**MÉTODOS DE PONTOS INTERIORES COM
MUDANÇA DE ESCALA PARA
PROGRAMAÇÃO LINEAR.**

Antonio Roberto Balbo.

Orientador: Prof. Dr. Marcos Nereu Arenales.

**Dissertação apresentada ao Instituto de Ciências
Matemáticas de São Carlos, da Universidade de São
Paulo, para obtenção do Título de Mestre em
Ciências de Computação e Matemática Computacional**

São Carlos

1991

... aos meus pais (Ladi e Avelino)
e a todos os espíritos de luz que
regem e iluminam o universo...

" A alma é divina e a obra é imperfeita.
Este padrão signala ao vento e aos céus
Que, da obra ousada, é minha a parte feita:
O por-fazer é só com Deus. "

(F. Pessoa)

AGRADECIMENTOS

Em primeiro lugar, quero deixar a minha gratidão para o professor Dr. Marcos Nereu Arenales ("Marcão"), pela colaboração, presteza, incentivo e principalmente pelo crédito ao meu esforço na realização deste trabalho.

As instituições CNPq e CAPES, e à Reitoria da UNESP, pelo auxílio financeiro na realização deste trabalho.

Ao amigo Kazuo Aoyanagi, pelo auxílio computacional.

Ao amigos, Valdemir Garcia Ferreira e Sandra de Oliveira Vendramini pelos momentos difíceis ultrapassados.

Aos professores Antonio Cesar da Costa Barros e Magda Kimico Kaibara, pelo auxílio teórico.

Ao departamento de matemática da UNESP de Bauru e aos professores deste, pela colaboração.

Aos professores Cristiano Lyra Filho e Ana Friedlander Martinez, por terem gentilmente concordado em participar como membros da banca examinadora e pelas sugestões dadas.

A todos aqueles, amigos e colegas que convivi neste percurso.

ÍNDICE.

INTRODUÇÃO.....	i
CAPÍTULO I - ELEMENTOS BÁSICOS.	
1- Preliminares.....	1
2- Problemas de Programação Linear.....	2
3- Direções factíveis e de descida.....	4
4- Limite do passo.....	5
5- Região de confiança.....	6
CAPÍTULO II - PROBLEMAS COM RESTRIÇÕES DO TIPO DESIGUALDADE:	
$Ax \leq b.$	
1- O Algoritmo Dual Afim.....	8
2- Interpretação geométrica do Algoritmo Dual Afim.....	15
CAPÍTULO III - PROBLEMAS COM RESTRIÇÕES CANALIZADAS:	
$b^- \leq Ax \leq b^+.$	
1- Algoritmo Dual Afim para restrições canalizadas.....	27
2- Um resultado numérico.....	34
CAPÍTULO IV - PROBLEMAS COM RESTRIÇÕES DE IGUALDADE.	
1- Algoritmo Primal Afim.....	36
2- Uma interpretação geométrica.....	40
3- Algoritmo dos Quadrados Mínimos Ponderados.....	42
4- Convergência dos algoritmos com restrições de igualdade..	44
5- A complexidade dos algoritmos com restrições de igualdade.....	48
6- O Algoritmo de Karmarkar.....	54
CAPÍTULO V - PROBLEMAS COM VARIÁVEIS CANALIZADAS.	
1- Algoritmo de Chandru/Kocher para variáveis canalizadas...	60
2- Algoritmos para outros problemas com variáveis canalizadas.....	66
APÊNDICE A - O MÉTODO SIMPLEX.....	71
BIBLIOGRAFIA.....	75

RESUMO.

Neste trabalho, revemos alguns métodos de pontos interiores para programação linear, com ênfase no método "dual-afim" de Adler, Karmarkar, Resende e Veiga, (1986), o qual tem se mostrado o mais promissor. Além disso, estendemos o método dual afim para explorar a estrutura de restrições canalizadas: $b^- \leq Ax \leq b^+$, como consequência imediata, com uma ligeira modificação do mesmo.

ABSTRACT.

In this work we review some interior points methods for linear programming, with emphasis on the affine scaling method of Adler, Karmarkar, Resende and Veiga, (1986), which has seemed to be the most promissor. Furthermore, we extend the affine scaling method to explore interval linear programming: $b^- \leq Ax \leq b^+$, as immediate consequence, upper bounds variables are obviously treated, with a slight modification of the affine scaling method.

INTRODUÇÃO.

Quando Karmarkar, em 1984, publicou o seu trabalho de pesquisa - " A new polynomial-time algorithm for linear programming " , denominado algoritmo " das transformações projetivas " , promoveu grande impacto na pesquisa operacional, pois abria-se novas perspectivas de campos de pesquisa.

Com relação ao algoritmo, Karmarkar provou que este tinha complexidade polinomial, afirmou que este era mais eficiente que o método Simplex de Dantzig mesmo nos piores casos de aplicação computacional e utilizou a estratégia de determinar uma solução ótima caminhando por pontos interiores à região factível.

Entretanto, implementações feitas por outros pesquisadores deste algoritmo não condiziam com a afirmação de Karmarkar em termos computacionais. Mas um algoritmo variante deste, [Adler et al, (1986)], trouxe resultados promissores, possibilitando comparações em desempenho com uma implementação avançada do método Simplex, o código MINO'S de Murtagh e Saunders, (1977).

Este algoritmo explora a estrutura de restrições no formato $Ax \leq b$, " formato dual ", se caracterizando por trabalhar com mudança de escala a cada iteração , definida pelos autores como "transformação afim". Daí a denominação de algoritmo " dual afim".

É importante observar que, até o momento, não se conseguiu demonstrar a complexidade polinomial do mesmo e provavelmente nem tenha complexidade polinomial.

Neste trabalho procuramos compilar alguns dos principais trabalhos de pesquisa de métodos de pontos interiores para programação linear.

O capítulo 1 fornece elementos básicos que serão utilizados no decorrer do trabalho.

O capítulo 2 é dedicado ao desenvolvimento algébrico do algoritmo " dual afim ", e a uma interessante interpretação geométrica do mesmo.

O capítulo 3 é reservado a uma extensão do algoritmo " dual afim " para a estrutura de restrições canalizadas: $b^- \leq Ax \leq b^+$.

A análise de algoritmos que exploram a estrutura do conjunto de restrições no formato, $Ax = b$, $x \geq 0$, " formato primal " , é feita no capítulo 4 e neste, destacamos:

- o algoritmo " primal afim ", publicado por Dikin em 1967 e mais recentemente por Cavalier e Soyster, entre outros, onde fazemos sua análise e interpretação geométrica;
- o algoritmo " dos quadrados mínimos ponderados ", publicado por Chandru e Kochar, em 1985, que nada mais é que o algoritmo anterior, visto com um outro enfoque.
- o algoritmo " afim polinomial ", publicado por Gonzaga, em 1989;
- o algoritmo de Karmarkar, publicado em 1984.

Finalmente, o capítulo 5 é dedicado a estrutura do conjunto de restrições no formato: $Ax = b$, $l \leq x \leq u$, ou seja, o problema de variáveis canalizadas, analisado entre outros por Chandru e Kochar, em 1986.

Após a publicação do método de Karmarkar, novos algoritmos de pontos interiores surgiram, como os algoritmos de " trajetória central" e os mais novos algoritmos tipo "primal-dual".

Nosso próximo passo, para dar continuidade a este trabalho de pesquisa, será analisar os algoritmos de "trajetória central", baseados em autores como Gonzaga, Vaidya e Renegar, entre outros, para futuramente analisarmos os algoritmos tipo "primal-dual", conforme [Ye, (1991)].

CAPÍTULO I

ELEMENTOS BÁSICOS.

Neste capítulo introduziremos alguns resultados básicos de álgebra linear, análise matemática e programação matemática que serão utilizados no decorrer do texto.

1 - PRELIMINARES.

Sejam o vetor $x \in \mathbb{R}^n$ e a matriz $A \in \mathbb{R}^{m \times n}$, denotamos:

x^T : vetor transposto de x ;

$$\|x\| = \sqrt{x^T x} = \left(\sum_{i=1}^n x_i^2 \right)^{1/2} : \text{norma euclídeana de } x;$$

$$\|x\|_1 = \sum_{i=1}^n |x_i|;$$

$$\|x\|_\infty = \max_{i=1..n} |x_i|;$$

$A^T \in \mathbb{R}^{n \times m}$: matriz transposta de A .

