Pelas considerações acima, $\|x_k + s_P - y_c\| \geq \beta$, e então a solução de (1.5.30) é dada por

$$s_k^Q(\Delta) = y_c - x_k + \beta \frac{(x_k + s_P - y_c)}{\|x_k + s_P - y_c\|},$$

o que completa a demonstração. □

Os Teoremas 1.5.2 e 1.5.6 mostram que o custo computacional de Algoritmo 1.5.1 é bastante pequeno. De fato, por (1.5.25) vemos que $s_k^Q(\Delta)$ pode ser calculada no Passo 4 do Algoritmo 1.5.1 usando $8n+O(1)$ flops e duas raízes quadradas. Além disso, por (1.5.9) e (1.5.10), vemos que o custo computacional de se calcular os minimizadores globais de (1.5.3) e (1.5.4) também é pequeno.

Para calcularmos $\overline{s}_k(\Delta)$ no Passo 2 do Algoritmo 1.3.1, vamos considerar o seguinte problema:

$$\begin{array}{ll} \min & \psi_k(s) \\ \text{s/a} & \|s + x_k\| \leq R \\ & \|s\| \leq \Delta \, . \end{array} \tag{1.5.31}$$

Claramente, uma solução de (1.5.31) satisfaz (1.3.2). As condições (1.3.2), (1.3.4) e o resultado de convergência estabelecido pelo Teorema 1.4.1, no entanto, nos mostram que não é preciso resolver (1.5.31) de maneira precisa para se obter a convergência do algoritmo. De qualquer forma, se B_k é uma boa aproximação para a matriz Hessiana, uma solução exata para (1.5.31) certamente melhora o desempenho do Algoritmo 1.3.1. Contrariamente a (1.5.2), (1.5.31) não é um problema convexo pois B_k não é necessariamente semi-definida positiva. Assim, (1.5.31) é um problema bem mais difícil de resolver que (1.5.2). Analogamente ao que fizemos para (1.5.2), vamos considerar os seguintes subproblemas relacionados com (1.5.31):

$$\begin{array}{ll} \min & \psi_k(s) \\ \text{s/a} & \|s + x_k\| \leq R \, , \end{array} \tag{1.5.32}$$

$$\begin{array}{ll} \min & \psi_k(s) \\ \text{s/a} & \|s\| \leq \Delta \end{array} \tag{1.5.33}$$

e

$$\begin{array}{ll} \min & \psi_k(s) \\ \text{s/a} & \|s + x_k\| = R \\ & \|s\| = \Delta \, . \end{array} \tag{1.5.34}$$

Tomando-se (1.5.32) - (1.5.34), a solução $\tilde{s}_k(\Delta)$ de (1.5.31) pode ser calculada usando-se o algoritmo a seguir, que essencialmente computa soluções globais e locais de (1.5.32), (1.5.33) e a solução global de (1.5.34). Celis, Dennis e Tapia [1984], Zhang [1988] e Williamson [1990] estudaram um problema semelhante a (1.5.31), onde ao invés da restrição $\|s + x_k\| \leq R$, uma restrição quadrática convexa geral é considerada. Zhang introduz um algoritmo para resolver este problema, mas trabalha com algumas hipóteses restritivas. A técnica introduzida aqui para resolver (1.5.31) fundamenta-se principalmente na caracterização de minimizadores locais-não globais de quadráticas em esferas, apresentada por Martínez [1994] e acreditamos ser inédita.

Algoritmo 1.5.7

Passo 1. Se $\|x_k\| + \Delta \leq R$,
defina $\tilde{s}_k(\Delta) \equiv$ minimizador global de (1.5.33) e pare.
Se $\|x_k\| + R \leq \Delta$,
defina $\tilde{s}_k(\Delta) \equiv$ minimizador global de (1.5.32) e pare.

Passo 2. Defina s_G^1 a solução global de (1.5.32) que minimiza $\|s\|$.
Se $\|s_G^1\| \leq \Delta$, defina $\tilde{s}_k(\Delta) = s_G^1$ e pare.

Passo 3. Defina s_G^2 a solução global de (1.5.33) que minimiza $\|s + x_k\|$.
Se $\|s_G^2 + x_k\| \leq R$, defina $\tilde{s}_k(\Delta) = s_G^2$ e pare.

Passo 4. Faça $\mathcal{C} = \phi$.
Se existe s_L^1 solução local-não global de (1.5.32) e $\|s_L^1\| \leq \Delta$, faça
$\mathcal{C} = \{s_L^1\}$.

Passo 5. Se existe s_L^2 solução local-não global de (1.5.33) e $\|s_L^2 + x_k\| \leq R$, faça
$\mathcal{C} \leftarrow \mathcal{C} \cup \{s_L^2\}$.

Passo 6. Calcule s^3 solução global de (1.5.34).
Faça $\mathcal{C} \leftarrow \mathcal{C} \cup \{s^3\}$.

Passo 7. Defina

$$\tilde{s}_k(\Delta) = \arg\min\{\psi_k(s) \mid s \in \mathcal{C}\}. \tag{1.5.35}$$

Encerramos esta seção mostrando que o Algoritmo 1.5.7 calcula um minimizador global de (1.5.31).

TEOREMA 1.5.8. O Algoritmo 1.5.7 está bem definido e calcula um minimizador global $\tilde{s}_k(\Delta)$ de (1.5.31).

Demonstração. Definimos

$$B_R = \{s \in \mathbb{R}^n \mid \|s + x_k\| \leq R\} \tag{1.5.36}$$

e

$$B_\Delta = \{s \in \mathbb{R}^n \mid \|s\| \leq \Delta\}. \tag{1.5.37}$$

Se $\|x_k\| + \Delta \leq R$ então $B_\Delta \subset B_R$ e portanto, qualquer minimizador global de (1.5.33) é minimizador global de (1.5.31). Se $\|x_k\| + R \leq \Delta$ então $B_R \subset B_\Delta$ e portanto resolver (1.5.32) é equivalente a resolver (1.5.31). Logo, se a direção $\tilde{s}_k(\Delta)$ é calculada no Passo 1, então $\tilde{s}_k(\Delta)$ é solução de (1.5.31).

O conjunto G_R das soluções globais de (1.5.32) é $G_R = V_R \bigcap B_R$, onde V_R é uma variedade afim (ver Gay [1981], Moré e Sorensen [1983], Sorensen [1982]). Frequentemente, V_R é um único ponto, mas deve-se considerar o caso geral. O caso em que V_R é uma variedade afim não-trivial é conhecido na literatura de região de confiança como *"hard case"*.

O problema

$$\begin{array}{ll} \min & \|s\| \\ \text{s/a} & s \in G_R \end{array} \tag{1.5.38}$$

considerado no Passo 2 do Algoritmo 1.5.7 é resolvido facilmente. Basta encontrar o ponto em V_R de norma mínima e então projetá-lo em B_R. Claramente, se existe $s \in G_R$ tal que $\|s\| \leq \Delta$, a solução de (1.5.38) tem que satisfazer estas mesmas condições. Este ponto é, portanto, uma solução global de (1.5.31) neste caso. Por argumentos análogos mostra-se que $\tilde{s}_k(\Delta)$ é uma solução global de (1.5.33) se for calculada pelo Passo 3.

Se as soluções globais tanto de (1.5.32) quanto de (1.5.33) forem infactíveis para (1.5.31) então a solução global de (1.5.31) deve ser uma solução local não-global de (1.5.32) ou de (1.5.33) ou ainda, uma solução global de (1.5.34). Soluções locais não-globais são caracterizadas por Martínez [1994], que mostra a existência de no máximo uma solução local-não global e apresenta um algoritmo para calculá-la ou detectar a inexistência de solução. Se é calculada uma solução local-não global de (1.5.32) (respectivamente (1.5.33)) que é factível para (1.5.31), esta solução é armazenada no Passo 4 (resp. 5), sendo uma candidata a solução global de (1.5.31).

A última possibilidade é a de que as duas restrições de (1.5.31) sejam ativas na solução. Neste caso, encontramos a solução global de (1.5.34). Agora, pelos Lemas 1.5.3 – 1.5.5, o problema (1.5.34) é equivalente a

$$\begin{array}{ll} \min & \psi_k(s) \\ \text{s/a} & s \in \mathcal{H} \\ & \|x_k + s - y_c\| = \beta \, . \end{array} \tag{1.5.39}$$

Após uma mudança de variáveis conveniente, (1.5.39) se reduz à minimização de uma quadrática (geral) numa esfera. Portanto, podemos encontrar um minimizador global de (1.5.39) através da caracterização de Gay [1981], Sorensen [1982] e Moré e Sorensen [1983], bastando armazenar uma solução global de (1.5.39).

No Passo 7 do algoritmo comparamos os valores de $\psi_k(s)$ nos pontos candidatos (no máximo três). Claramente ao menor destes três valores corresponde um minimizador global de (1.5.31). □

OBSERVAÇÃO. Usando-se a caracterização para minimizadores globais de funções quadráticas em bolas euclidianas dada por Gay [1981] e Sorensen [1982], é possível definir vários algoritmos para se executar os Passos 1, 2 e 3 do Algoritmo 1.5.7. A maneira mais direta é usar a decomposição espectral de B_k, seguida pela resolução de uma única

equação não-linear. É necessário definir tolerâncias adequadas para se detectar *"hard-case"* e para se decidir pela parada no processo iterativo não-linear. Moré e Sorensen [1983], no entanto, desenvolveram um algoritmo em que, ao invés da decomposição espectral, são necessárias apenas algumas fatorações de Cholesky de modificações diagonais de B_k. A decomposição espectral de B_k é essencialmente tudo o que se precisa para o cálculo de minimizadores locais não globais (Martínez [1994]), embora também seja possível desenvolver algoritmos alternativos baseados em fatorações de Cholesky.

1.6 IMPLEMENTAÇÃO COMPUTACIONAL

Escrevemos um programa baseado no Algoritmo 1.3.1 para resolver o problema (1.2.1) quando D é dado por (1.5.1) e $f \in C^2$. Vamos detalhar algumas características desta implementação.

(i) **O raio inicial Δ^k.** Optamos por usar Δ_{min} grande de modo que o algoritmo possa dar passos grandes quando longe da solução. Assim, tomamos

$$\Delta^k \equiv \Delta_{min} = 2R \tag{1.6.1}$$

como raio inicial para todo $k \in I\!N$. Também efetuamos testes com as escolhas $\Delta_{min} = 10^{-4}$ e $\Delta^k = \max\{\Delta_{min}, 4\Delta_{k-1}\}$. Com estas escolhas, economizamos algumas avaliações de função em casos críticos. Apesar disto, detectamos que a probabilidade de se obter minimizadores globais aumenta quando usamos (1.6.1).

(ii) **Os parâmetros M e γ.** Usamos M grande ($M \geq \max\{\|\nabla^2 f(x)\| \mid x \in D, 10^4\}$ e γ pequeno ($\gamma = 0.1$), de forma que a condição (1.3.2) fosse facilmente satisfeita quando $\overline{s}_k(\Delta)$ é uma solução aproximada para (1.5.31).

(iii) **O parâmetro de decréscimo suficiente θ.** Usamos $\theta = 10^{-4}$.

(iv) **A escolha de Δ_{novo}** (ver 1.3.5). Escolhemos $\Delta_{novo} = \dfrac{\|\overline{s}_k(\Delta)\|}{2}$ em (1.3.5), de maneira que (1.3.5) vale se $\tau_1 = \tau_2 = 1/2$.

(v) **Cálculo de $s_k^Q(\Delta)$.** O subproblema (1.3.1) é resolvido pelo Algoritmo 1.5.1, usando-se as fórmulas (1.5.6), (1.5.9), (1.5.10), (1.5.26) e (1.5.25).

(vi) **Cálculo de $\overline{s}_k(\Delta)$.** Conforme já mencionamos, decidimos obter $\overline{s}_k(\Delta)$ como uma solução aproximada de (1.5.31). Para tanto, minimizadores locais e globais de (1.5.32), (1.5.33) e (1.5.34) devem ser considerados, como vimos na Seção 1.5. Pelo Algoritmo 1.5.7 e pelo Teorema 1.5.8, estes problemas têm a forma

$$\begin{array}{ll} \min & q(s) \\ \text{s/a} & \|s\| \leq \Delta \end{array} \tag{1.6.2}$$

onde q é uma quadrática. Obtemos minimizadores globais de (1.6.2) usando a decomposição espectral de $\nabla^2 q$, a caracterização de Gay–Moré–Sorensen para minimizadores globais e o procedimento iterativo de Hebden e Moré (Hebden [1973], Moré [1978, 1983]). A obtenção de minimizadores locais-não globais de (1.6.2) é feita usando-se o algoritmo de Martínez [1994]. Nestes algoritmos utilizamos tolerâncias pequenas como critérios de parada, de forma que são obtidas soluções praticamente exatas de (1.5.32), (1.5.33) e (1.5.34). Vamos chamar de $\widehat{s}_k(\Delta)$ a solução aproximada de (1.5.31) obtida por este procedimento. Se, devido a erros das aproximações, $\widehat{s}_k(\Delta)$ é ligeiramente infactível para (1.5.31), projetamos $\widehat{s}_k(\Delta)$ na região factível usando uma variação do Algoritmo 1.5.1 que consiste em usar $\|s-\widehat{s}_k(\Delta)\|^2$ como função objetivo. Denominamos esta projeção de $\overline{s}_k(\Delta)$. Finalmente, testamos a salvaguarda $\psi_k(\overline{s}_k(\Delta)) \leq \gamma Q_k(s_k^Q(\Delta))$. Se esta desigualdade não é satisfeita, substituímos $\overline{s}_k(\Delta)$. Isto nunca aconteceu nos experimentos numéricos.

(vii) **Escolha de B_k.** Testamos diferentes escolhas para B_k:

(a) (NULL): $B_k = 0$ para todo $k \in I\!N$.

(b) (HESS): B_k é uma aproximação para $\nabla^2 f(x_k)$ usando diferenças finitas, conforme sugerido em Dennis e Schnabel [1983], pp. 104-106.

(c) (BFGS0): $B_0 = 0$ e B_{k+1} é gerada pela fórmula BFGS para todo $k \in I\!N$ (Dennis e Schnabel [1983], pp. 198-203). Assim, usando a convenção $0/0 = 0$,

$$B_{k+1} = B_k + \frac{y_k y_k^T}{y_k^T s_k} - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} \tag{1.6.3}$$

onde

$$s_k = x_{k+1} - x_k \tag{1.6.4}$$

e

$$y_k = g(x_{k+1}) - g(x_k). \tag{1.6.5}$$

Se $|y_k^T s_k| < 10^{-6}\|y_k\|\|s_k\|$ ou $|s_k^T B_k s_k| < 10^{-6}\|s_k\|\|B_k s_k\|$ deixamos $B_{k+1} = B_k$.

(d) (BFGS1): Idêntico a (c), exceto que $B_0 = I$.

(e) (SRO0): $B_0 = 0$ e B_{k+1} é gerada pela correção simétrica de posto um, para $k \in I\!N$ (Dennis e Schnabel [1983], p. 211; Conn. Gould e Toint [1988b]). Então,

$$B_{k+1} = B_k + \frac{(y_k - B_k s_k)(y_k - B_k s_k)^T}{(y_k - B_k s_k)^T s_k}. \tag{1.6.6}$$

Se $|(y_k - B_k s_k)^T s_k| \leq 10^{-6}\|y_k - B_k s_k\|\|s_k\|$, deixamos $B_{k+1} = B_k$.

(f) (SRO1): Idêntico a (e), exceto que $B_0 = I$.

(g) (PSB0): $B_0 = 0$ e B_{k+1} é gerada pela fórmula Powell-Symmetric-Broyden para $k \in I\!N$ (Dennis e Schnabel [1983] pp. 195-198).

(h) (PSB1): Idêntico a (g), exceto que $B_0 = I$.

Nos casos (b) – (h), se $\Delta_k \leq 10^{-4}$, fazemos $B_{k+1} = 0$.

1.7 EXPERIMENTOS NUMÉRICOS.

Desenvolvemos um programa em FORTRAN 77 usando precisão dupla, baseado na implementação descrita na Seção 1.6. Os testes foram feitos num VAX 785 com sistema operacional VMS. Trabalhamos com três conjuntos de experimentos: problemas-testes clássicos, problemas de regularização e um problema de empacotamento.

(a) Problemas-testes clássicos. Utilizamos como funções testes o conjunto sugerido por Moré, Garbow e Hillstrom [1981]. Estas funções são da forma $f(x) = \sum_{i=1}^{m} f_i^2(x)$, onde $f_i : I\!R^n \to I\!R$. Para cada função consideramos o problema (1.2.1), onde D é dado por (1.5.1), com dois valores diferentes para R. O raio *"grande"* é dado por $4\|x_*\|$ e o *"pequeno"*, por $0.25\|x_*\|$, onde x_* é um minimizador irrestrito de f. Na Tabela 1.7.1 são dados os raios efetivamente usados em cada teste. Os números das funções correspondem à numeração de Moré, Garbow e Hillstrom.

Número	"Nome"	n	m	R grande	R pequeno
1	ROS2	2	2	5.7	0.35
2	FR	2	2	25.6	1.6
7	HV	3	3	4.	0.25
18	EXP6	6	13	15.	1.
9	GAUSS	3	15	2.	0.125
3	POW	2	2	10.	2.5
12	BOX	3	10	4.	0.25
25	VD	10	12	12.65	0.79
20	WAT	6	31	4.	0.25
23	PEN I	4	5	4.	0.25
24	PEN II	4	8	4.	0.25
4	BROWN	2	3	10^6	10.
16	BD	4	20	20.	4.
26	TRIG	10	10	12.	1.
21	ROS10	10	10	12.65	0.79
22	EPOW	12	12	8.	0.5
5	BEALE	2	3	12.16	0.76
14	WOOD	4	6	8.	0.5

Tabela 1.7.1 : Problemas-testes clássicos

Para cada problema usamos dois pontos iniciais, um no interior da bola (origem, quando f está bem definida neste ponto ou $(0.01, \ldots, 0.01)^T$ caso contrário) e outro na fronteira, $x_0 \equiv (1, -1, \ldots, (-1)^{n+1})^T R/\sqrt{n}$.

Utilizamos os seguintes critérios de parada:

$$C_1 : \left\| \frac{\max\{1, \|x_k\|\}}{\max\{1, |f(x_k)|\}} \nabla f(x_k) \right\|_\infty \leq 10^{-4}$$

e

$$C_2 : |\|x_k\| - R| \leq 10^{-3} \text{ e } \frac{x_k^T \nabla f(x_k)}{\|x_k\| \|\nabla f(x_k)\|} \leq -0.9999.$$

C_1 é o critério de convergência sugerido por Dennis e Schnabel [1983] para minimização sem restrições. C_2 estabelece que o ponto x_k está próximo da fronteira e o cosseno do ângulo entre o gradiente da função e o gradiente da restrição está próximo de -1. Em um pequeno número de casos, observamos que para pontos próximos de soluções do problema, só foi possível obter decréscimo em f com valores muito pequenos para Δ_k, onde nem C_1 nem C_2 foram satisfeitos.

Nestes casos, fizemos uma iteração adicional substituindo B_k pela matriz nula $n \times n$. Se, com esta modificação, Δ_k continua muito pequeno, também declaramos convergência. Chamamos de C_3 este critério de convergência:

$$C_3 : B_k = 0 \text{ e } \Delta_k \leq 10^{-4}.$$

É importante notar que quando ocorre C_3, estamos num ponto em que não conseguimos decréscimo da função quando minimizamos sua aproximação linear em uma região de confiança muito pequena, o que é um sinal muito provável da proximidade de um minimizador local.

Finalmente, também paramos quando um número grande de avaliações de função não é suficiente para se obter convergência:

$$E : \text{ atingiu-se o limite de 200 avaliações de função.}$$

Nos casos C_1, C_2 e C_3, x_k está certamente próximo a um ponto estacionário. Assim, declaramos convergência quando qualquer destes critérios é satisfeito.

Os resultados são apresentados nas Tabelas 1.7.2 - 1.7.5 e estão organizados da seguinte maneira:

Tabela 1.7.2: R pequeno e x_0 no interior.

Tabela 1.7.3: R pequeno e x_0 na fronteira.

Tabela 1.7.4: R grande e x_0 no interior.

Tabela 1.7.5: R grande e x_0 na fronteira.

Nestas tabelas indicamos por CP o critério de parada (C_1, C_2, C_3 ou E) e por IT e AF, respectivamente, o número de iterações e avaliações de função efetuadas.

	Escolhas para B_k																							
	NULL			HESS			BFGS0			BFGS1			SRO0			SRO1			PSB0			PSB1		
Função	CP	IT	AF	CP	IT	AF	CP	IT	AF	CP	IT	AF	CP	IT	AF	CP	IT	AF	CP	IT	AF	CP	IT	AF
ROS2	C2	20	102	C2	4	6	C2	4	6	C2	4	6	C2	4	6	C2	4	6	C2	5	13	C2	5	13
FR	C2	14	65	C2	3	4	C2	5	8	C2	5	8	C2	5	8	C2	4	7	C2	4	8	C2	5	9
HV	C2	12	41	C2	5	22	C2	6	9	C2	6	9	C2	6	12	C2	5	8	C2	6	7	C2	6	7
EXP6	C2	24	93	C2	8	12	C2	5	7	C2	5	7	C2	5	7	C2	5	7	C2	5	7	C2	5	7
GAUSS	C2	6	11	C2	4	5	C2	4	5	C2	4	5	C2	4	5	C2	4	5	C2	4	5	C2	4	5
POW	C3	5	70	C2	20	56	C2	42	132	C3	6	48	C3	7	37	C3	6	18	C3	17	128	C3	6	18
BOX	C3	20	192	C1	2	3	C1	10	74	C1	4	9	C1	5	15	C1	3	8	C1	5	15	C1	3	9
VD	C2	1	2	C2	1	2	C2	1	2	C2	1	2	C2	1	2	C2	1	2	C2	1	2	C2	1	2
WAT	C2	6	9	C2	2	3	C2	4	5	C2	4	5	C2	4	5	C2	4	5	C2	4	5	C2	4	5
PEN I	C2	1	2	C2	1	2	C2	1	2	C2	2	3	C2	1	2	C2	2	3	C2	1	2	C2	2	3
PEN II	C2	1	2	C2	1	2	C2	1	2	C2	1	2	C2	1	2	C2	1	2	C2	1	2	C2	1	2
BROWN	C2	1	2	C2	1	2	C2	1	2	C2	1	2	C2	1	2	C2	1	2	C2	1	2	C2	1	2
BD	C2	1	2	C2	2	3	C2	1	2	C2	1	2	C2	1	2	C2	1	2	C2	1	2	C2	1	2
TRIG	C2	7	8	C2	2	3	C2	3	4	C2	3	4	C2	3	4	C2	3	4	C2	3	4	C2	3	4
ROS10	C2	23	119	C2	4	6	C2	4	6	C2	4	6	C2	4	6	C2	4	6	C2	9	55	C2	5	13
EPOW	C2	2	3	C2	1	2	C2	2	3	C2	1	2	C2	2	3	C2	2	3	C2	2	3	C2	2	3
BEALE	C2	4	5	C2	3	4	C2	3	4	C2	3	4	C2	3	4	C2	3	4	C2	3	4	C2	3	4
WOOD	C2	23	62	C2	2	3	C2	12	39	C2	11	26	C2	9	19	C2	8	13	C2	8	11	C2	8	11

Tabela 1.7.2: Resultados numéricos. R pequeno, x_0 no interior.

	Escolhas para B_k																							
	NULL			HESS			BFGS0			BFGS1			SRO0			SRO1			PSB0			PSB1		
Função	CP	IT	AF	CP	IT	AF	CP	IT	AF	CP	IT	AF	CP	IT	AF	CP	IT	AF	CP	IT	AF	CP	IT	AF
ROS2	C2	15	67	C2	3	4	C2	8	15	C2	13	19	C2	13	17	C2	13	17	C2	11	14	C2	11	13
FR	C2	14	61	C2	3	4	C2	4	5	C2	4	5	C2	4	5	C2	4	5	C2	5	10	C2	5	10
HV	C2	13	48	C2	4	9	C2	4	5	C2	4	5	C3	3	9	C3	3	9	C2	5	16	C2	5	16
EXP6	E	41	200	C2	6	11	C2	33	104	C2	37	120	C2	14	36	C2	5	33	C2	15	19	C2	11	18
GAUSS	C2	3	6	C2	2	3	C2	3	4	C2	3	4	C2	3	4	C2	3	4	C3	3	9	C2	4	6
POW	C3	6	72	C3	31	93	C3	7	20	C3	7	20	C3	7	20	C3	7	20	C3	7	20	C3	7	20
BOX	C3	6	45	C1	3	4	C3	10	82	C1	6	7	C1	5	9	C1	5	6	C1	6	10	C1	6	7
VD	C2	1	2	C2	1	2	C2	1	2	C2	1	2	C2	1	2	C2	1	2	C2	1	2	C2	1	2
WAT	C2	5	8	C2	2	3	C2	4	5	C2	4	5	C2	4	5	C2	4	5	C2	5	6	C2	5	6
PEN I	C2	1	2	C2	1	2	C2	1	2	C2	1	2	C2	1	2	C2	1	2	C2	1	2	C2	1	2
PEN II	C2	11	12	C2	2	3	C2	6	7	C2	6	7	C2	5	8	C2	5	8	C2	6	7	C2	6	7
BROWN	C2	1	2	C2	1	2	C2	1	2	C2	1	2	C2	1	2	C2	1	2	C2	1	2	C2	1	2
BD	C2	2	3	C2	2	3	C2	2	3	C2	2	3	C2	2	3	C2	2	3	C2	2	3	C2	2	3
TRIG	C2	13	14	C2	2	3	C2	4	5	C2	4	5	C2	4	5	C2	4	5	C2	4	5	C2	4	5
ROS10	C2	16	74	C2	3	4	C2	9	17	C2	9	16	C2	23	119	C2	24	63	C2	15	27	C2	22	93
EPOW	C2	2	3	C2	2	3	C2	2	3	C2	2	3	C2	2	3	C2	2	3	C2	2	3	C2	2	3
BEALE	C2	3	4	C2	2	3	C2	3	4	C2	1	2	C2	3	4	C2	1	2	C2	3	4	C2	1	2
WOOD	C2	23	62	C2	2	3	C2	8	22	C2	6	11	C2	8	10	C2	8	10	C2	6	10	C2	6	10

Tabela 1.7.3: Resultados numéricos. R pequeno, x_0 na fronteira.

	Escolhas para B_k																							
	NULL			HESS			BFGS0			BFGS1			SRO0			SRO1			PSB0			PSB1		
Função	CP	IT	AF	CP	IT	AF	CP	IT	AF	CP	IT	AF	CP	IT	AF	CP	IT	AF	CP	IT	AF	CP	IT	AF
ROS2	E	22	200	C1	13	19	C1	19	83	C1	19	39	C1	17	49	C1	20	48	C1	20	81	C1	23	119
FR	E	20	200	C3	5	13	C3	14	112	C3	7	21	C1	8	20	C1	8	20	C3	12	27	C3	12	27
HV	E	22	200	C1	10	88	C1	20	154	C1	24	35	C1	22	59	C1	23	54	E	177	200	C1	148	159
EXP6	E	63	200	E	60	200	C3	19	118	C3	10	13	C1	86	144	C1	9	10	C1	20	41	C1	13	14
GAUSS	E	31	200	C1	7	13	C1	21	138	C1	11	16	C1	12	25	C1	10	21	C1	62	67	C1	20	24
POW	C3	6	99	E	61	200	E	82	200	C3	6	28	C3	7	67	C3	6	28	C3	18	148	C3	6	28
BOX	E	16	200	C1	4	8	C3	11	150	C1	4	9	C1	8	27	C1	3	8	C1	11	84	C1	3	9
VD	C3	5	55	C3	18	87	C1	16	113	C1	15	18	C3	15	21	C1	13	21	E	171	200	C3	29	157
WAT	E	28	200	C1	10	11	E	24	200	C1	33	50	C1	25	59	C1	28	46	E	161	200	E	64	200
PEN I	C3	1	21	C3	1	7	C3	1	22	C3	3	19	C3	1	22	C3	3	19	C3	1	22	C3	3	19
PEN II	C3	9	117	C1	8	16	E	22	200	C1	6	12	C1	19	61	C1	6	12	C1	14	57	C1	6	12
BROWN	C1	1	2	C1	20	127	C1	1	2	C1	1	2	C1	1	2	C1	1	2	C1	1	2	C1	1	2
BD	E	27	200	C3	6	11	C3	15	144	C3	18	73	C3	9	17	C3	9	17	C3	13	25	C3	25	28
TRIG	E	20	200	C1	10	11	E	21	200	C1	17	28	C1	31	108	C1	32	94	E	140	200	E	180	200
ROS10	E	15	200	C1	13	19	E	20	200	C1	39	168	E	25	200	C1	42	134	E	24	200	E	26	200
EPOW	E	27	200	C1	12	14	E	24	200	C1	33	95	C1	35	171	C1	29	79	E	44	200	C1	28	114
BEALE	E	18	200	C1	7	16	C1	11	71	C1	8	18	C1	14	30	C1	14	30	C1	14	42	C1	14	39
WOOD	E	23	200	C1	7	10	E	23	200	C3	15	74	C1	25	70	C1	20	53	C1	19	65	C1	19	66

Tabela 1.7.4: Resultados numéricos. R grande, x_0 no interior.

	Escolhas para B_k																							
	NULL			HESS			BFGS0			BFGS1			SRO0			SRO1			PSB0			PSB1		
Função	CP	IT	AF	CP	IT	AF	CP	IT	AF	CP	IT	AF	CP	IT	AF	CP	IT	AF	CP	IT	AF	CP	IT	AF
ROS2	E	19	200	C1	19	24	C1	36	69	C1	31	37	C1	41	94	C1	38	72	C1	38	64	C1	46	153
FR	E	22	200	C1	15	16	C3	14	32	C1	11	13	C1	11	13	C1	11	13	C1	61	63	C1	61	63
HV	E	21	200	C1	10	18	C3	20	188	C1	12	26	C1	15	23	C1	15	26	C1	17	37	C3	13	39
EXP6	E	66	200	C1	72	85	C3	52	179	C1	68	70	C1	64	102	C1	74	136	E	199	200	E	199	200
GAUSS	E	30	200	C1	13	27	C1	23	40	C1	22	95	C1	13	31	C1	15	30	C1	12	52	C1	12	34
POW	C3	8	110	C3	38	56	C3	9	23	C3	6	32	C3	9	23	C3	6	32	C3	9	23	C3	6	32
BOX	E	24	200	C1	10	40	C1	16	88	C1	16	17	C1	15	36	C1	15	23	C1	29	30	C1	28	29
VD	C3	6	66	C3	18	76	C1	16	91	C1	20	25	C3	16	28	C3	19	33	C3	181	33	C3	15	27
WAT	E	32	200	C1	6	7	E	30	200	C1	38	57	C1	32	78	C1	33	77	E	135	200	E	142	200
PEN I	C3	2	23	C1	45	68	C3	2	24	C3	4	21	C3	2	24	C3	4	21	C3	2	24	C3	4	21
PEN II	E	21	200	C1	10	11	C3	14	129	C1	13	21	C1	15	42	C1	15	37	C1	19	43	E	182	200
BROWN	C3	2	38	E	35	200	C1	27	125	C3	4	33	C3	4	33	C3	4	33	C3	4	33	C3	4	33
BD	E	34	200	C3	8	11	C3	20	153	C3	18	69	C3	14	22	C3	14	22	C3	127	134	C3	127	134
TRIG	C2	16	67	C2	8	10	C2	12	34	C2	13	28	C2	11	28	C2	11	62	C2	22	34	C2	20	29
ROS10	E	24	200	C1	18	22	E	28	200	E	70	200	E	52	200	E	46	200	E	37	200	E	36	200
EPOW	E	28	200	C1	15	16	E	23	200	C1	35	77	C1	34	117	C1	27	61	E	115	200	E	142	200
BEALE	E	19	200	C1	13	32	C1	13	34	C3	8	13	C3	8	13	C3	8	13	C3	8	13	C3	8	13
WOOD	E	27	200	C1	10	11	C1	48	171	C1	35	59	C1	58	185	C1	63	171	E	77	200	E	107	200

Tabela 1.7.5: Resultados numéricos. R grande, x_0 na fronteira.

Na atual implementação de nosso método, encontramos uma solução exata para o problema (1.5.31). Por se tratar de um problema relativamente caro, o método se torna atraente apenas quando a avaliação da função f é muito cara. Por esta razão, o principal critério para se avaliar a eficiência das diferentes escolhas para B_k deve ser o número efetuado de avaliações de função.

Através destes experimentos numéricos, concluímos o seguinte:

(a) Para R pequeno, obtivemos minimizadores na fronteira da bola na grande maioria dos testes. Quando x_0 é interior, observamos pouca diferença entre o desempenho de HESS e dos melhores métodos secantes (BFGS1 ou SRO1). De fato, neste caso, os pontos estacionários irrestritos estão afastados da região de interesse e portanto a aproximação linear de f domina amplamente o comportamento da função. Apesar disto, a curvatura na direção de ∇f é suficientemente importante para fazer com que BFGS1 e SRO1 sejam mais eficientes que a implementação onde $B_0 = 0$. Um monitoramento mais cuidadoso de B_k (ver Contreras e Tapia [1991], Oren e Luenberger [1974], Oren e Spedicato [1974], Shanno e Phua [1978]) deve proporcionar um desempenho ainda melhor para BFGS e SRO.

(b) Quando R é pequeno mas x_0 está na fronteira, o desempenho de HESS é superior ao dos melhores métodos secantes. Neste caso, a maioria das iterações gera pontos na fronteira e termos de segunda ordem verdadeiros tornam-se mais importantes. Em outras palavras, a informação contida na aproximação diagonal $B_0 = I$ é muito pobre neste caso.

(c) Os casos em que R é grande não são muito relevantes para o nosso estudo. Nestes casos, como a solução está no interior da bola, o comportamento geral dos algoritmos tende a ser o mesmo que no caso irrestrito.

Com o objetivo de colocar nossos resultados computacionais num contexto, resolvemos o mesmo conjunto de testes usando o algoritmo de Programação Quadrática Seqüencial (PQS), com Lagrangiano Aumentado como função de mérito, de Gill, Murray, Saunders e Wright [1992]. Na implementação do algoritmo PQS usamos as verdadeiras Hessianas do Lagrangiano, com uma fatoração de Cholesky modificada para garantir matrizes definidas positivas. Os resultados são apresentados na Tabela 1.7.6. Para o algoritmo PQS o critério de convergência C significa:

$$\left|[(\nabla f(x_k))_i + 2\lambda(x_k)_i]\frac{|(x_k)_i| + 1}{|f(x_k)| + 1}\right| \leq 10^{-4}$$

para $i = 1, \ldots, n$ onde λ é a estimativa final para o multiplicador de Lagrange. O critério

de convergência C^* significa:

$$\left|[(\nabla f(x_k))_i + 2\lambda(x_k)_i]\frac{\max\{|(x_k)_i|, tip\ x_i\}}{\max\{|f(x_k)|, tip\ f\}}\right| \leq 10^{-4},$$

para $i = 1, 2$, onde $tip\ x = (10^6, 10^{-6})^T$ e $tip\ f = 10^6$.

