

Sobre Abertura e Densidade de Polítopos Não-Degenerados
e uma Família de Algoritmos Primais de Pontos Interiores
para Programação Convexa

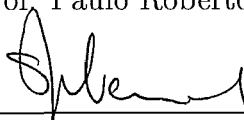
Alfredo Wagner Martins Pinto

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

Aprovada por:



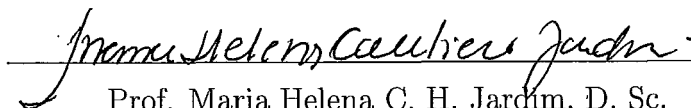
Prof. Paulo Roberto Oliveira, Dr. Ing.



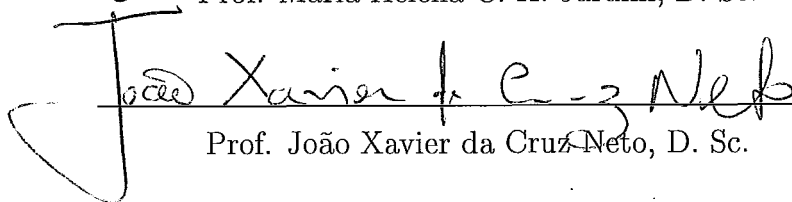
Prof. Suzana S. de Makler, D. Sc.



Prof. Regina Sandra Burachik, D. Sc.



Prof. Maria Helena C. H. Jardim, D. Sc.



Prof. João Xavier da Cruz Neto, D. Sc.

RIO DE JANEIRO, RJ - BRASIL
OUTUBRO DE 2000

PINTO, ALFREDO WAGNER MARTINS

Sobre Abertura e Densidade de Politopos
Não-Degenerados e uma Família de Algorit-
mos Primais de Pontos Interiores para Pro-
gramação Convexa [Rio de Janeiro] 2000

VII, 33 p. 29,7 cm (COPPE/UFRJ, D.Sc.,
Engenharia de Sistemas e Computação, 2000)

Tese – Universidade Federal do Rio de Ja-
neiro, COPPE

1 - Programação Matemática

I. COPPE/UFRJ II. Título (série)

*À minha mãe, Terezinha e
à minha família*

Agradecimentos

Em primeiro lugar quero agradecer a Paulo Roberto Oliveira, pela orientação, estímulo, amizade e principalmente pela confiança na minha capacidade.

Aos amigos Xavier e Orizon, que me incentivaram a iniciar este curso, e sempre me apoiaram. E mais uma vez ao Xavier pelas inúmeras sugestões para melhoria deste trabalho.

Aos amigos que neste longo período dividiram moradia comigo. Bôca, Piozão, Marlon, Tulião, Anderson, Marco, Silvano, Jurandir, Pierre, Paulinho, e em particular a Gilvan companheiro de longa data e sempre presente nas horas boas e nas incertas.

Aos funcionários da COPPE sistemas, pela presteza, eficiência e camaradagem.

As amigas Sheila Campos e Vera Lúcia Prudência, que tiveram uma colaboração inestimável na digitação deste trabalho.

Aos colegas do Departamento de Matemática e do Instituto de Ciências Exatas da Universidade Federal do Amazonas, por todo apoio dado.

A Moema, pelos bons momentos que compartilhamos, que tornaram meus dias no Rio mais agradáveis. E pelo meu aumento de produtividade.

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

**SOBRE ABERTURA E DENSIDADE DE POLITOPOS
NÃO-DEGENERADOS E UMA FAMÍLIA DE ALGORITMOS
PRIMAIS DE PONTOS INTERIORES PARA PROGRAMAÇÃO
CONVEXA.**

Alfredo Wagner Martins Pinto
Outubro/ 2000

Orientador: Paulo Roberto Oliveira.

Programa: Engenharia de Sistemas e Computação.

É feito o estudo da topologia do conjunto dos politopos. Mostra-se que o conjunto dos politopos não-degenerados é denso, sendo aberto e denso no subconjunto dos politopos limitados. É estudado também uma família de algoritmos de pontos interiores para programação convexa, onde é mostrado a convergência fraca no caso C^1 .

Abstract of Thesis presented to COOPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science(D.Sc.).

**ABOUT OPENING AND DENSITY OF NONDEGENERATE
POLYTOPES AND A FAMILY OF ALGORITHMS PRIMAL OF
INTERIOR POINTS FOR CONVEX PROGRAMMING.**

Alfredo Wagner Martins Pinto
October/2000

Advisor: Paulo Roberto Oliveira.

Department: Systems and Computation Engineering.

It is made the study of the topology of the set of the polytopes. It is shown that the set the nondegenerate polytopes is dense, being open and dense in the subset of bounded polytopes. It is also studied a family of interior point algorithms for convex programming, where the weak convergence is shown in the C^1 case.

Conteúdo

1	Introdução	1
2	Topologia no Conjunto de Problemas	3
2.1	Introdução	3
2.2	Problemas Não Degenerados e Problemas Degenerados	4
2.3	Topologia de Hausdorff no Conjunto dos Problemas	13
3	Algoritmo Afim Potencial para Programação Convexa	19
3.1	Introdução	19
3.2	O Algoritmo para (3.1.1)	20
3.3	Convergência do Algoritmo 3.1	22
3.4	O Algoritmo para (3.1.2)	30
	Bibliografia	31

Capítulo 1

Introdução

Muitos algoritmos, já foram estudados em programação matemática, para solucionar problemas com restrições lineares. Como por exemplo, o algoritmo afim escala de Dikin para o Problema de Programação Linear (PPL), o qual ele mesmo prova convergência em [8] com a hipótese de não degenerescência. Diversas provas de convergência foram dadas para este algoritmo no caso degenerado, podemos citar a dada por Monteiro, Tsuchiya e Wang [24]. Saigal [27] prova convergência para uma generalização do algoritmo de Dikin para o PPL, nos casos não-degenerado e degenerado. Algoritmos semelhantes foram estudados para programação convexa e quadrática. Podemos citar o algoritmo estudado por Sun [28], onde é provado a convergência sem hipótese de não-degenerescência.

O ponto crucial é que a prova de convergência dos algoritmos é muito mais simples no caso não-degenerado. Isto nos leva a pensar em como se distribuem os problemas não-degenerados e degenerados no “Universo dos Problemas”, onde se entende por “problema” ao par (A, b) tal que $A \in \mathbb{R}^{m \times n}$ é de posto m , $b \in \mathbb{R}^m$ é não nulo e o conjunto viável $\mathcal{P}(A, b)$ é não vazio, isto é gostaríamos de saber qual a relação topológica do conjunto dos problemas não-degenerados no espaço dos problemas. Daí vem a motivação para o nosso trabalho o qual está dividido em duas partes que podem ser lidas independentemente.

No capítulo 2 estudamos a “Topologia do Conjunto dos Problemas”, encarando o espaço dos problemas de duas maneiras. Na primeira consideramos o conjunto dos problemas com a topologia de $\mathbb{R}^{m \times n} \times \mathbb{R}^m$, onde provamos que o conjunto dos problemas não-degenerados é denso no conjunto dos problemas. E é aberto e denso no conjunto dos problemas limitados. Na segunda maneira introduzimos a “Topologia de Hausdorff” no espaço de problemas e mostramos que nessa topologia o conjunto dos problemas não-degenerados é aberto sem a hipótese de limitação.

No capítulo 3 propomos o algoritmo afim escala potencial estudado por Saigal [27] para programação linear, no caso de programação convexa. Utilizamos técnicas de geometria Riemanniana, estudadas por Xavier [6], para determinar a direção de busca linear, que é o simétrico do gradiente da função objetivo considerada intrinsecamente como função definida no interior do conjunto viável. Observamos que a direção de busca proposta por Eggermont [9] e por Gonzaga [12] são casos particulares da direção que propomos, para determinados valores de um parâmetro r (respectivamente $r = 1$ e $r = 2$). Obtivemos resultados de convergência semelhantes aos obtidos por Iussem [16] para o algoritmo proposto por Eggermont, e aos obtidos por Gonzaga [12], para o caso afim escala. A estrutura do algoritmo que estudamos é essencialmente a mesma do algoritmo analisado por Iussem [16], porém algumas das técnicas de demonstração de convergência fraca são semelhantes as utilizadas por Gonzaga.

Capítulo 2

Topologia no Conjunto de Problemas

2.1 Introdução

A principal motivação para o estudo de “problemas não-degenerados” no contexto de programação matemática, está no fato de que em geral a demonstração de convergência de diversos algoritmos é bem mais simples do que para problemas degenerados, exemplos clássicos são os algoritmos afim escala para o problema de programação linear, ou mesmo no caso do algoritmo afim escala generalizando para o **PPL** também.

Mostraremos que o conjunto dos problemas não-degenerados, é denso no conjunto dos problemas \mathcal{T} . Mostraremos também que o conjunto dos problemas não degenerados e limitados é um aberto em \mathcal{T} , e em particular é aberto e denso em \mathcal{L} , subconjunto de \mathcal{T} formado pelos problemas limitados. O que concluímos é que os problemas não-degenerados estão por “toda a parte” de \mathcal{T} , e que eles são a “maioria” dos problemas de \mathcal{L} .

Vamos também introduzir no conjunto dos problemas a Topologia de Hausdoff, e mostrar que nesta topologia o conjunto das problemas não-degenerados é aberto, sem a hipótese de limitação.

2.2 Problemas Não Degenerados e Problemas Degenerados

Consideremos o problema de programação não linear

$$\begin{aligned} & \text{minimizar} && f(x) \\ & \text{sujeito a} && Ax = b \\ & && x \geq 0 \end{aligned} \tag{2.2.1}$$

onde A é uma matriz $m \times n$, e $b \in \mathbb{R}^m$

A estrutura do conjunto viável de (2.2.1)

$$P(A, b) = \{ x \in \mathbb{R}^n ; Ax = b \text{ e } x \geq 0 \}$$

é essencial na análise dos algoritmos desenvolvidos para (2.2.1). A hipótese fundamental que se faz sobre A , é que tenha posto m , pois caso contrário existiriam restrições redundantes. É um fato conhecido, que o conjunto de matrizes de posto m é aberto no conjunto das matrizes $m \times n$. Neste caso o universo que consideraremos é o conjunto aberto de $\mathbb{R}^{m \times n} \times \mathbb{R}^m$

$$\mathcal{U} = \{ (A, b); A \text{ tem posto } m \text{ e } b \in \mathbb{R}^m \}$$

Definição 2.2.1 Chamaremos de problema ao par $(A, b) \in \mathcal{U}$, tal que $P(A, b) \neq \emptyset$. O conjunto $\mathcal{T} = \{ (A, b) \in \mathcal{U}; P(A, b) \neq \emptyset \}$ será chamado conjunto dos problemas. O subconjunto de \mathcal{T} formado pelos problemas (A, b) tais que existe $x > 0$ em $P(A, b)$ será denotado por \mathcal{A} . Se $P(A, b)$ é limitado, diremos que o problema (A, b) é limitado. Denotaremos por \mathcal{L} o subconjunto de \mathcal{T} formado pelos problemas limitados.