A matriz $A \in \mathbb{R}^{m \times n}$, será confundida com a transformação linear $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$, $T(x) = Ax$, e denotamos:

$N(A) = \{x \in \mathbb{R}^n \text{ tal que } Ax = 0\}$: o sub-espaço núcleo de A ;

$\text{Im}(A^T) = \{y \in \mathbb{R}^n \text{ tal que } \exists \psi \in \mathbb{R}^m \text{ e } A^T \psi = y\}$: o sub-espaço imagem de A^T .

Note que $N(A)$ e $\text{Im}(A^T)$ são sub-espaços do $\mathbb{R}^n$. Descrevemos abaixo os principais resultados entre $N(A)$ e $\text{Im}(A^T)$.

1- São ortogonais os sub-espaços $N(A)$ e $\text{Im}(A^T)$, ou seja, $\forall x \in N(A)$ e $\forall y \in \text{Im}(A^T)$ então $x^T y = 0$.

2- A dimensão ($\dim$) dos sub-espaços $N(A)$ e $\text{Im}(A^T)$ é:

$$\dim(N(A)) = n-m \quad \text{e} \quad \dim(\text{Im}(A^T)) = m.$$

3- $\mathbb{R}^n$ é a soma direta dos sub-espaços $N(A)$ e $\text{Im}(A^T)$, ou seja, $\mathbb{R}^n = N(A) \oplus \text{Im}(A^T)$, logo $\forall x \in \mathbb{R}^n$, então x pode ser decomposto em $x = x_p + x_{-p}$ onde $x_p \in N(A)$ e $x_{-p} \in \text{Im}(A^T)$.

As proposições dadas a seguir, serão úteis no capítulo 4 quando definirmos a direção de busca para os algoritmos primais.

Proposição 1: Seja $A \in \mathbb{R}^{m \times n}$ e $\text{posto}(A) = m$, então:

A matriz $(AA^T) \in \mathbb{R}^{m \times m}$ é não singular.

Proposição 2: $\forall x \in \mathbb{R}^n$, $x \neq 0$ e considerando a decomposição $x = x_p + x_p^-$ onde $x_p \in N(A)$ e $x_p^- \in \text{Im}(A^T)$, então a expressão algébrica de x_p e x_p^- é dada por:

$$x_p = (I - A^T(A A^T)^{-1}A)x \text{ e}$$

$$x_p^- = (A^T(A A^T)^{-1}A)x$$

O elemento x_p é chamado projeção ortogonal de x no núcleo de A .

2 - Problemas de Programação Linear.

Consideremos o problema de programação linear:

$$\begin{aligned} &\text{minimizar } c^T x \\ &\text{sujeito a } \begin{cases} Ax = b \\ x \geq 0 \end{cases} \end{aligned} \quad (1)$$

onde $A \in \mathbb{R}^{m \times n}$; $\text{posto}(A) = m$ e seu problema dual é dado por:

$$\begin{aligned} &\text{maximizar } b^T y \\ &\text{sujeito a } A^T y \leq c. \end{aligned} \quad (2)$$

Consideremos o problema de programação linear:

$$\begin{aligned} &\text{minimizar } c^T x \\ &\text{sujeito a } Ax \leq b \end{aligned} \quad (3)$$

onde $A \in \mathbb{R}^{m \times n}$ e $\text{posto}(A) = n$, $m \geq n$, que é equivalente ao problema:

$$\begin{aligned} &\text{minimizar } c^T x \\ &\text{sujeito a } \begin{cases} Ax + z = b \\ z \geq 0 ; \quad z \in \mathbb{R}^m. \end{cases} \end{aligned} \quad (4)$$

A teoria de programação linear adota o problema (1) como formato padrão, referido aqui como problema em "formato primal".

Por se caracterizar pelas restrições de desigualdade e ter uma certa "semelhança" com (2), o problema (3) será referido como problema em "formato dual".

Os algoritmos, a serem vistos, serão desenvolvidos de acordo com a especificação do formato do problema.

Uma solução é dita "factível" se satisfizer as restrições do problema de programação linear e será ótima se minimizar (maximizar) a função objetivo do problema.

Encontramos na programação linear um teorema importante que relaciona as soluções ótimas de (1) e de (2) respectivamente:

Teorema 1: (folgas complementares).

(x^*, y^*) são soluções ótimas em (1) e (2) respectivamente, se e somente se as três condições abaixo são satisfeitas:

- i) x^* é factível em (1), isto é, $Ax^* = b$ e $x^* \geq 0$.
- ii) y^* é factível em (2), isto é, $A^T y^* \leq c$.
- iii) $(x^*) (c - A^T y^*) = 0$ (condição das folgas complementares).

Este teorema nos será útil nos capítulos 2, 4 e 5.

É preciso que especifiquemos também o conjunto viável para cada problema, já que direções de busca e soluções factíveis estão relacionados com este conjunto.

O conjunto viável do problema (1) denotaremos por S:

$$S = \{x \in R^n \text{ tal que } Ax=b \text{ e } x \geq 0\}.$$

O conjunto viável do problema (3) por P:

$$P = \{x \in R^n \text{ tal que } Ax \leq b\}.$$

O conjunto viável do problema (4) denotaremos por P':

$$P' = \{x \in R^n; z \in R_+^m \text{ tal que } Ax + z = b\}.$$

Como os algoritmos a serem vistos trabalham com pontos interiores à região viável, definamos o conjunto de pontos interiores de S e P, respectivamente. O conjunto dos pontos interiores de S é definido por:

$$\text{int}(S) = \{x \in R^n \text{ tal que } Ax=b \text{ e } x > 0\}.$$

O conjunto dos pontos interiores de P é definido por:

$$\text{int}(P) = \{x \in R \text{ tal que } Ax < b\}.$$

3 - Direções factíveis e de descida.

Definição 1: Dizemos que um vetor $d \in \mathbb{R}^n$ é uma direção factível para $x \in S$ ($x \in P$), se existe $\rho > 0$ tal que $x + \varepsilon d \in S$, $\forall \varepsilon \in [0, \rho]$. ($x + \varepsilon d \in P$, $\forall \varepsilon \in [0, \rho]$)

A seguir caracterizaremos direções factíveis para (1) e (4)..

Proposição 4: Se uma direção $d \in \mathbb{R}^n$, é factível para um ponto $\tilde{x} \in S$ do problema (1), então $d \in N(A)$.

Prova:

d : direção factível $\Rightarrow \exists \varepsilon > 0$ tal que $A(\tilde{x} + \varepsilon d) = b$.

Desde que $A\tilde{x} = b$ e $\varepsilon > 0 \Rightarrow Ad = 0$, (com $\varepsilon > 0$, suficientemente pequeno). $\square$

Corolário 4.1: Se o vetor $d' \in \mathbb{R}^{m+n}$; (problema (4))

$$d' = \begin{bmatrix} d_x \\ d_z \end{bmatrix}$$

onde $d_x \in \mathbb{R}^n$ e $d_z \in \mathbb{R}^m$, é uma direção factível para um ponto $\tilde{x}_z \in \mathbb{R}^{m+n}$;

$$\tilde{x}_z = \begin{bmatrix} \tilde{x} \\ \tilde{z} \end{bmatrix}, \quad \tilde{z} \geq 0,$$

então $d' \in N(A, I)$ ($I \in \mathbb{R}^{m \times m}$ é a matriz identidade).

No apêndice A, caracterizaremos as direções factíveis adotadas pelo método Simplex.

A seguir definiremos a condição para que uma direção factível seja de descida. Para isto é preciso que façamos a seguinte definição:

Definição 2: (derivada direcional).

Seja $S \subset \mathbb{R}^n$; S aberto; $f: S \rightarrow \mathbb{R}$. Dado um ponto $\tilde{x} + \lambda h \in S$, define-se a derivada direcional na direção h por:

$$\nabla_h f(\tilde{x}) = \lim_{\lambda \rightarrow 0} \frac{f(\tilde{x} + \lambda h) - f(\tilde{x})}{\lambda}$$

Além disso é válido a seguinte proposição:

Proposição 5: $\nabla_h f(\tilde{x}) = (\nabla f(\tilde{x}))^T h$, onde $\nabla f(\tilde{x})$ é o vetor gradiente de f em $\tilde{x}$, formado pelas derivadas parciais : $\frac{\partial f(\tilde{x})}{\partial \tilde{x}_j}$

Definição 3: Dizemos que uma direção factível $h \in R^n$ é de descida para uma função $f: R^n \rightarrow R$ em $\tilde{x}$ se, e somente se, $\exists \delta > 0$ tal que $f(\tilde{x} + \epsilon h) \leq f(\tilde{x})$, $\forall \epsilon \in [0, \delta]$.