Observamos que o algoritmo de região de confiança tem um desempenho claramente superior ao do PQS quando R é pequeno (solução na fronteira). Para R grande, em que a solução é irrestrita, os desempenhos são semelhantes.

	R pequeno x_0 no interior						R pequeno x_0 na fronteira						R grande x_0 no interior						R grande x_0 na fronteira					
função	RC			PQS			RC			PQS			RC			PQS			RC			PQS		
	CP	IT	AF	CP	IT	AF	CP	IT	AF	CP	IT	AF	CP	IT	AF	CP	IT	AF	CP	IT	AF	CP	IT	AF
ROS2	C_2	4	6	C	5	8	C_2	3	4	C	4	5	C_1	13	19	C	13	19	C_1	19	24	C	21	28
FR	C_2	3	4	C	10	16	C_2	3	4	C	4	5	C_3	5	13	C	5	10	C_1	15	16	C	15	16
HV	C_2	5	22	C	15	18	C_2	4	9	C	6	7	C_1	10	88	C	13	14	C_1	10	18	C	10	14
EXP6	C_2	8	12	C	28	29	C_2	6	11	C	41	42	E	60	200	C	50	51	C_1	72	85	E	76	124
GAUSS	C_2	4	5	C	8	14	C_2	2	3	C	9	12	C_1	7	13	C	8	12	C_1	13	27	C	78	80
POW	C_2	20	56	C	11	12	C_3	34	93	C	49	77	E	61	200	C	11	12	C_3	38	56	C	50	51
BOX	C_1	2	3	C	4	7	C_1	3	4	C	12	13	C_1	4	8	C	5	7	C_1	10	40	C	36	40
VD	C_2	1	2	C	22	38	C_2	1	2	C	25	61	C_3	18	87	C	18	20	C_3	18	76	C	15	16
WAT	C_2	2	3	C	16	29	C_2	2	3	C	16	31	C_1	10	11	C	11	12	C_1	6	7	C	6	7
PENI	C_2	1	2	C	6	7	C_2	1	2	C	1	2	C_3	1	7	C	13	14	C_1	45	68	C	52	88
PENII	C_2	1	2	C	6	7	C_2	2	3	C	6	10	C_1	8	16	C	8	9	C_1	10	11	C	23	24
BROWN	C_2	1	2	C*	8	21	C_2	1	2	C*	8	15	C_1	20	127	C*	15	112	E	35	200	C*	53	129
BD	C_2	2	3	C	11	16	C_2	2	3	C	11	17	C_3	6	11	C	7	8	C_3	8	11	C	8	9
TRIG	C_2	2	3	C	6	9	C_2	2	3	C	6	8	C_1	10	11	C	10	11	C_2	8	10	C	51	58
ROSI0	C_2	4	6	C	5	8	C_2	3	4	C	4	5	C_1	13	19	C	13	19	C_1	18	22	C	21	28
EPOW	C_2	1	2	C	10	23	C_2	2	3	C	9	13	C_1	12	14	C	11	21	C_1	15	16	C	15	16
BEALE	C_2	3	4	C	7	10	C_2	2	3	C	5	6	C_1	7	16	C	19	44	C_1	13	32	C	23	40
WOOD	C_2	2	3	C	5	7	C_2	2	3	C	5	6	C_1	7	10	C	8	10	C_1	10	11	C	11	12

Tabela 1.7.6: Resultados comparativos: nosso algoritmo de região de confiança – RC (usando Hessianas verdadeiras) × o algoritmo PQS de Gill, Murray, Saunders e Wright.

(b) **Problemas de regularização.** Um outro conjunto de problemas-testes foi gerado como se segue (ver Vogel [1990]). Consideramos a equação integral

$$F(x)(t) \equiv \int_0^1 log\left[\frac{(t-\tau)^2+0.04}{(t-\tau)^2+(0.2-x(\tau))^2}\right] d\tau = y(t), \tag{1.7.1}$$

com as condições de fronteira $x(0) = x(1) = 0$. Tomamos como solução verdadeira a função:

$$x(t) = c_1 \exp(d_1(t-p_1)^2) + c_2 \exp(d_2(t-p_2)^2) + c_3 t + c4\ ,$$

onde $c_1 = -0.1$, $c_2 = -0.075$, $d_1 = -40$, $d_2 = -60$, $p_1 = 0.4$, $p_2 = 0.67$, e c_3 e c_4 são escolhidos de forma que $x(0) = x(1) = 0$. Dado y, o problema de encontrar $x(t)$ que satisfaça (1.7.1) aproximadamente é mal-posto e para ser resolvido é preciso usar regularização (ver Tikhonov e Arsenin [1977]). A abordagem adotada por Vogel para resolver (1.7.1) consiste em substituir esta equação por

$$\begin{array}{ll} \min & \|F(x)-y\|^2 \\ \text{s/a} & |x|^2 \leq \beta^2 \end{array} \tag{1.7.2}$$

onde $|x|^2 = \int_0^1 x'(t)^2 dt$.

Após discretização, (1.7.2) se converte num problema de dimensão finita do tipo (1.2.1). Além disso, fazendo-se uma mudança óbvia de variáveis, (1.7.2) se transforma num problema do tipo

$$\begin{array}{ll} \min & f(x) \\ \text{s/a} & \|x\|^2 \leq \beta^2\ . \end{array} \tag{1.7.3}$$

Vogel usou um algoritmo de região de confiança para resolver (1.7.3) com a aproximação $F'(x_k)^T F'(x_k)$ (Gauss-Newton) para a matriz Hessiana. Assim, seus subproblemas são convexos. Usamos nosso algoritmo para resolver (1.7.3), tomando a Hessiana verdadeira do problema. Desta forma, os subproblemas não são necessariamente convexos e precisamos utilizar o Algoritmo 1.5.7 para resolvê-los. Na Tabela 1.7.7 comparamos os resultados do nosso algoritmo de região de confiança usando Hessianas verdadeiras com a aproximação de Vogel. IT e AF denotam, respectivamente, o número de iterações e avaliações de função efetuados. Observamos que, em geral, o algoritmo de região de confiança com subproblemas convexos utilizou o dobro do número de iterações demandadas pelo nosso algoritmo com Hessianas verdadeiras. Portanto, parece vantajoso permitir subproblemas não convexos em métodos de região de confiança para resolver problemas com restrições do tipo bolas, originadas de regularização.

β	Hessianas verdadeiras		Aproximação de Vogel (GAUSS-NEWTON)	
	IT	AF	IT	AF
0.200	5	6	13	14
0.250	8	9	14	15
0.275	8	9	14	15
0.300	9	10	14	15
0.325	10	11	15	16
0.400	10	13	15	16
0.500	18	23	16	17

Tabela 1.7.7: Resultados comparativos do algoritmo de região de confiança: usando Hessianas verdadeiras e usando a aproximação de Vogel.

(c) **Um problema de empacotamento.** Um problema aberto clássico em Geometria Combinatória é o seguinte (ver Conway e Sloane [1988], Capítulo 1): existem 25 pontos na esfera unitária de $\mathbb{R}^4$ tais que a distância entre qualquer par deles é maior ou igual a 1? Esta pergunta pode ser formulada como um problema contínuo de otimização global que consiste na minimização de

$$f(y_1,\ldots,y_{25}) = \sum_i (1 - \|y_i\|^2)^2 + \sum_{\substack{i,j \\ i\neq j}} [(1 - \|y_i - y_j\|^2)_+]^2$$

onde $z_+ = \max\{0, z\}$. Chamamos de x o vetor em $\mathbb{R}^{100}$ cujas componentes são $(y_1^T,\ldots,y_{25}^T)^T$. É natural impor a restrição $\|x\|^2 \leq 25$ para a minimização de f. Fizemos uma série de testes com diferentes pontos iniciais, gerados aleatoriamente de modo que $\|y_i^0\| = 1, i = 1,\ldots,25$ e comparamos o desempenho do nosso método de região de confiança com o algoritmo PQS. Os resultados são apresentados na Tabela 1.7.8, onde IT, AF e DIST indicam, respectivamente, o número de iterações efetuadas, o número de avaliações de função e a menor distância entre dois pontos distintos obtidos como aproximação final pelos métodos.

	Região de Confiança				PQS			
semente	IT	AF	$f(x_*)$	DIST	IT	AF	$f(x_*)$	DIST
2	22	78	0.33	0.92	500	500	8.1	0.30
3	35	89	0.31	0.91	500	500	9.9	0.29
7	44	149	0.27	0.91	500	500	16.1	0.25
9	31	75	0.28	0.92	500	500	7.3	0.26
11	62	313	1.03	0.67	500	500	8.1	0.23
13	27	73	0.27	0.91	500	500	9.1	0.17
17	38	118	0.27	0.91	500	500	13.7	0.30
19	32	88	0.27	0.91	500	500	14.	0.27
67	46	153	0.27	0.91	500	500	10.	0.14
991	27	70	0.31	0.91	500	500	7.3	0.27
1001	26	92	2.27	0.92	500	500	11.	0.35

Tabela 1.7.8: Resultados comparativos com o problema das 25 esferas.

1.8 OBSERVAÇÕES FINAIS

A implementação de algoritmos para resolver problemas como (1.2.1) sem linearização das restrições depende da possibilidade de se resolver eficientemente subproblemas não triviais. Por muito tempo, os únicos subproblemas considerados em otimização prática foram os sistemas de equações lineares. Esta situação começou a mudar com a introdução dos métodos de região de confiança nos anos 80. Celis, Dennis e Tapia [1984] e Powell e Yuan [1990] introduziram algoritmos de programação não-linear baseados num subproblema difícil para o qual ainda não se conseguiu uma solução completamente satisfatória (Zhang [1988], Dennis, Martínez, Tapia e Williamson [1990], Williamson [1990]). O subproblema considerado neste Capítulo, no caso em que D é uma bola euclidiana, não é tão difícil quanto o de Celis–Dennis–Tapia, mas tem claramente um custo computacional bem maior que o de se resolver um sistema linear. Assim, no atual estágio de desenvolvimento, nosso método deve ser usado apenas se o tempo computacional das avaliações de função domina os cálculos.

Vamos discutir brevemente a dificuldade de se resolver o subproblema de região de confiança associado à minimização de uma quadrática arbitrária ψ em domínios diferentes de bolas euclidianas. Quando D é uma esfera, a dificuldade de se minimizar ψ na interseção de D com uma bola é essencialmente a mesma que a do Algoritmo 1.5.7, mas menos casos precisam ser considerados, conforme veremos na Seção 2.4. Quando D é

o complemento de uma bola euclidiana, a complexidade do subproblema de região de confiança é exatamente a mesma da do Algoritmo 1.5.7, e os mesmos casos devem ser considerados. Deve-se fazer adaptações naturais deste algoritmo para se lidar com o caso em que o domínio D é o complemento da união finita de bolas disjuntas. Pesquisas recentes (Stern e Wolkowicz [1993], Moré [1993]) parecem indicar um aumento no número de casos em que seremos capazes de minimizar quadráticas em domínios não lineares.

CAPÍTULO 2

MÉTODOS DE REGIÃO DE CONFIANÇA PARA MINIMIZAÇÃO COM RESTRIÇÕES DE IGUALDADE

2.1 INTRODUÇÃO

A minimização de uma função diferenciável com restrições não lineares de igualdade (MRI) é um problema clássico em programação não-linear (ver, por exemplo, Fletcher [1987], Luenberger [1984], Gill, Murray e Wright [1981]). Existem duas maneiras tradicionais de se lidar com este problema: os métodos de Penalização (e Lagrangiano Aumentado) e Programação Quadrática Sequencial (PQS). Em Penalização, as restrições são incorporadas na função objetivo, de tal forma que a resolução do problema original se converte na resolução de uma seqüência de subproblemas de minimização irrestrita. Na abordagem PQS, as restrições não lineares são substituídas por aproximações lineares e também se resolve uma seqüência de subproblemas onde cada um consiste em minimizar uma função quadrática com restrições lineares (ver Dennis, El-Alem e Maciel [1992], El-Alem [1991], Celis, Dennis e Tapia [1984], Powell e Yuan [1991], etc.).

Tanto Penalização quanto PQS pressupõem a impossibilidade de se lidar com as restrições não lineares em sua forma original. Existem boas razões para considerarmos as restrições não lineares como algo intratável e portanto optarmos pela linearização ou pela incorporação destas na função objetivo. De fato, métodos de busca linear, em geral, não se aplicam diretamente em sua forma mais simples, pois o conjunto factível não precisa obrigatoriamente conter semi-retas. Além disso, os subproblemas de região de confiança

(Moré [1983]) podem se tornar mais difíceis que o problema original.

Por outro lado, pesquisas recentes na resolução de problemas de minimização com função objetivo quadrática e com uma única restrição quadrática vem ampliando a possibilidade de se resolver eficientemente os subproblemas de região de confiança associados ao problema MRI sem nenhuma modificação nas restrições originais. Gay [1981], Sorensen [1982] e Moré e Sorensen [1983] resolveram o problema de se encontrar um minimizador global de uma quadrática geral com uma restrição quadrática convexa. Mais recentemente, Martínez [1994] caracterizou os minimizadores locais deste mesmo problema. Esta caracterização é importante porque quando o conjunto factível é a interseção de duas regiões, os minimizadores locais da função objetivo em cada uma destas duas regiões são candidatos naturais a minimizadores globais do problema original. Tanto Stern e Wolkowicz [1993] quanto Moré [1993] caracterizaram os minimizadores globais de uma quadrática geral com uma restrição quadrática não necessariamente convexa. Além disso, uma pesquisa intensa tem sido dedicada ao problema de se minimizar uma quadrática geral na interseção de uma bola com um cilindro, conhecido como problema CDT (Celis, Dennis e Tapia [1984]).

As considerações acima em conjunto com as observações do Capítulo 1 nos fazem acreditar que em breve seremos capazes de resolver efetivamente problemas de minimização com função objetivo quadrática e região factível dada pela interseção de uma hipersuperfície quadrática com uma bola euclidiana. De fato, usando os resultados clássicos de Gay, Moré e Sorensen e a caracterização de Martínez, podemos resolver satisfatoriamente o problema de minimizar uma quadrática na interseção de uma esfera com uma bola. Este subproblema de região de confiança aparece na minimização de uma função não-linear arbitrária com uma restrição de igualdade quadrática e estritamente convexa. Acreditamos, portanto, na viabilidade de se considerar métodos de região de confiança para problemas com restrições gerais, que devem ser preservadas sem modificação, e de tal maneira que o formato da região de confiança também seja independente das restrições.

Neste capítulo provamos convergência superlinear para uma subclasse dos métodos introduzidos no Capítulo 1 para o problema MRI. Apresentamos também resultados de convergência global de segunda ordem e convergência quadrática local quando se trabalha com Hessianas verdadeiras no algoritmo. Descrevemos ainda algoritmos específicos para o caso em que a região factível é uma esfera. Uma aplicação para este caso é detalhada no Capítulo 5, onde consideramos o Problema do Vetor Inicial em codificação.

A organização deste capítulo é a seguinte: na Seção 2.2 introduzimos um *algoritmo local* para resolver o problema MRI, que está bem definido numa vizinhança de um ponto que satisfaz as condições suficientes de segunda ordem para um minimizador local. Pro-

vamos convergência local e convergência superlinear se as aproximações para a Hessiana satisfazem uma condição do tipo Dennis–Moré (Dennis e Moré, [1974]). Finalmente, sob as hipóteses de Dennis–Moré, também mostramos que as iterações do algoritmo local poduzem descenso suficiente na função objetivo. O ingrediente principal para as provas desta seção é a teoria dos métodos quase-Newton de ponto fixo, introduzida recentemente por Martínez [1992]. Na Seção 2.3 apresentamos o algoritmo de região de confiança RCMRI, cuja convergência global para pontos estacionários de primeira ordem segue dos resultados apresentados no Capítulo 1. Com o uso de matrizes Hessianas verdadeiras neste algoritmo, provamos que todo ponto de acumulação é estacionário de segunda ordem. Finalmente, mostramos que numa vizinhança de um ponto que satisfaz condições suficientes de segunda ordem os algoritmos local e RCMRI coincidem, e portanto o algoritmo RCMRI também tem propriedades de convergência local. Na Seção 2.4 definimos uma implementação do algoritmo RCMRI para o caso de uma restrição esférica. Para encerrar, a Seção 2.5 é destinada a algumas observações finais.

2.2 O MÉTODO LOCAL.

Nesta seção definimos uma *algoritmo local* para resolver o problema de minimização com restrições de igualdade (MRI). Em outras palavras, introduzimos um método que está bem definido numa vizinhança de uma solução apropriada, provamos convergência do método para pontos iniciais suficientemente próximos desta solução e apresentamos condições para a obtenção de convergência superlinear.

Definimos o problema MRI como se segue:

$$\begin{array}{ll} \min & f(x) \\ \text{s/a} & h(x) = 0 \end{array} \tag{2.2.1}$$

onde $f : I\!R^n \to I\!R, h : I\!R^n \to I\!R^m$ e $f, h \in C^2(I\!R^n)$. Denotamos por $h'(x) \in I\!R^{m\times n}$ a matriz Jacobiana de $h(x)$ e definimos $S = \{x \in I\!R^n \mid h(x) = 0\}$. $\|\cdot\|$ denota uma norma arbitrária em $I\!R^n$.

O método local para resolver (2.2.1) é definido pelo Algoritmo 2.2.1 a seguir:

Algoritmo 2.2.1 (Local)

Seja $x_0 \in S$ uma aproximação inicial para a solução de (2.2.1).

Dados $x_k \in \mathbb{R}^n$, $B_k \in \mathbb{R}^{n\times n}$, $B_k = B_k^T$, calculamos x_{k+1} como a solução y de:

$$\begin{array}{ll} \min & \frac{1}{2}(y-x_k)^T B_k(y-x_k) + g_k^T(y-x_k) \\ \text{s/a} & h(y) = 0 \end{array} \tag{2.2.2}$$

onde $g_k \equiv g(x_k)$ e $g \equiv \nabla f$.

A solução de (2.2.2) existe e é única apenas sob circunstâncias especiais, que serão estudadas posteriormente. O Algoritmo 2.2.1 pode ser interpretado como um método quase-Newton de ponto fixo no sentido de Martínez [1992]. Dados $x \in \mathbb{R}^n$ e $B \in \mathbb{R}^{n\times n}$ simétrica, definimos $\Phi(x, B)$ como a solução de

$$\begin{array}{ll} \min & \frac{1}{2}(y-x)^T B(y-x) + g(x)^T(y-x) \\ \text{s/a} & h(y) = 0\,. \end{array} \tag{2.2.3}$$

Assim, o Algoritmo 2.2.1 pode ser escrito como

$$x_{k+1} = \Phi(x_k, B_k). \tag{2.2.4}$$

Analogamente a Martínez [1992], denotamos por $\Phi'(x, B)$ a matriz Jacobiana com relação a x. No lema a seguir calculamos esta matriz.

LEMA 2.2.2. Suponhamos que para algum $x \in \mathbb{R}^n$, $B = B^T$, (2.2.3) tem uma única solução y, onde posto $(h'(y)) = m$, $\mu \in \mathbb{R}^m$ é o vetor dos multiplicadores de Lagrange correspondente e

$$z^T \left[B + \sum_{i=1}^{m} \mu_i \nabla^2 h_i(y) \right] z > 0 \tag{2.2.5}$$

para todo $z \in \mathcal{N}(h'(y))$ (espaço nulo de $h'(y)$), $z \neq 0$. Então

$$\Phi'(x, B) = P \left[P^T (B + \sum_{i=1}^{m} \mu_i \nabla^2 h_i(y)) P \right]^{-1} P^T (B - \nabla^2 f(x)), \tag{2.2.6}$$

onde $P \in \mathbb{R}^{n\times(n-m)}$ é uma matriz cujas colunas formam uma base para $\mathcal{N}(h'(y))$.

Demonstração. Se $y \in \mathbb{R}^n$ é uma solução de (2.2.3), pelas condições de otimalidade de Lagrange temos que

$$\begin{aligned} B(y-x) + g(x) + h'(y)^T\mu &= 0 \\ h(y) &= 0\,. \end{aligned} \tag{2.2.7}$$

(2.2.7) é um sistema com $n+m$ equações não-lineares nas variáveis x, y, B e μ. Como posto $(h'(y)) = m$ e por (2.2.5), a matriz $\begin{pmatrix} B + \sum_{i=1}^{m}\mu_i\nabla^2 h_i(y) & h'(y)^T \\ h'(y) & 0 \end{pmatrix}$ é não singular. Assim, podemos aplicar o Teorema da Função Implícita em (2.2.7), obtendo, pela derivação em relação a x, que:

$$\begin{pmatrix} B + \sum_{i=1}^{m}\mu_i\nabla^2 h_i(y) & h'(y)^T \\ h'(y) & 0 \end{pmatrix} \begin{pmatrix} \Phi'(x,B) \\ C \end{pmatrix} = \begin{pmatrix} B - \nabla^2 f(x) \\ 0 \end{pmatrix}, \tag{2.2.8}$$

onde C é a matriz das derivadas de μ em relação a x. Portanto,

$$\left[B + \sum_{i=1}^{m}\mu_i\nabla^2 h_i(y)\right]\Phi'(x,B) + h'(y)^T C = B - \nabla^2 f(x) \tag{2.2.9}$$

e

$$h'(y)\Phi'(x,B) = 0. \tag{2.2.10}$$

Por (2.2.10), existe $M \in \mathbb{R}^{(n-m)\times n}$ tal que

$$\Phi'(x,B) = PM. \tag{2.2.11}$$

Substituindo (2.2.11) em (2.2.9) e pré-multiplicando por P^T, obtemos

$$P^T\left[B+\sum_{i=1}^{m}\mu_i\nabla^2 h_i(y)\right]PM = P^T(B-\nabla^2 f(x)).$$

Logo,

$$M=\left\{P^T\left[B+\sum_{i=1}^{m}\mu_i\nabla^2 h_i(y)\right]P\right\}^{-1}P^T(B-\nabla^2 f(x)). \tag{2.2.12}$$

Desta forma, (2.2.6) segue de (2.2.11) e (2.2.12). □

Hipóteses Locais Gerais. Vamos supor que $x_* \in \mathbb{R}^n$ é uma solução de (2.2.1) onde $h'(x_*)$ tem posto completo e valem as condições suficientes de segunda ordem para um minimizador local, isto é,

$$z^T G_* z > 0 \tag{2.2.13}$$

para todo $z \in \mathcal{N}(h'(x_*)), z \neq 0$, onde $G_* = \nabla^2 f(x_*) + \sum_{i=1}^{m}\mu_i^*\nabla^2 h_i(x_*)$ e $\mu^* \in \mathbb{R}^m$ é o vetor dos multiplicadores de Lagrange associado a (2.2.1) e x_*.

Pelo Teorema de Função Implícita, estas hipóteses garantem a existência de $\Phi(x,B)$ e $\Phi'(x,B)$ numa vizinhança $\Omega \times \mathcal{D}$ de $(x_*, \nabla^2 f(x_*))$. Além disso, podemos assumir que $x_* = \Phi(x_*, B)$ para todo $B \in \mathcal{D}$ e portanto por (2.2.6),

$$\Phi'(x_*,B) = P_*\left[P_*^T\left(B+\sum_{i=1}^{m}\mu_i(B)\nabla^2 h_i(x_*)\right)P_*\right]^{-1}P_*^T(B-\nabla^2 f(x_*)), \tag{2.2.14}$$

onde $\mu(B) \in \mathbb{R}^m$ é o vetor dos multiplicadores do problema (2.2.3) e P_* é uma matriz cujas colunas formam uma base para $\mathcal{N}(h'(x_*))$.

A continuidade de $\Phi'(x_*, B)$ com relação a B em $\mathcal{D}$ é garantida por argumentos elementares, com uma possível restrição em $\mathcal{D}$. Também assumimos que existem $L, p > 0$ tais que

$$\|\Phi'(x, B) - \Phi'(x_*, B)\| \leq L\|x - x_*\|^p \tag{2.2.15}$$

para todo $x \in \Omega, B \in \mathcal{D}$. Claramente,

$$\Phi'(x_*, \nabla^2 f(x_*)) = 0. \tag{2.2.16}$$

Esta discussão nos permite provar o seguinte teorema de convergência local.

TEOREMA 2.2.3. Suponhamos que f, h, x_* satisfazem as Hipóteses Locais Gerais. Dado $r \in (0,1)$, existem $\varepsilon = \varepsilon(r)$ e $\delta = \delta(r)$ tais que se $\|x - x_*\| \leq \varepsilon$ e $\|B - \nabla^2 f(x_*)\| \leq \delta$, então

$$\|\Phi(x, B) - x_*\| \leq r\|x - x_*\|. \tag{2.2.17}$$

Além disso, se $\|x_0 - x_*\| \leq \varepsilon$ e $\|B_k - \nabla^2 f(x_*)\| \leq \delta$ para todo $k = 0, 1, 2, \ldots$, a seqüência gerada pelo Algoritmo 2.2.1 está bem definida, converge para x_* e satisfaz

$$\|x_{k+1} - x_*\| \leq r\|x_k - x_*\|$$

para todo $k = 0, 1, 2, \ldots$.

Demonstração. O resultado segue de (2.2.15), (2.2.16) e (2.2.6) como conseqüência do Teorema 3.1 de Martínez [1992]. □

A propriedade estabelecida pelo lema a seguir será utilizada na prova de que o Algoritmo 2.2.1 produz descenso suficiente na função objetivo de (2.2.1).

LEMA 2.2.4. Suponhamos válidas as hipóteses do Teorema 2.2.3. Se $\mu^k \in \mathbb{R}^m$ é o vetor dos multiplicadores de Lagrange associados a (2.2.2), então existem $c_1, c_2 > 0$, $k_0 \in \mathbb{N}$ tais que $\|\mu^k\| \leq c_1$ e

$$\frac{(x_{k+1} - x_k)^T (B_k + \sum_{i=1}^{m} \mu_i^k \nabla^2 h_i(x_k))(x_{k+1} - x_k)}{\|x_{k+1} - x_k\|^2} \geq c_2 \tag{2.2.18}$$

para todo $k \geq k_0$.

Demonstração. Segue do Teorema 2.2.3, de (2.2.13), da continuidade dos multiplicadores de Lagrange e do fato de que $h(x_k) = 0$ para todo $k \in \mathbb{N}$. □

No próximo teorema apresentamos uma condição do tipo Dennis–Moré para convergência superlinear da seqüência gerada pelo Algoritmo 2.2.1. A condição do tipo Dennis–Moré associada à convergência superlinear de algoritmos PQS (Boggs, Tolle e Wang [1982]) envolve o efeito da aproximação da Hessiana do Lagrangeano no incremento. É interessante observar que quando não aproximamos as restrições pelos seus modelos lineares, a condição para convergência superlinear está associada com as aproximações para a Hessiana da função objetivo.

TEOREMA 2.2.5. Suponhamos válidas as hipóteses do Teorema 2.2.3. Se

$$\lim_{k \to \infty} \frac{\|(B_k - \nabla^2 f(x_*))(x_{k+1} - x_k)\|}{\|x_{k+1} - x_k\|} = 0 \tag{2.2.19}$$

então

$$\lim_{k\to\infty} \frac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|} = 0. \tag{2.2.20}$$

Demonstração. Por argumentos elementares de continuidade, (2.2.19) e (2.2.6) implicam em

$$\lim_{k\to\infty} \frac{\|[\Phi'(x_*, B_k) - \Phi'(x_*, \nabla^2 f(x_*))](x_{k+1} - x_k)\|}{\|x_{k+1} - x_k\|} = 0. \tag{2.2.21}$$

Portanto, (2.2.20) segue do Teorema 4.2 de Martínez [1992]. □

O teorema a seguir estabelece a ordem de convergência do Algoritmo 2.2.1 usando-se os Hessianos verdadeiros (versão Newton).

TEOREMA 2.2.6. Suponhamos válidas as hipóteses do Teorema 2.2.3. Se, para todo $k \in \mathbb{N}$, $B_k = \nabla^2 f(x_k)$, então existe $c > 0$ tal que

$$\|x_{k+1} - x_*\| \leq c\|x_k - x_*\|^{p+1}, \tag{2.2.22}$$

onde p satisfaz a hipótese (2.2.15).

Demonstração. O resultado segue do Teorema 4.3 de Martínez [1992]. □

O último resultado desta seção tem por objetivo fundamentar as propriedades de convergência global do método. Em linhas gerais, estabelece que numa vizinhança apropriada de x_*, se a condição de Dennis–Moré é válida, então uma propriedade de descenso suficiente é satisfeita.

TEOREMA 2.2.7. Suponhamos válidas as Hipóteses Locais Gerais, $f, h \in C^3(\mathbb{R}^n), \alpha \in (0,1)$. Suponhamos ainda que $\{x_k\}$ é uma seqüência arbitrária de pontos que satisfazem as restrições de (2.2.1) e convergem a x_*, $\{B_k\}$ é uma seqüência de matrizes tal que $\Phi(x_k, B_k)$ está bem definida para todo $k \in \mathbb{N}$ e

$$\lim_{k\to\infty} \frac{\|[B_k - \nabla^2 f(x_*)]s_k\|}{\|s_k\|} = 0. \tag{2.2.23}$$

Então, existe $k_0 \in \mathbb{N}$ tal que para todo $k \geq k_0$,

$$f(x_k + s_k) \leq f(x_k) + \alpha\psi_k(s_k), \tag{2.2.24}$$

onde $s_k = \Phi(x_k, B_k) - x_k$ e $\psi_k(s) = g_k^T s + \frac{1}{2}s^T B_k s$ para todo $s \in \mathbb{R}^n$.

Demonstração. Pelas condições de otimalidade de primeira ordem para (2.2.2), existe $\mu^k \in \mathbb{R}^m$ tal que

$$B_k s_k + g_k + h'(y_k)^T \mu^k = 0 \tag{2.2.25}$$

e

$$h(y_k) = 0 \tag{2.2.26}$$

para todo $k \in \mathbb{N}$, onde $y_k = x_k + s_k$. Por (2.2.25),

$$g_k^T s_k = -(\mu^k)^T h'(y_k)s_k - s_k^T B_k s_k. \tag{2.2.27}$$

Pela fórmula de Taylor temos, para $i = 1, 2, \ldots, m$, que

$$h_i(x_k) = h_i(y_k) - h_i'(y_k)s_k + \frac{1}{2}s_k^T \nabla^2 h_i(y_k)s_k + o(\|s_k\|^2). \tag{2.2.28}$$

Como $h_i(x_k) = h_i(y_k) = 0$, de (2.2.28) segue que

$$-h_i'(y_k)s_k + \frac{1}{2}s_k^T \nabla^2 h_i(y_k)s_k + o(\|s_k\|^2) = 0.$$

Logo,

$$\sum_{i=1}^{m} \mu_i^k \left[-h_i'(y_k)s_k + \frac{1}{2}s_k^T \nabla^2 h_i(y_k)s_k + o(\|s_k\|^2)\right] = 0,$$

isto é,

$$(\mu^k)^T h'(y_k)s_k = \frac{1}{2}s_k^T \left(\sum_{i=1}^{m} \mu_i^k \nabla^2 h_i(y_k)\right) s_k + (\mu^k)^T o(\|s_k\|^2). \tag{2.2.29}$$

Por (2.2.27) e (2.2.29) temos

$$g_k^T s_k = -s_k^T \left[B_k + \frac{1}{2}\sum_{i=1}^{m} \mu_i^k \nabla^2 h_i(y_k)\right] s_k - (\mu^k)^T o(\|s_k\|^2). \tag{2.2.30}$$

Pela fórmula de Taylor temos

$$f(y_k) = f(x_k) + g_k^T s_k + \frac{1}{2}s_k^T \nabla^2 f(x_k)s_k + o(\|s_k\|^2). \tag{2.2.31}$$

Então, por (2.2.30) e pela limitação de $\|\mu^k\|$,

$$f(y_k) = f(x_k) - \frac{1}{2}s_k^T \left[2B_k - \nabla^2 f(x_k) + \sum_{i=1}^{m} \mu_i^k \nabla^2 h_i(y_k)\right] s_k + o(\|s_k\|^2). \tag{2.2.32}$$

Mas, pela condição (2.2.23), $\|(B_k - \nabla^2 f(x_k))s_k\| = o(\|s_k\|)$.

Logo, de (2.2.32) segue que

$$f(y_k) = f(x_k) - \frac{1}{2}s_k^T \left[B_k + \sum_{i=1}^{m} \mu_i^k \nabla^2 h_i(y_k)\right] s_k + o(\|s_k\|^2). \tag{2.2.33}$$

Portanto, como $h \in C^3(\mathbb{R}^n)$ e pela limitação de $\|\mu^k\|$,

$$\begin{aligned} f(y_k) &= f(x_k) - \frac{\overline{\alpha}}{2}s_k^T \left[B_k + \sum_{i=1}^{m} \mu_i^k \nabla^2 h_i(x_k)\right] s_k \\ &\quad - \frac{(1-\overline{\alpha})}{2}s_k^T \left[B_k + \sum_{i=1}^{m} \mu_i^k \nabla^2 h_i(x_k)\right] s_k + o(\|s_k\|^2) \end{aligned} \tag{2.2.34}$$

onde $\overline{\alpha} \in (\alpha, 1)$.