Na teoria de pontos interiores, para problemas com restrições lineares os algoritmos estudados são para problemas que estão em \mathcal{A} . Vamos mostrar agora que os problemas que estão em \mathcal{A} são a “maioria”.

Teorema 2.2.1 \mathcal{A} é aberto e denso em \mathcal{T} .

Demonstração:

Sejam $(\tilde{A}, \tilde{b}) \in \mathcal{A}$ e $y > 0$ em $P(\tilde{A}, \tilde{b})$. Vamos considerar J um subconjunto de índices de $1, \dots, n$ tal que, a submatriz $m \times m$ \tilde{A}_J de \tilde{A} formada pelas colunas

com índices em J seja invertível. Tal matriz existe pois o posto de \tilde{A} é m . Sem perda de generalidade podemos considerar $J = \{1, \dots, m\}$. Tome $\bar{J} = \{1, \dots, n\} \setminus J$ e consideraremos da mesma forma $A_{\bar{J}}$.

Então podemos escrever $Ax = b$ como

$$A_J x_J + A_{\bar{J}} x_{\bar{J}} = b$$

onde

$$x = \begin{pmatrix} x_J \\ x_{\bar{J}} \end{pmatrix}$$

Como \tilde{A}_J é invertível existe $r > 0$ tal que se $\|A - \tilde{A}\| \leq r$ temos $\|A_J^{-1}\| \leq \frac{1}{r}$ (ver [19] página 254). Neste caso a função

$$f : \overline{B(\tilde{A}, r)} \times \overline{B(\tilde{b}, 1)} \longrightarrow \mathbb{R}^m$$

$$(A, b) \longmapsto A_J^{-1}(b - A_{\bar{J}} y_{\bar{J}})$$

onde $y_{\bar{J}}$ é dado no início da demonstração é fixo, é contínua.

Observe que $f(\tilde{A}, \tilde{b}) = (\tilde{A}_J)^{-1}(\tilde{b} - \tilde{A}_{\bar{J}} y_{\bar{J}}) = y_J > 0$, pois

$$\tilde{A}_J y_J + \tilde{A}_{\bar{J}} y_{\bar{J}} = \tilde{b} \Rightarrow y_J = (\tilde{A}_J)^{-1}(\tilde{b} - \tilde{A}_{\bar{J}} y_{\bar{J}})$$

Então pela continuidade de f existe uma vizinhança V de (\tilde{A}, \tilde{b}) tal que $f(A, b) > 0 \forall (A, b) \in V$. Temos então que se $(A, b) \in V$,

$$x = \begin{pmatrix} A_J^{-1} \cdot (b - A_{\bar{J}} y_{\bar{J}}) \\ y_{\bar{J}} \end{pmatrix} > 0$$

e também temos

$$Ax = A_J \cdot A_J^{-1} \cdot (b - A_{\bar{J}} y_{\bar{J}}) + A_{\bar{J}} y_{\bar{J}} = b$$

Deste modo \mathcal{A} é aberto.

Densidade

Sejam $(\tilde{A}, \tilde{b}) \in \mathcal{T}$, e $y \geq 0$ tal que $\tilde{A}y = \tilde{b}$. Tomemos $d > 0$ tal que $\tilde{A}d \neq 0$. Então $(\tilde{A}, \tilde{b} + \frac{1}{k}\tilde{A}d) \in \mathcal{A} \forall k \in \mathbb{N}$, pois $x^* = y + \frac{1}{k}d > 0$ está em $P(\tilde{A}, \tilde{b} + \frac{1}{k}\tilde{A}d)$. E claramente $\lim_{k \rightarrow \infty} (\tilde{A}, \tilde{b} + \frac{1}{k}\tilde{A}d) = (\tilde{A}, \tilde{b})$.

De acordo com o que vimos acima, questões relativas a abertura e densidade, de conjuntos de \mathcal{T} , podem ser tratados em \mathcal{A} .

Para o que trataremos a seguir, é importante estabelecer algumas noções e definições:

Se $J = \{j_1, \dots, j_l\}$, com $j_1 < j_2 < \dots < j_l$ é um subconjunto de elementos, de $\{1, 2, \dots, n\}$, e $x \in \mathbb{R}^n$, denotaremos por x_J o vetor de \mathbb{R}^l , $x_J = (x_{j_1}, \dots, x_{j_l})$, onde x_{j_i} é a j_i -ésima coordenada do vetor x . Feito isto, se $\bar{J} = \{1, \dots, n\} \setminus J$, A_J é a submatriz de A formada pelas colunas correspondentes, aos elementos de J , e $A_{\bar{J}}$ é da mesma forma com relação aos elementos de \bar{J} . Podemos escrever Ax como $A_Jx_J + A_{\bar{J}}x_{\bar{J}}$.

Definição 2.2.2 *Um elemento $v \in P(A, b)$ é chamado vértice de $P(A, b)$, se o conjunto de vetores, formado pelas colunas de A correspondentes aos índices das componentes não nulas de v é linearmente independente.*

Da definição vemos que um vértice tem no máximo m coordenadas não nulas, visto que posto de $A = m$.

Definição 2.2.3 *Um vértice v é dito não-degenerado, se o número de coordenadas não nulas de v é exatamente m . Caso contrário o vértice é dito degenerado.*

Definição 2.2.4 *Um problema $(A, b) \in \mathcal{T}$, é dito não-degenerado, se todos os seus vértices são não-degenerados. Caso contrário o problema é dito degenerado.*

Vamos denotar por $\mathcal{ND} = \{(A, b) \in \mathcal{T}; (A, b) \text{ é não-degenerado}\}$, o conjunto dos problemas não-degenerados.

Em seguida vamos dar uma caracterização para \mathcal{ND} , que faz com que, a demonstração de convergência, de diversos algoritmos de pontos interiores sejam mais fáceis.

Em primeiro lugar vamos estabelecer algumas notações:

- $(A_{rr})_{m \times n}$ é o conjunto de todos os arranjos do conjunto $\{1, \dots, n\}$ tomados m a m .
- $C_{m,n}$ é o conjunto de todas as “combinações ordenadas” do conjunto $\{1, \dots, n\}$, tomados m a m . No sentido que se $J = \{j_1, \dots, j_m\}$ pertence a $C_{m,n}$, $j_1 < j_2 < \dots < j_m$.
- X é a matriz $\text{diag}(x)$, matriz tal que $x_{ij} = \delta_{ij}x_i$, onde x_i é a i -ésima coordenada do vetor x . Da mesma forma X^r , é a matriz tal que $x_{ij} = \delta_{ij}(x_i)^r$.
- $\mathcal{N}(A)$ é o espaço nulo da matriz A .

Lema 2.2.1 *Sejam A uma matriz $m \times n$ e $X = \text{diag}(x)$. Então vale a fórmula*

$$\det(AXA^T) = \sum_{J \in C_{m,n}} (\det A_J)^2 x_{j_1} \dots x_{j_m} \quad (2.2.2)$$

onde $J = \{j_1, \dots, j_m\}$. Em particular $\det(AA^T) = \sum_{J \in C_{m,n}} (\det A_J)^2$

Demonstração:

Vamos estabelecer as seguintes notações:

$A = (a_1 \dots a_n)$, a_i é a i -ésima coluna de A

$$A^T = \begin{pmatrix} a_1^T \\ \vdots \\ a_n^T \end{pmatrix}$$

$[c_1 \dots c_m] = \det(C)$, onde $C = (c_1 \dots c_m)$

Feito isto, teremos

$$\begin{aligned} AXA^T &= (a_1 \dots a_n) \begin{pmatrix} x_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & x_n \end{pmatrix} \begin{pmatrix} a_1^T \\ \vdots \\ a_n^T \end{pmatrix} = (a_1 \dots a_n) \begin{pmatrix} x_1 a_1^T \\ \vdots \\ x_n a_n^T \end{pmatrix} = \\ &= \sum_{i=1}^n x_i a_i a_i^T \end{aligned}$$

Se denotarmos $a_i = \begin{pmatrix} A_{1i} \\ \vdots \\ A_{mi} \end{pmatrix}$ teremos que $a_i a_i^T = (A_{1i} a_i \dots A_{mi} a_i)$

Então

$$\begin{aligned} AXA^T &= \left(\sum_{i=1}^n x_i A_{1i} a_i \quad \dots \quad \sum_{i=1}^n x_i A_{mi} a_i \right) \text{ e} \\ \det(AXA^T) &= \left[\sum_{i=1}^n x_i A_{1i} a_i \quad \dots \quad \sum_{i=1}^n x_i A_{mi} a_i \right] \end{aligned} \quad (2.2.3)$$

Aplicando a n linearidade da função determinante e o fato que $[a_{j_1}, \dots, a_{j_m}] = 0$ sempre que existirem dois índices $j_i = j_r$ podemos escrever (2.2.3) como

$$\det(AXA^T) = \sum_{J \in (A_{rr})_{n,m}} x_{j_1} \dots x_{j_m} \cdot A_{1j_1} \dots A_{mj_m} \cdot [a_{j_1} \dots a_{j_m}] \quad (2.2.4)$$

onde $J = \{j_1, \dots, j_m\}$. Se $J = \{j_1, \dots, j_m\}$ com $j_1 < \dots < j_m$ e σ é qualquer permutação de J teremos, $x_{\sigma_1} \dots x_{\sigma_m} = x_{j_1} \dots x_{j_m}$ e $[a_{\sigma_1}, \dots, a_{\sigma_m}] = (\text{signal } \sigma) \cdot [a_{j_1}, \dots, a_{j_m}]$ onde $\text{signal } \sigma = 1$ se σ for uma permutação par e -1 se σ for uma permutação ímpar de J . Podemos então escrever (2.2.4) como

$$\det(AXA^T) = \sum_{J \in C_{n,m}} x_{j_1} \dots x_{j_m} [a_{j_1} \dots a_{j_m}] \cdot \left(\sum_{\sigma} \text{signal } \sigma \right) A_{1\sigma_1} \dots A_{m\sigma_m} \quad (2.2.5)$$

De acordo com [14], página 193 $[a_{j_1} \dots a_{j_m}] = \sum_{\sigma} (\text{signal } \sigma) A_{1\sigma_1} \dots A_{m\sigma_m}$

Portanto

$$\det(AXA^T) = \sum_{J \in C_{n,m}} (\det A_J)^2 x_{j_1} \dots x_{j_m}$$

□

Vamos agora dar uma caracterização dos problemas não-degenerados. Para isto vamos definir as funções:

$$\begin{aligned} f_A &: \mathbb{R}^n \longrightarrow \mathbb{R} \\ x &\longmapsto \det(AXA^T) = \sum_{J \in C_{n,m}} (\det A_J)^2 x_{j_1} \dots x_{j_m} \\ g &: \mathcal{T} \longrightarrow \mathbb{R} \\ (A, b) &\longmapsto \inf\{f_A(x); x \in P(A, b)\} \end{aligned}$$

As propriedades da função f_A enunciadas a seguir, decorrem diretamente de sua definição:

- (1) $f_A(x) \geq 0 \quad \forall x \in \mathbb{R}_+^n$. Segue direto da fórmula (2.2.2).