Proposição 6: Se $f(x) = c^T x$, uma direção h é de descida se e somente se:

$$\nabla_h f(x) = c^T h < 0, \quad \forall x \in R^n.$$

4 - Limite do passo.

Quando perturbamos x em uma direção factível d gerando um novo ponto $x + \epsilon d$, $\epsilon > 0$, dizemos que ϵ é o tamanho do passo na direção d definido na seguinte proposição:

Proposição 8: Sejam $\tilde{x} \in S$, e $d \geq 0$, então:
 $\tilde{x} + \epsilon d$ é viável, $\forall \epsilon > 0$, caso contrário, o maior valor de ϵ tal que $\tilde{x} + \epsilon d \in S$ é dado por:

$$\epsilon = \min_{i=1..n} \{-\tilde{x}_i/d_i \text{ tal que } d_i < 0\}.$$

Considerando o problema (4), se $d_z \geq 0$, então $\tilde{z} + \epsilon d_z$ é factível $\forall \epsilon > 0$, caso contrário o maior valor de ϵ tal que $\tilde{z} + \epsilon d_z \in P$ é dado por:

$$\epsilon = \min_{i=1..m} \{-\tilde{z}_i/d_{z_i} \text{ tal que } d_{z_i} < 0\},$$

Note que estes valores para o passo fazem com que os novos pontos estejam sobre a fronteira da região viável.

Os algoritmos de pontos interiores utilizarão a estratégia heurística de caminhar $\alpha \epsilon$, com $\alpha \in (0,1)$, para garantir que o novo ponto seja interior à região viável.

5 - Região de confiança.

Dado um ponto $\tilde{x} \in \mathbb{R}^n$, definimos como região de confiança em $\tilde{x}$ uma vizinhança $V(\tilde{x}) \subset \mathbb{R}^n$ e estaremos interessados em resolver:

$$\begin{aligned} &\text{minimizar } c^T x \\ &\text{sujeito a } x \in V(\tilde{x}) \end{aligned} \quad (5)$$

Se $V(\tilde{x})$ é tal que:

$$V(\tilde{x}) = B_\varepsilon(\tilde{x}) = \{x \in \mathbb{R}^n \text{ tal que } \|x - \tilde{x}\| \leq \varepsilon\},$$

(bola centrada em $\tilde{x}$ com raio ε), $\varepsilon > 0$, então uma solução para (5) a partir de $\tilde{x}$, já bem conhecida, será:

$$x = \tilde{x} + \varepsilon d = \tilde{x} + \varepsilon (-c/\|c\|), \quad (6)$$

o que pode ser interpretado como sendo o oposto do vetor gradiente a melhor direção para se perturbar $\tilde{x}$ quando a região de confiança é definida por uma bola.

Se considerarmos um novo problema:

$$\begin{aligned} &\text{minimizar } c^T x \\ &\text{sujeito a } x \in V(\tilde{x}) = B_\varepsilon(\tilde{x}) \cap \{Ax = b\}, \end{aligned} \quad (7)$$

a solução (6) pode ser reconsiderada, desde que tomemos $d \in N(A)$ e assim:

$$x = \tilde{x} + \varepsilon (-c_p/\|c_p\|), \quad (8)$$

onde,

$$c_p = (I - A^T(AA^T)^{-1}A)c,$$

é o vetor gradiente projetado em $N(A)$.

Agora, se tivermos uma vizinhança $V(\tilde{x})$ definida por um elipsóide muito particular:

$$V(\tilde{x}) = E(\tilde{x}) = \{x \in \mathbb{R}^n \text{ tal que } \|H^{-1}(x - \tilde{x})\| = \varepsilon\},$$

onde $H = \text{diag}(\eta_i)$, $\eta_i > 0$, $i = 1, \dots, n$; basta que efetuemos uma mudança de escala na variável x ,

$$y = H^{-1}x, \quad (y_i = \eta_i^{-1}x_i), \quad \tilde{y} = H^{-1}\tilde{x},$$

e o problema (5) será equivalente a:

$$\begin{aligned} &\text{minimizar } c^T y \\ &\text{sujeito a } \|y - \tilde{y}\| = \epsilon, \end{aligned}$$

cujas solução única é:

$$y = \tilde{y} + \epsilon \left(- (Hc) / (\|Hc\|) \right).$$

Assim, a direção oposta ao gradiente precedida pela mudança de escala, fornece a melhor direção para se perturbar $\tilde{x}$ quando a vizinhança é definida por $E(\tilde{x})$.

Se reconsiderarmos:

$$\begin{aligned} &\text{minimizar } c^T x \\ &\text{sujeito a } x \in V(\tilde{x}) = E(\tilde{x}) \cap \{Ax = b\}, \end{aligned}$$

este problema é transformado pela mudança de escala em:

$$\begin{aligned} &\text{minimizar } c^T Hy \\ &\text{sujeito a } \begin{cases} AHy = b \\ \|y - \tilde{y}\| = \epsilon, \end{cases} \end{aligned}$$

cujas solução direta de (8) é dada por:

$$y = \tilde{y} + \epsilon \left(-(Hc_p) / (\|Hc_p\|) \right),$$

onde Hc_p é o vetor gradiente escalado e projetado em $N(AH)$.

Observe que os resultados anteriores podem ser estendidos para $V(\tilde{x}) \cap \{Ax \leq b\}$, desde que $Ax \leq b$ é equivalente a $Ax + z = b$, $z \geq 0$. Isto será visto com maiores detalhes no capítulo 2, seção 2.

A estratégia de se fazer mudança de escala a cada iteração será usual nos métodos a serem vistos, observando que: se tomarmos $\eta_1 = (\tilde{x}_1)$; $\tilde{x}_1 > 0$, então $\tilde{y} = H^{-1}\tilde{x} = e$, onde "e" denotará o vetor de coordenadas unitárias, importante para a análise destes métodos e $(y - e)^T(y - e)$ será a bola de centro "e", resultados muito utilizados a partir do capítulo 2.

CAPÍTULO II

PROBLEMAS COM RESTRIÇÕES DO TIPO DESIGUALDADE: $Ax \leq b$.

1 - O Algoritmo dual afim.

Neste capítulo revemos o método apresentado por Adler, Kar-markar, Resende e Veiga, inicialmente seguindo um procedimento basicamente algébrico, para depois deduzí-lo com uma abordagem mais geométrica.

1.1 - Preliminares.

Consideremos o programa linear na formatação dual:

$$\begin{array}{ll} \text{minimizar} & c^T x \\ \text{sujeito a} & Ax \leq b; \end{array} \quad (1)$$

onde $A \in \mathbb{R}^{m \times n}$, $\text{posto}(A) = n$, $m \geq n$, ou seja, as colunas da matriz A são linearmente independentes.

Introduzindo-se as variáveis de folga, temos:

$$\begin{array}{ll} \text{minimizar} & c^T x \\ \text{sujeito a} & \begin{cases} Ax + z = b \\ z \geq 0. \end{cases} \end{array} \quad (2)$$

Seja $\bar{x}$ um ponto interior factível para (1), ou seja, $A\bar{x} < b$, logo $\bar{z} = b - A\bar{x} > 0$ é um ponto factível para (2). Além disso, pela proposição 5, capítulo 1, o conjunto das direções factíveis de (2) é determinado pela resolução do seguinte sistema:

$$Ad_x + d_z = 0. \quad (3)$$

1.2 - A mudança de escala.

Considere a matriz diagonal formada pelo vetor das variáveis de folgas $\bar{z}$:

$$\bar{Z} = \begin{bmatrix} \bar{z}_1 & & 0 \\ & \ddots & \\ 0 & & \bar{z}_m \end{bmatrix}$$

A matriz $\bar{Z}$ é usada para definirmos a seguinte mudança de escala (mudança de variável):

$$w \in \mathbb{R}^m \text{ tal que } w = \bar{Z}^{-1}z.$$

Esta mudança de variável é particularmente interessante pois, $\bar{z}_1$ próximo a zero é transformado em 1, ou seja, a solução factível $(\bar{x}, \bar{z})$ é transformado em $(\bar{x}, e)$. Basta observarmos que:

$$\bar{w} = \bar{Z}^{-1}\bar{z} = e.$$

Assim teremos o seguinte problema equivalente a (2):

$$\begin{aligned} &\text{minimizar } c^T x \\ &\text{sujeito a } \begin{cases} Ax + \bar{Z}w = b \\ w \geq 0 \end{cases} \end{aligned} \quad (4)$$

1.3 - Determinação da direção de busca.