Mas, pelo Lema 2.2.4, existem $c_2 > 0$ e $k_0 \in \mathbb{N}$ tais que para todo $k \geq k_0$,

$$\frac{(1-\overline{\alpha})}{2}s_k^T \left[B_k + \sum_{i=1}^{m} \mu_i^k \nabla^2 h_i(x_k)\right] s_k \geq c_2\|s_k\|^2. \tag{2.2.35}$$

Como $-c_2\|s_k\|^2 + o(\|s_k\|^2) < 0$ para k suficientemente grande, concluímos que, para k suficientemente grande,

$$f(y_k) \leq f(x_k) - \frac{\overline{\alpha}}{2}s_k^T \left[B_k + \sum_{i=1}^{m} \mu_i^k \nabla^2 h_i(x_k)\right] s_k. \tag{2.2.36}$$

Agora, por (2.2.30),

$$-\frac{1}{2}s_k^T \left[B_k + \sum_{i=1}^{m} \mu_i^k \nabla^2 h_i(x_k)\right] s_k = g_k^T s_k + \frac{1}{2}s_k^T B_k s_k + o(\|s_k\|^2). \tag{2.2.37}$$

Assim, por (2.2.36) e (2.2.37)

$$\begin{aligned} f(y_k) &\leq f(x_k) - \frac{\alpha}{2} s_k^T \left[B_k + \sum_{i=1}^{m} \mu_i^k \nabla^2 h_i(x_k) \right] s_k - \frac{\overline{\alpha} - \alpha}{2} s_k^T \left[B_k + \sum_{i=1}^{m} \mu_i^k \nabla^2 h_i(x_k) \right] s_k \\ &= f(x_k) + \alpha [g_k^T s_k + \frac{1}{2} s_k^T B_k s_k + o(\|s_k\|^2)] - \frac{\overline{\alpha} - \alpha}{2} s_k^T \left[B_k + \sum_{i=1}^{m} \mu_i^k \nabla^2 h_i(x_k) \right] s_k \\ &= f(x_k) + \alpha [g_k^T s_k + \frac{1}{2} s_k^T B_k s_k] - \frac{\overline{\alpha} - \alpha}{2} s_k^T \left[B_k + \sum_{i=1}^{m} \mu_i^k \nabla^2 h_i(x_k) \right] s_k + o(\|s_k\|^2) \end{aligned}$$

e (2.2.24) segue do Lema 2.2.4. □

2.3 O ALGORITMO RCMRI – RESULTADOS DE CONVERGÊNCIA.

Nesta seção introduzimos um algoritmo de região de confiança para resolver o problema MRI (2.2.1). A sigla RCMRI corresponde a Regiões de Confiança para Minimização com Restrições de Igualdade. Assumimos que $f, h \in C^3(\mathbb{R}^n)$.

Algoritmo 2.3.1 (RCMRI)

Seja $x_0 \in \mathbb{R}^n$ uma aproximação inicial, $h(x_0) = 0$. Sejam $\tau_1, \tau_2, \alpha, \Delta_{min}, \Delta^0$ tais que $0 < \tau_1 \leq \tau_2 < 1, \alpha \in (0,1), \Delta_{min} > 0, \Delta^0 \geq \Delta_{min}$. Dados $x_k \in \mathbb{R}^n$ tal que $h(x_k) = 0, \Delta^k \geq \Delta_{min}$ e $B_k \in \mathbb{R}^{n \times n}, B_k = B_k^T$, os passos para se obter x_{k+1} e Δ_k são os seguintes:

Passo 1. Faça $\Delta \leftarrow \Delta^k$

Passo 2. Calcule $\overline{s}_k(\Delta)$, uma solução global de

$$\begin{array}{ll} \min & \psi_k(s) \equiv \frac{1}{2} s^T B_k s + g_k^T s \\ \text{s/a} & h(x_k + s) = 0 \\ & \|s\| \leq \Delta \, . \end{array} \qquad (2.3.1)$$

Se $\psi_k(\bar{s}_k(\Delta)) = 0$, parar.

Passo 3. Se

$$f(x_k + \bar{s}_k(\Delta)) \leq f(x_k) + \alpha\psi_k(\bar{s}_k(\Delta)), \tag{2.3.2}$$

defina $x_{k+1} = x_k + \bar{s}_k(\Delta)$, $\Delta_k = \Delta$,

escolha $\Delta^{k+1} \geq \Delta_{min}$ e $B_{k+1} \in I\!R^{n\times n}$, $B_{k+1} = B_{k+1}^T$.

senão, faça $\Delta \leftarrow \Delta_{novo}$,

onde $\Delta_{novo} \in [\tau_1\|\bar{s}_k(\Delta)\|, \tau_2\Delta]$

volte para o Passo 2.

O restante desta seção é dedicado à prova de que todo ponto limite de uma seqüência gerada pelo Algoritmo 2.3.1 satisfaz condições de otimalidade.

DEFINIÇÃO 2.3.2. Dados $x \in I\!R^n$ tal que $h(x) = 0$ e $b > 0$, dizemos que $\gamma : [-b, b] \to I\!R^n$ é uma *curva factível passando por* x se

a) $h(\gamma(t)) = 0$ para todo $t \in [-b, b]$;

b) $\gamma \in C^3([-b, b])$, $\gamma'(0) \neq 0$;

c) $\gamma(0) = x$.

TEOREMA 2.3.3. Se x_* é um minimizador local de (2.2.1) então, para toda curva factível passando por x_*, temos

$$g(x_*)^T\gamma'(0) \equiv (f \circ \gamma)'(0) = 0 \tag{2.3.3}$$

e

$$(f \circ \gamma)''(0) \geq 0. \tag{2.3.4}$$

Demonstração. Trivial pois 0 é um minimizador local de $f \circ \gamma : [-b, b] \to I\!R$. □

O Teorema 2.3.3 motiva a seguinte definição:

DEFINIÇÃO 2.3.4. Dizemos que $x_* \in S$ é um *ponto estacionário* de (2.2.1) se para toda curva factível γ passando por x_* temos (2.3.3) e (2.3.4) satisfeitos.

No Teorema 2.3.5 a seguir mostramos que o Algoritmo 2.3.1 pára apenas se o ponto é estacionário.

TEOREMA 2.3.5. Se $B_k = \nabla^2 f(x_k)$ e o Algoritmo 2.3.1 pára no Passo 2 ($\psi_k(\overline{s}_k(\Delta) = 0$), então x_k é um ponto estacionário de (2.2.1).

Demonstração. Seja γ uma curva factível passando por x_k. Como $\psi_k(0) = 0 = \psi_k(\overline{s}_k(\Delta))$, segue que 0 é uma solução de (2.3.1). Sendo 0 um ponto interior da região factível de (2.3.1), temos $(\psi_k \circ \gamma)'(0) = 0$ e $(\psi_k \circ \gamma)''(0) \geq 0$. É fácil ver que estas duas condições implicam em (2.3.3) e (2.3.4). □

O próximo teorema estabelece que se o Algoritmo 2.3.1 não pára no Passo 2, então a k-ésima iteração é concluída num tempo finito.

TEOREMA 2.3.6. Se x_k não é um ponto estacionário de (2.2.1) e $B_k = \nabla^2 f(x_k)$, então x_{k+1} está bem definido pelo Algoritmo 2.3.1.

Antes de demonstrarmos o Teorema 2.3.6, precisamos de algumas definições e lemas técnicos.

DEFINIÇÃO 2.3.7. Dada γ, uma curva factível passando por x, definimos, para $\Delta \geq 0$,

$$\begin{aligned} \tau_+(\gamma,\Delta) &= \min\{t \in [0,b] \mid \|\gamma(t)-\gamma(0)\| = \Delta\}, \\ \tau_-(\gamma,\Delta) &= \max\{t \in [-b,0] \mid \|\gamma(t)-\gamma(0)\| = \Delta\}. \end{aligned}$$

LEMA 2.3.8. Vamos supor que $\gamma_k : [-b,b] \to \mathbb{R}^n, \gamma : [-b,b] \to \mathbb{R}^n, b > 0$, $\gamma_k, \gamma \in C^3([-b,b])$ para todo $k \in \mathbb{N}, \gamma'(0) \neq 0$ e $\lim_{k\to\infty} \|\gamma_k - \gamma\|_3 = 0$, onde

$$\|\beta\|_3 = \max\{\|\beta(t)\|, \|\beta'(t)\|, \|\beta''(t)\|, \|\beta'''(t)\| \mid t \in [-b,b]\}.$$

Então existem $c_3, c_4, \overline{\Delta} > 0, k_0 \in \mathbb{N}$ tais que $\tau_+(\gamma_k,\Delta), \tau_-(\gamma_k,\Delta), \tau_+(\gamma,\Delta), \tau_-(\gamma,\Delta)$ estão bem definidos e

$$\begin{array}{rcl} c_3\Delta \leq & \tau_+(\gamma_k,\Delta) & \leq c_4\Delta \\ c_3\Delta \leq & |\tau_-(\gamma_k,\Delta)| & \leq c_4\Delta \\ c_3\Delta \leq & \tau_+(\gamma,\Delta) & \leq c_4\Delta \\ c_3\Delta \leq & |\tau_-(\gamma,\Delta)| & \leq c_4\Delta \end{array} \tag{2.3.5}$$

para todo $\Delta \in [0,\overline{\Delta}], k \geq k_0$.

Demonstração. Segue fazendo-se uma pequena adaptação na prova do Lema 1.2.5. □

LEMA 2.3.9. Se $\overline{x}$ é um ponto regular de $S = \{x \in \mathbb{R}^n \mid h(x) = 0\}$ (Luenberger [1984], p. 314) então para toda curva factível $\gamma : [-b,b] \to S$ passando por $\overline{x}$ e para toda seqüência $\{x_k\}_{k=1}^{\infty} \subset S$ convergindo a $\overline{x}$ existem $b_1 \in (0,b)$ e $\gamma_k : [-b_1,b_1] \to S$ $(k \in \mathbb{N})$ seqüência de curvas factíveis passando por x_k tais que

$$\lim_{k\to\infty} \|\gamma_k - \gamma\|_3 = 0. \tag{2.3.6}$$

Demonstração. Este resultado é uma conseqüência do Teorema 1.2.7. □

Demonstração do Teorema 2.3.6. Como x_k não é um ponto estacionário, existe $\gamma : [-b, b] \to S$ curva factível passando por x_k tal que

$$(f \circ \gamma)'(0) = g(x_k)^T \gamma'(0) < 0 \tag{2.3.7}$$

ou

$$(f \circ \gamma)'(0) = 0, \quad (f \circ \gamma)''(0) < 0. \tag{2.3.8}$$

Se ocorre (2.3.7), a demonstração segue análoga à do Teorema 1.3.3. Resta considerar a possibilidade (2.3.8). Assim, temos

$$(f \circ \gamma)''(0) = \gamma'(0)^T \nabla^2 f(x_k) \gamma'(0) + g(x_k)^T \gamma''(0) = \quad a \quad < 0. \tag{2.3.9}$$

Seja $\overline{\Delta} > 0$ tal que $\tau_+(\Delta) \equiv \tau_+(\gamma, \Delta)$ e $\tau_-(\Delta) \equiv \tau_-(\gamma, \Delta)$ estão bem definidos e sejam $c_3, c_4 > 0$ tais que vale (2.3.5) para todo $\Delta \in [0, \overline{\Delta}]$. Do Passo 2 do Algoritmo 2.3.1 temos

$$\begin{aligned} \psi_k(\overline{s}_k(\Delta)) &\leq \psi_k(\gamma(t) - x_k) \\ &= \frac{1}{2}(\gamma(t) - x_k)^T \nabla^2 f(x_k)(\gamma(t) - x_k) + g(x_k)^T(\gamma(t) - x_k) \end{aligned} \tag{2.3.10}$$

onde $t = \tau_+(\Delta)$ ou $t = \tau_-(\Delta)$.

Agora, $\gamma(t) = \gamma(0) + t\gamma'(0) + \dfrac{t^2}{2}\gamma''(0) + o(t^2)$, portanto

$$\psi_k(\overline{s}_k(\Delta)) \leq \frac{t^2}{2}(\gamma'(0)^T\nabla^2 f(x_k)\gamma'(0) + g(x_k)^T\gamma''(0)) + tg(x_k)^T\gamma'(0) + o(t^2). \quad (2.3.11)$$

Mas, por (2.3.8), $g(x_k)^T\gamma'(0) = 0$ e por (2.3.5) temos

$$\frac{\psi_k(\overline{s}_k(\Delta))}{\Delta^2} \leq \frac{c_4^2\psi_k(\overline{s}_k(\Delta))}{t^2} \leq \frac{c_4^2}{2}(\gamma'(0)^T\nabla^2 f(x_k)\gamma'(0) + g(x_k)^T\gamma''(0)) + \frac{o(t^2)}{t^2}.$$

Por (2.3.9) segue que

$$\frac{\psi_k(\overline{s}_k(\Delta))}{\Delta^2} \leq \frac{ac_4^2}{2} + \frac{o(t^2)}{t^2}. \quad (2.3.12)$$

Então por (2.3.12) temos

$$\limsup_{\Delta\to 0} \frac{\psi_k(\overline{s}_k(\Delta))}{\Delta^2} \leq \frac{ac_4^2}{2} < 0 \quad (2.3.13)$$

e portanto existe $\overline{\overline{\Delta}} > 0$ tal que

$$\frac{\psi_k(\overline{s}_k(\Delta))}{\Delta^2} \leq c_5 = \frac{ac_4^2}{4} < 0 \quad (2.3.14)$$

para todo $\Delta \in (0, \overline{\overline{\Delta}}]$.

Definimos, para $\Delta > 0$,

$$\rho(\Delta) = \frac{f(x_k + \overline{s}_k(\Delta)) - f(x_k)}{\psi_k(\overline{s}_k(\Delta))}. \quad (2.3.15)$$

Então, se $\Delta \in (0, \overline{\overline{\Delta}}]$, por (2.3.14) e (2.3.15) temos

$$\begin{aligned} |\rho(\Delta) - 1| &= \left| \frac{f(x_k + \overline{s}_k(\Delta)) - f(x_k) - \psi_k(\overline{s}_k(\Delta))}{\psi_k(\overline{s}_k(\Delta))} \right| \\ &= \frac{o(\|\overline{s}_k(\Delta)\|^2)}{|\psi_k(\overline{s}_k(\Delta))|} = \frac{o(\Delta^2)}{\Delta^2} \frac{\Delta^2}{|\psi_k(\overline{s}_k(\Delta))|} \leq \frac{1}{|c_5|} \frac{o(\Delta^2)}{\Delta^2}. \end{aligned}$$

Portanto

$$\lim_{\Delta \to 0} \rho(\Delta) = 1. \tag{2.3.16}$$

Em decorrência de (2.3.16), após um número finito de reduções no raio de confiança, a condição (2.3.2) é satisfeita e então x_{k+1} está bem definido. □

Apresentamos a seguir o resultado de convergência global do Algoritmo 2.3.1.

TEOREMA 2.3.10. Suponhamos que a seqüência $\{x_k\}$ é gerada pelo Algoritmo 2.3.1 com $B_k \equiv \nabla^2 f(x_k), x_* \in S$ é regular e $\lim_{k \in I\!K_1} x_k = x_*$ onde $I\!K_1$ é um subconjunto infinito de $I\!N$. Então x_* é um ponto estacionário do problema (2.2.1).

Demonstração. Vamos considerar duas possibilidades:

$$\inf_{k \in I\!K_1} \Delta_k = 0 \tag{2.3.17}$$

ou

$$\inf_{k \in I\!K_1} \Delta_k > 0. \tag{2.3.18}$$

Assumimos inicialmente a validade de (2.3.17). Então existe $I\!K_2$, subconjunto infinito de $I\!K_1$, tal que

$$\lim_{k \in I\!K_2} \Delta_k = 0. \tag{2.3.19}$$

Logo, existe $k \in I\!N$ tal que $\Delta_k < \Delta_{min}$ para todo $k \geq k_2, k \in I\!K_2$. Mas, a cada iteração k tentamos inicialmente o raio $\Delta^k \geq \Delta_{min}$. Portanto, para todo $k \in I\!K_3 \equiv \{k \in I\!K_2 \mid k \geq k_2\}$, existem $\overline{\Delta}_k$ e $\overline{s}_k(\overline{\Delta}_k)$ tais que $\overline{s}_k(\overline{\Delta}_k)$ é uma solução global de:

$$\begin{array}{ll} \min & \psi_k(s) \\ \text{s/a} & h(x_k + s) = 0 \\ & \|s\| \leq \overline{\Delta}_k \end{array} \tag{2.3.20}$$

e

$$f(x_k + \overline{s}_k(\overline{\Delta}_k)) > f(x_k) + \alpha\psi_k(\overline{s}_k(\overline{\Delta}_k)). \tag{2.3.21}$$

Pela atualização do raio de confiança no Algoritmo (2.3.1), para $k \in I\!K_3$ temos:

$$\Delta_k > \tau_1 \|\overline{s}_k(\overline{\Delta}_k)\|. \tag{2.3.22}$$

Desta forma, por (2.3.19) e (2.3.22)

$$\lim_{k \in I\!K_3} \|\overline{s}_k(\overline{\Delta}_k)\| = 0. \tag{2.3.23}$$

Suponhamos que x_* não seja estacionário. Então existem $b > 0, \gamma : [-b, b] \to S$ uma curva factível passando por x_* tal que

$$(f \circ \gamma)'(0) = g(x_*)^T \gamma'(0) < 0 \tag{2.3.24}$$

ou

$$(f \circ \gamma)'(0) = g(x_*)^T \gamma'(0) = 0 \tag{2.3.25}$$

e

$$(f \circ \gamma)''(0) \equiv \gamma'(0)^T \nabla^2 f(x_*)\gamma'(0) + g(x_*)^T \gamma''(0) = a < 0. \tag{2.3.26}$$

Se ocorre (2.3.24), a prova segue a mesma estrutura do Teorema 1.4.1, onde foram consideradas condições de estacionaridade de primeira ordem. Assim, devemos nos deter ao caso em que ocorrem (2.3.25) e (2.3.26).

Como x_* é regular e $\lim_{k \in I\!K_3} = x_*$, pelo Lema 2.3.9 existem $b' \in (0, b)$ e $\gamma_k : [-b', b'] \to S$ $(k \in I\!K_3)$ seqüência de curvas factíveis passando por x_k tais que

$$\lim_{k \in I\!K_3} \|\gamma_k - \gamma\|_3 = 0. \tag{2.3.27}$$

Por (2.3.27) e pelo Lema 2.3.8, existem $k_3 \in I\!N$, $\overline{\Delta} > 0$ tais que $\tau_+(\gamma_k, \Delta), \tau_-(\gamma_k, \Delta), \tau_+(\gamma, \Delta)$ e $\tau_-(\gamma, \Delta)$ estão bem definidos para todo $k \in I\!K_4 \equiv \{k \in I\!K_3 \mid k \geq k_3\}, \Delta \in [0, \overline{\Delta}]$. Além disso, (2.3.5) vale para todo $k \in I\!K_4, \Delta \in [0, \overline{\Delta}]$. Seja $k_4 \in I\!N$ tal que

$$\|\overline{s}_k(\overline{\Delta}_k)\| \leq \overline{\Delta} \tag{2.3.28}$$

para todo $k \in I\!K_5 \equiv \{k \in I\!K_4 \mid k \geq k_4\}$. Há duas possibilidades para a definição de t_k:

$$\begin{aligned} & t_k = \tau_+(\gamma_k, \|\overline{s}_k(\overline{\Delta}_k)\|) \\ \text{ou} \quad & \\ & t_k = \tau_-(\gamma_k, \|\overline{s}_k(\overline{\Delta}_k)\|) \, . \end{aligned} \tag{2.3.29}$$

Faremos abaixo a escolha conveniente. De qualquer forma, pelo Lema 2.3.8, para todo $k \in I\!K_5$ temos

$$c_3 \|\overline{s}_k(\overline{\Delta}_k)\| \leq |t_k| \leq c_4 \|\overline{s}_k(\overline{\Delta}_k)\|. \tag{2.3.30}$$

Agora, como $\overline{s}_k(\overline{\Delta}_k)$ é minimizador global de (2.3.20),

$$\begin{aligned}\psi_k(\overline{s}_k(\overline{\Delta}_k)) \;\leq\; & \psi_k(\gamma_k(t_k)-\gamma_k(0) = \frac{t_k^2}{2}\Big[\frac{\gamma_k(t_k)-\gamma_k(0)}{t_k}\Big]^T\nabla^2 f(x_k)\Big[\frac{\gamma_k(t_k)-\gamma_k(0)}{t_k}\Big] \\ & + t_k g(x_k)^T\Big[\frac{\gamma_k(t_k)-\gamma_k(0)}{t_k}\Big]. \qquad (2.3.31)\end{aligned}$$

Mas, pelo teorema de Taylor, como por (2.3.6) as derivadas terceiras de γ_k são limitadas, temos

$$\frac{\gamma_k(t_k)-\gamma_k(0)}{t_k} = \gamma_k'(0) + \frac{t_k}{2}\gamma_k''(0) + \frac{o(t_k^2)}{t_k}. \qquad (2.3.32)$$

De (2.3.31) e (2.3.32) segue que

$$\psi_k(\overline{s}_k(\overline{\Delta}_k)) \leq \frac{t_k^2}{2}\Big[\gamma_k'(0)^T\nabla^2 f(x_k)\gamma_k'(0) + g(x_k)^T\gamma_k''(0)\Big] + t_k g(x_k)^T\gamma_k'(0) + o(t_k^2). \qquad (2.3.33)$$

Trocando t_k por τt_k, onde $\tau = \pm 1$ é escolhido de modo que

$$\tau t_k g(x_k)^T\gamma_k'(0) \leq 0, \qquad (2.3.34)$$

segue de (2.3.30), (2.3.33) e (2.3.34) que

$$\begin{aligned}\frac{\psi_k(\overline{s}_k(\overline{\Delta}_k))}{\|\overline{s}_k(\overline{\Delta}_k)\|^2} \;&\leq\; \epsilon_4^2\frac{\psi_k(\overline{s}_k(\overline{\Delta}_k))}{t_k^2} \\ &\leq\; \frac{c_4^2}{2}(\gamma_k'(0)^T\nabla^2 f(x_k)\gamma_k'(0) + g(x_k)^T\gamma_k''(0)) + \frac{o(t_k^2)}{t_k^2}. \qquad (2.3.35)\end{aligned}$$

Desta forma, por (2.3.23), (2.3.26) e (2.3.27),

$$\limsup_{k\in I\!K_5} \frac{\psi_k(\overline{s}_k(\overline{\Delta}_k))}{\|\overline{s}_k(\overline{\Delta}_k)\|^2} \leq \frac{c_4^2}{2}(\gamma_k'(0)^T\nabla^2 f(x_k)\gamma_k'(0) + g(x_k)^T\gamma_k''(0)) = \frac{ac_4^2}{2} < 0.$$

Portanto, existe $k_5 \in I\!N$ tal que para todo $k \in I\!K_6 \equiv \{k \in I\!K_5 | k \geq k_5\}$ temos

$$\frac{\psi_k(\overline{s}_k(\overline{\Delta}_k))}{\|\overline{s}_k(\overline{\Delta}_k)\|^2} \leq c_6 \equiv \frac{ac_4^2}{4} < 0. \tag{2.3.36}$$

Definimos, para $k \in I\!K_6$,

$$\overline{\rho}_k = \frac{f(x_k + \overline{s}_k(\overline{\Delta}_k)) - f(x_k)}{\psi_k(\overline{s}_k(\overline{\Delta}_k))}. \tag{2.3.37}$$

Então, por (2.3.36),

$$\begin{aligned} |\overline{\rho}_k - 1| &= \left|\frac{f(x_k + \overline{s}_k(\overline{\Delta}_k)) - f(x_k) - \psi_k(\overline{s}_k(\overline{\Delta}_k))}{\psi_k(\overline{s}_k(\overline{\Delta}_k))}\right| \\ &= \frac{o(\|\overline{s}_k(\overline{\Delta}_k)\|^2)}{|\psi_k(\overline{s}_k(\overline{\Delta}_k))|} = \frac{o(\|\overline{s}_k(\overline{\Delta}_k)\|^2)}{\|\overline{s}_k(\overline{\Delta}_k)\|^2} \frac{\|\overline{s}_k(\overline{\Delta}_k)\|^2}{|\psi_k(\overline{s}_k(\overline{\Delta}_k))|} \leq \frac{1}{|c_6|} \frac{o(\|\overline{s}_k(\overline{\Delta}_k)\|^2)}{\|\overline{s}_k(\overline{\Delta}_k)\|^2}. \end{aligned}$$

Logo,

$$\lim_{k\in K_6} \overline{\rho}_k = 1. \tag{2.3.38}$$

Como (2.3.38) contradiz (2.3.21), x_k é estacionário neste caso.

Suponhamos agora a validade de (2.3.18). Como $\lim_{k\in\mathbb{K}_1} x_k = x_*$ e $\{f(x_k)\}$ é estritamente decrescente, temos que

$$\lim_{k\in\mathbb{K}_1}(f(x_{k+1}) - f(x_k)) = 0. \tag{2.3.39}$$

Mas, por (2.3.2),

$$f(x_{k+1}) \leq f(x_k) + \alpha\psi_k(\bar{s}_k(\Delta_k)). \tag{2.3.40}$$

Logo, de (2.3.39) e (2.3.40) segue que

$$\lim_{k\in\mathbb{K}_1} \psi_k(\bar{s}_k(\Delta_k)) = 0. \tag{2.3.41}$$

Definimos $\underline{\Delta} = \inf_{k\in\mathbb{K}_1} \Delta_k > 0$ e s_* solução global de

$$\begin{array}{ll} \min & \frac{1}{2}s^T\nabla^2 f(x_*)s + g(x_*)^T s \\ \text{s/a} & h(x_* + s) = 0 \\ & \|s\| \leq \underline{\Delta}/2\ . \end{array} \tag{2.3.42}$$

Seja $k_6 \in \mathbb{K}_1$ tal que para todo $k \in \mathbb{K}_7 \equiv \{k \in \mathbb{K}_1 \mid k \geq k_6\}$ temos

$$\|x_k - x_*\| \leq \underline{\Delta}/2. \tag{2.3.43}$$

Definimos, para $k \in \mathbb{K}_7$,

$$\hat{s}_k = x_* + s_* - x_k. \tag{2.3.44}$$

Por (2.3.42) e (2.3.43) temos, para todo $k \in \mathbb{K}_7$,

$$\|\hat{s}_k\| \leq \underline{\Delta} \leq \Delta_k. \tag{2.3.45}$$

Além disso,

$$x_k + \hat{s}_k = x_* + s_* \in S. \tag{2.3.46}$$

De (2.3.45), (2.3.46) e (2.3.1) segue que

$$\psi_k(\overline{s}_k(\Delta_k)) \leq \psi_k(\hat{s}_k) \tag{2.3.47}$$

para todo $k \in I\!K_7$. Logo, por (2.3.41), (2.3.44) e (2.3.47),

$$\begin{aligned} \frac{1}{2}s_*^T\nabla^2 f(x_*)s_* + g(x_*)^T s_* &= \lim_{k\in I\!K_7}\left[\frac{1}{2}\hat{s}_k^T\nabla^2 f(x_k)\hat{s}_k + g(x_k)^T\hat{s}_k\right] \\ &\geq \lim_{k\in I\!K_7}\psi_k(\overline{s}_k(\Delta_k)) = 0. \end{aligned}$$

Portanto, 0 é minimizador de (2.3.42). Assim, x_* é estacionário neste caso e a prova está completa. □

TEOREMA 2.3.11. Suponhamos válidas as hipóteses do Teorema 2.3.10. Se x_* é um ponto limite de $\{x_k\}$ que satisfaz as Hipóteses Locais Gerais da Seção 2.2, então toda seqüência $\{x_k\}$ converge para x_* e existe $c > 0$ tal que vale (2.2.22).

Demonstração. Como x_* satisfaz as condições suficientes para um minimizador local estrito, existe $\varepsilon_1 > 0$ tal que x_* é o único ponto limite de $\{x_k\}$ no conjunto $\{x \in S \mid \|x - x_*\| \leq \varepsilon_1\}$. Seja $\varepsilon_2 \in (0, \varepsilon_1)$. Por (2.2.17), existe $\varepsilon_3 \in (0, \varepsilon_2)$ tal que

$$\|\Phi(x, \nabla^2 f(x)) - x\| < \varepsilon_1 - \varepsilon_2 \tag{2.3.48}$$

sempre que $\|x - x_*\| \leq \varepsilon_3$. Definimos

$$\underline{m} = \min\{f(x) \mid x \in S, \varepsilon_3 \leq \|x - x_*\| \leq \varepsilon_1\} \text{ e}$$

$$U = \{x \in S \mid \|x - x_*\| < \varepsilon_1 \text{ e } f(x) < \underline{m}\}.$$

Claramente U é um conjunto aberto, $x_* \in U$ e $\|x - x_*\| < \varepsilon_3$ para todo $x \in U$. Como x_* é um ponto limite de $\{x_k\}$, existe $k_0 \in I\!N$ tal que $x_{k_0} \in U$. Agora, por (2.3.48) e pela definição do Algoritmo 2.3.1,

$$\|x_{k_0+1} - x_{k_0}\| \leq \|\Phi(x_{k_0}, \nabla^2 f(x_{k_0})) - x_{k_0}\| \leq \varepsilon_1 - \varepsilon_2. \tag{2.3.49}$$

Desta forma, $\|x_{k_0+1} - x_*\| \leq \|x_{k_0} - x_*\| + \|x_{k_0+1} - x_{k_0}\| < \varepsilon_1$.

Pela definição do algoritmo, $f(x_{k_0+1}) < \underline{m}$, portanto $x_{k_0+1} \in U$. Por um argumento de indução, podemos mostrar que $x_k \in U$ para todo $k \geq k_0$. Assim, a seqüência converge para x_*.

Agora, por (2.2.17), $\lim_{k\to\infty} \|\Phi(x_k, \nabla^2 f(x_k)) - x_k\| = 0$. Então, existe $k_1 \in I\!N$ tal que $\|\Phi(x_k, \nabla^2 f(x_k)) - x_k\| < \Delta_{min}$ para todo $k \geq k_1$.

Portanto, para $k \geq k_1$, o primeiro ponto $\overline{s}_k(\Delta)$ a ser testado no Passo 3 do Algoritmo 2.3.1 é tal que $\|\overline{s}_k(\Delta)\| = \|\Phi(x_k, \nabla^2 f(x_k)) - x_k\|$.

Mas, pelo Teorema 2.2.7, existe $k_2 > k_1$ tal que este incremento tentativo satisfaz (2.3.2) para todo $k \geq k_2$. Isto significa que o Algoritmo 2.3.1 coincide com o Algoritmo Local 2.2.1 para todo $k \geq k_2$. Assim, o resultado desejado segue do Teorema 2.2.6. □

2.4 REGIÕES DE CONFIANÇA EM ESFERAS EUCLIDIANAS.

Nesta seção estudamos o caso em que

$$S = \{x \in I\!R^n \mid \|x\| = R\} \tag{2.4.1}$$

onde $R > 0$ e $\|\cdot\|$ é a norma euclidiana em $I\!R^n$.

Para calcularmos $\overline{s}_k(\Delta)$ no Passo 2 do Algoritmo 2.3.1, vamos considerar o seguinte problema:

$$\begin{array}{ll} \min & \psi_k(s) \\ \text{s/a} & \|s + x_k\| = R \\ & \|s\| \leq \Delta \, . \end{array} \tag{2.4.2}$$

Analogamente ao que fizemos na Seção 1.5, vamos considerar os seguintes subproblemas relacionados com (2.4.2):

$$\begin{array}{ll} \min & \psi_k(s) \\ \text{s/a} & \|s + x_k\| = R \end{array} \tag{2.4.3}$$

e

$$\begin{array}{ll} \min & \psi_k(s) \\ \text{s/a} & \|s + x_k\| = R \\ & \|s\| = \Delta \, . \end{array} \tag{2.4.4}$$

Com base nos subproblemas (2.4.3) e (2.4.4), o algoritmo a seguir calcula $\overline{s}_k(\Delta)$ solução global de (2.4.2) usando essencialmente soluções globais e locais de (2.4.3) e, se necessário, a solução global de 2.4.4. A sigla MINESF significa Minimização em Esferas.

Algoritmo 2.4.1 (MINESF)

Passo 1. Defina s_G a solução global de (2.4.3) que minimiza $\|s\|$.
Se $\|s_G\| \leq \Delta$, defina $\overline{s}_k(\Delta) = s_G$ e pare.

Passo 2. Faça $\mathcal{C} = \phi$
Se existe s_L solução local-não global de (2.4.3) e $\|s_L\| \leq \Delta$, faça

$\mathcal{C} = \{s_L\}$.

Passo 3. Calcule $\hat{s}$ solução global de (2.4.4).
Faça $\mathcal{C} \leftarrow \mathcal{C} \cup \{\hat{s}\}$.

Passo 4. Defina

$$\overline{s}_k(\Delta) = \arg\ \min\{\psi_k(s) | s \in \mathcal{C}\}.$$

TEOREMA 2.4.2. O Algoritmo 2.4.1 está bem definido e calcula um minimizador global $\overline{s}_k(\Delta)$ de (2.4.2).

Demonstração. Segue do Teorema 1.5.8, considerando-se que o Algoritmo 2.4.1 é uma versão simplificada do Algoritmo 1.5.7. □

2.5 OBSERVAÇÕES FINAIS.

Neste capítulo introduzimos um método de região de confiança para problemas de minimização com restrições de igualdade (MRI), onde não se utiliza linearização das restrições. Na verdade, a abordagem deste método dá continuidade à do Algoritmo RCARB, introduzido no Capítulo 1, para problemas com restrições arbitrárias. Isto porque no caso do problema MRI, além de provarmos convergência global de segunda ordem, também apresentamos resultados de convergência local usando a teoria dos métodos quase-Newton de ponto fixo introduzida recentemente por Martínez [1992]. A classe de problemas para a qual esta nova abordagem se aplica, no entanto, ainda é limitada pela dificuldade de se resolver os subproblemas. De qualquer maneira, tendo em vista as considerações apresentadas no Capítulo 1, propusemos um algoritmo para resolver problemas com uma restrição do tipo esfera euclidiana. Tais problemas aparecem de maneira natural na regularização de problemas mal-postos (Vogel [1990]) e ainda em formulações para o Problema do Vetor Inicial na teoria de codificação, conforme detalhado no Capítulo 5 deste trabalho.

CAPÍTULO 3

MÉTODOS DE REGIÃO DE CONFIANÇA PARA MINIMIZAÇÃO COM VARIÁVEIS CANALIZADAS

3.1 INTRODUÇÃO

Neste capítulo consideramos o problema

$$\begin{array}{ll} \min & f(x) \\ \text{s/a} & \ell \leq x \leq u \end{array} \tag{3.1.1}$$

onde $f : \mathbb{R}^n \to \mathbb{R}$ é diferenciável no conjunto factível $\{x \in \mathbb{R}^n \mid \ell \leq x \leq u\}$, que não precisa ser limitado. Estamos particularmente interessados no caso em que n é grande. Problemas de grande porte com variáveis canalizadas como (3.1.1) aparecem com freqüência em aplicações. Além disso, a importância de se desenvolver algoritmos eficientes para (3.1.1) vem sendo reforçada nos últimos anos por Conn, Gould e Toint [1988a-b, 1989, 1990]. Estes autores mostraram que praticamente qualquer problema de programação não linear de grande porte pode ser resolvido eficientemente através de técnicas de Lagrangiano Aumentado, desde que se tenha à disposição um bom método para se resolver (3.1.1).

Propomos um algoritmo de região de confiança para resolver (3.1.1) (ver Dennis e Schnabel [1983], Fletcher [1987], Moré [1983], Sorensen [1982]). A cada iteração de nosso algoritmo consideramos o problema de minimizar uma quadrática (não necessariamente

convexa) na caixa resultante da interseção da caixa original ($\ell \leq x \leq u$) com a região de confiança definida pela norma infinito.