- (2) $f_A(\lambda x) = \lambda^m f_A(x)$, $\forall \lambda \in \mathbb{R}$ e $x \in \mathbb{R}^n$. Segue do mesmo fato.
- (3) Se x tem no máximo m coordenadas não nulas, e $\tilde{J} = \{\tilde{j}_1, \dots, \tilde{j}_m\} \in C_{n,m}$ contém o conjunto de índices dessas coordenadas, então $f_A(x) = (\det A_{\tilde{J}})^2 x_{\tilde{j}_1} \dots x_{\tilde{j}_m}$. Com efeito, se $J \neq \tilde{J}$ está em $C_{n,m}$, x_J tem alguma coordenada nula, e o resultado segue direto da fórmula (2.2.2).
- (4) Se x tem no máximo $m - 1$ coordenadas não nulas então $f_A(x) = 0$. Com efeito, se $\tilde{J} \in C_{n,m}$ é um conjunto de índices, que contém o conjunto de índices de coordenadas não nulas de x , $x_{\tilde{J}}$ tem pelo menos uma coordenada nula. E o resultado segue de (3).
- (5) Se v é um vértice não degenerado de $P(A, b)$, $f_A(v) > 0$. Com efeito, se \tilde{J} é o conjunto de índices correspondentes às coordenadas não nulas de v , segue que $\det A_{\tilde{J}} \neq 0$, e por (3) $f_A(v) = (\det A_{\tilde{J}})^2 v_{\tilde{j}_1} \dots v_{\tilde{j}_m} > 0$.
- (6) Se v é um vértice degenerado de $P(A, b)$, $f_A(v) = 0$. É consequência direta de (4), pois v tem no máximo $m - 1$ coordenadas não nulas.
- (7) Um vértice v é degenerado, se e somente se, $f_A(v) = 0$.
- (8) Se $x \geq 0$ e $y \geq 0$ $f_A(x + y) \geq f_A(x) + f_A(y)$. Com efeito, $(x_{j_1} + y_{j_1}) \dots (x_{j_m} + y_{j_m}) = x_{j_1} \dots x_{j_m} + y_{j_1} \dots y_{j_m} + \delta_J$, onde $\delta_J \geq 0$.

Lema 2.2.2 *Seja $\{v_1, \dots, v_r\}$ o conjunto de vértices de $P(A, b)$. Então $x \in P(A, b) \iff x = \sum_{i=1}^r \lambda_i v_i + d$, onde $d \geq 0$ pertence a $\mathcal{N}(A)$, $\lambda_i \geq 0$ e $\sum_{i=1}^r \lambda_i = 1$.*

Demonstração:

ver [3]. □

Teorema 2.2.2 **Caracterização dos Problemas não Degenerados**

$(A, b) \in \mathcal{T}$ é não degenerado se, e somente se, $g(A, b) > 0$.

Demonstração:

Se $x \in P(A, b)$, pelo lema (2.2.2), $x = \sum_{i=1}^r \lambda_i v_i + d$, onde v_i é vértice de $P(A, b)$, $\lambda_i \geq 0$, $\sum_{i=1}^r \lambda_i = 1$ e $d \geq 0$ está em $\mathcal{N}(A)$. Pelas propriedades (2) e (8) temos que

$$f_A(x) = f_A\left(\sum_{i=1}^r \lambda_i v_i + d\right) \geq \sum_{i=1}^r \lambda_i^m f_A(v_i) + f_A(d) \geq \sum_{i=1}^r \lambda_i^m f(v_i),$$

ou seja

$$f_A(x) \geq \left(\sum_{i=1}^r \lambda_i^m\right) \min\{f_A(v_i) ; i = 1, \dots, r\}$$

O somatório $\sum_{i=1}^r \lambda_i^m$ atinge um mínimo $\mu > 0$ no conjunto compacto $\{(\lambda_1, \dots, \lambda_r); \lambda_i \geq 0 \text{ e } \sum_{i=1}^r \lambda_i = 1\}$. Portanto podemos escrever

$$f_A(x) \geq \mu \min\{f_A(v_i); i = 1, \dots, r\}, \forall x \in P(A, b)$$

Segue então da definição de ínfimo que

$$g(A, b) \geq \mu \min\{f_A(v_i); i = 1, \dots, r\}. \text{ Por outro lado,}$$

$$g(A, b) \leq f_A(v_i), \forall i = 1, \dots, r. \text{ Considerando então que}$$

$$\begin{aligned} 0 &\leq \mu \min\{f_A(v_i); i = 1, \dots, r\} \leq g(A, b) \\ &\leq \min\{f_A(v_i); i = 1, \dots, r\} \end{aligned}$$

Então $g(A, b) = 0$ se, e somente se, $\min\{f_A(v_i); i = 1, \dots, r\} = 0$, por (7) se, e somente se, $P(A, b)$ tem um vértice degenerado. \square

Corolário 2.2.1 (A, b) é não-degenerado se, e somente se, $(A X A^T)^{-1}$ existe para todo $x \in \mathcal{P}(A, b)$.

Demonstração:

Com efeito, pelo teorema (2.2.2) (A, b) é não-degenerado se, e somente se,

$$0 < g(A, b) \leq \det(A X A^T), \forall x \in \mathcal{P}(A, b).$$

E isto ocorre se, e somente se, $(A X A^T)^{-1}$ existe para todo $x \in \mathcal{P}(A, b)$. \square

O corolário (2.2.1) nos diz que o problema (A, b) é não-degenerado se, e somente se, a função $(A X A^T)^{-1}$ é contínua em $\mathcal{P}(A, b)$. Este fato é que torna as demonstrações de convergência dos algoritmos mais fáceis no caso não-degenerado.

Teorema 2.2.3 *O conjunto \mathcal{ND} dos problemas não degenerados, é denso em \mathcal{T} .*

Demonstração:

Sejam $(\tilde{A}, \tilde{b}) \in \mathcal{T}$, com $\tilde{b} \neq 0$. $I(\tilde{A}) = \{J \in C_{n,m}; \det(\tilde{A}_J) \neq 0\}$, $V(A, b) = \{J \in I(\tilde{A}); A_J^{-1}b \geq 0\}$ e $V\mathcal{ND}(\tilde{A}, \tilde{b}) = \{J \in V(\tilde{A}, \tilde{b}); \tilde{A}_J^{-1}\tilde{b} > 0\}$

Observemos que $V(\tilde{A}, \tilde{b})$, é o conjunto dos índices correspondentes aos vértices de $P(\tilde{A}, \tilde{b})$, pois se $\tilde{A}_{\bar{J}}^{-1}\tilde{b} \geq 0$, $v = \begin{pmatrix} \tilde{A}_{\bar{J}}^{-1}\tilde{b} \\ 0_{\bar{J}} \end{pmatrix}$ é um vértice de $P(\tilde{A}, \tilde{b})$, onde $0_{\bar{J}}$ é o vetor com índices em $\bar{J} = \{1, 2, \dots, n\} \setminus J$. Reciprocamente se v é vértice, e $\{j_1 < j_2 < \dots < j_l\}$, são os índices correspondentes às coordenadas não nulas de v , então $\{j_1, j_2, \dots, j_l\} \subset J$, para algum $J \in I(\tilde{A})$, visto que as colunas de \tilde{A} correspondentes a $\{j_1, j_2, \dots, j_l\}$ são L I.

$$\text{Assim } \tilde{A}v = \tilde{A}_J v_J + \tilde{A}_{\bar{J}} 0_{\bar{J}} = b \Rightarrow 0 \leq v_J = \tilde{A}_{\bar{J}}^{-1}b.$$

Estabelecido isto, fica claro que $V\mathcal{N}\mathcal{D}$ é o conjunto dos índices correspondentes aos vértices não degenerados.