Para melhorarmos uma solução corrente $\bar{x}$, o problema (4) terá papel central na determinação da direção adotada. Para isto, introduziremos agora as direções factíveis a serem adotadas, as quais denotamos por:

$$d = \begin{bmatrix} d_x \\ d_w \end{bmatrix}$$

Note que uma direção factível para o problema (4) deve pertencer ao núcleo de $(A, \bar{Z})$, ou seja,

$$Ad_x + \bar{Z} d_w = 0 \Rightarrow d_w = -\bar{Z}^{-1}Ad_x. \quad (5)$$

Além disso, estamos presos ao problema (1) e uma possível direção factível d_x para $\bar{x}$ será interessante se for uma direção de descida, ou seja, pela proposição 6, capítulo 1, $c^T d_x < 0$. (6)

Note de (5) que d_w a ser escolhida adiante deve pertencer ao espaço imagem de $[\bar{Z}^{-1}A]$, caso contrário uma direção infactível seria gerada (a estratégia normal dos métodos de pontos interiores é a de projetar o gradiente da função objetivo com respeito às variáveis que sofreram mudança de escala no espaço das direções factíveis). Admitindo que $d_w \in \text{Im}[\bar{Z}^{-1}A]$, podemos expressar d_x em função de d_w . Antes porém enunciamos a seguinte proposição:

Proposição 1: A matriz $[A^T \bar{Z}^{-2} A] \in R^{n \times n}$ é simétrica e definida positiva.

Prova:

A matriz $[A^T \bar{Z}^{-2} A]$ é simétrica pois $[A^T \bar{Z}^{-2} A]^T = [A^T \bar{Z}^{-2} A]$.

Seja $v \in R^n$ não nulo, então

$v^T [A^T \bar{Z}^{-2} A] v = (v^T [A^T \bar{Z}^{-1}]) \cdot ([\bar{Z}^{-1} A] v) = \| [\bar{Z}^{-1} A] v \|^2 > 0$, pois $\bar{Z}^{-1} A v$ é não nulo, uma vez que é combinação não nula de vetores l.i., desde que $\text{posto}(A) = n$. $\square$

Logo, podemos expressar d_x em função de d_w usando-se a seguinte estratégia:

$$\begin{aligned} \bar{Z}^{-1} A d_x + d_w &= 0 \Rightarrow \\ A^T \bar{Z}^{-1} (\bar{Z}^{-1} A d_x + d_w) &= 0 \Rightarrow \\ d_x &= - [A^T \bar{Z}^{-2} A]^{-1} A^T \bar{Z}^{-1} d_w. \end{aligned} \quad (7)$$

De (7), obtemos:

$$c^T d_x = - \{c [A^T \bar{Z}^{-2} A]^{-1} A^T \bar{Z}^{-1}\} d_w = - \{\bar{Z}^{-1} A [A^T \bar{Z}^{-2} A]^{-1} c\}^T d_w$$

e tem o valor mais negativo, quando

$$d_w = \bar{Z}^{-1} A [A^T \bar{Z}^{-2} A]^{-1} c \quad (8)$$

Note que (8) é uma escolha possível para d_w pois pertence ao espaço imagem de $[\bar{Z}^{-1} A]$. Substituindo (8) em (7) obtemos:

$$d_x = - [A^T \bar{Z}^{-2} A]^{-1} c. \quad (9)$$

Desta forma, (5) está satisfeito.

Mostremos que d_x definida em (9) é uma direção de descida, ou seja, satisfaz (6):

$c^T d_x = -c^T [A^T \bar{Z}^{-2} A]^{-1} c < 0$, pois, pela proposição 1, $[A^T \bar{Z}^{-2} A]$ é definida positiva.

Já que a cada iteração dependemos do ponto $\bar{z}$ para determinarmos $\bar{x}$, o novo ponto $\bar{z} + \varepsilon d_z$, $\varepsilon > 0$, é obtido desde que tomemos $d_z \in R^m$ tal que $d_z = -A d_x$, diretamente de (3).

O tamanho do passo, $\varepsilon > 0$, está definido na seção seguinte e é consequência direta das condições de fronteira do problema (2), analisadas na seção 4, capítulo 1.

1.4 - Determinação do passo.

A partir dos pontos $\bar{x}$ e $\bar{z}$, factíveis para (2), pré-determinados, a nova solução a ser adotada é dada por:

$$\begin{bmatrix} \bar{x} \\ \bar{z} \end{bmatrix} + \varepsilon \begin{bmatrix} d_x \\ d_z \end{bmatrix}; \quad \varepsilon > 0,$$

e $d_z = -Ad_x$ (direção factível para o problema (2)).

Para preservarmos a factibilidade do novo ponto a ser determinado, devemos ter:

$\bar{z} + \varepsilon d_z > 0$, já que $\bar{x} + \varepsilon d_x$ é irrestrito, então,

$$\varepsilon < - [\bar{z}_1 / (d_z)_1] \text{ se } (d_z)_1 < 0.$$

Portanto calculamos o valor limite:

$$\bar{\varepsilon} = \min_{i=1..m} \{ - [\bar{z}_i / (d_z)_i] \text{ se } (d_z)_i < 0 \}, \text{ seguindo a}$$

proposição 7, capítulo 1.

Uma heurística adotada para o tamanho do passo é a de caminharmos $\alpha\bar{\varepsilon}$, com $0 < \alpha < 1$, para garantir que o novo ponto seja interior à região viável.

1.5 - Algoritmo.

Após as considerações anteriores, podemos enunciar o seguinte algoritmo:

(FASE 1)

Encontre x^0 tal que $Ax^0 < b$ e $z^0 = b - Ax^0$, $k = 0$.

Escolha $\alpha \in \mathbb{R}$ tal que $0 < \alpha < 1$.

(FASE 2)

Enquanto { critério de parada não satisfeito } faça:

Calcule a direção:

$$d_x = -[A^T Z_k^{-2} A]^{-1} c, \text{ onde } Z_k = \text{diag}[z_1^k, \dots, z_m^k]$$

$$d_z = -A d_x.$$

Calcule o passo:

$$\bar{\epsilon} = \min_{i=1..m} \{ -z_i^k / (d_z)_i \mid \text{tal que } (d_z)_i < 0 \};$$

$$\epsilon = \alpha \bar{\epsilon}.$$

Atualize a solução:

$$x^{k+1} = x^k + \epsilon d_x;$$

$$z^{k+1} = z^k + \epsilon d_z \quad \text{ou} \quad z^{k+1} = b - Ax^{k+1};$$

$$k \leftarrow k + 1.$$

1.6 - Obtenção de um ponto inicial (FASE I).

Um processo para determinar uma solução inicial x^0 discutido em Adler e outros é dado a seguir e caracteriza a fase 1 do algoritmo acima.

Escolha $x^0 = (\|b\| / \|Ac\|)c$.

Se $z^0 = b - Ax^0 > 0$, x^0 é um ponto interior inicial. Caso contrário uma variável artificial x_a^0 , onde

$$x_a^0 = -2 * \min \{ z_i^0 ; i = 1 \dots m \}$$

é adicionada e (x^0, x_a^0) é um ponto interior da seguinte fase I do problema abaixo:

$$\text{minimizar } c^T x - M x_a$$

$$\text{sujeito a } Ax - e x_a \leq b;$$

M suficientemente grande

A fase I termina se para uma iteração k , $x_a^k < 0$, assim $Ax^0 < b$ e x^0 é um ponto inicial interior.

Se o algoritmo satisfizer o critério de término regular (seção 1.7, abaixo) e $x_a^k > 0$ então o problema inicial é infactível.

Se $x_a^k \rightarrow 0$ então $x^k \rightarrow x^*$, e a solução ótima é encontrada diretamente na fase I.

1.7 - Critério de parada.

Um critério de parada utilizado em [Adler et al, (1986)], útil ao algoritmo e apenas mencionado acima, foi:

"Se

$$|c^T x^{k+1} - c^T x^k| / \max\{1, |c^T x^k|\} < \text{tol}$$

então o processo iterativo é interrompido, onde tol é uma tolerância positiva e pequena, escolhida a priori."

1.8 - Aplicação ao formato primal.

O algoritmo desenvolvido no formato dual na seção 1.5, pode ser aplicado a problemas no formato primal, (problemas com restrições de igualdade e variáveis não negativas) semelhante à forma padrão, usualmente preferida.