Conforme é usual nos métodos de região de confiança atuais, não é necessário resolver precisamente o subproblema para se obter convergência global do algoritmo principal (ver Conn, Gould e Toint [1988a]). Na verdade, uma condição mais fraca relacionando o valor do modelo quadrático com a solução de um subproblema auxiliar bastante simples é suficiente para provar que todo ponto limite é estacionário. A solução do subproblema auxiliar faz o papel do *"ponto de Cauchy"* ou do *"ponto de Cauchy aproximado"* considerado em outros métodos de região de confiança para problemas com restrições (ver Conn, Gould e Toint [1988a]). Apesar de sua simplicidade, mostramos que este ponto auxiliar não só se constitui num ingrediente necessário para a prova de convergência do algoritmo, como também possui a propriedade de identificar o conjunto ativo correto, sob as mesmas condições que o ponto de Cauchy e o ponto de Cauchy aproximado o fazem.

Nos trabalhos de Conn, Gould e Toint [1988a-b] é enfatizado que o ponto de Cauchy aproximado tem boas propriedades de identificação na prática. Assim, no método de região de confiança proposto por eles para resolver problemas com restrições do tipo caixa é imposto que as restrições que são ativas no ponto de Cauchy aproximado também devem ser ativas no ponto tentativo considerado a cada iteração. Esta exigência (necessária para os resultados de identificação mas não para a convergência) simplifica o processo de se encontrar um ponto tentativo, pois viabiliza a aplicação do algoritmo de Gradientes Conjugados restrito ao conjunto ativo definido pelo ponto de Cauchy aproximado, truncando-se o processo se alguma restrição é violada.

Nossa abordagem difere da de Conn, Gould and Toint no aspecto discutido acima. Na verdade, não podemos afirmar que nosso *"ponto auxiliar"* tenha boas propriedades de identificação nas iterações iniciais. Neste sentido, a imposição de que se mantenha o conjunto ativo do ponto auxiliar não é razoável em nosso caso, sobretudo quando a aproximação para a solução ainda é ruim. Em conseqüência deste fato, precisamos de um algoritmo mais poderoso para a minimização de quadráticas (não necessariamente convexas) em caixas, que seja capaz de incorporar ou descartar restrições eficientemente. É por esta razão que investimos no desenvolvimento de um algoritmo eficiente para a minimizar uma quadrática com variáveis canalizadas. Este algoritmo generaliza um método introduzido por Friedlander e Martínez [1994] e combina técnicas de gradiente projetado com uma estratégia de restrições ativas (ver também Dembo e Tulowitzki [1987], Moré e Toraldo [1989, 1991]). Apresentamos um resultado de terminação finita sem hipóteses de não-degeneração e verificamos experimentalmente que apenas algumas iterações são suficientes para identificar um grande número de restrições ativas corretas.

Este capítulo está organizado da seguinte maneira: na Seção 3.2 descrevemos o algoritmo principal (BOX), provamos que está bem definido e que todo ponto limite é um ponto de Karush-Kuhn-Tucker para (3.1.1). Na Seção 3.3 mostramos que, assumindo-se uma condição de complementaridade estrita, o conjunto das restrições ativas é identificado após um número finito de iterações. As Seções 3.4, 3.5 e 3.6 podem ser lidas independentemente do restante deste capítulo. Na Seção 3.4 introduzimos um novo algoritmo (QUACAN) para minimizar uma quadrática (não necessariamente convexa) com variáveis canalizadas. A prova da terminação finita é feita na Seção 3.5. O algoritmo QUACAN é capaz de encontrar a solução global no caso convexo, ou, no caso geral, um ponto em que o gradiente projetado é pequeno dentro de uma tolerância pré-estabelecida. Na Seção 3.6 descrevemos algumas características importantes relativas à implementação do algoritmo QUACAN. Dedicamos a Seção 3.7 à descrição dos experimentos numéricos. Finalmente, na Seção 3.8 fazemos algumas observações finais.

3.2 O ALGORITMO BOX - RESULTADOS DE CONVERGÊNCIA.

Vamos considerar o problema (3.1.1) com f continuamente diferenciável para todo $x \in \mathbb{R}^n$ tal que $\ell \leq x \leq u$. Denotamos $g(x) = \nabla f(x)$.

Nesta seção e na próxima, $\|\cdot\| = \|\cdot\|_2$ e $\|\cdot\|_\#$ denota uma norma arbitrária em $\mathbb{R}^n$, bem como a norma matricial subordinada. Dada uma matriz simétrica B, denotamos por $\lambda_1(B) \leq \lambda_2(B) \leq \ldots \leq \lambda_n(B)$ os autovalores de B.

Sejam $\tau_1, \tau_2, \theta, \Delta_{min}$ e γ tais que $0 < \tau_1 \leq \tau_2 < 1, \theta \in (0,1], \Delta_{min} > 0$ e $\gamma \in (0,1]$. No início do algoritmo de região de confiança temos um ponto inicial factível arbitrário x_0, uma matriz simétrica B_0 $n \times n$ (aproximação para a Hessiana), uma matriz não-singular D_0 $n \times n$ (matriz de escalamento) e um raio inicial $\Delta^0 \geq \Delta_{min}$. O papel das matrizes de escalamento neste algoritmo é não só conferir um caráter mais geral à região de confiança adotada como também permitir um ajuste na ordem de grandeza de variáveis com magnitudes muito diferentes. Dados um ponto factível x_k, B_k simétrica, D_k não singular e $\Delta^k \geq \Delta_{min}$, os passos para se obter x_{k+1} e Δ_k são dados pelo seguinte algoritmo:

Algoritmo 3.2.1 (BOX)

Passo 1. Faça $\Delta \leftarrow \Delta^k$ e calcule $M_k > 0$ tal que

$$\lambda_n(B_k) \leq M_k \tag{3.2.1}$$

Passo 2. Calcule uma solução global $z_k^Q(\Delta)$ para

$$\begin{array}{ll} \min & Q_k(z) \equiv \frac{1}{2} M_k \|z\|^2 + g_k^T z \\ \text{s/a} & \ell \leq x_k + z \leq u \\ & \|D_k z\|_\# \leq \Delta \,, \end{array} \tag{3.2.2}$$

onde $g_k = g(x_k)$. Se $Q_k(z_k^Q(\Delta)) = 0$, parar.

Passo 3. Calcule $\overline{z}_k(\Delta)$ tal que

$$\left.\begin{array}{c} \psi_k(\overline{z}_k(\Delta)) \leq \gamma Q_k(z_k^Q(\Delta)) \\ \ell \leq x_k + \overline{z}_k(\Delta) \leq u \\ \|D_k \overline{z}_k(\Delta)\|_\# \leq \Delta \end{array}\right\}, \tag{3.2.3}$$

onde

$$\psi_k(z) = \frac{1}{2} z^T B_k z + g_k^T z \tag{3.2.4}$$

para todo $z \in \mathbb{R}^n$.

Passo 4. Se

$$f(x_k + \overline{z}_k(\Delta)) \leq f(x_k) + \theta \psi_k(\overline{z}_k(\Delta)) \tag{3.2.5}$$

então defina $x_{k+1} = x_k + \overline{z}_k(\Delta), \Delta_k = \Delta$,

escolha $\Delta^{k+1} \geq \Delta_{min}$, $B_{k+1} \in \mathbb{R}^{n \times n}$, $B_{k+1} = B_{k+1}^T$ e retorne

senão, $\Delta \leftarrow \Delta_{novo}$ onde

$$\Delta_{novo} \in [\tau_1 \|D_k \overline{z}_k(\Delta)\|_\#, \tau_2 \Delta] \tag{3.2.6}$$

e volte para o Passo 2.

Observamos que no Passo 3 existe $\overline{z}_k(\Delta)$ satisfazendo (3.2.3) já que $z_k^Q(\Delta)$ é uma escolha possível.

LEMA 3.2.2. Se o Algoritmo 3.2.1 pára no Passo 2 ($Q_k(z_k^Q(\Delta)) = 0$) então x_k é um ponto de Karush-Kuhn-Tucker (Luenberger [1984], p. 314) para o problema (3.1.1).

Demonstração. Se $Q_k(z_k^Q(\Delta)) = 0$ então 0 é um minimizador de (3.2.2). Logo, 0 satisfaz as condições de Karush-Kuhn-Tucker para (3.2.2). Mas isto claramente equivale a dizer que x_k satisfaz as condições de Karush-Kuhn-Tucker para (3.1.1) e portanto a prova está completa. □

TEOREMA 3.2.3. O Algoritmo 3.2.1 está bem definido. Em outras palavras, se o processo não pára no Passo 2 (com $Q_k(z_k^Q(\Delta)) = 0$) então x_{k+1} pode ser calculado repetindo-se os Passos 2–4 um número finito de vezes.

Demonstração. Se x_k não é um ponto estacionário (Karush-Kuhn-Tucker) para (3.1.1), existe
$d \in \mathbb{R}^n, d \neq 0$ tal que d é uma direção factível e de descida. Assim,

$$\ell \leq x_k + \lambda d \leq u \tag{3.2.7}$$

para todo $\lambda \in [0, 1]$ e

$$g_k^T d < 0. \tag{3.2.8}$$

Logo, para $\Delta > 0$ suficientemente pequeno, por (3.2.2) temos que:

$$\begin{aligned} Q_k(z_k^Q(\Delta)) &\leq Q_k\Big(\frac{\Delta d}{\|D_k d\|_\#}\Big) \\ &= \frac{1}{2} M_k \frac{\|d\|^2}{\|D_k d\|_\#^2} \Delta^2 + \frac{\Delta g_k^T d}{\|D_k d\|_\#}. \end{aligned}$$

Mas $\|D_k\|_\| d\|_\# \geq \|D_k d\|_\# \geq \dfrac{\|d\|_\#}{\|D_k^{-1}\|_\#}$. Portanto,

$$Q_k(z_k^Q(\Delta)) \leq \frac{1}{2} \frac{M_k \|d\|^2 \Delta^2 \|D_k^{-1}\|_\#^2}{\|d\|_\#^2} + \frac{\Delta g_k^T d}{\|D_k d\|_\#}$$

e

$$\frac{Q_k(z_k^Q(\Delta))}{\Delta} \leq \frac{1}{2} \frac{M_k \|D_k^{-1}\|_\#^2 \Delta}{\|d\|_\#^2} \|d\|^2 + \frac{g_k^T d}{\|D_k\|_\# \|d\|_\#}.$$

Desta forma,

$$\limsup_{\Delta \to 0} \frac{Q_k(z_k^Q(\Delta))}{\Delta} \leq \frac{g_k^T d}{\|D_k\|_\# \|d\|_\#} < 0. \tag{3.2.9}$$

Assim, por (3.2.3),

$$\limsup_{\Delta \to 0} \frac{\psi_k(\overline{z}_k(\Delta))}{\Delta} \leq \frac{\gamma g_k^T d}{\|D_k\|_\# \|d\|_\#} < 0. \tag{3.2.10}$$

Seja $\overline{\Delta} > 0$ tal que

$$\frac{\psi_k(\overline{z}_k(\Delta))}{\Delta} \leq \frac{\gamma g_k^T d}{2\|D_k\|_\# \|d\|_\#} = c_1 < 0 \tag{3.2.11}$$

para todo $\Delta \in (0, \overline{\Delta}]$.

Definimos, para todo $\Delta \in (0, \overline{\Delta}]$,

$$\rho(\Delta) = \frac{f(x_k + \overline{z}_k(\Delta)) - f(x_k)}{\psi_k(\overline{z}_k(\Delta))}. \tag{3.2.12}$$

Então, por (3.2.11) e (3.2.12),

$$\begin{aligned} |\rho(\Delta) - 1| &= \left| \frac{f(x_k + \overline{z}_k(\Delta)) - f(x_k) - \psi_k(\overline{z}_k(\Delta))}{\psi_k(\overline{z}_k(\Delta))} \right| \\ &\leq \left| \frac{f(x_k + \overline{z}_k(\Delta)) - f(x_k) - g_k^T \overline{z}_k(\Delta)}{c_1 \Delta} \right| + \left| \frac{\overline{z}_k(\Delta)^T B_k \overline{z}_k(\Delta)}{2c_1 \Delta} \right| \\ &\leq \left| \frac{f(x_k + \overline{z}_k(\Delta)) - f(x_k) - g_k^T \overline{z}_k(\Delta)}{c_1 \|D_k \overline{z}_k(\Delta)\|_\#} \right| + \frac{\|D_k^{-1}\|^2 M_k c \Delta}{2|c_1|}, \end{aligned} \tag{3.2.13}$$

onde c é uma constante que depende da relação entre $\|\cdot\|$ e $\|\cdot\|_\#$.

Desta forma, pela diferenciabilidade de f,

$$\lim_{\Delta \to 0} \rho(\Delta) = 1. \tag{3.2.14}$$

A equação (3.2.14) nos diz que após um número finito de reduções no raio de confiança (3.2.6), a condição (3.2.5) é verificada. Portanto, x_{k+1} está bem definido. □

Observação. É interessante notar que se $z_u = -g_k/M_k$, então $\|z - z_u\|^2 = \frac{2}{M_k} Q_k(z) + \|z_u\|^2$. Assim, $z_k^Q(\Delta)$ é a projeção euclidiana de z_u na região factível de (3.2.2). O cálculo desta projeção é trivial em muitas situações práticas, por exemplo, se $\|\cdot\|_\# = \|\cdot\|_\infty$ e D_k é diagonal. Isto sugere que $\|\cdot\|_\infty$ seja a escolha mais natural para $\|\cdot\|_\#$.

A convergência global do Algoritmo 3.2.1 é provada no teorema a seguir.

TEOREMA 3.2.4. Vamos supor que $\{x_k\}$ é uma seqüência infinita gerada pelo Algoritmo 3.2.1, $\mathbb{K}_1$ é um conjunto infinito de índices tal que $\lim_{k \in \mathbb{K}_1} x_k = x_*$ e $M_k, \|D_k\|_\#, \|D_k^{-1}\|_\#$ e $|\lambda_1(B_k)|$ são limitados para $k \in \mathbb{K}_1$. Então x_* é um ponto estacionário (Karush-Kuhn-Tucker) para (3.1.1).

Demonstração. Devemos considerar duas possibilidades:

$$\inf_{k \in \mathbb{K}_1} \Delta_k = 0 \tag{3.2.15}$$

ou

$$\inf_{k \in \mathbb{K}_1} \Delta_k > 0. \tag{3.2.16}$$

Assumimos inicialmente que vale (3.2.15). Então existe $\mathbb{K}_2$ subconjunto infinito de $\mathbb{K}_1$ tal que

$$\lim_{k \in \mathbb{K}_2} \Delta_k = 0. \tag{3.2.17}$$

Assim, existe $k_2 \in \mathbb{K}_2$ tal que $\Delta_k < \Delta_{min}$ para todo $k \geq k_2, k \in \mathbb{K}_2$. Mas, a cada iteração k, tentamos inicialmente um raio $\Delta^k \geq \Delta_{min}$. Então, para todo $k \in \mathbb{K}_3 \equiv \{k \in \mathbb{K}_2 \mid k \geq k_2\}$ existem $\overline{\Delta}_k, z_k^Q(\overline{\Delta}_k)$ e $\overline{z}_k(\overline{\Delta}_k)$ tais que $z_k^Q(\overline{\Delta}_k)$ é uma solução de

$$\begin{array}{ll} \min & Q_k(z) \\ \text{s/a} & \ell \leq x_k + z \leq u \\ & \|D_k z\|_\# \leq \overline{\Delta}_k \,, \end{array} \tag{3.2.18}$$

$$\psi_k(\overline{z}_k(\overline{\Delta}_k)) \leq \gamma Q_k(z_k^Q(\overline{\Delta}_k)), \tag{3.2.19}$$

$$f(x_k + \overline{z}_k(\overline{\Delta}_k)) > f(x_k) + \theta\psi_k(\overline{z}_k(\overline{\Delta}_k)) \tag{3.2.20}$$

e, por (3.2.6),

$$\Delta_k \geq \tau_1 \|D_k \overline{z}_k(\overline{\Delta}_k)\|_{\#}. \tag{3.2.21}$$

Desta forma, por (3.2.17) e (3.2.21),

$$\lim_{k \in I\!K_3} \|D_k \overline{z}_k(\overline{\Delta}_k)\|_{\#} = 0. \tag{3.2.22}$$

Suponhamos que x_* não seja um ponto estacionário de (3.1.1). Então existe $d \in I\!R^n, d \neq 0$ tal que

$$\ell \leq x_* + \lambda d \leq u \tag{3.2.23}$$

para todo $\lambda \in [0,1]$, e

$$g(x_*)^T d < 0. \tag{3.2.24}$$

(3.2.23) implica na existência de $k_3 \in I\!K_3, k_3 \geq k_2$, tal que

$$\ell \leq x_k + \lambda \frac{d}{2} \leq u \tag{3.2.25}$$

para todo $k \in I\!K_4 \equiv \{k \in I\!K_3 \mid k \geq k_3\}, \lambda \in [0,1]$.

Vamos definir para todo $k \in I\!K_4$,

$$d_k = \frac{\|D_k \overline{z}_k(\overline{\Delta}_k)\|_{\#} d}{\|D_k d\|_{\#}}. \tag{3.2.26}$$

Como $\|D_k d\|_{\#} \geq \|d\|_{\#} / \|D_k^{-1}\|_{\#}$ e $\|D_k^{-1}\|_{\#}$ é limitado para $k \in I\!K_1$, por (3.2.22) e (3.2.25) existe $k_4 \in I\!K_4$ tal que

$$\ell \leq x_k + d_k \leq u \tag{3.2.27}$$

para $k \in I\!K_5 \equiv \{k \in I\!K_4 \mid k \geq k_4\}$.

Claramente, $\|D_k d_k\|_\# = \|D_k \overline{z}_k(\overline{\Delta}_k)\|_\# \leq \overline{\Delta}_k$. Então, por (3.2.2), (3.2.3) e (3.2.26),

$$\begin{aligned}\psi_k(\overline{z}_k(\overline{\Delta}_k)) &\leq \gamma Q_k(x_k^Q(\overline{\Delta}_k)) \leq \gamma Q_k(d_k) = \gamma\Big[\frac{1}{2}M_k\|d_k\|^2 + g_k^T d_k\Big] \\ &\leq \gamma\Big[\frac{1}{2}M_k\|D_k^{-1}\|^2\|D_k d_k\|^2 + \frac{\|D_k\overline{z}_k(\overline{\Delta}_k)\|_\#}{\|D_k d\|_\#} g_k^T d\Big].\end{aligned}$$

para todo $k \in I\!K_5$. Então,

$$\begin{aligned}\frac{\psi_k(\overline{z}_k(\overline{\Delta}_k))}{\|D_k\overline{z}_k(\overline{\Delta}_k)\|_\#} &= \frac{\psi_k(\overline{z}_k(\overline{\Delta}_k))}{\|D_k d_k\|_\#} \\ &\leq \gamma\Big[\frac{1}{2}M_k\frac{\|D_k^{-1}\|^2\|D_k d_k\|^2}{\|D_k d_k\|_\#} + \frac{g_k^T d}{\|D_k d\|_\#}\Big], \qquad (3.2.28)\end{aligned}$$

para todo $k \in I\!K_5$.

Desta forma, pela limitação de M_k e $\|D_k^{-1}\|_\#$, pela equivalência das normas em dimensão finita, pela continuidade de $g(x)$ e por (3.2.22) e (3.2.24), concluímos que existem $c_2 < 0$ e $k_5 \in I\!K_5$ tais que

$$\frac{\psi_k(\overline{z}_k(\overline{\Delta}_k))}{\|D_k\overline{z}_k(\overline{\Delta}_k)\|_\#} \leq c_2 < 0 \qquad (3.2.29)$$

para todo $k \in I\!K_6 \equiv \{k \in I\!K_5 \mid k \geq k_5\}$.

Então, pela limitação de $\|D_k\|_\#$, existe $c_3 < 0$ tal que para todo $k \in I\!K_6$,

$$\frac{\psi_k(\overline{z}_k(\overline{\Delta}_k))}{\|\overline{z}_k(\overline{\Delta}_k)\|_\#} \leq c_3 < 0. \qquad (3.2.30)$$

Definimos, para todo $k \in I\!K_6$,

$$\overline{\rho}_k = \frac{f(x_k + \overline{z}_k(\overline{\Delta}_k)) - f(x_k)}{\psi_k(\overline{z}_k(\overline{\Delta}_k))}. \qquad (3.2.31)$$

Temos que

$$\overline{\rho}_k - 1 = a_k + b_k$$

onde

$$a_k = \frac{f(x_k + \overline{z}_k(\overline{\Delta}_k)) - f(x_k) - g_k^T \overline{z}_k(\overline{\Delta}_k)}{\psi_k(\overline{z}_k(\overline{\Delta}_k))}$$

e

$$b_k = -\frac{1}{2} \frac{\overline{z}_k(\overline{\Delta}_k)^T B_k \overline{z}_k(\overline{\Delta}_k)}{\psi_k(\overline{z}_k(\overline{\Delta}_k))}.$$

Agora, por (3.2.30),

$$\begin{aligned} |a_k| &\leq \frac{|f(x_k + \overline{z}_k(\overline{\Delta}_k)) - f(x_k) - g_k^T \overline{z}_k(\overline{\Delta}_k)|}{|c_3| \|\overline{z}_k(\overline{\Delta}_k)\|_\#} \\ \text{e} & \\ |b_k| &\leq \frac{1}{2} \frac{\max\{|\lambda_1(B_k)|, M_k\}}{\|\overline{z}_k(\overline{\Delta}_k)\|^2} |c_3| \|\overline{z}_k(\overline{\Delta}_k)\|_\#. \end{aligned}$$

Então, pelo Teorema do Valor Médio, $\lim_{k \in \mathbb{K}_6} a_k = 0$. Além disso, por (3.2.30) e pela limitação de $|\lambda_1(B_k)|$ e M_k, $\lim_{k \in \mathbb{K}_6} b_k = 0$. Portanto $\overline{\rho}_k - 1$ tende para 0, o que contradiz (3.2.20). Desta forma, a possibilidade (3.2.15) não pode ocorrer se x_* não for estacionário.

Suponhamos agora a validade de (3.2.16). Como $\lim_{k \in \mathbb{K}_1} x_k = x_*$ e $f(x_k)$ é estritamente decrescente, temos

$$\lim_{k \in \mathbb{K}_1} [f(x_{k+1}) - f(x_k)] = 0. \tag{3.2.32}$$

Mas, por (3.2.2), (3.2.3) e (3.2.5),

$$\begin{aligned} f(x_{k+1}) &\leq f(x_k) + \theta \psi_k(\overline{z}_k(\Delta_k)) \\ &\leq f(x_k) + \theta \gamma Q_k(z_k^Q(\Delta_k)). \end{aligned}$$

Portanto,

$$\lim_{k\in I\!K_1} Q_k(z_k^Q(\Delta_k)) = 0. \tag{3.2.33}$$

Definimos $\underline{\Delta} = \inf_{k\in I\!K_1} \Delta_k > 0$. Sejam $M, R > 0$ tais que $M_k \leq M$ e $\|D_k\|_\# \leq R$ para todo $k \in I\!K_1$. Seja z_* uma solução do problema

$$\begin{array}{ll} \min & g(x_*)^T z + \frac{M}{2}\|z\|^2 \\ \text{s/a} & \ell \leq x_* + z \leq u \\ & \|z\|_\# \leq \underline{\Delta}/2R\,. \end{array} \tag{3.2.34}$$

Seja $k_6 \in I\!K_1$ tal que para todo $k \in I\!K_7 \equiv \{k \in I\!K_1 \mid k \geq k_6\}$,

$$\|x_k - x_*\|_\# \leq \underline{\Delta}/2R. \tag{3.2.35}$$

Definimos, para todo $k \in I\!K_7$,

$$\hat{z}_k = x_* - x_k + z_*. \tag{3.2.36}$$

De (3.2.34) e (3.2.35) temos

$$\begin{array}{rcl} \|D_k\hat{z}_k\|_\# & \leq & \|D_k\|_\#\|x_* - x_k + z_*\|_\# \\ & \leq & R[\|x_k - x_*\|_\# + \|z_*\|_\#] \leq \underline{\Delta} \leq \Delta_k \end{array} \tag{3.2.37}$$

para todo $k \in I\!K_7$.

Além disso, por (3.2.36), $x_k + \hat{z}_k = x_* + z_*$ e então, por (3.2.34),

$$\ell \leq x_k + \hat{z}_k \leq u. \tag{3.2.38}$$

Assim, por (3.2.37), (3.2.38) e (3.2.2),

$$Q_k(z_k^Q(\Delta_k)) \leq Q_k(\hat{z}_k) \tag{3.2.39}$$

para todo $k \in I\!K_7$.

Agora, por (3.2.36), pela definição de M, por (3.2.39) e (3.2.33),

$$g(x_*)^T z_* + \frac{M}{2}\|z_*\|^2 = \lim_{k \in I\!K_7} g_k^T \hat{z}_k + \frac{M}{2}\|\hat{z}_k\|^2 \geq$$
$$\geq \lim_{k \in I\!K_7} Q_k(\hat{z}_k) \geq \lim_{k \in I\!K_7} Q_k(z_k^Q(\Delta_k) = 0.$$

Desta forma, 0 é um minimizador de (3.2.34) e portanto x_* é um ponto estacionário de (3.1.1), o que completa a prova. □

3.3 IDENTIFICAÇÃO DAS RESTRIÇÕES ATIVAS

Uma característica esperada em qualquer algoritmo para resolver problemas de minimização com restrições é que haja identificação das restrições ativas em um número finito de iterações. Quando isso ocorre, o algoritmo acaba se reduzindo a um método de minimização irrestrita e então teoremas de convergência local (superlinear ou quadrática) podem ser aplicados. Nesta seção provamos um resultado de identificação relativo ao Algoritmo 3.2.1. Assumimos ao longo desta seção que $\|\cdot\|_\# = \|\cdot\|_\infty$ e que as matrizes D_k são diagonais.

Para todo $x \in I\!R^n$ tal que $\ell \leq x \leq u$, definimos $A(x) \subset \{1, 2, \ldots, 2n\}$, o conjunto de índices das restrições ativas em x, por

$$\left.\begin{array}{rl} & i \in A(x) \Leftrightarrow [x]_i = \ell_i \\ \text{e} & \\ & n+i \in A(x) \Leftrightarrow [x]_i = u_i \end{array}\right\}, \tag{3.3.1}$$

onde $[x]_i$ denota a i-ésima componente do vetor x.

O principal resultado desta seção é estabelecido pelo Teorema 3.3.5, para o qual os Lemas 3.3.1 e 3.3.4 são resultados preparatórios.

LEMA 3.3.1. Vamos supor que $\{x_k\}$ é uma seqüência infinita gerada pelo Algoritmo 3.2.1, $I\!K_1$ é um conjunto infinito de índices tal que $\lim_{k \in I\!K_1} x_k = x_*$ e

$M_k, \|D_k\|_\#, \|D_k^{-1}\|_\#, |\lambda_1(B_k)|$ são limitados. Suponhamos ainda que $[x_*]_i = \ell_i$ e $\dfrac{\partial f}{\partial x_i}(x_*) > 0$ (ou $[x_*]_i = u_i$ e $\dfrac{\partial f}{\partial x_i}(x_*) < 0$). Então, existe $k_1 \in I\!K_1$ tal que

$$[x_k + z_k^Q(\Delta_k)]_i = [x_*]_i \tag{3.3.2}$$

para todo $k \in I\!K_1, k \geq k_1$.

Demonstração. Para todo $k \in I\!K_1$ definimos $\overline{g}_k \in I\!R^n$ por

$$\left.\begin{array}{l} [\overline{g}_k]_i = -\dfrac{\partial f}{\partial x_i}(x_k) \\ \text{e} \\ [\overline{g}_k]_j = 0 \ \text{ se } \ j \neq i \end{array}\right\}. \tag{3.3.3}$$

Chamamos de y_k a projeção ortogonal de $x_k + \overline{g}_k/M_k$ na caixa $\{x \in I\!R^n \mid \ell \leq x \leq u\}$. Como $\dfrac{\partial f}{\partial x_i}(x_*) \neq 0$, por (3.3.3) e pela limitação de M_k existe $k_2 \in I\!K_1$ tal que

$$\left.\begin{array}{l} [y_k]_i = [x_*]_i \\ \text{e} \\ [y_k]_j = [x_k]_j, i \neq j \end{array}\right\}, \tag{3.3.4}$$

para todo $k \geq k_2, k \in I\!K_1$.

Vamos considerar dois casos:
(a) Existe $k_3 \in I\!K_1$ tal que

$$\|D_k(y_k - x_k)\|_\# \leq \Delta_k \tag{3.3.5}$$

para todo $k \geq k_3, k \in I\!K_1$.

(b) Existe um conjunto infinito $I\!K_2 \subset I\!K_1$ tal que

$$\|D_k(y_k - x_k)\|_\# > \Delta_k \tag{3.3.6}$$

para todo $k \in I\!K_2$.

Vamos analisar inicialmente a possibilidade (a). Se $k \geq k_2$, temos que $[x_k - g_k/M_k]_i \leq \ell_i \ (\geq u_i)$. Então, como $z_k^Q(\Delta_k)$ é a projeção de $-g_k/M_k$ na região factível do problema (3.2.2), que neste caso é uma caixa, segue que $[x_k + z_k^Q(\Delta_k)]_i = \ell_i \ (u_i)$ para todo $k \geq k_2$. Assim, o resultado desejado vale com $k_1 = k_2$.

Consideremos agora o caso (b). Pela convergência de x_k para $k \in I\!K_1$ e por (3.3.4) temos que

$$\lim_{k \in I\!K_1} \|y_k - x_k\|_\# = 0. \tag{3.3.7}$$

Agora, pela limitação de $\|D_k\|_\#$,

$$\lim_{k \in I\!K_1} \|D_k(x_k - y_k)\|_\# = 0. \tag{3.3.8}$$

Logo, por (3.3.6) e (3.3.8), $\lim_{k \in I\!K_1} \Delta_k = 0$. Chamemos de $\Delta^k = \Delta_1^k > \Delta_2^k > \ldots > \Delta_{m(k)}^k = \Delta_k$ a seqüência de raios de confiança tentada a cada iteração k. Como $\Delta^k \geq \Delta_{min}$ e $\Delta_k \to 0$, existe $k_4 \in I\!K_1$ tal que $m(k) > 1$ para todo $k \geq k_4, k \in I\!K_1$. Chamemos de $v(k) \in \{1, 2, \ldots, m(k)\}$ o primeiro índice tal que

$$\|D_k(y_k - x_k)\|_\# > \Delta_{v(k)}^k. \tag{3.3.9}$$

Desta forma,

$$\Delta_{v(k)-1}^k \geq \|D_k(y_k - x_k)\|_\# > \Delta_{v(k)}^k. \tag{3.3.10}$$

Claramente, por (3.3.8),

$$\lim_{k \in I\!K_1} \Delta_{v(k)}^k = 0. \tag{3.3.11}$$

Mas, por (3.2.6), $\Delta_{v(k)}^k \geq \tau_1 \|D_k \overline{z}_k(\Delta_{v(k)-1}^k)\|_\#$. Então, por (3.3.10),

$$\|D_k(y_k - x_k)\|_\# \geq \frac{\tau_1}{\|D_k^{-1}\|_\#} \|\overline{z}_k(\Delta_{v(k)-1}^k)\|_\#. \tag{3.3.12}$$

Logo, escrevendo $\overline{z}_k = \overline{z}_k(\Delta_{v(k)-1}^k)$ para simplificar a notação, pela limitação de $\|D_k^{-1}\|_\#$ existe $c_1 > 0$ tal que

$$\|D_k(y_k - x_k)\|_\# \geq c_1 \|\overline{z}_k\|_\# \tag{3.3.13}$$

para todo $k \geq k_4, k \in I\!K_1$. Claramente, (3.3.8) e (3.3.13) implicam que $\lim_{k \in I\!K_1} \|\overline{z}_k\|_\# = 0$.

Agora, por (3.3.4), para todo $k \geq k_2, k \in I\!K_1$ temos

$$\begin{aligned} Q_k(y_k - x_k) &= \frac{1}{2}M_k\|y_k - x_k\|^2 + g_k^T(y_k - x_k) \\ &= \frac{1}{2}M_k\|y_k - x_k\|^2 + [g_k]_i[y_k - x_k]_i. \end{aligned} \tag{3.3.14}$$

Assim, pela limitação de M_k, existe $k_5 \geq k_4$ tal que

$$Q_k(y_k - x_k) \leq -\frac{1}{2}\frac{\partial f}{\partial x_i}(x_*)\|y_k - x_k\|_\# \tag{3.3.15}$$

para todo $k \geq k_5, k \in I\!K_1$.

Então, por (3.3.10) e (3.2.3), existe $c_2 > 0$ tal que

$$\psi_k(\overline{z}_k) \leq -c_2\|y_k - x_k\|_\# \tag{3.3.16}$$

para todo $k \geq k_5, k \in I\!K_1$. Pela equivalência das normas em $I\!R^n$ e pela limitação de $\|D_k\|_\#$, existe $c_3 > 0$ tal que

$$\psi_k(\overline{z}_k) \leq -c_3\|D_k(y_k - x_k)\|_\#$$

para todo $k \geq k_5, k \in I\!K_1$. Então, por (3.3.13),

$$|\psi_k(\overline{z}_k)| \geq c_1c_3\|\overline{z}_k\|_\# = c_4\|\overline{z}_k\|_\# \tag{3.3.17}$$

para todo $k \geq k_5, k \in I\!K_1$.

Definimos agora $\rho_k = \dfrac{f(x_k + \overline{z}_k) - f(x_k)}{\psi_k(\overline{z}_k)}$. Para $k \geq k_5, k \in I\!K_1$ temos

$$\rho_k - 1 = \frac{f(x_k + \overline{z}_k) - f(x_k) - \psi_k(\overline{z}_k)}{\psi_k(\overline{z}_k)} = a_k + b_k$$

onde

$$a_k = \frac{f(x_k + \overline{z}_k) - f(x_k) - g_k^T\overline{z}_k}{\psi_k(\overline{z}_k)}$$

e

$$b_k = -\frac{1}{2}\frac{\overline{z}_k^T B_k \overline{z}_k}{\psi_k(\overline{z}_k)}.$$

Por (3.3.17), pela equivalência das normas em $\mathbb{R}^n$ e pelo Teorema do Valor Médio, temos

$$|a_k| \leq \frac{|f(x_k + \overline{z}_k) - f(x_k) - g_k^T \overline{z}_k|}{c_4 \|\overline{z}_k\|_\#} \xrightarrow[k \in \mathcal{K}_1]{} 0.$$

Além disso, por (3.3.17), pela limitação de $|\lambda_1(B_k)|$ e M_k e pela equivalência das normas em $\mathbb{R}^n$,

$$|b_k| \leq \frac{1}{2} \frac{\max\{|\lambda_1(B_k)|, M_k\} \|\overline{z}_k\|^2}{c_4 \|\overline{z}_k\|_\#} \xrightarrow[k \in \mathcal{K}_1]{} 0.$$

Das duas últimas desigualdades segue que ρ_k tende para 1. Isto contradiz o fato de que $\overline{z}_k$ tenha sido rejeitado no Passo 4 do Algoritmo 3.2.1. Portanto (3.3.6) é falsa e a prova do lema está completa. □

HIPÓTESE 3.3.2. Se x_* é um ponto estacionário de primeira ordem do problema (3.1.1) (Karush-Kuhn-Tucker) e $\dfrac{\partial f}{\partial x_i}(x_*) = 0$, então $\ell_i < [x_*]_i < u_i$.