Vamos considerar $I(\tilde{A}) = \{J_1, J_2, \dots, J_l\}$, onde $l = \#I(\tilde{A})$, e

$$\tilde{A}_{J_k}^{-1} = \begin{pmatrix} h_{J_{k1}}^T \\ \vdots \\ h_{J_{km}}^T \end{pmatrix}. \text{ Definamos}$$

$H_{J_{ki}} = \{x \in \mathbb{R}^m ; h_{J_{ki}}^T x = 0\}$. $H_{J_{ki}}$ é fechado e possui interior vazio, pois é um hiperplano em \mathbb{R}^m . Desta forma

$\bigcup_{J_{ki}} H_{J_{ki}}$, $k \in \{1, \dots, l\}$, $i \in \{1, 2, \dots, m\}$ é fechado e possui interior vazio. Portanto $W = \mathbb{R}^m \setminus \bigcup_{J_{ki}} H_{J_{ki}}$ é aberto e denso em \mathbb{R}^m . Segue que existe $d \in W$. Obviamente se $d \in W$

$$h_{J_{ki}}^T d \neq 0 \quad \forall J_{ki}.$$

$W_1 = \{d \in \mathbb{R}^n ; \tilde{A}_{J_1}^{-1}d > 0\}$, é aberto, portanto $W_1 \cap W \neq \emptyset$. Suponhamos sem perda de generalidade que $\tilde{A}_{J_1}^{-1}b \geq 0$. Tomemos então $d \in W_1 \cap W$ e $\delta = \frac{1}{2} \min\left\{\frac{h_{J_{ki}}^T \tilde{b}}{|h_{J_{ki}}^T d|} ; h_{J_{ki}}^T \tilde{b} \neq 0\right\}$. Com esta escolha $\tilde{A}_{J_k}^{-1}(\tilde{b} + td) = \tilde{A}_{J_k}^{-1}\tilde{b} + t\tilde{A}_{J_k}^{-1}d$, não possui coordenada nula, $\forall t \in (0, \delta]$, e além disso $\tilde{A}_{J_1}^{-1}(\tilde{b} + td) = \tilde{A}_{J_1}^{-1}\tilde{b} + t\tilde{A}_{J_1}^{-1}d > 0$. Segue que os vértices de $(\tilde{A}, \tilde{b} + td)$ são todos não-degenerados, e $P(\tilde{A}, \tilde{b} + td) \neq \emptyset$. Portanto $\mathcal{N}\mathcal{D}$ é denso em \mathcal{T} . \square

Teorema 2.2.4 *O conjunto dos problemas não degenerados e limitados é aberto em \mathcal{T} .*

Demonstração:

Seja (\tilde{A}, \tilde{b}) não degenerado e limitado. Suponha por absurdo que exista uma sequência de problemas degenerados (A^k, b^k) tal que $\lim A^k = \tilde{A}$ e $\lim b^k = \tilde{b}$. Seja v^k um vértice degenerado de (A^k, b^k) . Então v^k possui no máximo $m - 1$

coordenadas não nulas e $A^k v^k = b^k$. Existem duas possibilidades para a seqüência $\{v^k\}$:

- (1) $\{v^k\}$ é limitada. Neste caso sem perda de generalidade podemos supor que $\lim v^k = v$, e como v^k possui no máximo $m - 1$ coordenadas não nulas, v também tem essa característica. Além disso

$$\lim A^k v^k = \lim b^k \Rightarrow \tilde{A}v = \tilde{b}$$

Pela propriedade (4) de f_A , $f_A(v) = \theta$, o que implica $g(\tilde{A}, \tilde{b}) = 0$. E o problema (\tilde{A}, \tilde{b}) seria degenerado contrariando a hipótese.

- (2) $\{v^k\}$ é ilimitada. Neste caso podemos supor $\lim \frac{v^k}{\|v^k\|} = d \geq 0$. E teremos $\lim A^k \frac{v^k}{\|v^k\|} = \lim \frac{b^k}{\|v^k\|} \Rightarrow \tilde{A}d = 0 \Rightarrow (\tilde{A}, \tilde{b})$ é ilimitado contrariando a hipótese.

□

A limitação no teorema (2.2.4) é necessária como veremos no seguinte exemplo:

Exemplo 2.2.1 *Tomemos*

$$A = \begin{pmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix} e b = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad (2.2.6)$$

(A, b) é ilimitado e possui um único vértice, que é não-degenerado à saber $(0 \ 1 \ 0 \ 1)^T$

$$A^k = \begin{pmatrix} \frac{1}{k} & 1 & -1 & 0 \\ \frac{1}{k} & 0 & -1 & 1 \end{pmatrix} e b^k = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad (2.2.7)$$

(A^k, b^k) possui dois vértices, à saber $(k \ 0 \ 0 \ 0)^T$ degenerado e $(0 \ 1 \ 0 \ 1)^T$ não degenerado. Obviamente $(A^k, b^k) \rightarrow (A, b)$. Temos então uma seqüência de problemas degenerados convergindo para um problema não degenerado.

2.3 Topologia de Hausdorff no Conjunto dos Problemas

A topologia Euclideana de um \mathbb{R}^p , de certa maneira não é a topologia ideal para o espaço dos problemas, no sentido que se (A, b) é um elemento de \mathcal{T} e Λ é uma matriz $m \times m$ invertível, $(\Lambda A, \Lambda b)$ é essencialmente o mesmo problema, visto que $P(A, b) = P(\Lambda A, \Lambda b)$. No entanto podemos escolher Λ de modo que a distância de (A, b) à $(\Lambda A, \Lambda b)$ seja tão grande quanto desejarmos, isto é teríamos dois problemas que são os “mesmos” a uma distância grande.

Uma maneira razoável de definir a distância entre dois problemas é através da distância entre seus conjuntos viáveis. Mas isto esbarraria no fato de termos conjuntos viáveis ilimitados. Desta forma, a maneira que consideramos ideal para medirmos a distância entre dois problemas é pela distância entre os fechos convexos de seus vértices, e a distância entre as compactificações de seus espaços nulos. Para tal vamos definir uma classe de equivalência no conjunto dos problemas, e em seguida definir uma topologia no espaço quociente. No final do capítulo mostraremos que neste espaço topológico o conjunto dos problemas não degenerados é aberto, sem a restrição de limitação.

Definição 2.3.5 *Sejam (A, b) e (\bar{A}, \bar{b}) dois elementos de \mathcal{A} . Diremos que $(A, b) \sim (\bar{A}, \bar{b})$, se existe uma matriz Λ $m \times m$ invertível tal que $A = \Lambda \bar{A}$ e $b = \Lambda \bar{b}$.*

É fácil ver que a relação assim definida é uma relação de equivalência.

Veremos agora algumas proposições que nos dão propriedades desta relação.

Proposição 2.3.1 *$(A, b) \sim (\bar{A}, \bar{b})$ se, e somente se, $P(A, b) = P(\bar{A}, \bar{b})$.*

Demonstração:

Suponhamos que $(A, b) \sim (\bar{A}, \bar{b})$. Então existe uma matriz Λ $m \times m$ invertível tal que

$$A = \Lambda \bar{A} \text{ e } b = \Lambda \bar{b},$$

se $x \in P(A, b)$, então $Ax = b$, o que implica $\Lambda^{-1}Ax = \Lambda^{-1}b$, ou seja

$$\bar{A}x = \bar{b} \implies x \in P(\bar{A}, \bar{b}).$$

De maneira análoga prova-se que se $x \in P(\bar{A}, \bar{b}) \implies x \in P(A, b)$.

Suponhamos agora que $P(A, b) = P(\bar{A}, \bar{b})$. Neste caso segue $\mathcal{N}(A) = \mathcal{N}(\bar{A})$, pois se $0 < x_o \in P(A, b)$ e $\eta \in \mathcal{N}(A)$, existe $\delta > 0$ tal que $x_o + t\eta \in P(A, b)$, $\forall t \in [0, \delta]$, conseqüentemente $\bar{b} = \bar{A}x_o = \bar{A}(x_o + t\eta)$, o que implica $\bar{A}\eta = 0$. Tomemos então $\{\eta_{m+1}, \dots, \eta_n\}$ como uma base de $\mathcal{N}(A)$, e $\{\xi_1, \dots, \xi_m, \eta_{m+1}, \dots, \eta_n\}$ uma base de \mathbb{R}^n . Teremos então que $(A\xi_1, \dots, A\xi_m)$ e $(\bar{A}\xi_1, \dots, \bar{A}\xi_m)$ são duas bases de \mathbb{R}^m . Seja $\Lambda : \mathbb{R}^m \rightarrow \mathbb{R}^m$ a transformação linear tal que $\Lambda(A\xi_i) = \bar{A}\xi_i$. Neste caso a transformação linear $\Lambda A : \mathbb{R}^n \rightarrow \mathbb{R}^m$, é tal que $(\Lambda A)\xi_i = \Lambda(A\xi_i) = \bar{A}\xi_i$ e $(\Lambda A)\eta_j = \Lambda(A\eta_j) = \Lambda 0 = 0$. Portanto $\Lambda A = \bar{A}$. Além disso, seja x tal que

$$Ax = b \text{ e } \bar{A}x = \bar{b}.$$

então

$$\Lambda Ax = \Lambda b$$

$$\bar{A}x = \Lambda b$$

$$\bar{b} = \Lambda b.$$

□

A proposição (2.3.1) nos garante que um problema limitado só pode ser equiivalente a um problema limitado.

Vamos fazer algumas notações para demonstrarmos as proposições seguintes.

$V(A, b)$ é o conjunto dos vértices de $P(A, b)$.

$C_o(K)$ é o fecho convexo do conjunto K .

$W(A, b) = C_o(V(A, b))$.

$ri(K)$ é o interior relativo do conjunto K .

Proposição 2.3.2 $P(A, b) = P(\bar{A}, \bar{b})$ se, e somente se $C_o(V(A, b)) = C_o(V(\bar{A}, \bar{b}))$ e $\mathcal{N}(A) = \mathcal{N}(\bar{A})$.

Demonstração:

Suponha que $P(A, b) = P(\bar{A}, \bar{b})$, pela proposição (2.3.1), existe Λ invertível tal que $\Lambda A = \bar{A}$ e $\Lambda b = \bar{b}$, assim os vértices de $P(A, b)$ e $P(\bar{A}, \bar{b})$ são os mesmos e $\mathcal{N}(A) = \mathcal{N}(\bar{A})$.

Suponhamos agora que $C_o(V(A, b)) = C_o(V(\bar{A}, \bar{b}))$ e $\mathcal{N}(A) = \mathcal{N}(\bar{A})$. Agora pelo lema (2.2.2)

$$x \in P(A, b) \iff x = v + d,$$

onde $v \in C_o(V(A, b))$, $d \geq 0$ e $Ad = 0$. E pelas hipóteses juntamente com o lema (2.2.2) se, e somente se, $x \in P(\overline{A}, \overline{b})$. \square

Espaços dos Problemas

Definição 2.3.6 *O espaço dos problemas é o conjunto*

$$\mathcal{P} = \mathcal{A} / \sim = \{ \overline{(A, b)} ; (A, b) \in \mathcal{A} \},$$

onde $\overline{(A, b)} = \{ (\Lambda A, \Lambda b) ; \Lambda \in \mathbb{R}^{m \times m} \text{ e } \det \Lambda \neq 0 \}$.

Se $\wp = \overline{(A, b)}$ é um elemento de \mathcal{P} , diremos que (A, b) é um *representante* de \wp . O conjunto de vértices de \wp será o conjunto de vértices de $P(A, b)$ para algum (e portanto para todo) representante de \wp , e será denotado por $V(\wp)$. Da mesma forma diremos que \wp é limitado se $P(A, b)$ for limitado para algum representante de \wp .