Uma aplicação a problemas no formato primal, conforme [Adler et al, (1986)], pode ser obtida da seguinte maneira:

O problema dual de (1) é:

$$\begin{array}{ll} \text{maximizar} & b^T y \\ \text{sujeito a} & \begin{cases} A^T y = c \\ y \geq 0 \end{cases} \end{array} \quad (10)$$

Podemos determinar uma solução $\bar{y}$ para (10) explorando diretamente a condição das folgas complementares do teorema 1, capítulo 1:

$$\begin{cases} Ax \leq b \\ A^T y = c \\ y \geq 0; y^T(b - Ax) = y^T z = 0 \end{cases} \Leftrightarrow \sum_{i=1}^m y_i z_i = 0.$$

Se definirmos $\Gamma_{yz} = Zy = (z_1 y_1, \dots, z_m y_m)^T$, então as folgas estão satisfeitas se $\|\Gamma_{yz}\| = 0$. O vetor Γ_{yz} é chamado de desvio das folgas complementares.

Para uma iteração corrente, com as soluções $\bar{x}$ e $\bar{z}$ já determinadas, uma tentativa de solução $\bar{y}$ de (9) pode ser obtida explorando-se o seguinte problema, analisado em [Chandru e Kochar, (1985)]:

$$\begin{cases} \text{minimizar } \|\Gamma_{yz}\|^2 = \sum_{i=1}^m (z_i y_i)^2 \\ \text{sujeito a } A^T \bar{y} = c; \end{cases}$$

e cuja solução é $\bar{y} = \bar{Z}^{-2} A(A^T \bar{Z}^{-2} A)^{-1} c$, ou seja minimiza o desvio das folgas com respeito a $\bar{y}$ relaxando-se a restrição de não negatividade $y \geq 0$.

Esta solução pode ser determinada em função da direção d_z :

$$\bar{y} = \bar{Z}^{-2} d_z;$$

ou seja, é a direção de busca de primeira ordem do algoritmo anterior, escalada por $\bar{Z}^{-2}$, [Todd e Burrell, (1986)].

Se para $\bar{y}$, as condições das folgas estão satisfeitas, isto é,

$$\begin{cases} \bar{z}_i \bar{y}_i = 0 \\ A^T \bar{y} = c; \end{cases}$$

temos uma possível tentativa de solução para (10), resta-nos saber se, a violação da restrição $y \geq 0$ afeta a otimalidade do problema.

A afirmação dada a seguir, extraída em [Adler et al, (1989)], trata da convergência do processo iterativo e foi demonstrada em [Chandru e Kochar, (1985)], a ser vista no capítulo 4 :

Teorema 1: " Assumindo a não degeneração do problema, se o problema (1) possui solução ótima, então existe uma subsequência de soluções tentativas $\bar{y}_k$ que converge para a solução ótima de (10). "

Sabemos que, se numa iteração corrente k , ocorrer:

$$\begin{cases} A^T y^k = c \\ y_i^k z_i^k = 0 \\ y^k \geq 0; \end{cases}$$

então x^k e y^k são ótimos respectivamente para (1) e (10), pelo teorema das folgas complementares.

Já que em métodos de pontos interiores somente soluções aproximadas são avaliadas, Adler e outros sugeriram o seguinte critério de parada:

"Se numa iteração corrente k ocorrer;

$$\begin{cases} y_i^k \geq -\varepsilon_1 \|y^k\| \\ |y_i^k z_i^k| \leq \varepsilon_1 \|y^k\| \|z^k\|; \end{cases} \quad i = 1, \dots, m$$

onde ε é uma pequena tolerância positiva, então o processo iterativo é interrompido e a solução corrente é assumida ótima".

2 - Interpretação geométrica do algoritmo dual afim.

Nesta seção deduzimos o método da seção anterior com uma motivação basicamente geométrica. Esta visão do método permitirá facilmente estendê-lo para problemas cujas restrições são canalizadas, objeto do capítulo 3.

2.1 - Preliminares.

Considere o problema (1) e o conjunto P :

$$P = \{x \in R^n \text{ tal que } Ax \leq b\}.$$

Considere a transformação afim, $T: P \rightarrow R_+^m$, que associa a cada $x \in P$ às variáveis de folgas das restrições de (1):

$$z = T(x) = b - Ax.$$

A seguir apresentamos algumas propriedades da transformação T .

P1: T é injetiva.

Prova: Se $T(x^1) = T(x^2) \Rightarrow b - Ax^1 = b - Ax^2 \Rightarrow A(x^1 - x^2) = 0$.

Como posto $(A) = n$, segue que $x^1 = x^2$. $\square$

Note que a transformação T não é sobrejetora (supondo $m > n$), isto é, existe $z \in R_+^m$ tal que $z \neq T(x)$, $\forall x \in R^n$. Entretanto, se restringirmos o contra-domínio de T temos:

P2: $T: P \rightarrow \text{Im}(T)$ é bijetora.

Assim, para qualquer que seja x factível, existe um único $z \in \text{Im}(T)$ tal que $T(x) = z$ e vice-versa, para qualquer que seja $z \in \text{Im}(T)$, existe um único x factível tal que $z = T(x)$.

P3: Se estendermos o domínio e contra-domínio de T : $T: R^n \rightarrow R^m$, então $\text{Im}(T) = \text{Im}(A) + \{b\}$, onde $\text{Im}(A)$ é o sub-espaco gerado pelas colunas da matriz A .

Ou seja, $\text{Im}(T)$, chamado sub-espaco afim, é a translação do sub-espaco $\text{Im}(A)$ pelo vetor b . A prova desta propriedade segue diretamente da definição de T .

P4: Sejam $z^1 = T(x^1)$ e $z^2 = T(x^2)$. Então $T(\lambda x^1 + (1 - \lambda)x^2) = \lambda z^1 + (1 - \lambda)z^2$.

Prova: Suponha que $x = \lambda x^1 + (1 - \lambda)x^2$, então

$$\begin{aligned} z &= T(x) = b - A(\lambda x^1 + (1 - \lambda)x^2) = \lambda b + (1 - \lambda)b - \lambda Ax^1 - (1 - \lambda)Ax^2 \\ &= \lambda(b - Ax^1) + (1 - \lambda)(b - Ax^2) \\ &= \lambda z^1 + (1 - \lambda)z^2. \end{aligned}$$

Suponha agora que $z = T(x) = \lambda z^1 + (1 - \lambda)z^2$, então

$$\begin{aligned} T(x) &= \lambda z^1 + (1 - \lambda)z^2 \\ &= \lambda T(x^1) + (1 - \lambda)T(x^2) \\ &= \lambda(b - Ax^1) + (1 - \lambda)(b - Ax^2) \\ &= T(\lambda x^1 + (1 - \lambda)x^2), \end{aligned}$$

e devido a T ser injetiva segue que

$$x = \lambda x^1 + (1 - \lambda)x^2. \quad \square$$

Definição 1: Se $x + \lambda r \in P$, $\forall \lambda \geq 0$, com x factível, dizemos que r é um raio da região de factibilidade do problema (1).

P5: Se r é um raio de P então $-Ar$ é um raio de $T(P)$ e $Ar \leq 0$.

Prova: $z = T(x + \lambda r) = b - A(x + \lambda r) = T(x) - \lambda Ar \in T(P)$,
 $\forall \lambda \geq 0$ e portanto $-Ar$ é um raio de $T(P)$. Além disso, desde
 que r é um raio de P segue que $A(x + \lambda r) \leq b$, $\forall \lambda \geq 0$. Assim
 $\lambda Ar \leq b - Ax$, $\forall \lambda \geq 0$. Logo $Ar \leq 0$. $\square$

Corolário das propriedades:

i) Pontos extremos e arestas de P , correspondem respectivamente
 a pontos extremos e arestas de $T(P)$.

2.2 - Interpretação geométrica da transformação T .

Considere $T: P \subset \mathbb{R}^2 \rightarrow \mathbb{R}_+^3$ (conforme figura 1)

figura 1.

Retornemos ao problema (1).