HIPÓTESE 3.3.3. Para todo $k = 0, 1, 2, \ldots$, se $[x_k + z_k^Q(\Delta_k)]_i = \ell_i$ (respectivamente $[x_k + z_k^Q(\Delta_k)]_i = u_i$) então $[x_{k+1}]_i \equiv [x_k + z_k^Q(\Delta_k)]_i = \ell_i$ (respectivamente $[x_{k+1}]_i \equiv [x_k + z_k^Q(\Delta_k)]_i = u_i$).

Claramente, a Hipótese 3.3.3 pode ser introduzida no Passo 3 do Algoritmo 3.2.1 sem necessidade de se modificar os resultados de convergência.

LEMA 3.3.4. Assumindo a validade das Hipóteses 3.3.2 e 3.3.3, suponhamos que $\{x_k\}$ é uma seqüência infinita limitada gerada pelo Algoritmo 3.2.1 e que M_k, $\|D_k\|_\#$, $\|D_k^{-1}\|_\#$ e $|\lambda_1(B_k)|$ são limitados para todo $k = 0, 1, 2, \ldots$. Então existe $k_0 \in \mathbb{N}$ tal que para todo $k \geq k_0$,

$$A(x_k) \subset A(x_{k+1}). \tag{3.3.18}$$

Demonstração. Vamos supor, por absurdo, que exista um conjunto infinito de índices $I\!K_1$ tal que

$$A(x_k) \not\subset A(x_{k+1}) \tag{3.3.19}$$

para todo $k \in I\!K_1$. Então, como o número de restrições é finito, existe $i \in \{1, 2, \ldots, 2n\}$ e um conjunto infinito de índices $I\!K_2 \subset I\!K_1$ tal que

$$i \in A(x_k) \tag{3.3.20}$$

mas

$$i \notin A(x_{k+1}) \tag{3.3.21}$$

para todo $k \in I\!K_2$.

Seja $\{x_k\}_{k \in I\!K_3}$ uma subseqüência convergente de $\{x_k\}_{k \in I\!K_2}$. Então $\lim_{k \in I\!K_3} x_k = x_*$ e (3.3.20) - (3.3.21) valem para todo $k \in I\!K_3$. Por (3.3.20) segue que $i \in A(x_*)$. Sem perda de generalidade, suponhamos que $i \leq n$ e então $[x_*]_i = \ell_i$. Pela Hipótese 3.3.2 temos que $\dfrac{\partial f}{\partial x_i}(x_*) > 0$. Então, pelo Lema 3.3.1, $[x_k + z_k^Q(\Delta_k)]_i = \ell_i$ para $k \in I\!K_3, k$ suficientemente grande. Portanto, pela Hipótese 3.3.3, $[x_{k+1}]_i = \ell_i$ para $k \in I\!K_3, k$ suficientemente grande, o que contradiz (3.3.21) e completa a prova. □

TEOREMA 3.3.5. Assumimos, como no Lema 3.3.4, que valem as Hipóteses 3.3.2 e 3.3.3, que $\{x_k\}$ é uma seqüência infinita limitada gerada pelo Algoritmo 3.2.1 e que $M_k, \|D_k\|_\#, \|D_k^{-1}\|_\#$ e $|\lambda_1(B_k)|$ são limitados. Seja x_* um ponto limite de $\{x_k\}$. Então existe $k_1 \in I\!N$ tal que $A(x_k) = A(x_*)$ para todo $k \geq k_1$.

Demonstração. Como o número de restrições é finito, pelo Lema 3.3.4 existem $k_1 \in I\!N, A \subset \{1, 2, \ldots, 2n\}$ tais que para todo $k \geq k_1$,

$$A(x_k) = A. \tag{3.3.22}$$

Como x_* é um ponto limite de $\{x_k\}$, temos claramente que $A \subset A(x_*)$. Vamos mostrar que

$$A(x_*) \subset A. \tag{3.3.23}$$

Suponhamos que (3.3.23) não se verifique. Então podemos assumir, sem perda de generalidade, que existe $i \in \{1,\ldots,n\}$ tal que $[x_*]_i = \ell_i$ mas $[x_k]_i > \ell_i$ para todo $k \geq k_1$. Seja $I\!K_1$ um conjunto infinito de índices tal que $\lim_{k \in I\!K_1} x_k = x_*$. Então, pelo Lema 3.3.1 e pelas Hipóteses 3.3.2 e 3.3.3, temos que $[x_{k+1}]_i = \ell_i$ para $k \in I\!K_1, k$ suficientemente grande. Isto contradiz o fato de que $[x_k]_i > \ell_i$ para todo $k \geq k_1$ e portanto a prova está completa. □

3.4 O ALGORITMO QUACAN

Na Seção 3.2 vimos que para encontrarmos um ponto tentativo apropriado para o Algoritmo 3.2.1 (BOX) precisamos calcular $\overline{z}_k(\Delta)$ satisfazendo (3.2.3), onde $z_k^Q(\Delta)$ é a solução do subproblema trivial (3.2.2). Claramente, poderíamos escolher $\overline{z}_k(\Delta) = z_k^Q(\Delta)$, comprometendo, contudo, a qualidade dos resultados práticos obtidos. Por isso, tomamos $\overline{z}_k(\Delta)$ como um minimizador aproximado de

$$\begin{array}{ll} \min & \psi_k(z) \\ \text{s/a} & \ell \leq x_k + z \leq u \\ & \|D_k z\|_\infty \leq \Delta \, . \end{array} \tag{3.4.1}$$

Se D_k é diagonal, as restrições de (3.4.1) formam uma caixa. Assim, escrevendo $z = x - x_k$ e com algum abuso de notação, (3.4.1) se reduz a um problema da forma:

$$\begin{array}{ll} \min & \psi_k(z) \\ \text{s/a} & z \in \Omega \, , \end{array} \tag{3.4.2}$$

onde $\psi(z) \equiv \frac{1}{2} z^T B z + b^T z$ e $\Omega = \{z \in I\!R^n \mid \ell \leq z \leq u, \ell < u\}$. A definição de ℓ e u não é a mesma em (3.4.1) e (3.4.2). Apesar disso, vamos usar as mesmas letras para simplificar a notação.

De agora em diante, vamos considerar o problema (3.4.2). Esta seção, assim como as próximas duas, podem ser lidas independentemente das Seções 3.2 e 3.3. A notação será ligeiramente modificada de forma que z_k denota a k-ésima componente da variável $z \in I\!R^n$ e z^k denota o vetor $z \in I\!R^n$ na k-ésima iteração. Nesta seção $\|\cdot\|$ é a norma 2 de vetores e matrizes. Denotamos por $P(z)$ a projeção ortogonal de $z \in I\!R^n$ em Ω. Definimos

$$\overline{g}(z) = -\nabla\psi(z) \equiv -(Bz + b). \tag{3.4.3}$$

Sejam L, K tais que $K < L$ e os autovalores de B estão contidos em $[K, L]$. Claramente,

$$\psi(w) - \psi(z) + \overline{g}(z)^T(w - z) = \frac{1}{2}(w - z)^T B(w - z) \tag{3.4.4}$$

para todo $z, w \in I\!R^n$. Então,

$$\frac{K}{2}\|w - z\|^2 \leq \psi(w) - \psi(z) + \overline{g}(z)^T(w - z) \leq \frac{L}{2}\|w - z\|^2 \tag{3.4.5}$$

para todo $z, w \in I\!R^n$.

Definimos uma *face aberta* de Ω como o conjunto $F_I \subset \Omega$ tal que $I \subset \{1, 2, \ldots, 2n\}$ e $F_I = \{z \in \Omega \mid z_i = \ell_i \text{ se } i \in I; z_i = u_i \text{ se } n + i \in I \text{ e } \ell_i < z_i < u_i \text{ caso contrário }\}$.

Chamamos $\overline{F}_I$ o fecho de $F_I, V(F_I)$ a menor variedade linear que contem $F_I, S(F_I)$ o subespaço paralelo a $V(F_I)$ e $\dim(F_I)$ a dimensão de $S(F_I)$. Desta forma, se F_I é não vazio, temos:

$$\begin{aligned} V(F_I) &= \{z \in I\!R^n \mid z_i = \ell_i \text{ se } i \in I, z_i = u_i \text{ se } n + i \in I\}, \\ S(F_I) &= \{z \in I\!R^n \mid z_i = 0 \text{ se } i \in I \text{ ou } n + i \in I\}, \\ \overline{F}_I &= \Omega \cap V(F_I) \\ \dim(F_I) &= n - \#I, \end{aligned}$$

onde $\#I$ denota o número de elementos de I.

Para todo $z \in \Omega$, definimos $\overline{g}_p(z) \in I\!R^n$ (gradiente projetado *"negativo"*) por:

$$\overline{g}_p(z)_i = \begin{cases} 0 & \text{se} \quad z_i = \ell_i \text{ e } \dfrac{\partial \psi}{\partial z_i}(z) \geq 0 \\ & \text{ou} \\ & \quad z_i = u_i \text{ e } \dfrac{\partial \psi}{\partial z_i}(z) \leq 0 \\ -\dfrac{\partial \psi}{\partial z_i}(z) & \text{caso contrário .} \end{cases} \tag{3.4.6}$$

Claramente, se z é uma solução de (3.4.2) então

$$\overline{g}_p(z) = 0. \tag{3.4.7}$$

Para todo $z \in \overline{F}_I$ definimos $\overline{g}_I(z) \in \mathbb{R}^n$ por

$$\overline{g}_I(z)_i = \begin{cases} 0 & \text{se } i \in I \text{ ou } n+i \in I \\ -\dfrac{\partial \psi}{\partial z_i}(z) & \text{caso contrário .} \end{cases} \tag{3.4.8}$$

Definimos também, para todo $z \in \overline{F}_I$,

$$\overline{g}_I^c(z)_i = \begin{cases} 0 & \text{se} \quad i \notin I \text{ e } n+i \notin I \\ 0 & \text{se} \quad i \in I \text{ e } \dfrac{\partial \psi}{\partial z_i}(z) \geq 0 \\ & \text{ou} \\ & \quad n+i \in I \text{ e } \dfrac{\partial \psi}{\partial z_i}(z) \leq 0 \\ -\dfrac{\partial \psi}{\partial z_i}(z) & \text{caso contrário .} \end{cases} \tag{3.4.9}$$

Claramente, para todo $z \in \overline{F}_I$ temos

$$\overline{g}_p(z) = \overline{g}_I(z) + \overline{g}_I^c(z). \tag{3.4.10}$$

Para todo $I \subset \{1,2,\ldots,2n\}$ tal que F_I é não-vazio definimos

$$\gamma_I = \min\{u_i - \ell_i \mid i \in I \text{ ou } n+i \in I\}, \tag{3.4.11}$$

e

$$\gamma = \min\{u_i - \ell_i \mid i \in \{1,\ldots,n\}\}. \tag{3.4.12}$$

Podemos agora introduzir um algoritmo para minimizar uma quadrática numa caixa:

Algoritmo 3.4.1 (QUACAN)

Seja $z^0 \in \Omega$ uma aproximação inicial arbitrária para a solução de (3.4.2) e seja $\varepsilon \geq 0$ tal que $\varepsilon > 0$ se $K < 0$.

Escolhemos $\delta > 0$ tal que, usando a convenção $0/0 = \infty$,

$$\left.\begin{array}{l} \delta < \min\left\{\dfrac{\varepsilon}{\sqrt{L\max\{0,-K\}}}, \sqrt{\dfrac{L}{\max\{0,-K\}}\gamma}\right\} \text{ se } L > 0 \\ \text{ou} \\ \delta < \max\left\{\sqrt{\dfrac{2\gamma\varepsilon}{-K}}, \sqrt{\dfrac{L}{K}\gamma}\right\} \text{ se } L \leq 0 \end{array}\right\}. \tag{3.4.13}$$

Suponhamos que $z^k \in F_I$ tenha sido calculado. Os passos que nos permitem decidir pela parada ou pelo cálculo de z^{k+1} são dados a seguir:

Passo 1. Se $\|\overline{g}_p(z^k)\| \leq \varepsilon$, parar.

Passo 2. Se alguma das condições (a) – (d) abaixo é valida, ir para o Passo 3. Caso contrário, ir para o Passo 4.

(a) $L > 0, K \leq 0$ e

$$\|\overline{g}_I(z^k)\| < \frac{1}{2}\left[\frac{\min\{\|\overline{g}_I^c(z^k)\|, L\gamma_I\}^2}{\delta L} + K\delta\right]. \tag{3.4.14}$$

(b) $K > 0, \|\overline{g}_I(z^k)\| \geq K\delta$ e vale (3.4.14).

(c) $K > 0, \|\overline{g}_I(z^k)\| < K\delta$ e

$$\|\overline{g}_I(z^k)\| < \sqrt{\frac{K}{L}}\min\{\|\overline{g}_I^c(z^k)\|, L\gamma_I\}. \tag{3.4.15}$$

(d) $L \leq 0$ e

$$\|\overline{g}_I(z^k)\| < \frac{\gamma_I\|\overline{g}_I^c(z^k)\|}{\delta} - \frac{L}{2\delta}\gamma_I^2 + \frac{K}{2}\delta. \tag{3.4.16}$$

Passo 3. Calcular $z^{k+1} \in \Omega - \overline{F}_I$ tal que

$$\psi(z^{k+1}) < \psi(z) \tag{3.4.17}$$

para todo $z \in V(F_I)$ tal que $\|z - z^k\| \leq \delta$.

Passo 4. Se $k = 0$ ou $z^{k-1} \notin F_I$ definir

$$d^k = \overline{g}_I(z^k).$$

Caso contrário, definir

$$d^k = \overline{g}_I(z^k) + \beta_k d^{k-1},$$

onde

$$\beta_k = \|\overline{g}_I(z^k)\|^2 / \|\overline{g}_I(z^{k-1})\|^2.$$

Passo 5. Se $\psi(z)$ é inferiormente ilimitada na semi-reta $L_k \equiv \{z^k + \lambda d^k, \lambda \geq 0\}$ ou se o minimizador y^k de $\psi(z)$ em L_k não pertence a F_I, então obter z^{k+1} como um ponto em $\overline{F}_I - F_I$ satisfazendo

$$\psi(z^{k+1}) < \psi(z^k). \tag{3.4.18}$$

Caso contrário, definir $z^{k+1} = y^k$.

Observação. Vamos esclarecer o significado do Algoritmo 3.4.1. Vejamos inicialmente que sempre podemos escolher δ estritamente positivo satisfazendo (3.4.13). Se a quadrática é convexa, podemos tomar δ arbitrariamente grande (Friedlander e Martínez [1994]). A idéia do algoritmo é que o fecho da face F_I seja abandonado apenas quando existir um ponto fora de $\overline{F}_I$ satisfazendo uma condição de decréscimo suficiente. Tal condição é dada por (3.4.17) e estabelece que o valor da função num novo ponto deve ser um limitante inferior para os valores de ψ numa bola de raio δ centrada em z^k e restrita a $V(F_I)$. Veremos na próxima seção que as condições (a) - (d) do Passo 2 garantem a existência de tal ponto. De fato, vamos mostrar que o minimizador de ψ no primeiro segmento factível gerado por $\overline{g}_I^c$ satisfaz a condição requerida, embora não sejamos obrigados a escolher tal ponto como z^{k+1}. Provaremos também que se $\|\overline{g}_I(z^k)\| = 0$ e o critério de parada (Passo 1) não é satisfeito, então valem as condições para o abandono da face. Por outro lado, se não se verifica o teste no Passo 2, são feitas iterações de Gradientes Conjugados dentro de F_I. Neste caso, se não existe minimizador ao longo da direção de descida ou se este não pertence a F_I, o novo ponto deve pertencer à fronteira da face atual.

3.5 TERMINAÇÃO FINITA DE QUACAN.

Nesta seção provaremos os seguintes teoremas:

TEOREMA 3.5.1. O Algoritmo 3.4.1 está bem definido.

TEOREMA 3.5.2. Se $\{z^k\}$ é uma seqüência gerada pelo Algoritmo 3.4.1 então $\{z^k\}$ converge a um ponto z^* satisfazendo $\|\overline{g}_p(z^*)\| \leq \varepsilon$ num número finito de iterações.

Para a prova dos Teoremas 3.5.1 e 3.5.2 vamos precisar do seguinte lema:

LEMA 3.5.3. Seja $z \in \overline{F}_I \subset \Omega$ tal que $\overline{g}_I^c(z) \neq 0$. Definimos

$$w_I^c(z) = \overline{g}_I^c(z)/\|\overline{g}_I^c(z)\|. \tag{3.5.1}$$

Então o segmento $(z, z + \gamma_I w_I^c(z)]$ está contido em $\Omega - \overline{F}_I$.

Demonstração. Segue diretamente das definições (3.4.9) e (3.4.11). Para maiores detalhes, ver Friedlander e Martínez [1994]. □

Demonstração do Teorema 3.5.1. Assumimos que $\|g_p(z^k)\| > \varepsilon$. Vamos analisar inicialmente o caso em que vale pelo menos uma das condições (a) ou (b) do Passo 2 do Algoritmo 3.4.1. Se $\|\overline{g}_I^c(z^k)\| > L\gamma_I$, de (3.4.14) segue que

$$\|\overline{g}_I(z^k)\| < \frac{1}{2}\Big[\frac{L\gamma_I^2}{\delta} + K\delta\Big]. \tag{3.5.2}$$

Então, por (3.4.5) e (3.4.14),

$$\begin{aligned} \psi(z^k + \gamma_I w_I^c(z^k)) &\leq \psi(z^k) - \gamma_I\|\overline{g}_I^c(z^k)\| + \frac{L}{2}\gamma_I^2 \\ &\leq \psi(z^k) - \frac{L}{2}\gamma_I^2 \\ &< \psi(z^k) - \|\overline{g}_I(z^k)\|\delta + \frac{K}{2}\delta^2. \end{aligned} \tag{3.5.3}$$

Se $\|\overline{g}_I^c(z^k)\| \leq L\gamma_I$, por (3.4.14) temos

$$\|\overline{g}_I(z^k)\| < \frac{1}{2}\Big[\frac{\|\overline{g}_I^c(z^k)\|^2}{\delta L} + K\delta\Big]. \tag{3.5.4}$$

Definimos $\rho = \|\overline{g}_I^c(z^k)\|/L$. Então, pelo Lema 3.5.3, $z^k + \rho w_I^c(z^k) \in \Omega - \overline{F}_I$. Além disso, por (3.5.4), (3.4.5) e (3.4.14),

$$\psi(z^k + \rho w_I^c(z^k)) \leq \psi(z^k) - \rho\|\overline{g}_I^c(z^k)\| + \frac{L}{2}\rho^2$$

$$
\begin{aligned}
&= \psi(z^k) - \frac{\|\overline{g}_I^c(z^k)\|^2}{L} + \frac{\|\overline{g}_I^c(z^k)\|^2}{2L} \\
&= \psi(z^k) - \frac{\|\overline{g}_I^c(z^k)\|^2}{2L} \\
&< \psi(z^k) - \|\overline{g}_I(z^k)\|\delta + \frac{K}{2}\delta^2. \qquad (3.5.5)
\end{aligned}
$$

Assim, mostramos que, se vale uma das condições (a) ou (b) do Passo 2, então existe $\overline{z} \in \Omega - \overline{F}_I$ satisfazendo

$$\psi(\overline{z}) < \psi(z^k) - \|\overline{g}_I(z^k)\|\delta + \frac{K}{2}\delta^2. \qquad (3.5.6)$$

Além disso, o minimizador de ψ em $[z^k, z^k + \gamma_I w_I^c(z^k)]$ satisfaz (3.5.6).

Agora, de (3.5.6) segue que, para todo $z \in \mathbb{R}^n$,

$$\psi(\overline{z}) - \psi(z) < -\|\overline{g}_I(z^k)\|\delta + \frac{K}{2}\delta^2 - [\psi(z) - \psi(z^k)]. \qquad (3.5.7)$$

Se $z, z^k \in V(F_I)$ então $z - z^k \in S(F_I)$. Então, por (3.4.5),

$$
\begin{aligned}
\psi(z) - \psi(z^k) + \overline{g}_I(z^k)^T(z - z^k) &= \psi(z) - \psi(z^k) - \nabla\psi(z^k)^T(z - z^k) \\
&\geq \frac{K}{2}\|z - z^k\|^2.
\end{aligned}
$$

Então, para todo $z \in V(F_I)$ temos

$$\psi(z) - \psi(z^k) \geq -\overline{g}_I(z^k)^T(z - z^k) + \frac{K}{2}\|z - z^k\|^2. \qquad (3.5.8)$$

Se vale (a), como $K \leq 0$, se $\|z - z^k\| \leq \delta$ temos

$$\psi(z) - \psi(z^k) \geq -\|\overline{g}_I(z^k)\|\delta + \frac{K}{2}\delta^2. \qquad (3.5.9)$$

Se vale (b) e $z \in V(F_I)$ é tal que $\|z - z^k\| \leq \delta$, por (3.5.8) temos

$$\psi(z) - \psi(z^k) \geq \min_{\|z-z^k\| \leq \delta} -\overline{g}_I(z^k)^T(z - z^k) + \frac{K}{2}\|z - z^k\|^2. \tag{3.5.10}$$

Como $\dfrac{\|\overline{g}_I(z^k)\|}{K} \geq \delta$, o minimizador de $-\overline{g}_I(z^k)^T(z-z^k)+\dfrac{K}{2}\|z-z^k\|^2$ na bola $\|z-z^k\| \leq \delta$ é $z^k+\dfrac{\overline{g}_I(z^k)}{\|\overline{g}_I(z^k)\|}\delta$. Então, por (3.5.10), a desigualdade (3.5.9) também vale neste caso. A partir de (3.5.9) e (3.5.7) concluímos que (3.4.17) vale sob as condições (a) ou (b).

Vamos analisar agora o caso em que vale a condição (c) do Passo 2. Se $\|\overline{g}_I^c(z^k)\| > L\gamma_I$ então, por (3.4.15) temos

$$\|\overline{g}_I(z^k)\| < \sqrt{\frac{K}{L}}L\gamma_I = \sqrt{KL}\gamma_I. \tag{3.5.11}$$

Então, por (3.4.5),

$$\begin{aligned} \psi(z^k + \gamma_I w_I^c(z^k)) &\leq \psi(z^k) - \gamma_I\|\overline{g}_I^c(z^k)\| + \frac{L}{2}\gamma_I^2 \\ &< \psi(z^k) - \frac{L}{2}\gamma_I^2 < \psi(z^k) - \frac{\|\overline{g}_I(z^k)\|^2}{2K}. \end{aligned} \tag{3.5.12}$$

Se $\|\overline{g}_I^c(z^k)\| \leq L\gamma_I$ temos, por (3.4.15), que

$$\|\overline{g}_I(z^k)\| < \sqrt{\frac{K}{L}}\|\overline{g}_I^c(z^k)\|. \tag{3.5.13}$$

Então, se $\rho = \|\overline{g}_I^c(z^k)\|/L$, de (3.4.5) segue que

$$\begin{aligned} \psi(z^k + \rho w_I^c(z^k)) &\leq \psi(z^k) - \frac{\|\overline{g}_I^c(z^k)\|^2}{L} + \frac{\|\overline{g}_I^c(z^k)\|^2}{2L} \\ &< \psi(z^k) - \frac{\|\overline{g}_I(z^k)\|^2}{2K}. \end{aligned} \tag{3.5.14}$$

Desta forma, provamos que se a condição (c) do Passo 2 é satisfeita então ou (3.5.12) ou (3.5.14) tem que ser verdadeira. Nos dois casos existe $\overline{z} \in \Omega - \overline{F}_I$ tal que

$$\psi(\overline{z}) < \psi(z^k) - \frac{\|\overline{g}_I(z^k)\|^2}{2K}. \tag{3.5.15}$$

Além disso, como nos casos (a) e (b), o minimizador de ψ em $[z^k, z^k + \gamma_I w_I^c(z^k)]$ satisfaz a condição (3.5.15).

Assim, para todo $z \in \mathbb{R}^n$ temos que

$$\psi(\overline{z}) - \psi(z) < -\frac{\|\overline{g}_I(z^k)\|^2}{2K} - [\psi(z) - \psi(z^k)]. \tag{3.5.16}$$

Mas, neste caso, $K > 0$ e $\frac{\|\overline{g}_I(z^k)\|}{K} < \delta$. Então, o minimizador de $-\overline{g}_I(z^k)^T(z - z^k) + \frac{K}{2}\|z - z^k\|^2$ na bola $\|z - z^k\| \leq \delta$ é $z^k + \frac{\overline{g}_I(z^k)}{K}$. Então, por (3.5.16) e (3.4.5),

$$\begin{aligned} \psi(\overline{z}) - \psi(z^k) &< -\frac{\|\overline{g}_I(z^k)\|^2}{2K} - [-\overline{g}_I(z^k)^T(z - z^k) + \frac{K}{2}\|z - z^k\|^2] \\ &\leq -\frac{\|\overline{g}_I(z^k)\|^2}{2K} - \left[-\frac{\|\overline{g}_I(z^k)\|^2}{K} + \frac{\|\overline{g}_I(z^k)\|^2}{2K}\right] = 0 \end{aligned} \tag{3.5.17}$$

para todo $z \in V(F_I)$ tal que $\|z - z^k\| \leq \delta$. Desta forma, a validade de (3.4.17) também está provada no caso (c).

Finalmente, vamos analisar o caso em que vale a condição (d) do Passo 2. Então, por (3.4.16) e (3.4.5),

$$\begin{aligned} \psi(z^k + \gamma_I w_I^c(z^k)) &\leq \psi(z^k) - \gamma_I \|\overline{g}_I^c(z^k)\| + \frac{L}{2}\gamma_I^2 \\ &< \psi(z^k) - \|\overline{g}_I(z^k)\|\delta + \frac{K}{2}\delta^2. \end{aligned} \tag{3.5.18}$$

Neste caso a situação é bastante semelhante ao caso (a). De fato, temos $K < 0$ e $\|\overline{g}_I(z^k)\| \geq K\delta$ e portanto, (3.5.7) segue de (3.5.18) e (3.4.17) é uma conseqüência de (3.5.18) e (3.5.7).

Até agora, mostramos que se vale qualquer uma das condições (a) – (d) do Passo 2, então z^{k+1} estará bem definido no Passo 3 se for escolhido de tal forma que, a longo

prazo, venha a ser o minimizador de ψ em $[z^k, z^k + \gamma_I w_I^c(z^k)]$.

Vamos assumir agora que $\|\overline{g}_p(z^k)\| > \varepsilon$ mas nenhuma das condições (a) - (d) seja satisfeita. Devemos considerar duas possibilidades:

$$\overline{g}_I(z^k) = 0 \tag{3.5.19}$$

ou

$$\overline{g}_I(z^k) \neq 0. \tag{3.5.20}$$

Vamos mostrar, por contradição, que (3.5.19) não pode valer. Se $K = 0$, (3.4.14) se verifica e então podemos excluir este caso. Se $K > 0$, temos $0 = \|\overline{g}_I(z^k)\| < K\delta$ e (3.4.15) vale trivialmente neste caso. Então, se vale (3.5.19), basta analisarmos o caso $K < 0$. Se $L > 0$ vale e (3.4.14) não se verifica, temos que

$$\frac{\min\{\|\overline{g}_I^c(z^k)\|, L\gamma_I\}^2}{\delta L} + K\delta \leq 0. \tag{3.5.21}$$

Então, ou

$$\frac{L\gamma_I^2}{\delta} + K\delta \leq 0 \tag{3.5.22}$$

ou

$$\frac{\|g_I^c(z^k)\|^2}{\delta L} + K\delta \leq 0. \tag{3.5.23}$$

Se vale (3.5.22), então $L\gamma_I^2 + K\delta^2 \leq 0$ e portanto $\delta^2 \geq -\frac{L\gamma_I^2}{K} \geq -\frac{L}{K}\gamma^2$, o que contradiz a escolha (3.4.13).

Agora, se valem (3.5.19) e (3.5.23), então por (3.4.10) temos que $\frac{\|\overline{g}_p(z^k)\|^2}{\delta L} + K\delta \leq 0$. Mas, como $\|\overline{g}_p(z^k)\| > \varepsilon$, temos que $\frac{\varepsilon^2}{\delta L} + K\delta < 0$. Portanto, $\delta^2 > -\frac{\varepsilon^2}{KL}$ e também obtemos uma contradição com (3.4.13).

Para concluir a análise de (3.5.19) resta considerar o caso $L \leq 0$. Como (3.4.16) não se verifica, temos que

$$\gamma_I \frac{\|\overline{g}_I^c(z^k)\|}{\delta} - \frac{L}{2\delta}\gamma_I^2 + \frac{K}{2}\delta \leq 0. \tag{3.5.24}$$

Mas, por (3.4.10) e (3.5.19), neste caso temos $\|\overline{g}_I^c(z^k)\| = \|\overline{g}_p(z^k)\| > \varepsilon$. Portanto, por (3.5.24), $\frac{\gamma_I \varepsilon}{\delta} - \frac{L\gamma_I^2}{2\delta} + \frac{K}{2}\delta < 0$. Em conseqüência, $\delta^2 > \frac{L\gamma_I^2}{K} - \frac{2\gamma_I \varepsilon}{K}$, o que nos leva a uma contradição com a escolha (3.4.13).

Desta forma, provamos que se $\|\overline{g}_p(z^k)\| > \varepsilon$ e nenhuma das condições (a) - (d) do Passo 2 do Algoritmo 3.4.1 se verifica, temos necessariamente que $\overline{g}_I(z^k) \neq 0$. Neste caso, o objetivo é encontrar z^{k+1} satisfazendo (3.4.18). Se $d^k = \overline{g}_I(z^k)$ então d^k é uma direção de descida e portanto (3.4.18) se verifica trivialmente. Vamos analisar o caso em que d^k é computada por

$$d^k = \overline{g}_I(z^k) + \beta_k d^{k-1}. \tag{3.5.25}$$

Como $z^k \in F_I$ e $z^{k-1} \in F_I$, segue que z^k deve ter sido obtido por

$$z^k = z^{k-1} - \frac{\nabla\psi(z^{k-1})^T d^{k-1}}{(d^{k-1})^T B d^{k-1}} d^{k-1}. \tag{3.5.26}$$

Então

$$\nabla\psi(z^k)^T d^{k-1} = 0. \tag{3.5.27}$$

De (3.5.25) e (3.5.27) obtemos que

$$\nabla\psi(z^k)^T d^k = -\|\overline{g}_I(z^k)\|^2 < 0.$$

Portanto, a direção d^k computada pelo Passo 5 é uma direção de descida. Logo, (3.4.18) pode ser obtido a partir de procedimentos bem conhecidos, o que completa a prova do Teorema 3.5.1. □

Os próximos lemas são usados na prova do Teorema 3.5.2.

LEMA 3.5.4. Se $\|\overline{g}_p(z^k)\| > \varepsilon$, então

(P1) z^{k+1} está bem definido e $\psi(z^{k+1}) < \psi(z^k)$.

(P2) Se $z^k \in F_I$ então uma, e somente uma, das seguintes propriedades é válida:

(i) $z^{k+1} \in F_I$. (3.5.28)

(ii) $z^{k+1} \in F_J \subset \overline{F}_I$, onde $\dim F_J < \dim F_I$. (3.5.29)

(iii) $z^{k+1} \notin \overline{F}_I$ e $\psi(z^{k+1}) < \psi(z)$ para todo $z \in V(F_I)$ tal que $\|z - z^k\| \leq \delta$.

Demonstração. A prova deste resultado segue diretamente do Teorema 3.5.1 e da definição do Algoritmo 3.4.1. □

LEMA 3.5.5. Existe $k_0 \in I\!N$ tal que para todo $k \geq k_0$, (3.5.28) ou (3.5.29) ocorre.

Demonstração. Suponhamos, por absurdo, que exista um conjunto infinito de índices $I\!K_1 \subset I\!N$ tal que para todo $k \in I\!K_1$ a possibilidade (iii) do Lema 3.5.4 seja válida. Como o número de faces distintas é finito, existe um conjunto infinito $I\!K_2 \subset I\!K_1$ e uma face F_I tal que $z^k \in F_I$ para todo $k \in I\!K_2$. Então, se $k \in I\!K_2$, temos que $\psi(z^{k+1}) < \psi(z)$ para todo $z \in F_I$ tal que $\|z - z^k\| \leq \delta$. Então, como $\psi(z^k)$ é estritamente decrescente, para dois valores diferentes $k_1, k_2 \in I\!K_2$, teremos necessariamente que $\|z^{k_1} - z^{k_2}\| > \delta > 0$. Mas isto é impossível pois F_I é limitado. Desta forma, a prova do resultado desejado está completa. □

Para completarmos os resultados necessários para a prova do Teorema 3.5.2, vamos considerar um algoritmo de direções conjugadas, analisando sua convergência.

Seja $\psi(z) = z^T Bz + b^T z$, onde $B \in I\!R^{n \times n}$ é uma matriz simétrica. Nenhuma outra hipótese adicional é feita sobre B. Consideremos o seguinte algoritmo:

Algoritmo 3.5.6.

Dados $z^0 \in I\!R^n, d^0 = g^0 = -\nabla\psi(z^0)$, tomar $k = 0$ e desde que $(d^k)^T Bd^k > 0$ e $g^k = \nabla\psi(z^k) \neq 0$, fazer:

$$\begin{aligned} \alpha_k &= -(g^k)^T d^k / (d^k)^T Bd^k \\ z^{k+1} &= z^k + \alpha_k d^k \\ \beta_k &= \|g^{k+1}\|^2 / \|g^k\|^2 \\ d^{k+1} &= -g^{k+1} + \beta_k d^k. \end{aligned}$$

Observamos que se B é positiva definida, o Algoritmo 3.5.6 é exatamente o algoritmo de Gradientes Conjugados. Sabemos que se B é positiva definida, a seqüência de pontos z^k gerada por este algoritmo obtem um ponto estacionário em, no máximo, n passos.

LEMA 3.5.7. Se o Algoritmo 3.5.6 não pára em m $(m < n)$ passos então $(d^m)^T g^m = -(g^m)^T g^m$. Portanto, d^m é uma direção de descida a partir de z^m.