Definição 2.3.7 *O núcleo unitário de \wp é o conjunto*

$$\mathcal{N}(\wp) = \{ d \in \mathcal{N}(A) ; \|d\| = 1 \},$$

onde (A, b) é um representante de \wp .

Claramente a definição acima independe da escolha do representante de \wp .

Vamos definir uma topologia em \mathcal{P} proveniente de uma métrica. Para isso veremos primeiro o que é a métrica de Hausdorff.

Definição 2.3.8 *Seja $\phi(\mathbb{R}^n)$ o conjunto de todos os compactos e não vazios de \mathbb{R}^n . Dados Y e Z em $\phi(\mathbb{R}^n)$, o número*

$$\rho(Y, Z) = \max \left\{ \sup_{y \in Y} d(y, Z), \sup_{z \in Z} d(z, Y) \right\},$$

é chamado *distância de Hausdorff* entre os conjuntos Y e Z .

Lema 2.3.3 *A função ρ definida acima, define uma métrica em $\phi(\mathbb{R}^n)$.*

Demonstração:

Ver em [17]. \square

Vamos denotar $B(Y, r) = \bigcup_{y \in Y} B(y, r)$, onde $B(y, r)$ é a bola aberta de centro y e raio r na norma euclidiana, como sendo a bola aberta de centro Y e raio r .

Métrica de Hausdorff em \mathcal{P}

Definição 2.3.9 *Sejam \wp e ℓ em \mathcal{P} , o número*

$$\rho_H(\wp, \ell) = \rho(C_o(V(\wp)), C_o(V(\ell))) + \rho(\mathcal{N}(\wp), \mathcal{N}(\ell))$$

será chamado distância de Hausdorff entre \wp e ℓ .

Proposição 2.3.3 *A função ρ_H definida acima define uma distância em \mathcal{P} .*

Demonstração:

Sejam \wp, ℓ e θ elementos de \mathcal{P} . (A, b) e (\bar{A}, \bar{b}) representantes de \wp e ℓ respectivamente.

$$\begin{aligned} \text{(i)} \quad \rho_H(\wp, \ell) &= \rho(C_o(V(\wp)), C_o(V(\ell))) + \rho(\mathcal{N}(\wp), \mathcal{N}(\ell)) = 0 \iff \\ &\rho(C_o(V(\wp)), C_o(V(\ell))) = 0 \text{ e } \rho(\mathcal{N}(\wp), \mathcal{N}(\ell)) = 0 \iff \\ &C_o(V(A, b)) = C_o(V(\bar{A}, \bar{b})) \text{ e } \mathcal{N}(\wp) = \mathcal{N}(\ell) \iff C_o(V(A, b)) = \\ &C_o(V(\bar{A}, \bar{b})) \text{ e } \mathcal{N}(A) = \mathcal{N}(\bar{A}). \text{ Pela proposição (2.3.2) se, e somente se, } \\ &P(A, b) = P(\bar{A}, \bar{b}), \text{ e pela proposição (2.3.1) se, e somente se, } \\ &(A, b) \sim (\bar{A}, \bar{b}) \iff \wp = \ell. \end{aligned}$$

(ii) Desigualdade triangular

$$\begin{aligned} \rho_H(\wp, \ell) &= \rho(V(\wp), V(\ell)) + \rho(\mathcal{N}(\wp), \mathcal{N}(\ell)) \\ &\leq \rho(V(\wp), V(\theta)) + \rho(V(\theta), V(\ell)) + \rho(\mathcal{N}(\wp), \mathcal{N}(\theta)) + \rho(\mathcal{N}(\theta), \mathcal{N}(\ell)) \\ &= \rho_H(\wp, \theta) + \rho_H(\theta, \ell) \end{aligned}$$

(iii) A simetria e o fato que $\rho_H(\wp, \ell) \geq 0$ são imediatos.

□

Podemos agora considerar \mathcal{P} como um espaço topológico com a topologia proveniente da métrica ρ_H .

Definição 2.3.10 Um elemento $\wp \in \mathcal{P}$ é dito degenerado, se (A, b) for um problema degenerado para algum representante de \wp . Caso contrário \wp é dito não-degenerado.

Vamos denotar por \mathcal{DP} o conjunto formado pelos elementos degenerados de \mathcal{P} , e por \mathcal{NDP} o conjunto formado pelos elementos não-degenerados de \mathcal{P} .

Para demonstração do teorema principal deste capítulo faremos uso do seguinte lema.

Lema 2.3.4 Se $\rho(Y, Z) < r$, então:

$$Z \subset B(Y, r) \text{ e}$$

$$Y \subset B(Z, r).$$

Demonstração:

Seja $z_o \in Z$, como $\rho(Z, Y) < r$

$$\max\left\{ \sup_{z \in Z} d(z, Y), \sup_{y \in Y} d(y, Z) \right\} < r.$$

Logo $\sup_{z \in Z} d(z, Y) < r$, o que implica $d(z_o, Y) < r$.

Y é fechado, portanto existe $\bar{y} \in Y$ tal que $\|z_o - \bar{y}\| = d(z_o, Y) < r$, deste modo teremos $z_o \in B(\bar{y}, r) \subset B(Y, r)$. De maneira análoga prova-se que $Y \subset B(Z, r)$. \square

Teorema 2.3.5 \mathcal{NDP} é aberto em \mathcal{P} .

Demonstração:

Seja $\wp \in \mathcal{NDP}$, e suponha que exista uma sequência \wp^k em \mathcal{DP} tal que $\rho_H(\wp^k, \wp) \rightarrow 0$. Sejam (A^k, b^k) representante de \wp^k e (A, b) representante de \wp .

Como $\rho_H(\wp^k, \wp) \rightarrow 0$ temos que $\rho(C_o(V(\wp^k)), C_o(V(\wp))) \rightarrow 0$

$$\max\left\{ \sup_{v \in C_o(V(\wp^k))} d(v, C_o(V(\wp))), \sup_{w \in C_o(V(\wp))} d(w, C_o(V(\wp^k))) \right\} \rightarrow 0 \quad (2.3.1)$$

Além disso, pelo lema (2.3.4), existe um k_o tal que $k \geq k_o$

$$C_o(V(\wp^k)) \subset B(C_o(V(\wp)), 1) \quad (2.3.2)$$

De (2.3.1) e (2.3.2), podemos concluir que se $w_k \in C_o(V(\wp^k))$ então $\{w_k\}$ é limitada e $d(w_k, C_o(V(\wp))) \rightarrow 0$.

Consideremos então a seqüência $v_k \in C_o(V(\wp^k))$ tal que v_k é um vértice degenerado de \wp^k .

Pelo que vimos acima esta seqüência é limitada, portanto vamos considera-la convergente. $v_k \rightarrow \bar{v}$.

Ora $d(v_k, C_o(V(\wp))) \rightarrow 0$, logo $\bar{v} \in C_o(V(\wp))$. O fato de v_k ser um vértice degenerado de \wp^k , nos diz que v_k possui no máximo $m - 1$ componentes não nulas. Segue que \bar{v} também possui no máximo $m - 1$ coordenadas não nulas.

Pela propriedade (4) da função f_A , se (A, b) é um representante de \wp , então $f_A(\bar{v}) = 0$, portanto $g(A, b) = 0$, e pelo teorema (2.2.2) \wp é degenerado, contrariando a hipótese. \square

Analisamos a questão da densidade de \mathcal{NDP} em \mathcal{P} , até o momento não chegamos a uma conclusão. Mas a análise de diversos expemplos nos leva a crer que \mathcal{NDP} é denso em \mathcal{P} .

Capítulo 3

Algoritmo Afim Potencial para Programação Convexa

3.1 Introdução

O objetivo deste capítulo é estudar algoritmos “Afim Escala Potencial”, para os problemas de programação convexa

$$\begin{aligned} & \text{minimizar } f(x) \\ & \text{sujeito a } Ax = b \\ & \quad x \geq 0 \end{aligned} \tag{3.1.1}$$

Considerando o problema não-degenerado, f é convexa e de classe C^1 , e $P(A, b) \neq \emptyset$. E o problema

$$\begin{aligned} & \text{minimizar } f(x) \\ & \text{sujeito a } x \geq 0 \end{aligned} \tag{3.1.2}$$

onde f é convexa de classe C^1 .

Problemas nestes formatos ocorrem com frequência em programação matemática.

Algoritmos afim potencial, generalizam o algoritmo proposto e analisado por Gonzaga em [12] para (3.1.1). E o algoritmo proposto por Eggermont em [9] para (3.1.2), e cuja análise de convergência foi estudada por Iussem em [16]. A generalização é no sentido que, as direções de busca propostas por Gonzaga (direção afim escala de Dikin), e por Eggermont são casos particulares da direção que propomos, para determinados valores de um parâmetro r (respectivamente $r = 2$ e $r = 1$).

Obtivemos os mesmos resultados que Gonzaga obteve para (3.1.1), e Iussem para (3.1.2), com o acréscimo que para (3.1.1) obtivemos convergência fraca.

O enfoque com o qual analisamos os problemas (3.1.1) e (3.1.2) é diferente dos enfoques dados por Gonzaga e Iussem, no sentido que encaramos \mathbb{R}_{++}^n como uma variedade Riemanniana com métrica $G(x) = X^{-r}$, $r \geq 1$ conforme notação estabelecida anteriormente. Consideramos $M = \{x \in \mathbb{R}_{++}^n; Ax = b\}$ como uma subvariedade de \mathbb{R}_{++}^n com métrica induzida. Além disso a direção de busca linear é o simétrico do gradiente da função f , considerada como uma função definida na respectiva variedade.

Algoritmos afim escala potencial já foram estudados por Saigal em [27] para Programação linear, onde é mostrado inclusive a convergência para o caso degenerado.

A ligação entre geometria e programação matemática data desde Luenberger [21]. Apartir daí diversos autores como Gabay [11], Bayer e Lagarias [1], Rapsák [26], Xavier e Oliveira [6], Ferreira [10], Udriste [30], entre outros estudaram algoritmos com este enfoque. A maioria dos algoritmos analisados pelos autores acima se utiliza de busca ao longo de geodésicas. O algoritmo que propomos não utiliza busca geodésica, e sim uma busca linear na direção do simétrico do gradiente da função f restrita a variedade, sendo portanto mais aplicável.