Suponha conhecido um ponto $\bar{x} \in \text{int}(P)$, e seja $V(\bar{x})$ uma vizinhança de $\bar{x}$ com $V(\bar{x}) \subset P$, tal que seja fácil resolver o seguinte problema:

$$\begin{cases} \text{minimizar} & c^T(\bar{x} + u) \\ \text{sujeito a} & \bar{x} + u \in V(\bar{x}). \end{cases} \quad (11)$$

Obs: Se $\bar{x}$ estivesse na fronteira de P , $\bar{x}$ não estaria no interior de $V(\bar{x}) \subset P$

Seja $\bar{u}$ uma solução do problema (11), então, desde que $\bar{x}$ é interior a P, segue que:

$$c^T(\bar{x} + \bar{u}) < c^T\bar{x} \Rightarrow c^T\bar{u} < 0,$$

ou seja, $\bar{u}$ é uma direção factível e de descida.

A factibilidade da direção segue do fato de que P é convexo e $\bar{x} + \varepsilon\bar{u} \in P$, $0 \leq \varepsilon \leq 1$.

Obs: Se $V(\bar{x})$ fosse uma bola então a solução do problema (11), $\bar{u}$, seria um múltiplo de $-c$, ou seja, o problema (11) indicaria a direção oposta do gradiente, independentemente da proximidade de $\bar{x}$ com a fronteira de P, conforme ilustra a figura 2.

figura 2.

Se $V(\bar{x}) = \{ \bar{x} + u \text{ tal que } \|u\| \leq \varepsilon \}$ então $\bar{u} = -\varepsilon c / \|c\|$

Assim, para a sequência: $x^{k+1} = x^k + \lambda_k u^k$, com $x^0 = \bar{x}$ e u^k solução do problema (11), usando $V(x^k)$ como bola e $\lambda_k > 0$, sendo o passo na direção de u^k , podemos ter $\{u^k\} \rightarrow 0$ sem que a sequência $\{x^k\}$ convirja para a solução ótima.

Como vimos, escolher $V(\bar{x})$ como uma bola, pode não ser uma boa escolha, pois, a proximidade da fronteira numa direção, obriga que todas as outras direções sejam também prejudicadas.

Uma escolha aparentemente mais razoável para $V(\bar{x})$ seria uma elipse, a qual, permite "diminuir" $V(\bar{x})$ numa determinada direção sem prejuízo de outras e mantendo o problema (11) fácil de ser resolvido, conforme ilustra a figura 3.

A proximidade de $\bar{x}$ com a fronteira numa direção, obrigando pequenos movimentos nesta direção, não inibe movimentos maiores em outras direções.

figura 3

Para explicitarmos $V(x)$ da forma anterior, recorremos ao espaço das variáveis de folga.

Seja $\bar{z} = T(\bar{x})$. Pela hipótese de $\bar{x} \in \text{int}(P)$, temos $\bar{z} > 0$.

Um elipsóide muito simples, centrado em $\bar{z}$, que "pune" apenas as direções dos eixos coordenados para os quais as componentes de $\bar{z}$ sejam próximas a zero (note que $\bar{z}_1$ próximo a zero, indica que $\bar{x}$ se aproxima da restrição $a_1 x \leq b_1$) e inteiramente contido em

R_+^m é :

$$E_z = \{ z \in R^m \text{ tal que } (z - \bar{z})^T \bar{z}^{-2} (z - \bar{z}) = 1 \}$$

com $\bar{z} = \text{diag}(\bar{z}_1, \dots, \bar{z}_m)$.

$$\left[E_z = \{ z \in R^m : \left[\frac{(z_1 - \bar{z}_1)}{\bar{z}_1} \right]^2 + \dots + \left[\frac{(z_m - \bar{z}_m)}{\bar{z}_m} \right]^2 = 1 \} \right]$$

Assim, $\forall z \in E_z$ temos que $z \in R_+^m$. Isto não implica contudo que z pertencente a E_z seja uma folga factível, isto é, pode acontecer que não exista $x \in R^n / z = T(x)$.

Definimos uma vizinhança de $\bar{z}$ por:

$$V(\bar{z}) = E_z \cap \text{Im}(T).$$

Uma vizinhança $V(\bar{x})$ inteiramente contida em P pode ser obtida tomando-se a imagem inversa de $V(\bar{z})$ pela transformação T :

$$V(\bar{x}) = T^{-1}(V(\bar{z})).$$

Para isto basta definir $z = T(x)$ em E_z :

$$(T(x) - T(\bar{x}))^T \bar{Z}^{-2} (T(x) - T(\bar{x})) = 1.$$

Como $T(x) - T(\bar{x}) = (b - Ax) - (b - A\bar{x}) = -A(x - \bar{x})$, segue:

$$V(\bar{x}) = \{ x \in R^n / (x - \bar{x})^T [A^T \bar{Z}^{-2} A] (x - \bar{x}) = 1 \}. \quad (12)$$

$V(\bar{x})$ define um elipsóide, porém não trivial, isto é, não tem seus eixos paralelos aos eixos coordenados como em E_z . Veremos que os autovetores de $[A^T \bar{Z}^{-2} A]$ definem as direções dos eixos e que $2/\sqrt{\lambda_i}$ é a amplitude do i -ésimo eixo do elipsóide, onde λ_i é o i -ésimo autovalor.

Para isto, consideramos o seguinte teorema da álgebra linear, ([Wilkinson, (1965)]):

Teorema 1: Seja $\bar{A}$ uma matriz simétrica e positiva definida, então:

- i) todos seus autovalores são positivos,
- ii) Existe uma base ortonormal de autovetores de $\bar{A}$.

Seja $Q = [q^1, \dots, q^n]$ uma matriz de autovetores de $[A^T \bar{Z}^{-2} A]$ satisfazendo a condição ii) do teorema 1, então:

$$Q^{-1} = Q^T.$$

Pela definição de autovalores e autovetores, temos:

$$[A^T \bar{Z}^{-2} A]Q = [\lambda_1 q^1, \dots, \lambda_n q^n] = Q \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$$

onde λ_i é o autovalor associado ao autovetor q^i , $i = 1, \dots, n$.

De outra forma,

$$A^T \bar{Z}^{-2} A = Q \Lambda Q^T,$$

com $\Lambda = \text{diag} [\lambda_1 \dots \lambda_n]$.

Considere agora a seguinte mudança de variável: $y = Q^T x$ ($x = Qy$), ou seja, dado um vetor x , podemos representá-lo na base $[q^1, \dots, q^n]$ e suas coordenadas serão $y_1 \dots y_n$.

Reescrevemos agora $V(\bar{x})$ no novo sistema de coordenadas y :

$$\begin{aligned} (x-\bar{x})^T A^T \bar{Z}^{-2} A (x-\bar{x}) &= (x-\bar{x})^T Q \Lambda Q^T (x-\bar{x}) \\ &= (Q^T x - Q^T \bar{x})^T \Lambda (Q^T x - Q^T \bar{x}) \\ &= (y-\bar{y})^T \Lambda (y-\bar{y}) = 1; \end{aligned}$$

$$\left[\left[\frac{(y_1 - \bar{y}_1)}{\sqrt{\lambda_1^{-1}}} \right]^2 + \dots + \left[\frac{(y_n - \bar{y}_n)}{\sqrt{\lambda_n^{-1}}} \right]^2 = 1 \right];$$

onde $\bar{y} = Q^T \bar{x}$ ($\bar{x} = Q\bar{y}$) é a representação de $\bar{x}$ no novo sistema de coordenadas.

A equação acima descreve um elipsóide (análogo a E_z), centrado em $\bar{y}$, com eixos paralelos aos eixos coordenados de y $\{q^1, \dots, q^n\}$, e de amplitude $2/\sqrt{\lambda_1}$, conforme ilustra a figura 4:

$\{q^1, q^2\}$: base de auto-vetores.
 $\{e^1, e^2\}$: base canônica.

figura 4.

Retornemos ao problema (11) com $V(\bar{x})$ definido pelo elipsóide (12), com $x = \bar{x} + u$:

$$\begin{aligned} &\text{minimizar } c^T(\bar{x} + u) \\ &\text{sujeito a } u^T [A^T \bar{Z}^{-2} A] u = 1. \end{aligned} \quad (13)$$

Com a mudança de variável $v = Q^T u$, o problema acima é equivalente a:

$$\begin{aligned} &\text{minimizar } c^T Q v + \{c^T \bar{x}\} \\ &\text{sujeito a } v^T \Lambda v = 1 \end{aligned} \quad (14)$$

Com uma nova mudança de variável, $\tilde{v} = \Lambda^{1/2} v$, teremos um novo problema equivalente a (13), onde o elipsóide é transformado numa bola:

$$\begin{aligned} &\text{minimizar } c^T Q \Lambda^{-1/2} \tilde{v} + \{ c^T \bar{x} \} \\ &\text{sujeito a } \tilde{v}^T \tilde{v} = 1 \end{aligned} \quad (15)$$

cuja solução é múltipla de:

$$\tilde{v} = - \Lambda^{-1/2} Q^T c.$$

Logo, a solução de (14) é múltipla de:

$$\bar{v} = - \Lambda^{-1} Q^T c.$$

Assim a solução de (13) é múltipla de:

$$\bar{u} = - Q \Lambda^{-1} Q^T c.$$

Desde que,

$$[A^T \bar{Z}^{-2} A]^{-1} = [Q \Lambda Q^T]^{-1} = Q \Lambda^{-1} Q^T,$$

e portanto a solução de (13) é múltipla de:

$$\bar{u} = - [A^T \bar{Z}^{-2} A]^{-1} c,$$

a qual é a direção usada em [Adler et al, (1986)], podendo ser obtida ao se resolver o problema (11) com $V(\bar{x})$ definido por (12).