Demonstração. Como o Algoritmo 3.5.6 não pára, temos necesariamente que $(d^k)^T B d^k > 0$. Então, este resultado é obtido da mesma maneira que no método de Gradientes Conjugados para B positiva definida, onde apenas o fato de que $(d^k)^T B d^k > 0$ para $k < m$ é usado. □

TEOREMA 3.5.8. O Algoritmo 3.5.6 obtém, em um número finito de passos, um ponto estacionário de $\psi(z)$ ou um ponto z e uma direção d tais que $\lim_{\lambda \to \infty} \psi(z - \lambda d) = -\infty$.

Demonstração. Se o Algoritmo 3.5.6 não pára em n passos, então pelo resultado clássico do método de Gradientes Conjugados devemos ter $g^n = 0$ e neste caso B é necessariamente positiva definida. Agora, se o algoritmo pára após m passos, com $m < n$, temos duas possibilidades:

a) $g^m = 0$ e então z^m é um ponto estacionário de $\psi(z)$.

b) $(d^m)^T B d^m \leq 0$.

Se vale (b), pelo Lema 3.5.7, $(d^m)^T g^m < 0$.

Logo, $\psi(z^m + \lambda d^m) = \psi(z^m) + \lambda (d^m)^T g^m + \dfrac{\lambda^2}{2} (d^m)^T B d^m$ e claramente $\lim_{\lambda \to \infty} \psi(z^m + \lambda d^m) = -\infty$.

Desta forma, a prova está completa. □

Finalmente, apresentamos a seguir a prova do Teorema 3.5.2.

Demonstração do Teorema 3.5.2. Vamos supor, por absurdo, que $\|\overline{g}_p(z^k)\| > \varepsilon$ para todo $k \in I\!N$. Seja $k_0 \in I\!N$ conforme estabelecido pelo Lema 3.5.5. Então, para todo $k \geq k_0$, vale (3.5.28) ou (3.5.29). Entretanto, não é possível ocorrer um número infinito

de iterações em que a propriedade (3.5.29) seja válida, pois a dimensão da face corrente decresce em cada iteração deste tipo. Desta forma, existem $\overline{k}_0 \in I\!N$ e $I \subset \{1, 2, \ldots, 2n\}$ tais que $z^k \in F_I$ para todo $k \geq \overline{k}_0$. Vamos supor, sem perda de generalidade, que $\overline{k}_0 = 0$ ou $z^{\overline{k}_0 - 1} \notin F_I$. Então a seqüência $z^{\overline{k}_0}, z^{\overline{k}_0 + 1}, \ldots$ é obtida por iterações de Gradientes Conjugados em $V(F_I)$ e a primeira direção de busca é $\overline{g}_I(z^{\overline{k}_0})$. Além disso, todas as direções de busca d^k são de descida e $(d^k)^T B d^k > 0$ para todas estas direções (caso contrário, pelo Passo 5, a iteração seguinte seria na fronteira de F_I). Assim, de acordo com o Teorema 3.5.8, existe $k \geq \overline{k}_0$ tal que $\overline{g}_I(z^k) = 0$. Mas, neste caso a definição de δ impõe que z^{k+1} seja computado no Passo 3 do Algoritmo 3.4.1, o que contradiz as hipóteses feitas anteriormente. Logo, a prova está completa. □

3.6 BUSCA NO CAMINHO POLIGONAL

Nesta seção especificamos como é feita efetivamente a implementação dos Passos 3 e 5 do Algoritmo 3.4.1 (QUACAN). Retomando, no Passo 3 requeremos que o novo ponto satisfaça (3.4.17), enquanto no Passo 5 exigimos a validade de (3.4.18). De qualquer forma, os dois passos são implementados usando-se uma estratégia de busca do tipo *"backtracking"* ao longo do caminho poligonal definido pela direção de busca. Existem apenas algumas diferenças de menor importância entre esta busca projetada e a estratégia adotada por Moré e Toraldo [1991], já que em nosso caso permitimos a ocorrência de Hessianas indefinidas ou singulares, o que implica na ilimitação da quadrática ψ ao longo da curva correspondente. Além disso, não usamos uma condição de Armijo para pararmos a busca, conforme Moré e Toraldo fazem.

Assumimos que z^k é a k-ésima aproximação para a solução do problema e que d^k é uma direção factível de descida. O algoritmo a seguir descreve o modo com que nossa busca projetada é feita.

Algoritmo 3.6.1 (Busca no Caminho Poligonal).

Passo 1. Calcular $\underline{\mu}_k = \max\{\mu \geq 0 \mid z^k + \mu d^k \in \Omega\}$ e

$$\begin{aligned} \overline{\mu}_k \;=\; & \max\{\mu \geq 0 \mid \exists i \in \{1, \ldots, n\} \text{ tal que } \ell_i = z_i^k + \mu d_i^k \\ & \text{ou } u_i = z_i^k + \mu d_i^k\}. \end{aligned}$$

Passo 2. Se $(d^k)^T B d^k \leq 0$, tomar $\overline{\lambda} = \overline{\mu}_k$.

Senão, calcular $\overline{\lambda} = -\dfrac{\nabla\psi(z^k)^T d^k}{(d^k)^T B d^k}$.

Se $\overline{\lambda} \leq \underline{\mu}_k$, fazer $\overline{z} = P(z^k + \overline{\lambda} d^k)$ e retornar.

Passo 3. Fazer $\overline{z} = P(z^k + \overline{\lambda} d^k)$.

Se alguma das condições (a) - (c) a seguir for válida, fazer $z^{k+1} = \overline{z}$ e retornar.

(a) Este algoritmo está sendo chamado do Passo 5 do Algoritmo 3.4.1 e $\psi(\overline{z}) < \psi(z^k)$.

(b) Este algoritmo está sendo chamado do Passo 3 do Algoritmo 3.4.1,

$$\|\overline{g}_I(z^k)\| \geq K\delta \text{ e } \psi(\overline{z}) < \psi(z^k) - \|\overline{g}_I(z^k)\|\delta + \frac{K}{2}\delta^2.$$

(c) Este algoritmo está sendo chamado do Passo 3 do Algoritmo 3.4.1,

$$\|\overline{g}_I(z^k)\| < K\delta \text{ e } \psi(\overline{z}) < \psi(z^k) - \frac{\|\overline{g}_I(z^k)\|^2}{2K}.$$

Passo 4. Computar λ_{novo}, minimizador da quadrática

$\phi(\lambda)$ tal que $\phi(0) = \psi(z^k), \phi'(0) = \nabla\psi(z^k)^T d^k$ e $\phi(\overline{\lambda}) = \psi(P(z^k + \overline{\lambda} d^k))$ (ver detalhes em Moré e Toraldo [1991]).

Se $\lambda_{novo} \leq 0.1\overline{\lambda}$ ou $\lambda_{novo} \geq 0.9\overline{\lambda}$,

fazer $\overline{\lambda} \leftarrow \overline{\lambda}/2$;

senão, tomar $\overline{\lambda} = \lambda_{novo}$.

Se $\overline{\lambda} \leq \underline{\mu}_k$, fazer $\overline{\lambda} \leftarrow \underline{\mu}_k, \overline{z} = P(z^k + \overline{\lambda} d^k)$ e retornar;

senão, ir para o Passo 3.

É fácil ver que se o Algoritmo 3.6.1 é chamado do Passo 5 do Algoritmo 3.4.1, então a direção de descenso (3.4.18) é satisfeita definindo-se $z^{k+1} = \overline{z}$. Isto se deve ao fato de que a direção conjugada d^k é uma direção de descida e a longo prazo o minimizador de ψ no primeiro segmento definido por d^k é um ponto tentativo.

Por outro lado, se o Algoritmo 3.6.1 é chamado do Passo 3 do Algoritmo 3.4.1, a situação é um pouco diferente. De fato, se definimos $d^k = g_I^c(z^k)$, conforme vimos na Seção 3.5, o minimizador de ψ no primeiro segmento determinado por d^k satisfaz (3.4.17). Desta forma, como o Algoritmo 3.6.1 acaba avaliando ψ neste ponto, segue que, com esta definição para d^k, obtemos em tempo finito um ponto satisfazendo (3.4.17) se $z^{k+1} = \overline{z}$. Cabe observar que (b) e (c) no Passo 3 do Algoritmo 3.6.1 são condições suficientes que garantem que $\overline{z}$ satisfaz (3.4.17). No entanto, a experimentação numérica bem como a tradição recomendam que seja dada uma chance ao gradiente projetado como uma direção que defina o caminho poligonal. Frequentemente, a utilização da direção $\overline{g}_p(z^k)$ no Algoritmo 3.6.1 produz um ponto satisfazendo (3.4.17). Mas, especialmente em situações de quase degenerescência, isto pode não ocorrer. Desta forma, não podemos basear nossa estratégia de abandono da face apenas no gradiente projetado, ou seja, a direção $\overline{g}_I^c(z^k)$ deveria também ser usada. Vamos, portanto, definir duas estratégias possíveis para o abandono da face no Passo 3 do Algoritmo 3.4.1. A primeira é baseada apenas na direção $g_I^c(z^k)$, enquanto a segunda toma $\overline{g}_I^c(z^k)$ apenas quando $\overline{g}_p(z^k)$ produz uma falha. Estas duas estratégias são descritas pelos Algoritmos 3.6.2 e 3.6.3 a seguir.

Algoritmo 3.6.2. Estratégia para abandonar a face baseada em $\overline{g}_I^c(z^k)$.

Definir $d^k = g_I^c(z^k)$.

Executar o Algoritmo 3.6.1 e definir $z^{k+1} = \overline{z}$.

Algoritmo 3.6.3. Estratégia para abandonar a face baseada em $\overline{g}_p(z^k)$ e $\overline{g}_I^c(z^k)$.

Definir $d^k = \overline{g}_p(z^k)$.

Executar o Algoritmo 3.6.1.

Se uma das condições (b) ou (c) no Passo 3 do Algoritmo 3.6.1 é satisfeita, definir $z^{k+1} = \overline{z}$ e retornar.

Senão, executar o Algoritmo 3.6.2.

Do esquema interpolador salvaguardado do Algoritmo 3.6.1, segue que uma possível falha no gradiente projetado como direção de busca é detectada em tempo finito. Neste caso, a saída $\overline{\lambda}$ do Algoritmo 3.6.1 é o minimizador de $\phi(\lambda)$ para $\lambda \in [0, \underline{\mu}_k]$ e as condições de descenso suficiente não se verificam provavelmente porque este intervalo é muito pequeno.

3.7 EXPERIMENTOS NUMÉRICOS

A implementação do Algoritmo 3.2.1 (BOX) foi feita através de um programa em linguagem FORTRAN (precisão dupla) usando subrotinas baseadas nos Algoritmos 3.4.1 (QUACAN), 3.6.1, 3.6.2 e 3.6.3. Vamos definir os parâmetros e procedimentos especiais utilizados.

Na implementação do Algoritmo 3.2.1, usamos $\|\cdot\|_{\#} = \|\cdot\|_{\infty}, \tau_1 = \tau_2 = 0.5, \theta = 10^{-4}, \gamma = 1, D_k = I$. Para calcularmos $\overline{z}_k(\Delta)$ no Passo 3 do Algoritmo 3.2.1, usamos o Algoritmo 3.4.1. O critério de parada para o Algoritmo 3.4.1 (Passo 1) foi $\|\overline{g}_p(z^k)\| \leq \varepsilon$, onde $\varepsilon = TOL\|\overline{g}_p(0)\|$ e $TOL \in (0,1)$ é um parâmetro de tolerância fornecido pelo usuário. Naturalmente, a condição $\psi_k(\overline{z}_k(\Delta)) \leq \gamma Q(z_k^Q(\Delta))$ também é exigida. Neste ponto nossa implementação difere da de Conn, Gould e Toint [1989]. O procedimento de parada adotado por Conn, Gould e Toint (CGT) relativamente ao subproblema pode ser escrito, usando nossa notação, da seguinte forma:

(a) Seja F_I a face da região factível do subproblema que contém z^{CP}, *"ponto de Cauchy aproximado"*. Aplicar Gradientes Conjugados no interior de F_I até que uma das condições (b) ou (c) a seguir seja satisfeita:

(b) $z^k \in \overline{F}_I - F_I$.

(c) $\|\overline{g}_I(z^k)\| \leq \eta_k\|\overline{g}_p(0)\|$.

Desta maneira, o ponto tentativo na abordagem CGT preserva as restrições ativas do ponto de Cauchy aproximado, mesmo quando se está longe da solução. Conn, Gould e Toint usam $\eta_k = \min\{0.1, \|\overline{g}_p(0)\|^{1/2}\}$ em (c).

As razões pelas quais usamos uma estratégia diferente são as seguintes:

(i) Acreditamos possuir uma técnica poderosa para minimizar uma quadrática numa caixa (Algoritmo 3.4.1 – QUACAN) que nos fornece bons pontos tentativos, não necessariamente na face definida por z^{CP}.

(ii) A escolha de η_k na estratégia CGT é dependente do escalamento de f. De fato, se multiplicarmos f por um número positivo suficientemente grande, podemos forçar a escolha $\eta_k = 0.1$ em praticamente todas as iterações. Também não vemos com bons olhos a idéia de se fazer $\eta_k \to 0$, pois não consideramos prático nem natural trabalhar com este

limite.

O Algoritmo 3.4.1 (QUACAN) também necessita da especificação de alguns parâmetros importantes. Os limitantes K e L para o espectro da matriz Hessiana do modelo quadrático são, numa opção padrão, estimados usando-se o Teorema de Gerschgorin. Também é possível deixar que o programa obtenha estes limitantes pelo método das potências. O usuário, porém, pode introduzir outras estimativas mais adequadas à estrutura do problema em questão. Naturalmente, M_k no Passo 2 do Algoritmo 3.2.1 coincide com L no Algoritmo 3.4.1. Em geral, recomendamos a escolha $\delta = 0.1 \times \Delta_k$, utilizada em nossos experimentos, embora o usuário possa adotar outras escolhas. Tomando-se δ muito grande, a tendência das iterações do Algoritmo 3.4.1 será permanecer na face definida pelo primeiro ponto. Por outro lado, com a escolha de δ muito pequeno o Algoritmo 3.4.1 tentará abandonar esta face o mais rápido possível. De qualquer forma, se o δ fornecido pelo usuário não satisfizer (3.4.13), nosso programa faz a modificação $\delta \leftarrow 0.99 \times \delta_{max}$, onde $(0, \delta_{max})$ é o intervalo de valores admissíveis para δ de acordo com (3.4.13).

Na primeira chamada do Algoritmo 3.4.1 dentro de uma iteração do Algoritmo principal (BOX), escolhemos $z^0 = z_k^Q(\Delta)$. Esta escolha é razoável pois garante que a condição (3.2.3) requerida para convergência seja satisfeita por todas as iterações do Algoritmo 3.4.1. Entretanto, tendo em vista uma eficiência maior, deve-se adotar diferentes estratégias fornecendo pontos iniciais melhores para QUACAN. Desta forma, quando é necessário se chamar o Algoritmo 3.4.1 mais de uma vez dentro da mesma iteração de BOX, a aproximação inicial z^0 é escolhida como a projeção da última saída de QUACAN na nova região factível. Uma vez obtido o incremento tentativo usando-se esta estratégia para inicializar QUACAN, testamos se a primeira desigualdade em (3.2.3) é satisfeita. Em caso negativo (extremamente improvável), definimos $z_k(\Delta) = z_k^Q(\Delta)$.

Finalmente, permite-se a cada chamada do Algoritmo 3.4.1 um número máximo de q iterações. Assim, uma vez feitas q iterações do Algoritmo 3.4.1, mesmo que o critério de convergência $\|\overline{g}_p(z^k)\| \leq \varepsilon$ não seja satisfeito, simplesmente escolhemos $\overline{z}_k(\Delta) = z^q$, já que z^q obviamente satisfaz (3.2.3).

Nestes testes usamos o critério de convergência

$$\|P(\nabla f(x_k))\| \leq 10^{-4} \max\{|f(x_k)|, 1\} / \max\{\|x_k\|, 1\},$$

onde P é o operador projeção na caixa $\ell \leq x \leq u$. O critério de parada efetivamente usado é dado pelo parâmetro de saída IER, cujos significados para os diferentes valores são os seguintes:

IER = 1: Convergência. A norma do gradiente projetado é suficientemente pequena, conforme descrição acima.

IER = 2: Convergência declarada porque $f(x_k) \leq 10^{-4}$.

IER = 3: Atingiu 500 avaliações de função.

IER = 4: Não se obteve descenso suficiente com $\Delta_k \leq 10^{-4}$.

Em todos os experimentos, testamos nosso algoritmo com diferentes valores para o parâmetro TOL, que define o critério de parada do Algoritmo 3.4.1. Para fins comparativos, também testamos a estratégia para o subproblema de Conn, Gould e Toint, descrita anteriormente , indicada por CGT em nossas tabelas.

Apresentamos a seguir a descrição dos problemas teste.

Problema 1 ("O Problema dos Aeroportos").

Consideremos m bolas disjuntas em $\mathbb{R}^2$, cujos centros são escolhidos aleatoriamente no quadrado $[-10, 10] \times [-10, 10]$ e cujos raios $r_i, i = 1, \ldots, m$ também são tomados aleatoriamente entre 0 e $\rho/2$, onde ρ é a distância mínima entre os diferentes centros. O problema consiste em encontrar um ponto (x_i, y_i) em cada bola tal que $\sum \|(x_i, y_i) - (x_j, y_j)\|_2$ é mínima, onde a soma envolve todos os pares (i, j) tais que $1 \leq i \leq m, 1 \leq j \leq m$ e $i \neq j$. Fazendo uma mudança de coordenadas, as incógnitas passam a ser as distâncias entre cada ponto (x_i, y_i) e o centro da i-ésima bola e os ângulos entre os segmentos que unem (x_i, y_i) ao centro da bola correspondente e o eixo das abcissas. Em todos os testes trabalhamos com as matrizes Hessianas verdadeiras. A função objetivo não é convexa e muitas das Hessianas que aparecem ao longo das iterações são indefinidas. Resolvemos este problema para diferentes valores de m.

A aproximação inicial requerida pelo Algoritmo 3.2.1 foi tomada como sendo os centros das bolas. Em nossos testes usamos $\Delta^0 = \Delta_{min} = 5$ e a atualização dos raios de confiança obedeceu à regra $\Delta^{k+1} = \max\{\Delta_{min}, 4\Delta_k\}$. Cabe observar que optamos por começar cada iteração usando um raio grande, de forma a permitir que o método desse passos grandes longe da solução. Naturalmente, tal decisão pode implicar que muitas vezes haja um desperdício de avaliações de função em pontos condenados ao fracasso. De qualquer forma, acreditamos que esta perda de eficiência é compensada pelo fato de que em algumas situações a obtenção de soluções globais se torna possível justamente porque é permitido obter um ponto tentativo longe do ponto atual.

Os resultados são apresentados na Tabela 3.7.1. Para cada teste explicitamos o número de bolas m (então a dimensão do problema é $n = 2m$), a estratégia de resolução para o subproblema: abordagem CGT ou TOL $\in \{1, 0.1, 0.01, 0.001, 0.0001\}$, o critério de parada IER, o número total de iterações efetuadas pelo Algoritmo 3.2.1 ITER, o número total de iterações efetuadas na resolução dos subproblemas ITSUB e o número total de avaliações de função efetuadas AFUN.

Problema 2 (Condições de Otimalidade em Programação Linear).

Consideremos o problema de Programação Linear (PPL):

$$\begin{array}{ll} \min & c^T x \\ \text{s/a} & Ax = b \\ & x \geq 0 \end{array}$$

onde $A \in \mathbb{R}^{m \times n}$. Uma solução $x \in \mathbb{R}^n$ deste problema, as variáveis duais correspondentes $y \in \mathbb{R}^m$ e as variáveis de folga duais $z \in \mathbb{R}^n$ compõem a solução (x, y, z) do seguinte problema de minimização com restrições de canalização:

$$\begin{array}{ll} \min & \|Ax - b\|^2 + \|c - A^T y - z\|^2 + (x^T z)^2 \\ \text{s/a} & x \geq 0, \quad z \geq 0. \end{array}$$

Geramos aleatoriamente problemas deste tipo, usando uma solução conhecida do PPL correspondente. A geração aleatória dos coeficientes obedeceu ao seguinte roteiro:

a) Os elementos de A foram gerados aleatoriamente entre -10 e 10.

b) Uma solução aleatória x^* do PPL foi gerada aleatoriamente entre 0 e 20, com $n - m$ componentes (escolhidas aleatoriamente) nulas. Problemas com degeneração primal foram gerados fazendo-se $n - 0.6m$ componentes (escolhidas aleatoriamente) valerem zero.

c) Os multiplicadores y^* foram escolhidos aleatoriamente entre -3 e 3.

d) As folgas duais z^* foram escolhidas aleatoriamente entre 0 e 10, respeitando-se entretanto a condição de complementaridade $x_i^* z_i^* = 0, i = 1, \ldots, n$.

e) O vetor gradiente c foi calculado utilizando-se a informação acima.

Para resolvermos os problemas tomamos uma aproximação inicial nula para x, y e z. Também usamos $\Delta^0 = \Delta_{min} = 10$.

Os resultados são apresentados na Tabela 3.7.2, onde destacamos: m, n: respectivamente, o número de linhas e colunas da matriz A; Zero: número de elementos nulos na solução primal; Estratégia, IER, ITER, ITSUB e AFUN: o mesmo que no problema dos Aeroportos.

É interessante observar que, para este problema, sempre obtivemos minimizadores globais. Tal comportamento se deve a uma propriedade teórica mais geral que apresentamos no Capítulo 4 deste trabalho.

M	Estratégia	IER	ITER	ITSUB	AFUN
5	CGT	1	9	70	14
	1.0	1	68	0	69
	0.1	1	5	10	8
	0.01	1	4	12	7
	0.001	1	4	16	7
	0.0001	1	4	17	7
10	CGT	1	8	92	10
	1.0	3	499	0	500
	0.1	1	6	15	7
	0.01	1	5	20	9
	0.001	1	5	31	9
	0.0001	1	5	40	9
15	CGT	1	22	161	25
	1.0	3	499	0	500
	0.1	1	8	45	9
	0.01	1	9	93	24
	0.001	1	10	132	25
	0.0001	1	11	180	26
20	CGT	1	7	160	9
	1.0	3	499	0	500
	0.1	1	9	57	14
	0.01	1	6	59	12
	0.001	1	10	202	25
	0.0001	1	9	243	25
25	CGT	1	11	189	15
	1.0	3	499	0	500
	0.1	1	7	40	9
	0.01	1	8	114	17
	0.001	1	5	62	8
	0.0001	1	5	76	8

Tabela 3.7.1: Experimentos com Problema 1

M	Estratégia	IER	ITER	ITSUB	AFUN
30	CGT	1	8	149	9
	1.0	3	499	0	500
	0.1	1	9	108	12
	0.01	1	4	39	7
	0.001	1	4	53	7
	0.0001	1	7	178	14
35	CGT	1	13	481	19
	1.0	3	499	0	500
	0.1	1	7	28	8
	0.01	1	6	59	9
	0.001	1	6	109	11
	0.0001	1	6	127	10
40	CGT	1	12	141	15
	1.0	3	499	0	500
	0.1	1	7	40	8
	0.01	1	5	52	8
	0.001	1	6	85	9
	0.0001	1	6	108	9
45	CGT	1	16	150	18
	1.0	3	499	0	500
	0.1	1	11	57	12
	0.01	1	5	70	8
	0.001	1	4	71	7
	0.0001	1	4	82	7
50	CGT	1	10	349	12
	1.0	3	499	0	500
	0.1	1	8	105	9
	0.01	1	4	51	7
	0.001	1	4	154	7
	0.0001	1	7	280	17

Tabela 3.7.1: Experimentos com Problema 1 (continuação)

M	N	ZERO	Estratégia	IER	ITER	ITSUB	AFUN
5	10	7	CGT	4	214	5372	330
			1.0	3	499	0	500
			0.1	2	8	129	9
			0.01	2	10	212	11
			0.001	2	7	170	8
			0.0001	2	19	475	20
5	10	5	CGT	3	293	7631	500
			1.0	3	499	0	500
			0.1	2	14	260	15
			0.01	2	16	351	17
			0.001	2	20	468	21
			0.0001	2	132	3284	133
5	20	17	CGT	3	304	13746	500
			1.0	3	499	10	500
			0.1	2	15	411	16
			0.01	2	10	359	11
			0.001	2	5	188	6
			0.0001	2	17	740	18
5	20	15	CGT	4	261	11015	380
			1.0	3	499	0	500
			0.1	2	20	472	21
			0.01	2	15	517	16
			0.001	2	17	607	18
			0.0001	2	112	5001	113
5	30	27	CGT	3	322	17814	501
			1.0	3	499	18	500
			0.1	2	15	469	16
			0.01	2	15	554	16
			0.001	2	5	255	6
			0.0001	2	5	255	6
5	30	25	CGT	4	284	15847	404
			1.0	3	499	12	500
			0.1	2	22	770	23
			0.01	2	19	772	20
			0.001	2	14	704	15
			0.0001	2	82	5314	83

Tabela 3.7.2: Experimentos com Problema 2

M	N	ZERO	Estratégia	IER	ITER	ITSUB	AFUN
10	20	14	CGT	2	179	7153	241
			1.0	3	499	50	500
			0.1	2	11	316	12
			0.01	2	8	319	9
			0.001	2	17	817	18
			0.0001	2	13	628	14
10	20	10	CGT	3	282	13155	500
			1.0	3	499	0	500
			0.1	2	66	2820	67
			0.01	2	24	1114	25
			0.001	2	19	929	20
			0.0001	2	160	7990	161
10	30	24	CGT	3	240	14149	502
			1.0	3	499	0	500
			0.1	2	20	922	21
			0.01	2	21	1375	22
			0.001	2	5	282	6
			0.0001	2	34	2380	35
10	30	20	CGT	3	291	15683	500
			1.0	3	499	0	500
			0.1	2	95	5122	96
			0.01	2	76	5185	77
			0.001	2	51	3451	52
			0.0001	2	311	21752	312
10	40	34	CGT	3	370	26005	500
			1.0	3	499	0	500
			0.1	2	29	1711	30
			0.01	2	21	1682	22
			0.001	2	5	364	6
			0.0001	2	6	512	7
10	40	30	CGT	3	325	23466	501
			1.0	3	499	0	500
			0.1	2	59	3204	60
			0.01	2	59	4533	60
			0.001	2	22	1814	23
			0.0001	2	125	11230	126

Tabela 3.7.2: Experimentos com Problema 2 (continuação)

M	N	ZERO	Estratégia	IER	ITER	ITSUB	AFUN
15	20	11	CGT	4	137	3189	255
			1.0	3	499	0	500
			0.1	2	6	160	7
			0.01	2	3	114	4
			0.001	2	4	180	5
			0.0001	2	6	309	7
15	20	5	CGT	4	114	5838	368
			1.0	3	499	0	500
			0.1	2	46	2143	47
			0.01	2	34	1791	35
			0.001	2	24	1265	25
			0.0001	2	124	6788	125
15	30	21	CGT	3	320	17676	500
			1.0	3	499	0	500
			0.1	2	11	479	12
			0.01	2	8	486	9
			0.001	2	5	356	6
			0.0001	2	33	2475	34
15	30	15	CGT	3	237	13750	500
			1.0	3	499	0	500
			0.1	2	59	3489	60
			0.01	2	68	4826	69
			0.001	2	78	5795	79
			0.0001	3	499	37403	500
15	40	31	CGT	3	237	16924	501
			1.0	3	499	0	500
			0.1	2	22	1393	23
			0.01	2	26	2250	27
			0.001	2	37	3456	38
			0.0001	2	23	2158	24
15	40	25	CGT	3	220	14771	506
			1.0	3	499	0	500
			0.1	2	76	5173	77
			0.01	2	62	5342	63
			0.001	2	65	5837	66
			0.0001	2	105	9952	106

Tabela 3.7.2: Experimentos com Problema 2 (continuação)

M	N	ZERO	Estratégia	IER	ITER	ITSUB	AFUN
20	30	18	CGT	3	288	12692	500
			1.0	3	499	0	500
			0.1	2	10	537	11
			0.01	2	8	549	9
			0.001	2	9	673	10
			0.0001	2	6	480	7
20	30	10	CGT	4	176	13210	470
			1.0	3	499	0	500
			0.1	3	499	37616	500
			0.01	3	499	39796	500
			0.001	3	499	39859	500
			0.0001	2	477	38148	478
20	35	23	CGT	3	278	15755	500
			1.0	3	499	0	500
			0.1	2	19	1285	20
			0.01	2	6	445	7
			0.001	2	4	330	5
			0.0001	2	3	270	4
20	35	15	CGT	3	230	18491	500
			1.0	3	499	0	500
			0.1	2	58	4471	59
			0.01	2	41	3607	42
			0.001	2	66	5871	67
			0.0001	3	499	44878	500
20	40	28	CGT	3	231	12899	500
			1.0	3	499	0	500
			0.1	2	47	3861	48
			0.01	2	43	4124	44
			0.001	2	50	4929	51
			0.0001	2	31	3092	32
20	40	20	CGT	3	198	14430	501
			1.0	3	499	0	500
			0.1	2	89	6730	90
			0.01	2	82	8044	83
			0.001	2	101	9999	102
			0.0001	2	244	24363	245

Tabela 3.7.2: Experimentos com Problema 2 (continuação)

3.8 OBSERVAÇÕES FINAIS

Neste capítulo apresentamos uma nova estratégia para minimizar funções diferenciáveis com variáveis canalizadas (Algoritmo BOX). De acordo com a filosofia atual dos métodos de região de confiança, não é necessário determinar a solução exata do subproblema para se obter convergência a um ponto estacionário de primeira ordem. Além disso, mostramos propriedades de convergência independentemente de computações prévias do *"ponto de Cauchy aproximado"*, em que se baseia a abordagem de Conn, Gould e Toint [1988a–b, 1989, 1990]. Para resolvermos o subproblema, usamos um novo método para minimizar uma quadrática com restrições de canalização. O fato dos resultados de convergência serem obtidos sem nenhuma alusão ao *"ponto de Cauchy aproximado"* torna possível definir estratégias naturais de inicialização para o algoritmo QUACAN. Em nossos experimentos testamos uma destas estratégias, mas muitas outras são possíveis. O método de minimização quadrática (Algoritmo QUACAN) combina buscas em caminhos poligonais com iterações de Gradientes Conjugados. Desta forma, a memória necessária é mínima, o que permite a resolução de problemas de grande porte. A principal diferença entre a nossa estratégia e a de Conn, Gould e Toint é que o método destes se baseia fortemente nas propriedades do *"ponto de Cauchy aproximado"*, enquanto nós procuramos explorar as excelentes propriedades de nosso algoritmo QUACAN. Cabe observar que tal política se mostrou compensadora através dos experimentos efetuados. De fato, em praticamente todos os testes nossa estratégia para o subproblema com TOL $= 0.1$ superou a estratégia CGT. De qualquer forma, uma pesquisa mais específica pode ser feita com o objetivo de detectar tipos de problema em que uma ou outra estratégia seja mais eficaz.

CAPÍTULO 4

UMA ESTRATÉGIA PARA MINIMIZAR FUNÇÕES CONVEXAS COM RESTRIÇÕES LINEARES

4.1 INTRODUÇÃO

Neste capítulo consideramos o problema:

$$\begin{array}{ll} \min & f(x) \\ \text{s/a} & Ax = b \\ & x \geq 0, \end{array} \tag{4.1.1}$$

onde $f \in C^2(\mathbb{R}^n)$ é convexa, $A \in \mathbb{R}^{m \times n}$ e o conjunto $\Omega \equiv \{x \in \mathbb{R}^n \mid Ax = b, x \geq 0\}$ é não vazio e limitado.

Estamos especialmente interessados no caso em que m e n são muito grandes e a estrutura da matriz A é tal que mesmo fatorações esparsas (ver Duff, Erisman e Reid [1986]) são inviáveis. Nestes casos, métodos de restrições ativas (ver Fletcher [1987]; Gill, Murray e Wright [1981], etc.) não podem ser aplicados a (4.1.1).

Um caso particular muito importante de (4.1.1) ocorre quando f é uma função quadrática convexa. De fato, se formos capazes de resolver eficientemente problemas quadráticos de grande porte da forma (4.1.1), então teremos o ingrediente essencial para o desenvolvimento de métodos eficientes baseados em Programação Quadrática Sequencial

(ver Dembo e Tulowitzki [1985]; Nickel e Tolle [1989]; Yabe, Yamaki e Takahashi [1991], etc.).

Nos últimos anos vem sendo reconhecido pela comunidade de otimizadores que a minimização de uma função com restrições simplesmente canalizadas é um problema cuja resolução eficiente para grande porte pode gerar algoritmos práticos bastante adequados para programação não linear geral (ver Conn, Gould e Toint [1988a-b, 1989, 1990, 1991]). Com base nesta filosofia, foram desenvolvidos algoritmos tanto para minimização geral com canalização quanto para minimizar quadráticas em caixas (ver Conn, Gould e Toint [1988, 1989]; Moré e Toraldo [1989, 1991]). Desta forma, com a disponibilidade de algoritmos e programas para minimização em caixas, torna-se natural que se tente reduzir problemas mais difíceis para este formato, mesmo às custas de se aumentar o número de variáveis. Esta é a idéia básica de nossa estratégia para minimizar funções convexas com restrições lineares. Convertemos o problema original (4.1.1) para o formato (3.1.1), resolvendo-o através do Algoritmo 3.2.1 (BOX). O interessante, porém, é que o problema (4.1.1) pode ser convertido em um problema canalizado equivalente ao original, que consiste em minimizar a norma Euclidiana do sistema não-linear obtido pelas condições de Karush-Kuhn-Tucker para (4.1.1), incluindo-se as condições de folgas complementares e para o qual pontos estacionários coincidem com minimizadores globais.

A organização deste capítulo é a seguinte: a prova do resultado de equivalência é feita na Seção 4.2. Na Seção 4.3 esta nova técnica é aplicada para estimar soluções de alguns sistemas de equações lineares com restrições lineares. Concluindo, algumas observações finais são feitas na Seção 4.4.

4.2 UM RESULTADO DE EQUIVALÊNCIA.

As condições de otimalidade de Karush-Kuhn-Tucker para (4.1.1) são:

$$\left.\begin{array}{ll} \nabla f(x)+ & A^T y - z = 0 \\ & Ax = b \\ & x^T z = 0 \\ & x \geq 0, z \geq 0. \end{array}\right\} \tag{4.2.1}$$

Devido à convexidade da função objetivo f, (4.2.1) são condições necessárias e suficientes para que $x \in \mathbb{R}^n$ seja um minimizador de (4.1.1). Os vetores $y \in \mathbb{R}^m$ e $z \in \mathbb{R}^n$ representam, respectivamente, as variáveis duais e as folgas duais complementares.