Os resultados e notações relativas a geometria utilizados nos parágrafos seguintes podem ser encontrados em [5].

3.2 O Algoritmo para (3.1.1)

Vamos considerar $M = \{x \in \mathbb{R}_{++}^n; Ax = b\} \neq \emptyset$, como sendo uma subvariedade de $(\mathbb{R}_{++}^n, X^{-r})$ com métrica induzida. Consideraremos sempre que o posto de A é m e que o problema (3.1.1) é não degenerado.

O espaço tangente a M é

$$T_x M = \{d \in \mathbb{R}^n; Ad = 0\}$$

para todo x . Portanto denotaremos por TM .

A projeção em TM segundo uma métrica G é dada por

$$P^G(x) = I - G^{-1}(x)A^T(AG^{-1}(x)A^T)^{-1}A \quad (3.2.1)$$

No caso específico em que $G(x) = X^{-r}$, teremos

$$P(x) = I - X^r A^T (AX^r A^T)^{-1} A \quad (3.2.2)$$

Portanto teremos o gradiente de f restrito a variedade (M, X^{-r}) como sendo

$$\nabla f^M(x) = P^G(x)G^{-1}(x)\nabla f(x) \quad (3.2.3)$$

ou seja

$$\nabla f^M(x) = P(x)X^r\nabla f(x) \quad (3.2.4)$$

Observe que $X^r\nabla f(x)$ é o gradiente de f em $(\mathbb{R}_{++}^n, X^{-r})$, e que para todo $d \in TM$ tal que $\|d\|_G = 1$

$$\langle d, \nabla f^M(x) \rangle_G \leq \|d\|_G \|\nabla f^M(x)\|_G = \|\nabla f^M\|_G$$

consequentemente $-\frac{\nabla f^M(x)}{\|\nabla f^M(x)\|_G}$ é a direção de descida máxima de f em x .

É um fato óbvio que $\nabla f^M(x) \in TM$.

Se f é convexa, então x^* é solução de (3.1.1) se, e somente se, existirem $y^* \in \mathbb{R}^m$, $0 \leq s^* \in \mathbb{R}^n$ tais que

$$\begin{aligned} A^T y^* + s^* &= \nabla f(x^*) \\ Ax^* &= b \\ x^* \geq 0 \text{ e } x_i^* s_i^* &= 0, \quad i = 1, 2, \dots, n \end{aligned} \quad (3.2.5)$$

$$(3.2.6)$$

Das Condições acima concluímos que:

$$y(x^*) = (AX^{*r}A^T)^{-1}AX^{*r}\nabla f(x^*)$$

$$s(x^*) = \nabla f(x^*) - A^T y(x^*)$$

$$X^* s^* = 0$$

Vamos então definir as seguintes funções em M .

$$y(x) = (AX^rA^T)^{-1}AX^r\nabla f(x) \quad (3.2.7)$$

$$s(x) = \nabla f(x) - A^T y(x) = (P(x))^T \nabla f(x) \quad (3.2.8)$$

$$d(x) = X^r s(x) = P(x)X^r \nabla f(x) = \nabla f^M(x) \quad (3.2.9)$$

Pela hipótese de não degenerescência, todas as funções definidas acima, podem ser estendidas continuamente para o fecho de M .

No que segue, usaremos a notação $\chi(v)$ para denotar a maior componente do vetor v .

Algoritmo 3.1. Dado $x^o \in M$, $\beta > 0$ e $0 < \delta < 1$

$$k = 0$$

Repita

$$y^k = y(x^k)$$

$$s^k = s(x^k)$$

$$d^k = d(x^k)$$

$$\mu^k = \chi(X_k^{-1}d^k, 0)$$

$$\alpha^k = \frac{\delta}{\beta + \mu^k}$$

$$t^k = \underset{t \in [0, \alpha^k]}{\operatorname{argmin}} f(x^k - td^k)$$

$$x^{k+1} = x^k - t^k d^k$$

$$k = k + 1$$

Em primeiro lugar observemos que $x^k \in M$, $\forall k \in \mathbb{N}$, com efeito

$$(X_k^{-1}x^{k+1})_i = 1 - t^k(x_i^k)^{-1}d_i^k \geq 1$$

se $d_i^k \leq 0$. Caso contrário

$$t^k(x_i^k)^{-1}d_i^k \leq \frac{\delta\mu^k}{\beta + \mu^k} < 1, \quad .$$

Logo $1 - t^k(x_i^k)^{-1}d_i^k > 0$

Portanto o algoritmo é de pontos interiores. Devemos tomar o parâmetro δ próximo de 1 e o parâmetro β próximo de zero, para que o intervalo de busca $[0, \alpha^k]$ seja o maior possível.

3.3 Convergência do Algoritmo 3.1

Neste parágrafo analisaremos a convergência do algoritmo 3.1. Vamos considerar sempre que $L_{x_o} = \{x \in \bar{M}; f(x) \leq f(x_o)\}$ é limitado.

Lema 3.3.5 *Sejam $x \in M$ e $d \in TM$. Então vale:*

$$(i) \quad \nabla f(x)^T d = s(x)^T d$$

$$(ii) \quad \nabla f(x)^T d(x) = \|X^{r/2} s(x)\|^2 = \|X^{-r/2} d(x)\|^2$$

Demonstração:

(i)

$$\begin{aligned} \text{Por (3.2.8), } s(x) &= \nabla f(x) - A^T y(x), \text{ assim} \\ s(x)^T d &= \nabla f(x)^T d - (A^T y(x))^T d \\ &= \nabla f(x)^T d - y(x)^T A d \end{aligned}$$

como $d \in TM$, $Ad = 0$ então

$$s(x)^T d = \nabla f(x)^T d$$

(ii)

Pelo item (i), $\nabla f(x)^T d(x) = s(x)^T d(x)$ e por (3.2.9)

$$\nabla f(x)^T d(x) = s(x)^T X^r s(x) = \|X^{r/2} s(x)\|^2$$

substituindo $s(x) = X^{-r} d(x)$

$$\nabla f(x)^T d(x) = \|X^{-r/2} d(x)\|^2$$

□

Lema 3.3.6 *A seqüência gerada pelo algoritmo satisfaz*

$$\lim_{k \rightarrow \infty} X_k s^k = 0$$

Demonstração:

Em primeiro lugar $X_k s^k \rightarrow 0$ se, e somente se, $X_k^{r/2} s^k \rightarrow 0$. Isto decorre do fato de L_{x_0} ser limitado, $x^k \in L_{x_0}$, $\forall k$, e de $s(x)$ ser contínua no fecho de M , pela hipótese de não degenerescência. Portanto x^k e s^k são limitadas. Vamos supor agora por absurdo que $\lim_{k \rightarrow \infty} X_k^{r/2} s^k \neq 0$. Neste caso existe uma subsequência x^{kj} e $\bar{\epsilon} > 0$ tal que $\|X_{kj}^{r/2} s^{kj}\| \geq \bar{\epsilon}$. Pela limitação de $\{x^k\}$, podemos supor sem perda de generalidade que $x^{kj} \rightarrow \bar{x} \in \bar{M}$, e pela hipótese de não degenerescência $s^{kj} \rightarrow s(\bar{x}) = \bar{s}$. Tomemos

$$L = \sup_{x \in L_{x_0}} \{ \chi(X^{-1}d(x), 0) \} = \sup_{x \in L_{x_0}} \{ \chi(X^{r-1}s(x), 0) \}$$

L é bem definido e finito. Teremos então na seqüência

$$\frac{\delta}{\beta} \geq \alpha_k = \frac{\delta}{\beta + \mu^k} \geq \frac{\delta}{\beta + L}$$

Deste modo pela definição da seqüência

$$\begin{aligned} f(x^{kj+1}) &\leq f(x^{kj} - \alpha d^{kj}), \quad \forall \alpha \in [0, \frac{\delta}{\beta + L}], \text{ ou seja} \\ f(x^{kj+1}) &\leq f(x^{kj}) - \alpha \nabla f(x^{kj})^T d^{kj} + r(x^{kj}, \alpha d^{kj}) \\ f(x^{kj+1}) - f(x_{kj}) &\leq -\alpha \nabla f(x^{kj})^T d^{kj} + r(x^{kj}, \alpha d^{kj}) \end{aligned} \quad (3.3.1)$$

Pela limitação de L_{x_0} e da monotocidade de $f(x^k)$, temos que $f(x^{kj+1}) - f(x^{kj}) \rightarrow 0$. Além disso pelo lema (3.3.5)

$$\nabla f(x^{kj})^T d^{kj} \rightarrow \|\bar{X}^{r/2} s(\bar{x})\|^2 \geq \bar{\epsilon}$$

Tomando agora o limite em (3.3.1) temos

$$0 \leq -\alpha \|\bar{X}^{r/2} s(\bar{x})\|^2 + r(\bar{x}, \alpha \bar{d}), \forall \alpha \in [0, \frac{\delta}{\beta + L}]$$

assim

$$\frac{r(\bar{x}, \alpha \bar{d})}{\alpha} \geq \|\bar{X}^{r/2} s(\bar{x})\|^2 \geq \bar{\epsilon}, \forall \alpha \in [0, \frac{\delta}{\beta + L}]$$

logo

$$\lim_{\alpha \rightarrow 0} \frac{r(\bar{x}, \alpha \bar{d})}{\alpha} \geq \bar{\epsilon},$$

contrariando a diferenciabilidade de f em \bar{x} . □

Corolário 3.3.2 *Na seqüência gerada pelo algoritmo vale*

$$\lim_{k \rightarrow \infty} d^k = 0$$

Demonstração:

$$d^k = X_k^r s^k, \text{ que tende a zero pelo lema anterior}$$

□

Corolário 3.3.3 $\lim_{k \rightarrow \infty} (x^{k+1} - x^k) = 0$

Demonstração:

$$\|x^{k+1} - x^k\| = \|t^k\| \|d^k\| \leq \frac{\delta}{\beta} \|d^k\|, \text{ que tende a zero pelo corolário anterior}$$

□

O lema (3.3.5) nos diz que qualquer ponto de acumulação \bar{x} , da seqüência $\{x^k\}$ satisfaz a complementariedade $\bar{X}s(\bar{x}) = 0$. Mostraremos agora que se f é convexa $\bar{s} = s(\bar{x}) \geq 0$ e portanto todo ponto de acumulação da seqüência é solução do problema (3.1.1).