O próximo exemplo ilustra o fato de que, nem sempre o elipsóide de $V(\bar{x})$ é o maior possível na região de factibilidade P.

Exemplo 1:

Consideremos o seguinte problema:

$$\begin{aligned} &\text{minimizar } -x_1 - x_2 \\ &\text{sujeito a } \begin{cases} x_2 \leq 2 \\ x_1 + 2x_2 \leq 2 \\ x_1 \leq 2 \end{cases} \end{aligned}$$

Logo: $z_1 = 2 - x_2$; $z_2 = 2 - x_1 - 2x_2$; $z_3 = 2 - x_1$.

Se tomarmos $\bar{x} = (0,0)^T$ então $\bar{z} = (2,2,2)^T$.

$$A^T \bar{Z}^{-2} A = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 5/4 \end{bmatrix}$$

Elipse nas variáveis x com centro $(0;0)$:

$$x^T [A^T \bar{Z}^{-2} A] x = 1$$

$$(1/2 x_1^2 + 5/4 x_2^2 + x_1 x_2 = 1)$$

Sejam $Q = [q^1, q^2]$: base de autovetores de $A^T \bar{Z}^{-2} A$;

λ_i : o i -ésimo autovalor associado a q^i , $i = 1, 2$,

então, $\lambda_1 = 1/4$, $\lambda_2 = 3/2$.

$$\text{se } \lambda_1 = 1/4 \Rightarrow q^1 = (2 / \sqrt{5} ; -1 / \sqrt{5})$$

$$\text{se } \lambda_2 = 3/2 \Rightarrow q^2 = (1 / \sqrt{5} ; 2 / \sqrt{5})$$

Elipse nas variáveis y ($y = P^T x$), com centro $(0;0)$:

$$y^T \Lambda y = 1, (\Lambda = \text{diag}(\lambda_i))$$

$$(y_1^2 \lambda_1 + y_2^2 \lambda_2 = 1)$$

Representação geométrica da elipse nas variáveis y , totalmen-
contida em $P' = \{y \in R^2 \text{ tal que } AQy \leq b\}$:

Representação geométrica da elipse nas variáveis x :

figura 6.

Observação:

Embora o elipsóide $(z - \bar{z})^T \bar{Z}^{-2} (z - \bar{z}) = 1$, seja o maior possível, centrado em $\bar{z}$ e inteiramente contido em $z \geq 0$ (veja a figura abaixo), quando restringimos $z \in T(P)$ (o que equivale às elipses das figuras 5 e 6) não temos a maior elipse contida na região de factibilidade P , nas variáveis x . Note que, a reta $x_1 = 2$ e $x_1 + 2x_2 = 2$ não interceptam a elipse, na figura 6.

figura 7

Vamos estender nossa interpretação geométrica agora à seguinte transformação composta:

$$W \circ T: P \rightarrow R_+^m \text{ onde } W \circ T(x) = \bar{Z}^{-1}(b - Ax).$$

Para isto consideremos a função $W: R^m \rightarrow R^m$, onde $W(z) = \bar{Z}^{-1}z$, e $T: P \rightarrow R_+^m$ tal que $T(x) = b - Ax$, assim,

$$\text{Im}(W \circ T) = \{w \in R_+^m \text{ tal que } w = W \circ T(x) = \bar{Z}^{-1}(b - Ax), x \in P\}.$$

Desde que, $T: P \rightarrow \text{Im}(T)$ e $W: R_+^m \rightarrow R_+^m$, são bijetivas, então vale a seguinte proposição:

Proposição 1: $W \circ T: P \rightarrow \text{Im}(W)$ é bijetiva.

A próxima proposição, decorrente da anterior, nos dá resultados importantes sobre a relação entre P e $\text{Im}(W \circ T)$.

Proposição 2: O elipsóide $E_x \subset P$ é transformado pela $W \circ T$ em $E_w \cap \text{Im}(W \circ T)$; $E_w = \{w \in R_+^m \text{ tal que } (w - e)^T (w - e) = 1\}$, ou seja,

$$E_x \subset P \xrightarrow[T]{T^{-1}} E_z \cap \text{Im}(T) \xrightarrow[W]{W^{-1}} E_w \cap \text{Im}(W \circ T).$$

Das proposições anteriores podemos concluir que são equivalentes os problemas:

$$\begin{aligned} & \text{minimizar } c^T [W \circ T]^{-1}(w) \\ & \text{sujeito a } \begin{cases} w \in \text{Im}(W \circ T) \subset R_+^m \\ (w - e)^T (w - e) = 1 \end{cases} \\ & \text{minimizar } c^T [T]^{-1}(z) \\ & \text{sujeito a } \begin{cases} z \in \text{Im}(T) \\ (z - \bar{z})^T \bar{Z}^{-2} (z - \bar{z}) = 1 \end{cases} \\ & \text{minimizar } c^T x \\ & \text{sujeito a } \begin{cases} Ax \leq b \\ (x - \bar{x})^T A^T \bar{Z}^{-2} A (x - \bar{x}) = 1 \end{cases} \end{aligned} \quad (16)$$

Como o elipsóide E_x está totalmente contido na região viável P , o problema (16) pode ser escrito por:

$$\begin{aligned} & \text{minimizar } c^T x \\ & \text{sujeito a } (x - \bar{x})^T A^T \bar{Z}^{-2} A (x - \bar{x}) = 1; \end{aligned}$$

que corresponde ao já estudado problema (11), com $V(\bar{x})$ definida por (12), analiticamente resolvido por mudar o sistema de coordenadas nas direções dos autovetores ortogonais de $[A^T \bar{Z}^{-2} A]$, levando a uma única solução múltipla de:

$$d_x = - [A^T \bar{Z}^{-2} A]^{-1} c.$$

Conclusão: Os resultados anteriores foram obtidos quando fizemos a mudança de escala nas variáveis de folga, definindo novas variáveis:

$$w = \bar{Z}^{-1}z.$$

Tal mudança de escala é interessante pois:

-transforma $\bar{z}$ em "e";

-transforma o elipsóide E_z na bola $(w - e)^T(w - e) = 1$.

Esta mudança de escala corresponde a uma transformação W ; $W: R_+^m \rightarrow R_+^m$; $W(z) = \bar{Z}^{-1}z$, e a composta $W \circ T: P \rightarrow \text{Im}(W \circ T)$ define o problema original nas variáveis escaladas. Além do mais, se substituirmos $w \geq 0$ por $(w - e)^T(w - e) = 1$, temos:

$$\begin{aligned} &\text{minimizar } c^T [W \circ T]^{-1}(w) \\ &\text{sujeito a } \begin{cases} w \in \text{Im}(W \circ T) \\ (w - e)^T(w - e) = 1; \end{cases} \end{aligned}$$

correspondendo ao problema (11), com $V(\bar{x})$ definida em (12), analisado neste capítulo, bastando substituir $w = W \circ T(x) = \bar{Z}^{-1}(b - Ax)$ e cuja solução nos leva à direção múltipla de;

$$\bar{u} = -[A^T \bar{Z}^{-2} A]^{-1} c,$$

que é a direção definida em [Adler et al, (1986)].

Outro fato importante que vimos é que, $V(\bar{x})$ pode não ser o maior elipsóide contido em P .

O algoritmo dual afim, analisado neste capítulo, quando implementado e comparado com uma implementação avançada do método simplex, o código MINO's ([Murtagh e Saunders, (1977)]), obteve bons resultados computacionais. Isto se deve ao seguinte fato, encontrado em [Adler et al, (1986)] :

"Mesmo que, d_x calculado aproximadamente em (9) seja infactível, um par de direções factíveis pode ainda ser obtido pelo cálculo de $d_z = -Ad_x$ ", pois o sistema (3) está satisfeito.