Agora, como f é contínua e o conjunto factível Ω é não vazio e limitado, o problema (4.1.1) admite um minimizador global x, com multiplicadores de Lagrange associados y e variáveis de folga duais z. Obviamente, os minimizadores globais de (4.1.1) coincidem com os minimizadores globais do seguinte problema:

$$\left.\begin{array}{ll} \min & F(x,y,z) \equiv \frac{1}{2}(\|\nabla f(x) + A^T y - z\|^2 + \|Ax - b\|^2 + (x^T z)^2) \\ \text{s/a} & x \geq 0, z \geq 0. \end{array}\right\} \tag{4.2.2}$$

Além disso, por (4.2.1), o valor da função objetivo de (4.2.2) é zero num minimizador global.

Agora, (4.2.2) é um problema de minimização com restrições simples, que pode ser resolvido, mesmo em situações de porte enorme, usando-se o Algoritmo 3.2.1 (BOX) ou a estratégia de Conn, Gould e Toint [1989]. Entretanto, a função objetivo de (4.2.2) não é convexa e algoritmos típicos para resolver problemas com restrições de canalização são convergentes apenas a pontos estacionários, não necessariamente minimizadores globais. No teorema a seguir mostramos que, felizmente neste caso, apesar da não convexidade da função F, pontos estacionários e minimizadores globais coincidem.

TEOREMA 4.2.1. Se $f \in C^2(\mathbb{R}^n)$ é convexa e o conjunto Ω é não vazio e limitado, então o problema (4.2.2) tem pelo menos um ponto estacionário (ponto de Karush-Kuhn-Tucker) e todo ponto estacionário de (4.2.2) é um minimizador global.

Demonstração. A primeira parte é trivial. Como Ω é limitado e f é contínua, o problema (4.1.1) tem um minimizador global. Este minimizador satisfaz (4.2.1) e portanto é um minimizador global de (4.2.2).

Suponhamos agora que (x, y, z) seja um ponto estacionário de (4.2.2). Então, existem $\gamma, \mu \in \mathbb{R}^n$ tais que:

$$A^T(Ax - b) + \nabla^2 f(x)(\nabla f(x) + A^T y - z) + (x^T z)z - \gamma = 0, \tag{4.2.3}$$

$$A(\nabla f(x) + A^T y - z) = 0, \tag{4.2.4}$$

$$-(\nabla f(x) + A^T y - z) + (x^T z)x - \mu = 0, \tag{4.2.5}$$

$$\gamma^T x = 0, \tag{4.2.6}$$

$$\mu^T z = 0, \tag{4.2.7}$$

$$x \geq 0, z \geq 0, \gamma \geq 0, \mu \geq 0. \tag{4.2.8}$$

Por (4.2.4) e (4.2.5) temos que

$$(x^T z)x - \mu \in \mathcal{N}, \tag{4.2.9}$$

onde $\mathcal{N}$ é o núcleo de A.

Desta forma, pré-multiplicando (4.2.3) por $(x^T z)x - \mu$ e usando (4.2.5), obtemos:

$$(x^T zx - \mu)^T \nabla^2 f(x)(x^T zx - \mu) + (x^T zx - \mu)^T (x^T zz - \gamma) = 0. \tag{4.2.10}$$

Como $\nabla^2 f(x)$ é semi-definida positiva, (4.2.10) implica em

$$(x^T zx - \mu)^T (x^T zz - \gamma) \leq 0.$$

Assim, por (4.2.6) e (4.2.7) segue que

$$(x^T z)^3 + \mu^T \gamma \leq 0. \tag{4.2.11}$$

Logo, por (4.2.8),

$$x^T z = 0 \tag{4.2.12}$$

e

$$\mu^T \gamma = 0. \tag{4.2.13}$$

Por (4.2.5) e (4.2.12),

$$-(\nabla f(x) + A^T y - z) = \mu \geq 0. \tag{4.2.14}$$

Mas, por (4.2.4), $-(\nabla f(x) + A^T y - z) \in \mathcal{N}$, e portanto, como Ω é limitado, (4.2.14) necesariamente implica em

$$-(\nabla f(x) + A^T y - z) = 0. \tag{4.2.15}$$

Então, por (4.2.3), (4.2.12) e (4.2.15),

$$A^T(Ax - b) = \gamma \geq 0. \tag{4.2.16}$$

Agora, (4.2.16) e (4.2.6) são as condições de otimalidade (necessárias e suficientes) para o seguinte problema quadrático convexo:

$$\begin{array}{ll} \min & \frac{1}{2}\|Ax - b\|^2 \\ \text{s/a} & x \geq 0. \end{array} \tag{4.2.17}$$

Assim, como Ω é não vazio, segue necessariamente que $Ax = b$. Esta igualdade, juntamente com (4.2.12) e (4.2.15) completa a prova. □

Observação. O problema

$$\begin{array}{ll} \min & \frac{1}{2}(\|\nabla f(x) + A^T y - z\|^2 + \|Ax - b\|^2 + x^T z) \\ \text{s/a} & x \geq 0, z \geq 0 \end{array} \tag{4.2.18}$$

é obviamente equivalente a (4.2.2). Apesar disso, é interessante notar que (4.2.18) pode ter pontos estacionários que não são minimizadores globais. De fato, basta considerarmos o problema de minimizar x sujeito a $0 \leq x \leq 2$ ou, no formato (4.1.1), minimizar x_1 sujeito a $x_1 + x_2 = 2$, $x_1 \geq 0, x_2 \geq 0$. O problema da forma (4.2.18) associado a este problema trivial admite o ponto estacionário $x = (2,0)^T$, $z = (0,0)^T$ que naturalmente não é um minimizador global de (4.2.18).

4.3 EXPERIMENTOS NUMÉRICOS.

Estamos interessados em encontrar estimadores robustos para parâmetros em problemas modelados por sistemas de equações lineares, sobredeterminados e de grande porte, com restrições lineares e variáveis canalizadas.

Assumimos que o sistema linear sobredeterminado é dado por

$$Hx = c, \tag{4.3.1}$$

e que as restrições são

$$Ax = b, \tag{4.3.2}$$

$$x \geq 0, \tag{4.3.3}$$

onde $A \in \mathbb{R}^{m \times n}, H \in \mathbb{R}^{\ell \times n}$ e o conjunto definido por (4.3.2) e (4.3.3) é não vazio e limitado. Usamos $m = n/2$ e $\ell = 2n$.

Para estimarmos os parâmetros $x_i, i = 1, \ldots, n$, poderíamos resolver o seguinte problema de programação quadrática:

$$\begin{array}{ll} \min & \frac{1}{2}\|Hx - c\|^2 \\ \text{s/a} & Ax = b \\ & x \geq 0. \end{array} \tag{4.3.4}$$

Sabe-se, no entanto, que a função objetivo quadrática como medidora de erro é muito sensível a valores extremos nas observações (*"outliers"*). Assim, é preferível escolher uma função que não tenha tal desvantagem, o que nos fez considerar o seguinte problema:

$$\begin{array}{ll} \min & \phi([Hx - c]_1) + \ldots + \phi([Hx - c]_\ell) \\ \text{s/a} & Ax = b \\ & x \geq 0, \end{array} \tag{4.3.5}$$

onde $\phi(t) = log(cosh(t))$ e $[Hx - c]_k, k = 1, \ldots, \ell$, é a k-ésima componente do vetor $Hx - c \in \mathbb{R}^\ell$ (ver Green [1990a-b] e Lange [1990]). Trata-se de uma função objetivo convexa e duas vezes diferenciável, o que nos permite aplicar as técnicas da Seção 4.2 e os algoritmos do Capítulo 3 (BOX-QUACAN) para resolver o problema (4.3.5).

Os problemas-teste foram gerados da seguinte forma:

a) O elemento a_{ij} da matriz A é $i - j$. O elemento h_{ij} de H é $(i - 3j)/(i + 3j)$.

b) Foi gerada uma solução x^* aleatoriamente entre 0 e 10.

c) O vetor b foi calculado pelo produto Ax^*.

d) Tomamos $c^* = Hx^*$ e $c_i = c_i^*(1 + r_i PERT/100)$, onde $PERT$ representa uma perturbação percentual sobre o vetor verdadeiro c^* e r_i é um número aleatório entre -1 e 1.

Foram gerados diferentes problemas, variando-se n e $PERT$. Cada problema foi convertido para o formato (4.2.2) e resolvido usando-se o Algoritmo 3.2.1 (BOX). No modelo quadrático (3.2.4) foram utilizados as Hessianas verdadeiras, a menos dos termos envolvendo as derivadas terceiras da função objetivo de (4.3.5), que foram desprezados. Tal simplificação se justifica completamente do ponto de vista da convergência local, já que todo termo de terceira ordem é multiplicado por um fator que se anula na solução. Assim, é esperado que o processo tenha um comportamento semelhante àquele em que as Hessianas completas e mais caras são usadas.

Como aproximação inicial para a solução de (4.2.2) tomamos $x_0 = 0, y_0 = 0, z_0 = 0$. O procedimento iterativo foi interrompido quando algum dos seguintes critérios foi atingido:

a) A norma do gradiente projetado de F em (4.2.2) é menor que $\dfrac{10^{-4}\max\{1, |F(x_k, y_k, z_k)|\}}{\max\{1, \|(x_k, y_k, z_k)\|\}}$.

b) $F(x_k, y_k, z_k) \leq 10^{-8}$.

c) $\Delta_k \leq 10^{-4}$.

Os resultados são apresentados na Tabela 4.3.1. Para cada teste exibimos n, $PERT$ e as seguintes informações:

ITER: Número de iterações efetuadas até a convergência.

AFUN: Número efetuado de avaliações de função.

V1: Valor da função objetivo do problema (4.2.2) no ponto $(x, y, z)_{FINAL}$.

V2: Valor da função objetivo do problema (4.3.5) no ponto $(x, y, z)_{FINAL}$.

LF: $\|Ax_{FINAL} - b\|$.

N	*PERT*	ITER	AFUN	V1	V2	LF
6	0	18	27	4.1E-9	1.7E-5	1.0E-7
	10	22	30	1.4E-7	4.0E-1	8.2E-6
	20	14	20	1.2E-5	1.5	8.6E-5
	30	13	18	8.3E-6	2.9	6.2E-5
	40	15	21	1.9E-6	4.4	2.4E-5
	50	16	22	8.4E-6	6.1	5.0E-5
12	0	23	34	5.0E-12	5.0E-6	6.5E-10
	10	23	30	2.2E-4	6.4	4.5E-6
	20	41	52	1.2E-5	17.58	7.0E-6
	30	45	61	9.2E-6	30.75	2.2E-6
	40	42	57	9.7E-5	45.36	1.2E-5
	50	33	39	1.2E-9	59.96	7.0E-7
18	0	38	78	2.2E-11	3.5E-6	2.6E-7
	10	45	66	5.3E-7	21.81	1.9E-6
	20	41	72	1.3E-5	54.91	1.3E-5
	30	56	72	8.2E-6	90.85	8.8E-7
	40	39	70	5.1E-4	129.20	3.3E-5
	50	70	97	2.3E-5	167.03	1.0E-6
24	0	31	50	6.6E-9	9.6E-5	2.3E-8
	10	31	44	1.8E-4	45.19	1.4E-6
	20	53	73	1.7E-4	108.71	7.7E-7
	30	74	101	1.1E-4	174.40	1.1E-7
	40	58	87	1.2E-4	240.93	1.3E-6
	50	56	79	3.1E-5	307.77	1.4E-6
30	0	52	92	1.5E-7	8.5E-4	2.6E-7
	10	95	128	2.1E-4	83.87	1.9E-6
	20	55	103	1.8E-2	200.46	3.2E-5
	30	83	94	2.9E-4	312.12	2.8E-7
	40	113	181	3.6E-4	428.58	1.8E-5
	50	91	152	2.5E-4	545.57	1.7E-5

Tabela 4.3.1: Experimentos Numéricos.

4.4 OBSERVAÇÕES FINAIS.

Neste capítulo mostramos que problemas de minimização com função objetivo convexa e restrições lineares podem ser resolvidos sem a utilização de fatorações de matrizes, desde que seja disponível um algoritmo eficiente para minimizar funções arbitrárias com restrições de canalização, mesmo que para tal algoritmo haja garantia apenas de convergência a pontos estacionários, como é padrão. Com a popularidade crescente de algoritmos para resolver problemas de grande porte com restrições do tipo caixas, cremos que a nossa abordagem é particularmente interessante para resolver problemas em que não é viável qualquer tipo de fatoração de matrizes. Acreditamos que esta seja a primeira estratégia para resolver problemas deste tipo que não faz uso de nenhum tipo de parâmetro penalizador. Programação quadrática convexa é apenas um caso particular dos tipos de problemas que podemos resolver, mas muitos outros problemas convexos podem ser relevantes em aplicações. Nossos experimentos numéricos parecem indicar que a abordagem proposta é viável, sobretudo porque dispomos de um algoritmo robusto (BOX) para resolver o problema (4.2.2). Esta técnica também pode ser aplicada na resolução de problemas de complementaridade linear sem nenhuma hipótese de limitação e em problemas de inequações variacionais (Friedlander, Martínez e Santos [1993, 1994]). Acreditamos que a utilização da técnica introduzida neste capítulo para a resolução dos subproblemas em programação quadrática sequencial de grande porte seja uma fonte promissora de pesquisas futuras.

CAPÍTULO 5

O PROBLEMA DO VETOR INICIAL EM CODIFICAÇÃO

5.1 INTRODUÇÃO

O canal gaussiano de comunicação é um modelo introduzido por Shannon [1948] no qual as mensagens a serem transmitidas são representadas por vetores em $\mathbb{R}^n$ com norma euclidiana unitária. O transmissor possui um conjunto finito de mensagens disponíveis, denominado *grupo de códigos*, que serão enviadas a um receptor através de um canal com ruídos obedecendo distribuição gaussiana. Em outras palavras, quando o vetor x é transmitido, o sinal recebido é representado pelo vetor $y = x + z$, que consiste do vetor enviado x acrescido de um vetor de ruídos z independente de x e cujas componentes são variáveis aleatórias obedecendo distribuição Gaussiana com média zero e variância conhecida. O receptor, que também possui o grupo de códigos, deve decidir qual foi a mensagem enviada. Por exemplo, se o grupo de códigos é $\{(1,0),(0,1)\}$ e o receptor recebe $y = (0.8, 0.1)$, a decisão com respeito à mensagem transmitida é feita escolhendo-se no grupo de códigos a mensagem possível que fica euclidianamente mais próxima da mensagem recebida, neste caso, $x = (1,0)$. Para evitar qualquer confusão por parte do receptor, é preciso que as mensagens no grupo de códigos estejam distantes entre si o mais possível. Assim, a escolha do grupo de códigos é crucial para a eficiência deste modelo de comunicação. Uma simplificação para esta escolha é supor que cada elemento do grupo de códigos é gerado pela multiplicação de um vetor inicial pelos elementos de um grupo de matrizes ortogonais previamente fixado. Neste sentido, o problema de encontrar o melhor vetor inicial e, portanto, o melhor código gerado por um certo grupo é conhecido na

teoria de codificação como o *problema do vetor inicial* (PVI) (ver Karlof [1989]). Embora já tenha sido resolvido em alguns casos especiais (Blake [1972], Djokovic e Blake [1972], Downey e Karlof [1980], Slepian [1968]), o PVI é um problema difícil cuja solução geral ainda não foi determinada.

Com base nos trabalhos de Karlof [1989] e Blake [1972], usamos grupos de matrizes ortogonais, mais especificamente grupos de permutações, para gerar os códigos. Em nossa abordagem, formulamos o PVI como um problema de programação não linear e utilizamos minimização em esferas (ver Seção 2.4) para resolvê-lo. Aplicamos nosso algoritmo para os grupos de permutação simétricos, em que originalmente o PVI possui até cerca de 1.800.000 restrições.

A organização deste capítulo é a seguinte: na Seção 5.2 apresentamos a formulação original do PVI bem como as formulações utilizadas em nossa abordagem. Na Seção 5.3 detalhamos os experimentos numéricos efetuados. Concluímos com a Seção 5.4, onde fazemos algumas observações finais.

5.2 FORMULAÇÕES PARA O PVI

Para viabilizar a decisão do receptor com respeito à mensagem enviada veremos que o PVI consiste em encontrar o código cuja distância mínima aos demais seja a maior possível. Inicialmente, o objetivo é achar $y_1, \ldots, y_p$, p vetores em $\mathbb{R}^n$ que solucionem o seguinte problema:

$$\max_{\substack{y_1,\ldots,y_p \in \mathbb{R}^n \\ \|y_i\|=1}} \min_{i \neq j} \|y_i - y_j\|, \tag{5.2.1}$$

onde $\|\cdot\| = \|\cdot\|_2$.

Agora, como $\|y_i - y_j\|^2 = 2 - 2y_i^T y_j$ quando $\|y_i\| = \|y_j\| = 1$, segue que maximizar a mínima distância é equivalente a minimizar o máximo produto interno.

Portanto, (5.2.1) pode ser reescrito como:

$$\min_{\substack{y_1,\ldots,y_p \in \mathbb{R}^n \\ \|y_i\|=1}} \max_{i \neq j} \; y_i^T y_j. \tag{5.2.2}$$

O problema (5.2.2) pode ser formulado da seguinte maneira:

$$\begin{array}{ll} \min\limits_{\substack{z\in\mathbb{R}\\ y_1,\ldots,y_p\in\mathbb{R}^n}} & z \\ \text{s/a} & \|y_i\| = 1, \qquad i = 1,\ldots,p \\ & y_i^T y_j \leq z, \ \text{ para todo } i \neq j\ . \end{array} \tag{5.2.3}$$

Quando existem muitas mensagens possíveis, (5.2.3) é um problema de programação não linear de porte enorme.

Agora, suponhamos que as mensagens y_i são da forma $y_i = G_i x$, com x fixo e $G_i \in \mathcal{G} = \{G_1, \ldots, G_p\}$, onde $\mathcal{G}$ é um grupo ortogonal de matrizes. O vetor x é denominado *vetor inicial.*

Assim, como G_i é ortogonal, $\|y_i\| = \|G_i x\| = \|x\|$ e sendo $\mathcal{G}$ um grupo, $y_i^T y_j = x^T G_i^T G_j x = x^T G_i^{-1} G_j x$, com $G_i^{-1} G_j \in \mathcal{G}$. Desta forma, podemos reescrever (5.2.3) como:

$$\begin{array}{ll} \min\limits_{\substack{z\in\mathbb{R}\\ x\in\mathbb{R}^n}} & z \\ \text{s/a} & \|x\| = 1 \\ & x^T G_j x \leq z, \ \ j = 1 \ldots p. \end{array} \tag{5.2.4}$$

É importante notar que $\mathcal{G}$ não deve conter a identidade nem a inversa de um elemento do grupo. Isto porque se $G_i = I$ para algum i, então as restrições $x^T x \leq z$ e $\|x\| = 1$ fariam com que $z = 1$ fosse solução, o que não é desejável. Agora, se existe $j \neq i$ tal que $G_j = G_i^{-1} = G_i^T$, como $x^T G_i x = x^T G_i^T x = x^T G_j x$, teríamos restrições redundantes. Assim, para que o problema fique bem definido, $\mathcal{G}$ é um grupo ortogonal de matrizes obtido eliminando-se a identidade e a inversa (quando distinta) de cada elemento do grupo ortogonal original.

Quando $\mathcal{G}$ é um grupo de permutações, hipótese assumida daqui em diante, sabe-se (ver, por exemplo, Karlof [1989]) que o vetor inicial ótimo $x^* \in \mathbb{R}^n$ satisfaz a propriedade

$$\sum_{i=1}^{n} x_i^* = 0. \tag{5.2.5}$$

O hiperplano ao qual pertence x^* pode ser interpretado como o conjunto dos pontos que mais se afastam do vetor $e = (1, 1, \ldots, 1) \in \mathbb{R}^n$, que permanece fixo para qualquer código do grupo de permutações $\mathcal{G}$.

Comparando-se os problemas (5.2.3) e (5.2.4), vemos que apesar do número de variáveis e da quantidade de restrições ter diminuído com a introdução do grupo $\mathcal{G}$, a ordem (número de elementos) p deste grupo ainda pode ser grande o suficiente para inviabilizar qualquer tratamento eficiente de (5.2.4) diretamente como um problema com restrições. Neste sentido, vamos reescrever (5.2.4) no formato (5.2.2) e analisar outras maneiras de abordar o PVI. Temos, portanto,

$$\min_{\|x\|=1} \max_{G\in\mathcal{G}} x^T G x. \tag{5.2.6}$$

Como $\max_{G\in\mathcal{G}} x^T G x$ não é diferenciável e $-1 \leq x^T G x \leq 1$ pois $\|x\| = \|Gx\| = 1$, temos $1 \leq x^T G x + 2 \leq 3$. Assim,

$$\begin{aligned} \max_{G\in\mathcal{G}} x^T G x &\sim \max_{G\in\mathcal{G}}(x^T G x + 2) \sim \max_{G\in\mathcal{G}} |x^T G x + 2| \\ &\sim \lim_{p\to\infty} \left[\sum_{G\in\mathcal{G}} |x^T G x + 2|^p\right]^{1/p}, \end{aligned}$$

onde $\sim$ denota a equivalência entre os problemas. Desta forma, propomos que (5.2.6) seja substituído por

$$\min_{\|x\|=1} \left[\sum_{G\in\mathcal{G}} |x^T G x + 2|^p\right]^{1/p}, \; p >> 1. \tag{5.2.7}$$

TEOREMA 5.2.1 Os pontos estacionários do problema (5.2.7) satisfazem a propriedade (5.2.5).

Demonstração. O problema (5.2.7) é equivalente a

$$\begin{aligned} \min \quad & f(x) = \left[\sum_{G\in\mathcal{G}} (x^T G x + 2)^p\right]^{1/p} \\ s/a \quad & x^T x = 1\,. \end{aligned} \tag{5.2.8}$$

Agora,

$$\nabla f(x) = \frac{1}{p}\left[\sum_{G\in\mathcal{G}}(x^TGx+2)^p\right]^{\frac{1}{p}-1}\sum_{G\in\mathcal{G}}p(x^TGx+2)^{p-1}(G+G^T)x.$$

Ou seja,

$$\nabla f(x) = \alpha\sum_{G\in\mathcal{G}}\beta_G(G+G^T)x \tag{5.2.9}$$

onde

$$\alpha = \frac{1}{p}\left[\sum_{G\in\mathcal{G}}(x^TGx+2)^p\right]^{\frac{1}{p}-1} \tag{5.2.10}$$

e

$$\beta_G = p(x^TGx+2)^{p-1}. \tag{5.2.11}$$

Seja $\overline{x}\in\mathbb{R}^n$ um ponto estacionário para o problema (5.2.8). Então,

$$\nabla f(\overline{x}) = \mu\overline{x}, \mu\in\mathbb{R} \tag{5.2.12}$$

e

$$\overline{x}^T\overline{x} = 1. \tag{5.2.13}$$

Pré-multiplicando (5.2.12) por $\overline{x}^T$, por (5.2.13) segue que:

$$\overline{x}^T\nabla f(\overline{x}) = \mu\overline{x}^T\overline{x} = \mu.$$

Então, de (5.2.9) temos

$$\mu = \alpha \sum_{G \in \mathcal{G}} \beta_G \overline{x}^T (G + G^T)\overline{x} = 2\alpha \sum_{G \in \mathcal{G}} \beta_G \overline{x}^T G \overline{x}. \tag{5.2.14}$$

Agora, pré-multiplicando (5.2.12) por $e^T = (1, 1, \ldots, 1) \in \mathbb{R}^n$, como G é um grupo de permutações $Ge = e$ e $G^T e = e$ para todo G e por (5.2.9) temos:

$$\begin{aligned} \mu e^T \overline{x} &= e^T \nabla f(\overline{x}) = \alpha \sum_{G \in \mathcal{G}} \beta_G e^T (G + G^T)\overline{x} \\ &= 2\alpha \sum_{G \in \mathcal{G}} \beta_G e^T \overline{x}. \end{aligned}$$

Ou seja,

$$(\mu - 2\alpha \sum_{G \in \mathcal{G}} \beta_G) e^T \overline{x} = 0. \tag{5.2.15}$$

Substituindo (5.2.14) em (5.2.15), segue que

$$[2\alpha \sum_{G \in \mathcal{G}} \beta_G (\overline{x}^T G \overline{x} - 1)] e^T \overline{x} = 0$$

Como $-1 \leq \overline{x}^T G \overline{x} \leq 1$ e $p >> 1$, de (5.2.10) e (5.2.11) segue que $\alpha > 0$ e para todo $G \in \mathcal{G}$, $\beta_G > 0$. Logo, como não podemos ter $\overline{x}^T G \overline{x} = 1$ para todo $G \in \mathcal{G}$, segue que $e^T \overline{x} = 0$, ou seja, a propriedade (5.2.5) é satisfeita para os pontos estacionários de (5.2.7). □

Uma outra maneira de contornar a não diferenciabilidade de $\max_{G \in \mathcal{G}} x^T G x$ em (5.2.6) é introduzir um parâmetro θ, que idealmente satisfaria $\theta = \max_{G \in \mathcal{G}} x^{*T} G x^*$, onde x^* é o vetor inicial ótimo. Consideramos, assim, o seguinte problema:

$$\min_{\|x\|=1} \frac{1}{2} \sum_{G \in \mathcal{G}} [(x^T G x - \theta)_+]^2, \tag{5.2.16}$$

onde $(x^T G x - \theta)_+ = \max\{0, x^T G x - \theta\}$.

TEOREMA 5.2.2: Se $\overline{x} \in I\!R^n$ é um ponto estacionário para (5.2.16) e θ é tal que $\theta < \overline{x}^T G \overline{x} < 1$ para algum $G \in \mathcal{G}$, então $\overline{x}$ satisfaz a propriedade (5.2.5).

Demonstração. O problema (5.2.16) é equivalente a

$$\begin{array}{ll} \min & f(x,\theta) = \dfrac{1}{2}\displaystyle\sum_{G\in\mathcal{G}}[(x^T G x - \theta)_+]^2 \\ \text{s/a} & x^T x = 1\,. \end{array} \tag{5.2.17}$$

Agora,

$$\nabla_x f(x,\theta) = \sum_{G\in\mathcal{G}}(x^T G x - \theta)_+ (G + G^T)x = \sum_{G\in\mathcal{G}} \gamma_G (G + G^T)x \tag{5.2.18}$$

onde

$$\gamma_G = (x^T G x - \theta)_+. \tag{5.2.19}$$

Como $\overline{x}$ é um ponto estacionário para (5.2.16), então também é estacionário para (5.2.17) e satisfaz

$$\nabla_x f(\overline{x},\theta) = \mu\overline{x}, \mu \in I\!R \tag{5.2.20}$$

e

$$\overline{x}^T\overline{x} = 1. \tag{5.2.21}$$

Pré-multiplicando (5.2.20) por $\overline{x}^T$, por (5.2.21) segue que

$$\overline{x}^T \nabla_x f(\overline{x},\theta) = \mu \overline{x}^T\overline{x} = \mu.$$

Então, de (5.2.18) temos

$$\mu = \sum_{G \in \mathcal{G}} \gamma_G \overline{x}^T (G + G^T) x = 2 \sum_{G \in \mathcal{G}} \gamma_G \overline{x}^T G \overline{x}. \tag{5.2.22}$$

Pré-multiplicando (5.2.20) por $e^T = (1, 1, \ldots, 1) \in I\!R^n$, como $\mathcal{G}$ é um grupo de permutações $Ge = G^T e = e$ para todo G, e por (5.2.18) temos:

$$\mu e^T \overline{x} = e^T \nabla f_x(\overline{x}, \theta) = \sum_{G \in \mathcal{G}} \gamma_G e^T (G + G^T) \overline{x} = 2 \sum_{G \in \mathcal{G}} \gamma_G e^T \overline{x}.$$

Ou seja,

$$(\mu - 2 \sum_{G \in \mathcal{G}} \gamma_G) e^T \overline{x} = 0. \tag{5.2.23}$$

Substituindo (5.2.22) em (5.2.23), temos:

$$2 \left[\sum_{G \in \mathcal{G}} \gamma_G (\overline{x}^T G \overline{x} - 1) \right] e^T \overline{x} = 0.$$

De (5.2.19), $\gamma_G \geq 0$. Além disso, $-1 \leq \overline{x}^T G \overline{x} \leq 1$, para todo $G \in \mathcal{G}$. Como por hipótese existe $G \in \mathcal{G}$ tal que $\theta < \overline{x}^T G \overline{x} < 1$, segue que $e^T \overline{x} = 0$. Em outras palavras, $\overline{x}$ satisfaz (5.2.5). □

Tanto no problema (5.2.7) quanto em (5.2.16), estamos substituindo um problema minimax por uma aproximação diferenciável. Ou seja, propomos duas maneiras de *"suavizar"* (5.2.6), recaindo em problemas de minimização em esferas.

Embora muitos problemas minimax provenham de se maximizar a mínima distância, como é o caso do PVI e do problema das 13 bolas (problema de empacotamento, seção 1.7 deste trabalho), a abordagem minimax também aparece em problemas de classificação (determinação de hiperplanos separadores), aproximação uniforme por polinômios e teoria dos jogos. Uma discussão mais detalhada destas aplicações é feita em Martínez, Santos e Santos [1993a].

5.3 EXPERIMENTOS NUMÉRICOS

Os grupos de códigos utilizados baseiam-se nos grupos simétricos (S_n) (ver, por exemplo, Hall [1959], Wielandt [1964]), que consistem em todas as permutações possíveis para n símbolos. Conforme descrito na Seção 5.2, obtem-se $\mathcal{G}$ excluindo-se de S_n a identidade e, para permutações com inversa distinta, escolhendo-se sempre a lexicograficamente menor. Assim, por exemplo,

$$S_3 = \left\{ \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} \right\}$$

e

$$\mathcal{G} = \left\{ \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \right\},$$

ou, em notação matricial,

$$S_3 = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \right\}$$

e

$$\mathcal{G} = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \right\}.$$

Na Tabela 5.3.1 a seguir apresentamos o número efetivo de códigos nos grupos testados. A quantidade de símbolos de cada código é dada por n, $|S_n|$ é a ordem do grupo simétrico S_n e $|\mathcal{G}|$ é o número de elementos no grupo de códigos $\mathcal{G}$.

n	$\lvert S_n \rvert$	$\lvert \mathcal{G} \rvert$
3	6	4
4	24	16
5	120	72
6	720	397
7	5040	2635
8	40320	20541
9	362880	182749
10	3628800	1819147

Tabela 5.3.1: Número de códigos presentes nos grupos de permutação testados.

Conforme apresentado em Blake [1972], quando o PVI é formulado em termos dos grupos simétricos, a solução ótima pode ser expressa como Gx^*, para algum $G \in S_n$ e o vetor $x^* \in I\!R^n$ dado por:

$$x^* = \begin{cases} \left(\frac{n-1}{2}\alpha, \frac{n-3}{2}\alpha, \ldots, \alpha, 0, -\alpha, \ldots, -\frac{n-3}{2}\alpha, -\frac{n-1}{2}\alpha\right)^T, & n \text{ ímpar} \\ \left(\frac{n-1}{2}\alpha, \frac{n-3}{2}\alpha, \ldots, \frac{\alpha}{2}, -\frac{\alpha}{2}, \ldots, -\frac{n-3}{2}\alpha, -\frac{n-1}{2}\alpha\right)^T, & n \text{ par} \end{cases} \tag{5.3.1}$$

onde

$$\alpha = \left[\frac{12}{(n-1)(n)(n+1)}\right]^{1/2}. \tag{5.3.2}$$

A título de comparação e validação dos resultados obtidos, apresentamos a seguir a Tabela 5.3.2 com os valores numéricos de θ^* e α. O valor α também pode ser interpretado como $\alpha = \min_{i \neq j} |x_i^* - x_j^*|$, x^* vetor inicial ótimo.

n	θ^*	α
3	0.50000000	0.70710678
4	0.80000000	0.44721360
5	0.90000000	0.31622777
6	0.94285714	0.23904572
7	0.96428571	0.18898224
8	0.97619048	0.15430335
9	0.98333333	0.12909944
10	0.98787879	0.11009638

Tabela 5.3.2: Valores numéricos exatos de θ^* e α

Para gerarmos as permutações de S_n que constituem $\mathcal{G}$, implementamos o algoritmo de Johnson e Trotter (Stanton & White [1986]), baseado em transposições adjacentes. Incorporamos este procedimento na rotina de avaliação da função objetivo, vetor gradiente e matriz Hessiana, de forma que os códigos do grupo $\mathcal{G}$ são gerados à medida que os cálculos vão sendo feitos, ou seja, sem qualquer armazenamento adicional.

Tanto (5.2.7) quanto (5.2.16) são problemas de minimização em esferas que foram resolvidos por uma implementação para o Algoritmo 2.3.1 (RCMRI) em que a resolução dos subproblemas é feita pelo Algoritmo 2.4.1 (MINESF) ($\alpha = 10^{-4}$ e $\tau_1 = \tau_2 = 0.5$). Esta implementação é uma variante simplificada do programa desenvolvido para os testes com minimização em bolas (Seções 1.6 e 1.7) e que utiliza rotinas do EISPACK (TRED2 e TQL2) para fazer a decomposição espectral das matrizes Hessianas dos modelos quadráticos.

Para (5.2.7), em que $f \in C^2(\mathbb{R}^n)$, utilizamos efetivamente a matriz hessiana $\nabla^2 f$ no modelo quadrático. Já para (5.2.16), como não existe $\nabla^2_{xx} f(\cdot, \theta)$ trabalhamos com a seguinte aproximação para a matriz Hessiana:

$$B(x,\theta) = \sum_{G \in \mathcal{H}} \left\{ (x^T G x - \theta)(G + G^T) + [(G + G^T)x][(G + G^T)x]^T \right\} \tag{5.3.3}$$

onde

$$\mathcal{H} = \left\{ G \in \mathcal{G} \mid x^T G x > \theta \right\} \tag{5.3.4}$$

ou $B(x,\theta) = 0$ se $\mathcal{H} = \emptyset$. Esta aproximação resulta de considerarmos a matriz obtida derivando-se o vetor $\nabla f(x,\theta)$ em relação a x, prevendo-se, porém, possíveis descontinuidades.

Nos testes com os problemas (5.2.7) e (5.2.16) tomamos vetores iniciais x_0 com componentes geradas aleatoriamente entre -1 e 1 e então normalizados de forma que $\|x_0\| = 1$. Para cada n ($3 \leq n \leq 10$) testamos 5 sementes diferentes, o que totalizou 80 experimentos.