Fixemos então \bar{x} um ponto de acumulação da seqüência gerada pelo algoritmo, e $\bar{s} = s(\bar{x})$. Vamos escrever

$$\bar{x} = \bar{x}_B \oplus \bar{x}_N,$$

onde $B \cup N$ é uma partição do conjunto $\{1, 2, \dots, n\}$ tal que $N = \{i; \bar{s}_i \neq 0\}$. Neste caso $\bar{x}_N = 0$.

Observemos que se \bar{x} e $\bar{\bar{x}}$ são pontos de acumulação da seqüência gerada pelo algoritmo, $f(\bar{x}) = f(\bar{\bar{x}})$, pois $f(x^k)$ é convergente. Portanto para que \bar{s} seja um multiplicador ótimo de (3.1.1), temos que provar que $\bar{s}_N > 0$.

Pelas considerações feitas acima definiremos o conjunto

$$\Omega = \{x \in \bar{M}; f(x) = f(\bar{x}) \text{ e } x_N = 0\}$$

Pela limitação de L_{x_0} , Ω é compacto. Observe que os pontos de acumulação da seqüência, tais que $x_N = 0$ estão necessariamente em Ω . Em seguida faremos uma “dilatação” de Ω , e mostraremos que existe uma dilatação de Ω tal que, se \bar{s}_N possui alguma componente negativa, a seqüência terá ponto de acumulação fora da dilatação. O que vai contrariar o corolário (3.3.3).

Uma γ -dilatação de Ω é o conjunto

$$\Omega_\gamma = \{x \in \bar{M}; \|x - y\| \leq \gamma, \text{ para algum } y \in \Omega\}$$

Observemos que se $x \in \bar{M}$ e $x \notin \Omega_\gamma$, $d(x, \Omega) > \gamma$

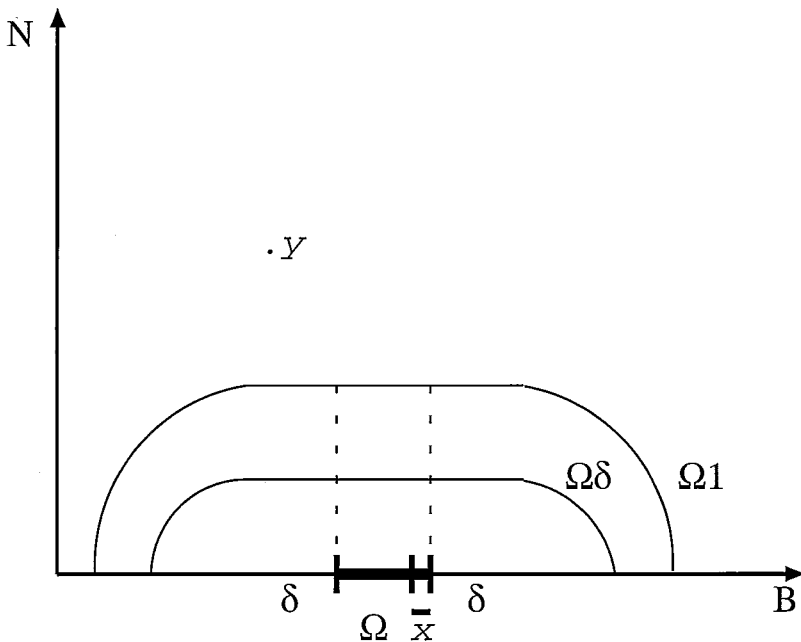


Figura 3.1: Descrição Pictórica

Lema 3.3.7 *O conjunto Ω é convexo*

Demonstração:

Sejam u e v em Ω e $t \in [0, 1]$. Então

$$(tu + (1-t)v)_N = tu_N + (1-t)v_N = 0.$$

Pela convexidade de f

$$f(tu + (1-t)v) \leq tf(u) + (1-t)f(v)$$

$$f(tu + (1-t)v) \leq tf(\bar{x}) + (1-t)f(\bar{x}) = f(\bar{x})$$

Observe que $tu + (1-t)v - \bar{x}$ está em TM , portanto pelo lema (3.3.5)

$$\nabla f(\bar{x})^T(tu + (1-t)v - \bar{x}) = s(\bar{x})^T(tu + (1-t)v - \bar{x}) =$$

$$\bar{s}_N(tu + (1-t)v - \bar{x})_N + \bar{s}_B(tu + (1-t)v - \bar{x})_B = 0$$

pois $s_B(\bar{x}) = 0$ e $(tu + (1-t)v - \bar{x})_N = 0$.

Usando novamente a convexidade de f teremos

$$f(tu + (1-t)v) \geq f(\bar{x}) + \nabla f(\bar{x})^T(tu + (1-t)v - \bar{x}) = f(\bar{x}).$$

Segue então que $f(tu + (1-t)v) = f(\bar{x})$ e portanto Ω é convexo. \square

Mostraremos agora um resultado de análise convexa que as vezes não se encontra na literatura.

Lema 3.3.8 *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ convexa de classe C^1 . Se f é constante no conjunto convexo Ω , então ∇f é constante em Ω .*

Demonstração:

O teorema é evidentemente verdadeiro, se Ω possui um único elemento. Suponhamos então que Ω possua mais de um elemento. Observemos os seguintes fatos:

Fato 1 - Se h é uma função convexa de classe C^1 e $\nabla h(\bar{x}) = 0$, então \bar{x} é minimizador global de h .

Fato 2 - Se $\bar{x} \in ri(\Omega)$ então $\nabla f^T(\bar{x})(y - \bar{x}) = 0, \forall y \in \Omega$.

Com efeito, seja $y \in \Omega$. Como $\bar{x} \in ri(\Omega)$, existe $\epsilon > 0$ tal que $(1-t)\bar{x} + ty \in M, \forall t \in [-\epsilon, 1]$. Definamos

$$g(t) = f((1-t)\bar{x} + ty)$$

Então g é constante em $[-\epsilon, 1]$, logo

$$0 = g'(t) = \nabla f^T((1-t)\bar{x} + ty)(y - \bar{x}), \quad \forall t \in (-\epsilon, 1)$$

em particular para $t = 0$

$$\nabla f^T(\bar{x})(y - \bar{x}) = 0$$

Fixemos $\bar{x} \in ri(\Omega)$. Seja

$$h(x) = f(x) - \nabla f^T(\bar{x})(x - \bar{x})$$

h é convexa e de classe C^1 e

$$\nabla h(x) = \nabla f(x) - \nabla f(\bar{x})$$

logo

$$\nabla h(\bar{x}) = 0$$

Pelo fato 1, \bar{x} é minimizador global de h .

Se $y \in \Omega$, $h(y) = f(y) - \nabla f^T(\bar{x})(y - \bar{x})$, e pelo fato 2 $h(y) = f(y) = f(\bar{x}) = h(\bar{x})$, portanto todo elemento de Ω é minimizador global de h .

Logo $\nabla h(y) = 0$, $\forall y \in \Omega$. Ou seja,

$$\nabla f(y) = \nabla f(\bar{x}), \quad \forall y \in \Omega$$

□

Lema 3.3.9 Para todo $x \in \Omega$, $s(x) = \bar{s} = s(\bar{x})$

Demonstração:

$$s(x) = (I - A^T(AX^r A^T)^{-1}AX^r) \nabla f(x)$$

mas em Ω , $\nabla f(x) = \nabla f(\bar{x}) = \bar{s} + A^T\bar{y}$, então

$$s(x) = (I - A^T(AX^r A^T)^{-1}AX^r)(\bar{s} + A^T\bar{y})$$

Note que $x \in \Omega$ implica $x_N = 0$ e como $\bar{s}_B = 0$, temos que $X^r\bar{s} = 0$. Logo

$$s(x) = \bar{s} + A^T\bar{y} - A^T(AX^r A^T)^{-1}AX^r\bar{s} - A^T(AX^r A^T)^{-1}AX^r A^T\bar{y}$$

$$s(x) = \bar{s} + A^T\bar{y} - A^T\bar{y}$$

$$s(x) = \bar{s}$$

Mostraremos agora a existência de uma dilatação Ω_δ de Ω , onde o sinal de $d_i(x)$ é o mesmo de \bar{s}_i , se $i \in N$. Seja então

$$\theta = \frac{1}{2} \min_{i \in N} |\bar{s}_i|$$

Lema 3.3.10 *Existe $\delta > 0$ tal que para todo $x \in \Omega_\delta$ vale:*

- (i) $\|s(x) - \bar{s}\| \leq \theta$
- (ii) Se $x_N > 0$, então $s_i(x)\bar{s}_i > 0$ e $d_i(x)\bar{s}_i > 0$, $\forall i \in N$

Demonstração:

- (i) Como o problema é não-degenerado $s(x)$ é uniformemente contínua no compacto Ω_1 , portanto existe $\delta > 0$ tal que se x e y estão em Ω_1 com $\|x - y\| \leq \delta$, então $\|s(x) - s(y)\| \leq \theta$. Logo se $x \in \Omega_\delta$, existe $y \in \Omega$ com $\|x - y\| \leq \delta$, conseqüentemente $\|s(x) - s(y)\| \leq \theta$ como s é constante em Ω $s(y) = s(\bar{x}) = \bar{s}$. Ou seja,

$$\|s(x) - \bar{s}\| \leq \frac{1}{2} \min_{i \in N} |\bar{s}_i|$$

- (ii) Se $i \in N$, pelo item anterior $|s_i(x) - \bar{s}_i| \leq \|s(x) - \bar{s}\| \leq \frac{1}{2} \min_{j \in N} \{|\bar{s}_j|\} \leq \frac{1}{2} |\bar{s}_i|$, isto significa que $s_i(x)\bar{s}_i > 0$. Como $d_i(x) = x_i^r s_i(x)$, segue que

$$d_i(x)\bar{s}_i > 0$$

□

Mostraremos agora que $\bar{s}_N > 0$. Para isso provaremos que os pontos de acumulação da seqüência que não estão em Ω estão fora de Ω_δ . Veremos em seguida que se alguma componente de \bar{s}_N for negativa a seqüência possui ponto de acumulação em Ω e ponto de acumulação fora de Ω_δ . Este fato irá de encontro com o lema (3.3.3).

Lema 3.3.11 *Seja \tilde{x} ponto de acumulação $\{x^k\}$. Então $\tilde{x} \in \Omega$ ou $\tilde{x} \notin \Omega_\delta$.*

Demonstração:

Suponha que exista um ponto de acumulação \tilde{x} de $\{x^k\}$ tal que $\tilde{x} \in \Omega_\delta$, $\tilde{x} \notin \Omega$, e $\tilde{s} = s(\tilde{x})$. Pela convergência de $f(x^k)$, temos $f(\tilde{x}) = f(\bar{x})$, como $\tilde{x} \notin \Omega$, existe $i \in N$ tal que $\tilde{x}_i > 0$. Pelo lema anterior $\tilde{s}_i\bar{s}_i > 0$ o que implica $\tilde{s} \neq 0$. Portanto $\tilde{x}_i\tilde{s}_i \neq 0$, contrariando a complementariedade. □

Lema 3.3.12 *Se $\bar{s} = s(\bar{x})$ possui uma componente negativa, então $\{x^k\}$ possui um ponto de acumulação $\tilde{x} \notin \Omega_\delta$.*

Demonstração:

Em primeiro lugar observe que, se todos os pontos de acumulação da seqüência estão em Ω_δ , pelo lema (3.3.11) eles estariam em Ω , que está contido no interior relativo de Ω_δ . Portanto a seqüência inteira estaria contida em Ω_δ a partir de um determinado k_o . Vamos assumir então que exista $i \in N$ tal que $(\bar{s}_N)_i < 0$, e que todo ponto de acumulação de $\{x^k\}$ esteja em Ω_δ e portanto em Ω . Então se $k \geq k_o$ teríamos $x^k \in \Omega_\delta$ e $s(x^k)_i \bar{s}_i > 0$ o que implica $(s(x^k))_i < 0$ e pelo lema (3.3.10) $(d^k)_i < 0$, e neste caso

$$x_i^{k+1} = x_i^k - t^k d_i^k \geq x_i^k \geq x_i^{k_o} > 0,$$

e portanto a seqüência x_i^k seria crescente, contrariando o fato que $\liminf x_i^k = 0$. \square

Lema 3.3.13 $\bar{s}_N > 0$

Demonstração:

Sabemos que $\bar{s}_B = 0$, suponhamos que exista $i \in N$, tal que $(\bar{s}_N)_i < 0$. Vamos mostrar neste caso que existe uma subseqüência $\{x^{kj}\}$ de $\{x^k\}$, tal que $x^{kj} \rightarrow \bar{u} \in \Omega$ e $x^{kj+1} \rightarrow \bar{v} \notin \Omega_\delta$. O que vai contrariar o corolário (3.3.3).

Pelo lema (3.3.12), se existe $i \in N$ tal que $(\bar{s}_N)_i < 0$ a seqüência $\{x^k\}$ possui ponto de acumulação fora de Ω_δ e ponto de acumulação \bar{x} em Ω . Logo existe uma subseqüência $\{x^{kj}\}$ tal que $x^{kj} \in \Omega_\delta$ e $x^{kj+1} \notin \Omega_\delta$, visto que Ω_δ é fechado. Podemos supor essas seqüências convergentes. Pelo lema (3.3.11), $x^{kj} \rightarrow u \in \Omega$ e $x^{kj+1} \rightarrow v \notin ri(\Omega_\delta)$. Isto implica que $\lim_{k \rightarrow \infty} \|x^{kj+1} - x^{kj}\| \geq \delta$ contrariando o corolário (3.3.3). \square

Teorema 3.3.6 *Seja $\{x^k\}$ a seqüência gerada pelo algoritmo (3.1). Se L_{x_o} é limitado, então vale:*

- (i) *Todo ponto de acumulação de $\{x^k\}$ é solução do problema (3.1.1)*
- (ii) $\lim_{k \rightarrow \infty} (x^{k+1} - x^k) = 0$

Demonstração:

A prova de (i) é consequência dos lemas (3.3.6) e (3.3.13). A prova de (ii) é o corolário (3.3.3). \square

3.4 O Algoritmo para (3.1.2)

O algoritmo para (3.1.2), tem essencialmente a mesma estrutura do algoritmo para (3.1.1). Apenas não precisaremos definir a função $y(x)$ de (3.2.8). E as funções $s(x)$ e $d(x)$ tomam a forma

$$s(x) = \nabla f(x)$$

$$d(x) = X^r s(x) = \nabla f^M(x)$$

Observe que $M = (\mathbb{R}_{++}^n, X^{-r})$ então $T_x M = \mathbb{R}^n$. E $\nabla f^M(x) = X^r \nabla f(x)$.

Algoritmo 3.2. Dado $x^0 \in M$, $\beta > 0$ e $0 < \delta < 1$

$$k = 0$$

Repita

$$s^k = s(x^k)$$

$$d^k = d(x^k)$$

$$\mu^k = \chi(X_k^{-1} d^k, 0)$$

$$\alpha^k = \frac{\delta}{\beta + \mu^k}$$

$$t^k = \underset{t \in [0, \alpha^k]}{\operatorname{argmin}} f(x^k - t d^k)$$

$$x^{k+1} = x^k - t^k d^k$$

$$k = k + 1$$

O algoritmo 3.2 tem a mesma estrutura do algoritmo 3.1, sendo até mais simples. Todas as provas feitas para a convergência do algoritmo 3.1, são facilmente adaptáveis para o algoritmo 3.2. Valendo assim os mesmos resultados.

Bibliografia

- [1] D. A. Bayer and J. C. Lagarias, The Nonlinear Geometry of Linear Programming I Affine and Projective Scaling Trajectories, Transactions of the American Mathematical Society, vol 314, No 2, August 1989.
- [2] D. A. Bayer and J. C. Lagarias, The Nonlinear Geometry of Linear Programming II Legendre Transform Coordinates and Central Trajectories, Transactions of the American Mathematical Society, vol 314, No 2, August 1989.
- [3] M. S. Bazaraa, H. D. Sherali and C. M. Shetty, Nonlinear Programming Theory and Algorithms, John Wiley and Sons, 1993.
- [4] M. P. do Carmo, Riemannian Geometry, Academic Press, 1991.
- [5] J. X. da Cruz Neto, Métodos Geodésicos na Programação Matemática, Tese de Doutorado, COPPE/UFRJ, Rio de Janeiro, 1995.
- [6] J. X. da Cruz Neto and P. R. Oliveira, Geodesic Methods in Riemannian Manifolds, UFRJ, Resarch Report 95 - 10, PESC/COPPE, 1995.
- [7] I. I. Dikin, Iterative Solution of problems of Linear and Quadratic Programming, Soviet Mathematics Doklady, 8: 647 - 675, 1967.
- [8] I.I. Dikin, On the Speed of a Iterative Process, Upravlyaemye Sistemi, 12: 54 - 60, 1974.
- [9] P. P. B. Eggermont, Multiplicative Iterative Algorithms for Convex Program, Linear Algebra Appl, 130, 25 - 42, 1990.
- [10] O. P. Ferreira, Programação Matemática em Variedades Riemannianas: Algoritmos Subgradiente e Ponto Proximal, Tese de Doutorado, COPPE/UFRJ, Rio de Janeiro, 1997.
- [11] D. Gabay, Minimizing a Differentiable Funtion over a Differential Manifolds, JOTA, volume 37, No 2, 1982.

- [12] C. C. Gonzaga and L. A. Carlos, A Primal Affine - Scalling Algorithms for Linearly Constrained Convex Programs, Tech. report ES - 230/90, COPPE/UFRJ, 21941 Rio de Janeiro, RJ, December, 1990.
- [13] O. Güler, Barrier Functions in Interior Point Methods, Mathematics of Operations Research, vol 21, No 4, November 1996.
- [14] K. Hoffman and R. Kunze, Álgebra Linear, Livros Técnicos e Científicos, 1979.
- [15] M. Iri, Integrability of Vector and Multivector Fields Associated with Interior Point Methods for Linear Programming, Mathematical Programming 52 (1991), 511 - 525.
- [16] A. N. Iussem, An Interior Point Multiplicative Methods for Optimization Under Positivity Constraints, Acta Mathematicae, 38, 163 - 184, 1995.
- [17] K. Kuratowski, Topology, volume 1, Academic Press, New York and London, 1966.
- [18] J. C. Lagarias, The Geometry III, Projective Legendre Transform Coordinates and Hilbert Geometry, Transactions of the American Mathematical Society, vol 320, No 1, July 1990.
- [19] E. L. Lima, Curso de Análise, vol 2, Projeto Euclides, IMPA, CNPq, 1981.
- [20] E. L. Lima, Espaços Métricos, Projeto Euclides, IMPA/CNPq, 1977.
- [21] D. G. Luenberger, The Gradient Projection Methods along Geodesics, Management Science 18 (1972), 620 - 631.
- [22] P. Mc Mullen and P. Shephard in collaboration with J. E. Reeve and A. A. Ball, Convex Polytopes and the Upper Bound Conjecture, Cambridge at the University Press, 1971.
- [23] D. C. Monteiro and T. Tsuchiya, Global Convergence of the Affine Scaling Algorithm for Convex Programming, SIAM J. Optim., vol 8, No 1, 26 - 58, 1998.
- [24] R. Monteiro, T. Tsuchiya and Waug, A Simplified Global Convergence Proof of Affine Scalling Algorithm, Annals of Operations Research, vol 47, 443 - 482, 1993.
- [25] F. Potra and Y. Ye, A Quadratically Convergent Algorithm for Solving Entropy Optimization Problems, SCAM J. Optimization, vol 3, No 4, 483 - 860, November 1993.

-
- [26] T. Rapsák and T. T. Thang, A Class of Polynomial Variable Metric Algorithms for Linear Optimization, *Mathematical Programming* 74, 319 - 331, 1996.
- [27] R. Saigal, The Primal Power Affine Scalling Method, Technical Report, No 93 - 21, Depart. of Industrial and Operations Engineering, University of Michigan, 1993.
- [28] J. Sun, A Convergence proof on Affine - Scolling Algorithm for Convex Quadratic Programming whithout Nonefeneracy Assumptions, Tecnical Report, Departament of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL 60208, USA 1990.
- [29] J. T. Tiel, *Convex Analysis*, John Wiley and Sons, 1984.
- [30] C. Udriste, *Convex Functions and Optimization Methods in Riemannian Geometry*, Kluwer Academic Publishers, 1994.
- [31] Y. Ye and E. Tse, An Extension of Karmarkar's Projective Algorithm for Convex Quadratic Programming.