Tendo em vista que a função objetivo de (5.2.7) é tal que $|\mathcal{G}|^{1/p} \leq f(x) \leq 3|\mathcal{G}|^{1/p}$, escolhemos p de forma que $|\mathcal{G}|^{1/p} < 1.6$, conforme especificado na Tabela 5.3.3 a seguir. Neste critério para a escolha de p procuramos estabelecer um compromisso entre limitar a função o mais estreitamente possível e ao mesmo tempo baratear os cálculos sem prejudicar a viabilidade de aproximar (5.2.6) por (5.2.7), conforme detalhamos na Seção 5.2 .

n	3	4	5	6	7	8	9	10
p	5	10	10	20	30	30	60	60

Tabela 5.3.3: Valores utilizados para p nos testes com o problema (5.2.7)

Os resultados dos testes com o problema (5.2.7) estão resumidos na Tabela 5.3.4, onde n é a quantidade de símbolos de cada código, SEM é a semente utilizada, $\theta = \max_{G \in \mathcal{G}} \overline{x}^T G \overline{x}$ e $d = \min_{i \neq j} |\overline{x}_i - \overline{x}_j|$, onde $\overline{x}$ é a solução obtida e ITER e AFUN são, respectivamente, os números de iterações e avaliações de função efetuados.

n	SEM	θ	d	$ITER$	$AFUN$
3	17	0.50000000	0.70710678	4	5
	29	0.50000000	0.70710678	6	8
	53	0.50000000	0.70710678	8	10
	89	0.50000000	0.70710678	8	10
	97	0.50000000	0.70710678	4	5
4	17	0.93035887	0.26389606	7	10
	29	0.93035848	0.26389680	7	10
	53	0.93035840	0.26389696	8	11
	89	0.93035433	0.26390467	9	14
	97	0.93035887	0.26389607	7	12
5	17	0.95975634	0.20060822	6	9
	29	0.95975633	0.20060824	9	14
	53	0.95975633	0.20060825	9	13
	89	0.95975633	0.20060824	9	14
	97	0.95975633	0.20060825	10	16
6	17	0.97280471	0.16491068	7	9
	29	0.97280495	0.16491015	10	16
	53	0.97280584	0.16490765	11	22
	89	0.97280584	0.16490766	8	13
	97	0.97278968	0.16495551	11	23
7	17	0.97851258	0.14658589	7	12
	29	0.97851259	0.14658586	11	23
	53	0.97851258	0.14658586	10	20
	89	0.97851258	0.14658589	10	13
	97	0.97851258	0.14658589	7	11
8	17	0.98593105	0.11861258	10	27
	29	0.98589945	0.11874576	8	18
	53	0.98594650	0.11854745	10	26
	89	0.98590077	0.11874016	14	31
	97	0.98589913	0.11874707	11	28
9	17	0.98905149	0.10463513	7	18
	29	0.98905145	0.10463531	10	25
	53	0.98905146	0.10463528	10	27
	89	0.98905145	0.10463531	11	24
	97	0.98905145	0.10463531	9	23
10	17	0.99277663	0.084990415	13	32
	29	0.99280973	0.084795434	11	30
	53	0.99277227	0.085016084	9	24
	89	0.99277656	0.084990810	15	35
	97	0.99280193	0.084841460	13	34

Tabela 5.3.4: Resultados dos testes com o problema (5.2.7)

O valor ótimo para d é dado por α em (5.3.1) - (5.3.2). Assim, comparando-se os valores α e d, nas Tabelas 5.3.2 e 5.3.4, respectivamente, vemos que o objetivo de determinar o vetor inicial ótimo foi atingido apenas para $n = 3$. Nos demais casos, embora o critério de parada garanta a obtenção de um ponto estacionário para o problema (5.2.7), a solução $\overline{x}$ encontrada não é a solução ótima do PVI quando $\mathcal{G}$ é formado a partir dos grupos simétricos. De qualquer forma, pelo fato de termos obtido sempre uma solução do tipo $G\overline{x}$ para algum G em $\mathcal{G}$, podemos conjecturar que encontramos o ótimo global do problema (5.2.7).

Para os testes com o problema (5.2.16), definimos:

$$\varphi(\theta) = \min_{\|x\|=1} f(x, \theta), \tag{5.3.5}$$

onde

$$f(x, \theta) = \frac{1}{2} \sum_{G \in \mathcal{G}} [(x^T G x - \theta)_+]^2. \tag{5.3.6}$$

Sob uma hipótese de diferenciabilidade para x, a função $\varphi : \mathbb{R} \to \mathbb{R}$ é diferenciável, conforme provaremos abaixo. É possível obter a diferenciabilidade de φ sob hipóteses de regularidade menos restritivas para a solução $x(\theta)$, conforme apresentado em Martínez, Santos e Santos [1993b]. Agora, como $-1 \leq x^T G x \leq 1$ para todo x na esfera unitária, e todo $G \in \mathcal{G}$, podemos restringir a análise para $\theta \in [-1, 1]$. Além disso, φ é identicamente nula para $\theta \geq \theta^* = \max_{G \in \mathcal{G}} (x^*)^T G x^*$, onde x^* é o vetor inicial ótimo. Desta forma, devemos encontrar θ^*, o menor zero de φ, pois então $x^* = \arg \min_{\|x\|=1} f(x, \theta^*)$ é a solução procurada.

Para provarmos que φ é diferenciável, assumimos que existe $\theta_0 < \theta_*$ tal que $x'(\theta)$ existe e é contínua para todo $\theta > \theta_0$. Nestas condições, para todo $\theta > \theta_0$ temos

$$\varphi'(\theta) = \frac{\partial f}{\partial \theta}(x(\theta), \theta) = -\sum_{G \in \mathcal{G}} (x(\theta)^T G x(\theta) - \theta)_+. \tag{5.3.7}$$

De fato, pela regra da cadeia e pela hipótese de diferenciabilidade,

$$\varphi'(\theta) = \nabla_x f(x(\theta), \theta)^T x'(\theta) + \frac{\partial f}{\partial \theta}(x(\theta), \theta). \tag{5.3.8}$$

Agora, para $x(t) \in \{x \in I\!R^n \mid \|x\| = 1\}$, dado $\theta > \theta_0$ temos

$$f(x(\theta), \theta) \leq f(x(t), \theta)$$

para todo $t > \theta_0$.

Portanto,

$$\frac{d}{dt} f(x(t), \theta)\Big|_{t=\theta} = 0.$$

Mas, $\frac{d}{dt} f(x(t), \theta)\Big|_{t=\theta} = \nabla_x f(x(\theta), \theta)^T x'(\theta)$.

Logo (5.3.7) segue de (5.3.8) e a prova está completa.

Agora, como $\varphi'(\theta^*) = 0$, sabe-se que a convergência do método de Newton para a raiz θ^* é apenas linear (ver por exemplo, Dennis e Schnabel [1983]).

Podemos, no entanto, acelerar esta convergência usando relaxação, isto é, introduzindo um parâmetro $\omega \geq 1$ no esquema iterativo de Newton:

$$\bar{\theta} = \theta - \omega \frac{\varphi(\theta)}{\varphi'(\theta)}. \tag{5.3.9}$$

Quando ω é igual à multiplicidade da raiz procurada, a convergência do método de Newton relaxado é quadrática (ver Ortega e Rheinboldt [1970]). Como em nosso caso $\varphi(\theta^*) = \varphi'(\theta^*) = 0$, utilizamos $\omega = 2$.

Assim, os testes com o problema (5.2.16) foram feitos com base no seguinte algoritmo:

ALGORITMO 5.3.1

Dados $0 < \gamma \leq 1, \varepsilon_1 > 0, \varepsilon_2 > 0, x_0 \in I\!R^n$ tal que $\|x_0\| = 1$;

Passo 1. Determinar $\theta_0 = \max_{G \in \mathcal{G}} x_0^T G x_0$

Passo 2. Obter intervalo inicial $[\theta_\ell, \theta_u] \subset [-1, \theta_0]$, de forma que

$$\varphi(\theta_\ell) > \varepsilon_2, \varphi(\theta_u) \leq \varepsilon_2 \text{ e } |\theta_u - \theta_\ell| = \gamma|\theta_0 + 1|$$

Passo 3. Aplicar Newton relaxado:
Calcular $\varphi'(\theta_\ell)$
Enquanto $(\varphi'(\theta_\ell) \neq 0)$ e $(\varphi(\theta_\ell) > \varepsilon_2)$ e $((\theta_u - \theta_\ell) > \varepsilon_1)$ fazer:

$$\theta_\ell \leftarrow \theta_\ell - 2\frac{\varphi(\theta_\ell)}{\varphi'(\theta_\ell)}$$

calcular $\varphi(\theta_\ell)$ e $\varphi'(\theta_\ell)$

Passo 4. Refinar intervalo final, obtendo $[\theta_\ell, \theta_u]$ tal que

$$\theta_u - \theta_\ell \leq \varepsilon_1, \varphi(\theta_\ell) > \varepsilon_2 \text{ e } \varphi(\theta_u) \leq \varepsilon_2.$$

Passo 5. $\theta^* = \theta_u$

$x^* = \arg\min_{\|x\|=1} f(x, \theta^*)$

$d^* = \min_{\|i \neq j\|} |x_i^* - x_j^*|$

Os resultados dos testes com o problema (5.2.16), utilizando $\gamma = 0.1, \varepsilon_1 = 10^{-3}$ e $\varepsilon_2 = 10^{-8}$, contrariamente ao ocorrido com o problema (5.2.7), demandaram um grande número de iterações do algoritmo de minimização em esferas e consequentemente um grande número de avaliações da função $f(x, \theta)$. Os valores obtidos para θ e d, no entanto, estão bastante próximos dos verdadeiros valores, conforme pode ser visto comparando-se a Tabela 5.3.5 a seguir com a Tabela 5.3.2. Fixada a quantidade de símbolos n no código, dentre os cinco testes feitos utilizando-se sementes distintas para obter o ponto inicial, apresentamos na Tabela 5.3.5 aqueles com o menor número de avaliações de função.

n	θ	d	$ITER$	$AFUN$
3	0.50000000	0.70710763	30	34
4	0.79999888	0.44720948	34	41
5	0.90024733	0.31583626	31	40
6	0.94377069	0.23712966	40	53
7	0.96433066	0.18885106	80	98
8	0.97619586	0.15426772	69	98
9	0.98333355	0.12909196	93	139
10	0.98775704	0.10996450	96	128

Tabela 5.3.5: Resumo dos resultados dos testes com problema (5.2.16)

Observamos na Tabela 5.3.5, que para $n = 4$ e $n = 10$ os valores obtidos para θ são ligeiramente melhores que o valor ótimo apresentado na Tabela 5.3.2, provavelmente devido às tolerâncias utilizadas. Com relação à propriedade (5.2.5), satisfeita pelo vetor inicial ótimo e pelos pontos estacionários dos problemas (5.2.7) e (5.2.16) (conforme Teoremas 5.2.1 e 5.2.2), para os testes com o problema (5.2.7), obtivemos $|\sum_{i=1}^{n} \overline{x}_i| \leq 10^{-11}$, enquanto para os testes com o problema (5.2.16), $10^{-4} \leq |\sum_{i=1}^{n} x_i^*| \leq 10^{-1}$. Já que os valores para θ e α obtidos pela resolução de (5.2.16) são muito melhores que os obtidos por (5.2.7), vemos que a satisfação da propriedade de (5.2.5) não é uma boa medida para a qualidade da solução encontrada.

Tendo em vista o alto custo da avaliação da função φ, definida em (5.3.5) uma terceira estratégia para resolver o problema do vetor inicial em codificação é combinar as duas abordagens já testadas da seguinte forma: utilizar o par $(\overline{x}, \overline{\theta})$ obtido através do problema (5.2.7) como inicializador de um processo de refinamento com base no problema (5.2.16), seguindo filosofia análoga à do Algoritmo 5.3.1. Além do método de Newton relaxado, utilizamos outras duas estratégias para obter o menor zero θ^* da função φ e poder comparar os resultados: bisseção e interpolação inversa (polinômio na forma de Newton usando diferenças divididas) com um grau máximo para o polinômio interpolador (número de ponto ≤ 10, grau ≤ 9).

O refinamento consiste de três etapas básicas: obtenção de um intervalo inicial $[\theta_\ell, \theta_u]$ onde $\theta_u - \theta_\ell < \varepsilon_0$, $\varphi(\theta_u) \leq \varepsilon_2$ e $\varphi(\theta_\ell) > \varepsilon_2$; refinamento propriamente dito e obtenção do intervalo final $[\theta_\ell, \theta_u]$, com $\theta_u - \theta_\ell < \varepsilon_1, \varphi(\theta_u) \leq \varepsilon_2$ e $\varphi(\theta_\ell) > \varepsilon_2$. Assim, $\theta^* = \theta_u$, $x^* = \arg \min_{\|x\|=1} f(x, \theta^*)$ e $d^* = \min_{i \neq j} |x_i^* - x_j^*|$.

Utilizamos as seguintes tolerâncias: $\varepsilon_0 = 10^{-1}$ (intervalo inicial), $\varepsilon_1 = 10^{-3}$ (intervalo refinado) e $\varepsilon_2 = 10^{-8}$ (tolerância para φ). Fixada a quantidade de símbolos do código ($n \geq 4$), os vetores ótimos obtidos pelo refinamento (provenientes das soluções de (5.2.7) com diferentes sementes para a geração do ponto inicial) diferiram apenas por uma permutação nos elementos. Os valores θ^*, d^*, ITER e AFUN ficaram praticamente constantes, e por isso os resultados do refinamento estão resumidos na Tabela 5.3.6 a seguir, cujo formato se assemelha às Tabelas 5.3.4 e 5.3.5, com o acréscimo da coluna TIPO para caracterizar a técnica de refinamento empregada: B para bisseção, I para interpolação inversa usando no máximo 10 pontos e N para o método de Newton relaxado.

n	$TIPO$	θ^*	d^*	$ITER$	$AFUN$
4	B	0.80023783	0.44694749	20	32
	I	0.79995736	0.44720522	23	38
	N	0.80015054	0.44705772	7	13
5	B	0.90017579	0.31595521	23	33
	I	0.89995236	0.31621610	30	45
	N	0.90015941	0.31597189	8	11
6	B	0.94310592	0.23855202	35	49
	I	0.94281073	0.23902473	40	72
	N	0.94286206	0.23903530	15	20
7	B	0.96443296	0.18859292	41	56
	I	0.96423824	0.18895711	81	141
	N	0.96451355	0.18840988	17	24
8	B	0.97638072	0.15368973	60	87
	I	0.97615251	0.15427498	91	172
	N	0.97621081	0.15424174	21	26
9	B	0.98357409	0.12816579	56	80
	I	0.98329961	0.12907078	87	164
	N	0.98333372	0.12909655	23	32
10	B	0.98812953	0.10891551	54	69
	I	0.98778570	0.10999563	64	107
	N	0.98793846	0.10987476	26	37

Tabela 5.3.6: Resultados do refinamento.

Comparando os valores para ITER e AFUN da Tabela 5.3.5 com o total destes mesmos valores considerando-se as Tabelas 5.3.4 e 5.3.6, vemos que mesmo o melhor teste com o problema (5.2.16) não supera o desempenho da estratégia combinada de resolver (5.2.7) e então refinar a solução obtida através do problema (5.2.16) pelo método de Newton. Além disso, obtivemos $|\sum_{i=1}^{n} x_I^B| \leq 10^{-8}$, $|\sum_{i=1}^{n} x_i^I| \leq 10^{-3}$ e $|\sum_{i=1}^{n} x_i^N| \leq 10^{-11}$, onde x^B, x^I e x^N são os vetores iniciais ótimos obtidos pelo refinamento da solução encontrada para (5.2.7) usando bisseção, interpolação e Newton, respectivamente.

Para que tenhamos uma idéia do tempo computacional gasto nestes experimentos, apresentamos na Tabela 5.3.7 o tempo utilizado (Sun Sparc Station 2) para os testes com $n = 9$ e 10.

	problema (5.2.7)	problema (5.2.16)	refinamento		
			B	I	N
$n = 9$	$30'$	$2h30'$	$35'$	$1h5'$	$15'$
$n = 10$	$2h20'$	$12h$	$6h$	$11h$	$4h$

Tabela 5.3.7: Tempo gasto (em média) nos testes mais dispendiosos

Os gráficos a seguir permitem a visualização comparativa dos resultados apresentados nas Tabelas 5.3.4, 5.3.5 e 5.3.6.

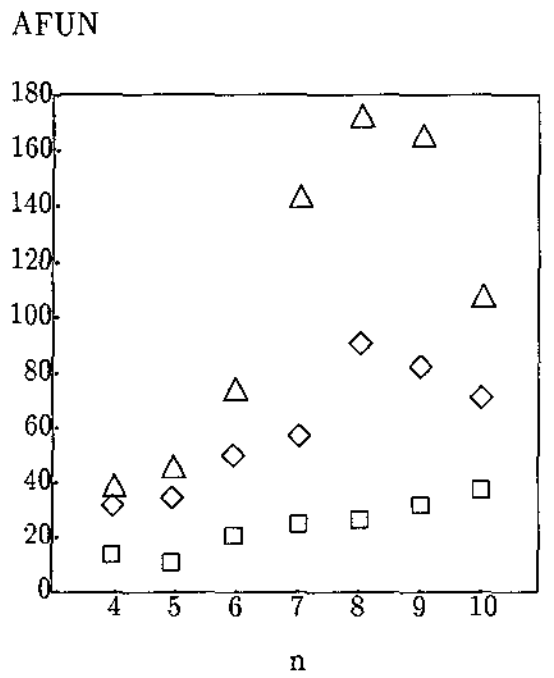

Gráfico 5.3.1: Número de avaliações de função (AFUN) empregado nas diferentes estratégias de refinamento (◇ = bisseção, □ = Newton, △ = interpolação) em função da dimensão (n) do problema.

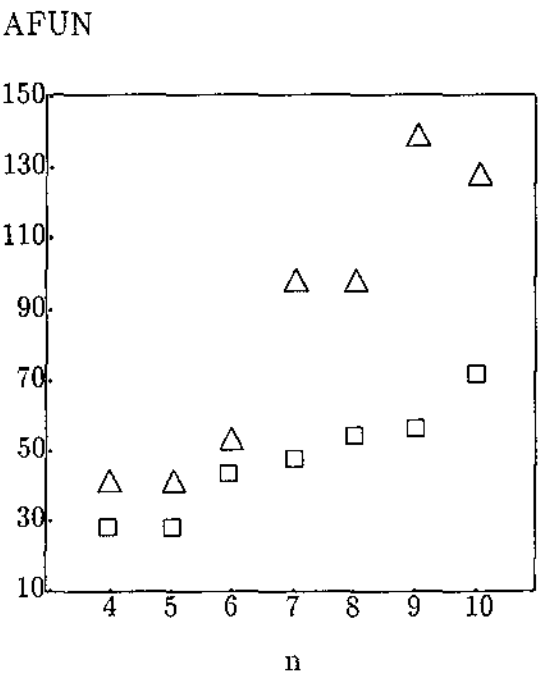

Gráfico 5.3.2: Número total de avaliações de função (AFUN) empregado na resolução do problema (5.2.16) (△) e no refinamento da solução obtida para (5.2.7) via método de Newton (□) em função da dimensão (n) do problema.

5.4 OBSERVAÇÕES FINAIS

Neste capítulo formulamos o problema do vetor inicial em codificação (PVI) do ponto de vista de minimização em esferas euclidianas, ou seja, como uma aplicação para o Algoritmo 2.3.1 (RCMRI) com subproblemas resolvidos pelo Algoritmo 2.4.1 (MINESF). Resolvemos problemas que originalmente apresentam até cerca de 1.800.000 restrições. Com o objetivo de validar simultaneamente os algoritmos e as abordagens propostas para

a resolução do PVI, optamos por trabalhar com o grupo de códigos proveniente dos grupos de permutação simétricos (S_n), com vetor inicial ótimo conhecido (Blake [1972]).

Neste sentido, obtivemos resultados bastante satisfatórios com o uso de uma estratégia híbrida, que consiste em inicialmente obter uma aproximação para o vetor ótimo através de uma das formulações e então refinar este vetor dentro da precisão desejada usando a outra formulação. As duas formulações propostas para o PVI reduzem-no a um problema com uma restrição esférica mas com função objetivo de avaliação cara. Desta forma, a estratégia híbrida mostrou-se eficiente tanto no que se refere à qualidade do vetor inicial ótimo determinado quanto em termos do esforço computacional despendido.

CONCLUSÕES

A implementabilidade da minimização em conjuntos arbitrários usando-se regiões de confiança está fortemente associada ao desenvolvimento de estratégias eficientes para se minimizar quadráticas em domínios não triviais. Neste trabalho consideramos com detalhes os casos em que os conjuntos factíveis são bolas ou esferas euclidianas bem como caixas provenientes de variáveis canalizadas. Baseados em pesquisas recentes (Stern e Wolcowicz [1993], Moré [1993]), acreditamos que, num futuro próximo, a resolução do subproblema de região de confiança será dominada num maior número de casos.

Nossa abordagem para minimização em caixas (algoritmo BOX) tem como alvo principal os problemas de grande porte, onde a fatoração de matrizes é indesejável. Neste sentido, o algoritmo QUACAN trabalha apenas com o produto de matriz por vetor e foi desenvolvido de modo a permitir a incorporação ou o descarte de muitas restrições a cada iteração. Com base no algoritmo BOX desenvolvemos uma estratégia para resolver problemas convexos com restrições lineares, que não necessita de parâmetros penalizadores. Tal estratégia pode ser vista como um ingrediente essencial para a obtenção de métodos de programação quadrática sequencial eficientes para grande porte, o que constitui uma fonte promissora de pesquisas futuras. Tal estratégia também se mostrou eficaz na resolução de problemas de complementariedade linear (Friedlander, Martínez e Santos [1993]) e inequações variacionais (Friedlander, Martínez e Santos [1994]).

A minimização em esferas foi utilizada como ferramenta na resolução do Problema do Vetor Inicial em codificação. Trabalhamos com problemas que originalmente apresentam até cerca de 1.800.000 restrições e que, quando reformulados do ponto de vista de minimização em esferas, recaem em problemas com função objetivo de avaliação bastante cara, e portanto indicados para o nosso algoritmo (MINESF), que procura poupar avaliações de função.

REFERÊNCIAS

BLAKE, I.F. Distance Properties of Group Codes for the Gaussian Channel. *SIAM Journal on Applied Mathematics*, **23**, (3), pp. 312–324, 1972.

BOGGS, P.T., Toole, J.W. & Wang, P. On the Lqcal Covergence of Quasi-Newton Methods for Constrained Optimization. *SIAM Journal on Control and Optimization*, **20**, pp. 161–171, 1982.

BURKE, J.V., Moré, J.J. & Toraldo, G. Convergence Properties of Trust Region Methods for Linear and Convex Constraints. *Mathematical Programming*, **47**, pp. 305–306, 1990.

CELIS, M.R., Dennis, J.E. & Tapia, R.A. A trust region strategy for nonlinear equality constrained optimization. In: Boggs, P.T., Byrd, R. & Schnabel, R. eds. *Numerical Optimization*, Philadelphia, SIAM, 1984, pp. 71–82.

CONN, A.R., Gould, N.I.M. and Toint, Ph. L. Global convergence of a class of trust region algorithms for optimization with simple bounds. *SIAM Journal on Numerical Analysis*, **25**, pp. 433–460, 1988(a). Ver também *SIAM Journal on Numerical Analysis*, **26**, pp. 764–767, 1989.

CONN, A.R., Gould, N.I.M & Toint, Ph.L. Testing a class of methods for solving minimization problems with simple bounds on the variables. *Mathematics of Computation*, **50**, pp. 399–430, 1988(b).

CONN, A.R., Gould, N.I.M. & Toint, Ph. L. A comprehensive description of LANCELOT. Technical Report, Hatfield Polytechnique Center, Hatfield, 1990.

CONN, A.R., Gould, N.I.M. & Toint, Ph.L. A globally convergent augmented Lagrangean algorithm for optimization with general constraints and simple bounds. *SIAM Journal on Numerical Analysis*, **28**, pp. 545–572, 1991.

CONTRERAS, M. & Tapia, R.A. Sizing the BFGS and DFP updates: a numerical study. Technical Report 91-19, Department of Mathematical Sciences, Rice University, Houston, Texas, 1991.

CONWAY, J.H. & Sloane, N.J.C. *Sphere Packings, Lattices and Groups*, Springer-Verlag, New York, 1988.

DEMBO, R.S. & Tulowitzki, U. Sequential truncated quadratic programming methods, In: Boggs, P.T., Byrd, R.H. & Schnabel, R.B. eds. *Numerical Optimization*, Philadelphia, SIAM, 1984, pp. 83–101.

DEMBO, R.S. & Tulowitzki, U. On the minimization of quadratic functions subject to box constraints. Working paper series B-17, School of Organization and Management, Yale University, 1987.

DENNIS, J.E., El-Alem, M.M. & Maciel, M.C. A global convergence theory for general trust-region based algorithms for equality constrained minimization. Technical Report 92–28, Department of Mathematical Sciences, Rice University, Houston, Texas, 1992.

DENNIS, J.E., Martínez, J.M., Tapia, R.A. & Williamson, K.A. An algorithm based on a convenient trust-region subproblem for nonlinear programming, Technical Report, Department of Mathematical Sciences, Rice University, Houston, Texas, 1990.

DENNIS, J.E. & Moré, J.J. A Characterization of Superlinear Convergence and Its Application to Quasi-Newton Methods, *Mathematics of Computation*, **28**, 126, pp. 546–560, 1974.

DENNIS, J.E. & Schnabel, R.B. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* Prentice Hall, Englewood Cliffs, New Jersey, 1983.

DJOKOVIC, D.Z. & Blake, I.F. An Optimization Problem for Unitary and Orthogonal Representations of Finite Groups. *Transactions of American Mathematical Society*, **164**, pp. 267–274, 1972.

DOWNEY, C. & Karlof, J. The analysis of optimal $[M, 3]$ group codes for the Gaussian channel. *Util. Math.*, **18**, pp. 51–70, 1980.

DUFF, I.S., Erisman, A.M. & Reid, J.K. *Direct Methods for Sparse Matrices*, Clarendon Press, Oxford, 1986.

EL-ALEM, M.M. A global convergence theory for the Celis-Dennis-Tapia trust region algorithm for constrained optimization. *SIAM Journal on Numerical Analysis*, **28**, pp. 266–290, 1991.

FLETCHER, R. *Practical methods of optimization.* John Wiley and Sons, Chichester, New York, Brisbane, Toronto and Singapore, 1987.

FRIEDLANDER, A. & Martínez, J.M. On the maximization of a concave quàdratic function with box constraints. *SIAM Journal on Optimization*, **4**, pp. 177–192, 1994.

FRIEDLANDER, A., Martínez, J.M. & Santos, S.A. A new trust region algorithm for bound contrained minimization. Relatório de Pesquisa RP 19/92, IMECC – UNICAMP, Campinas, 1992. A aparecer em *Journal of Applied Mathematics & Optimization.*

FRIEDLANDER, A., Martínez, J.M. & Santos, S.A. On the resolution of linearly constrained convex minimization problems. *SIAM Journal on Optimization*, **4**, pp. 331–339, 1994.

FRIEDLANDER, A., Martínez, J.M. & Santos, S.A. Resolution of linear complementarity problems using minimization with simple bounds. Relatório de Pesquisa RP 66/93, IMECC – UNICAMP, Campinas, 1993.

FRIEDLANDER, A., Martínez, J.M. & Santos, S.A. A new strategy for solving variational inequalities in bounded polytopes. Relatório de Pesquisa RP 02/94, IMECC – UNICAMP, Campinas, 1994.

GAY, D.M. Computing optimal locally constrained steps. *SIAM Journal on Scientific and Statistical Computing*, **2**, pp. 186–197, 1981.

GILL, P.E., Murray, W., Saunders, M.A. & Wright, M.H. Some theoretical properties of an augmented Lagrangean merit function. In: Pardalos, P.M. ed. *Advances in Optimization and Parallel Computing*, Amsterdam, Elsevier, pp. 127–143, 1992.

GILL, P.E., Murray, W. & Wright, M.H. *Practical Optimization*, Academic Press, London and New York, 1981.

GREEN, P. On the use of the EM algorithm for penalized likelihood estimation. *Journal of the Royal Statistical Society B*, **52**, pp. 443–452, 1990. +

GREEN, P. Bayesian reconstruction for emission tomography data using a modified EM algorithm. *IEEE Transactions in Medical Imaging MI - 9*, pp. 84–94, 1990.

HALL, M. *The Theory of Groups.* The MacMillan Co., New York, 1959.

HEBDEN, M.D. An algorithm for minimization using exact second derivatives. Technical Report TP 515, AERE Harwell, Harwell, Oxfordshire, 1973.

HEIKENSCHLOSS, M. Mesh independence for nonlinear least squares problems with norm constraints, Technical Report 90-18, Department of Mathematical Sciences, Rice University, Houston, Texas, 1990.

KARLOF, J. Permutation Codes for the Gaussian Channel. *IEEE Transactions on Information Theory*, **35**, (4), pp. 726-732, 1989.

LANGE, K. Convergence of EM image reconstruction algorithms with Gibbs smoothing. *IEEE Transactions in Medical Imaging MI - 9*, pp. 439-446, 1990.

LUENBERGER, D.G. *Linear and nonlinear programming.* 2 ed. Addison Wesley, Massachusetts, California, London, Amsterdam, Ontario, Sydney, 1984.

MARTÍNEZ, J.M. Fixed-Point Quasi-Newton Methods. *SIAM Journal on Numerical Analysis*, **5**, pp. 1413-1434, 1992.

MARTÍNEZ, J.M. Local minimizers of quadratic functions on Euclidean balls and spheres. *SIAM Journal on Optimization*, **4**, pp. 159-176, 1994.

MARTÍNEZ, J.M. & Santos, S.A. A trust region strategy for minimization on arbitrary domains. Relatório Técnico nº 63, IMECC - UNICAMP, Campinas, 1991. A aparecer em *Mathematical Programming.*

MARTÍNEZ, J.M., Santos, L.T. & Santos, S.A. Problemas minimax e aplicações. Relatório de Pesquisa RP 35/93, IMECC - UNICAMP, Campinas, 1993.

MARTÍNEZ, J.M., Santos, L.T. & Santos, S.A. A Minimax method with application to the Initial Vector Coding Problem. Relatório de Pesquisa RP 60/93, IMECC - UNICAMP, Campinas, 1993.

MORÉ, J.J. The Levenberg-Marquardt algorithm: implementation and theory. In: Watson, G.A. ed. *Numerical Analysis, Dundee 1977, Lecture Notes in Mathematics*, **630**, Berlin, Springer-Verlag, 1978, pp. 105-116.

MORÉ, J.J. Recent developments in algorithms and software for trust region methods. In: Bachem, A., Grötschel, M. & Korte, B. eds., *Mathematical Programming: State of the Art*, Berlin, Springer-Verlag, 1983, pp. 258-287.

MORÉ, J.J. Generalizations of the Trust Region Problem. Preprint MCS-P349-0193, Argonne National Laboratory, Argone, Illinois, 1993.

MORÉ, J.J., Garbow, B.S. & Hillstrom, K.E. Testing unconstrained optimization software. *ACM Transactions on Mathematical Software*, **7**, pp. 17-41, 1981.

MORÉ, J.J. & Sorensen, D.C. Computing a trust region step, *SIAM Journal on Scientific and Statistical Computing*, **4**, pp. 553–572, 1983.

MORÉ, J.J. & Toraldo, G. Algorithms for bound constrained quadratic programming problems. *Numerische Mathematik*, **55**, pp. 377–400, 1989.

MORÉ, J.J. & Toraldo, G. On the solution of large quadratic programming problems with bound constraints. *SIAM Journal on Optimization*, **1**, pp. 93–113, 1991.

NICKEL, R.H. & Tolle, J.W. A sparse sequential quadratic programming algorithm. *Journal on Optimization Theory and Applications*, **60**, pp. 453–473, 1989.

OREN, S.S. & Luenberger, D.G. Self-scaling variable metric (SSVM) algorithms. *Management Science*, **20**, pp. 845–862, 1974.

OREN, S.S. & Spedicato, E. Optimal conditioning of self scaling variable metric algorithms. *Mathematical Programming*, **10**, pp. 70–90, 1976.

ORTEGA, J.M. & Rheinboldt, W.C. *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, New York, San Francisco, London, 1970.

POWELL, M.J.D. & Yuan, Y. A trust region algorithm for equality constrained optimization. *Mathematical Programming*, **49**, pp. 189–211, 1990.

SHANNO, D.F. & Phua, K. Matrix conditioning and nonlinear optimization. *Mathematical Programming*, **14**, pp. 149–160, 1978.

SHANNON, C. A mathematical theory of communication, *Bell Syst. Tech. J.*, **27**, 3 and 4, July and October, 1948.

SLEPIAN, D. Group Codes for the Gaussian Channel. *Bell Syst. Tech. J.*, **17**, pp. 575–602, 1968.

SORENSEN, D.C. Newton' s method with a model trust region modification. *SIAM Journal on Numerical Analysis*, **19**, pp. 409–426, 1982.

STANTON, D. & White, D. *Constructive Combinatorics*, Undergraduate Texts in Mathematics, Springer-Verlag, New York, Berlin, Heidelberg, Tokyo, 1986.

STERN, R.J. & Wolkowicz, H. Indefinite trust region subproblems and non-symmetric eigenvalue perturbations. Technical Report SOR 93–1, Department of Civil Engineering and Operations Research, Princeton University, Princeton, New Jersey, 1993.

TIKHONOV, A.N. & Arsenin, V.Y. *Solutions of ill-posed problems*, John Wiley and Sons, New York, Toronto, London, 1977.

TOINT, Ph.L. Global convergence of a class of trust region methods for non-convex minimization in Hilbert space. *IMA Journal on Numerical Analysis*, **8**, pp. 231–252, 1988.

VARDI, A. A trust region algorithm for equality constrained minimization: convergence properties and implementation. *SIAM Journal on Numerical Analysis*, **22**, pp. 575–591, 1985.

VAVASIS, S. Approximation algorithms for indefinite quadratic programming. Technical Report 91–1228, Department of Computer Science, Cornell University, Ithaca, New York, 1991.

VOGEL, C.R. A constrained least-squares regularization method for nonlinear ill-posed problems. *SIAM Journal on Control and Optimization*, **28**, pp. 34–49, 1990.

WIELANDT, H. *Finite Permutation Groups.* Academic Press, New York, London, 1964.

WILLIAMSON, K.A. A robust trust region algorithm for nonlinear programming. Ph. D. Thesis, Department of Mathematical Sciences, Rice University, Houston, Texas, 1990.

YABE, H., Yamaki, N. & Takahashi, S. Global convergence of sequential inexact QP method for constrained optimization. *SUT Journal of Mathematics*, **27**, pp. 127–138, 1991.

ZHANG, Y. Computing a Celis-Dennis-Tapia trust-region step for equality constrained optimization. Technical Report 88–16, Department of Mathematical Sciences, Rice University, Houston, Texas, 1988.