CAPÍTULO III

PROBLEMAS COM RESTRIÇÕES CANALIZADAS: $b^- \leq Ax \leq b^+$.

1 - Algoritmo dual afim para restrições canalizadas.

No capítulo 2, um método de pontos interiores foi deduzido explorando-se a estrutura de restrições no formato "dual", $Ax \leq b$.

Neste capítulo uma extensão daquele método é desenvolvida para a estrutura de restrições canalizadas.

1.1 - Preliminares.

Considere o problema com restrições canalizadas:

$$\begin{cases} \text{minimizar } c^T x \\ \text{sujeito a } b^- \leq Ax \leq b^+ \end{cases}$$

onde $A \in R^{m \times n}$ e $\text{po}(A) = n$.

Naturalmente este problema pode ser visto como um caso particular do anterior, bastando reescrevê-lo na seguinte forma equivalente:

$$\begin{aligned} &\text{minimizar } c^T x \\ &\text{sujeito a } \begin{cases} Ax \leq b^+ \\ -Ax \leq -b^- \end{cases} \end{aligned}$$

e então tratá-lo como anteriormente.

Entretanto a aplicação direta do algoritmo dual afim neste problema equivalente, deve levar a um procedimento ineficiente.

Podemos, ainda, explorar as particularidades das restrições (note que a matriz A é repetida) e trabalhar implicitamente com a duplicação das restrições, aproveitando-se de cálculos repetidos que apareceriam na determinação de $\bar{u}$. Entretanto a exploração direta das particularidades podem ser bem mais facilmente percebidas seguindo-se a linha geométrica adotada na dedução do algoritmo dual afim, no capítulo 2.

1.2 - Variáveis de folga e mudança de escala.

Definindo-se as variáveis de folga como a imagem da transformação linear; $Z: \mathbb{R}^n \rightarrow \mathbb{R}^m$ tal que

$$z = Z(x) = Ax,$$

a região de factibilidade;

$$P = \{x \in \mathbb{R}^n \text{ tal que } b^- \leq Ax \leq b^+\} \subset \mathbb{R}^n,$$

corresponde no $\mathbb{R}^m$ ao conjunto $Z(P)$ definido por;

$$Z(P) = \{z \in \mathbb{R}^m \text{ tal que } b^- \leq z \leq b^+, z \in \text{Im}(Z)\}$$

onde;

$$\text{Im}(Z) = \{z \in \mathbb{R}^m \text{ tal que } z = Ax, \text{ para algum } x \in \mathbb{R}^n\}$$

é o subespaço do $\mathbb{R}^m$ gerado pelas colunas de A .

Note que $Z(P)$ é a intersecção do hiperparalelepípedo;

$$\{z \in \mathbb{R}^m \text{ tal que } b^- \leq z \leq b^+\}$$

com o subespaço $\text{Im}(Z)$ e $Z: P \rightarrow Z(P)$ é uma bijeção uma vez que Z é bijectiva.

Veja a figura abaixo:

figura 1

O maior elipsóide centrado em $\bar{z}$ e inteiramente contido no hiperparalelepípedo cujos eixos são paralelos aos eixos coordenados é dado por:

$$E_z = \{z \in \mathbb{R}^m \text{ tal que } (z - \bar{z})^T \bar{Z}^{-2}(z - \bar{z}) = 1\},$$

onde $\bar{Z} = \text{diag}(\bar{z}_1')$ com:

$$\bar{z}_1' = \begin{cases} \bar{z}_1 - b_1^- & \text{se } \bar{z}_1 \leq (b_1^- + b_1^+)/2 \\ \bar{z}_1 - b_1^+ & \text{se } \bar{z}_1 \geq (b_1^- + b_1^+)/2 \end{cases}$$

Se definirmos $f \in \mathbb{R}^m$ tal que :

$$f_i = \begin{cases} b_i^- & \text{se } \bar{z}_i \leq (b_i^- + b_i^+)/2 \\ b_i^+ & \text{se } \bar{z}_i \geq (b_i^- + b_i^+)/2; \end{cases}$$

então $\bar{Z} = \text{diag}(\bar{z}_i - f_i)$.

Note que f é o vértice do hiperparalelepípedo mais próximo de $\bar{z}$.

Considere agora a seguinte transformação afim;

$$W: \mathbb{R}^m \rightarrow \mathbb{R}^m \text{ tal que } w = W(z) = \bar{Z}^{-1}(z - f).$$

Esta transformação define um novo sistema de variáveis w_i que além de promover uma mudança de escala em z , tem como origem a imagem do vértice de f , (ver figura 2).

figura 2

Esta transformação pode também inverter a orientação do eixo coordenado (na figura 6, $\bar{z}'_1 = \bar{z}_1 - b_1^+ < 0$), uma vez que podemos ter $\bar{z}'_1 < 0$.

Proposição 1: O elipsóide E_z é transformado na esfera de centro "e" e raio unitário pela transformação W .

Prova:

Inicialmente temos que;

$$w = W(z) = \bar{Z}^{-1}(z - f), \text{ então } z = W^{-1}(w) = \bar{Z}w + f,$$

$$\bar{z} = W^{-1}(e) = \bar{Z}e + f.$$

Substituindo as expressões acima em E_z temos:

$$1 = (z - \bar{z})^T \bar{Z}^{-2} (z - \bar{z}) = (w - e)^T (w - e).$$

□

Observe que o hiperparalelepípedo; $b^- \leq z \leq b^+$ corresponde nas variáveis w_1 ao seguinte hiperparalelepípedo:

$$0 \leq w \leq \text{diag}(|\bar{z}'_1|)^{-1}(b^+ - b^-).$$

A esfera $(w - e)^T(w - e) = 1$ está inteiramente contida no hiperparalelepípedo transformado: $0 \leq w \leq \text{diag}(|\bar{z}'_1|)^{-1}(b^+ - b^-)$, pois esta corresponde ao elipsóide $(z - \bar{z})^T \bar{Z}^{-2}(z - \bar{z}) = 1$, o qual está inteiramente contido no hiper-paralelepípedo; $b^- \leq z \leq b^+$.

Isto pode ser observado na seguinte proposição:

Proposição 2: O hiperparalelepípedo transformado;

$$0 \leq w \leq \text{diag}(|\bar{z}'_1|)^{-1}(b^+ - b^-),$$

contém o hipercubo;

$$0 \leq w \leq 2e.$$

prova:

Basta mostrar que $[(b^+ - b^-)/|\bar{z}'_1|] \leq 2$.

Para mostrar isto consideremos os dois casos abaixo:

i) se $\bar{z}_1 \geq (b_1^+ + b_1^-)/2$, ($\bar{z}'_1 = \bar{z}_1 - b_1^+$), temos que:

$$2\bar{z}_1 \geq b_1^- + b_1^+ \Rightarrow 2\bar{z}_1 - 2b_1^+ \geq b_1^- - b_1^+ \Rightarrow$$

$$2(\bar{z}_1 - b_1^+) \geq b_1^- - b_1^+ \Rightarrow 2 \leq (b_1^+ - b_1^-)/(\bar{z}_1 - b_1^+).$$

ii) se $\bar{z}_1 \leq (b_1^+ + b_1^-)/2$, ($\bar{z}'_1 = \bar{z}_1 - b_1^-$), temos que;

$$2\bar{z}_1 - 2b_1^- \leq b_1^+ - b_1^- \Rightarrow 2 \leq (b_1^+ - b_1^-)/(\bar{z}_1 - b_1^-).$$

Logo, dos itens i) e ii) podemos concluir que:

$$[(b_1^+ - b_1^-)/|\bar{z}'_1|] \geq 2.$$

□

Desde que $W \circ Z: P \rightarrow \text{Im}(W \circ Z)$ é uma transformação bijetiva, a região de factibilidade P (nas variáveis x_1) corresponde ao seguinte conjunto (nas variáveis w_1):

$$\text{Im}(W \circ Z) \cap \{w \in \mathbb{R}^m \text{ tal que } 0 \leq w \leq \text{diag}(|\bar{z}'_1|)^{-1}(b^+ - b^-)\}$$

1.3 - Determinação da direção de busca.

Se restringirmos as soluções à esfera $(w - e)^T(w - e) = 1$ temos o seguinte problema:

$$\begin{aligned} &\text{minimizar} && c^T [W_0 Z]^{-1}(w) \\ &\text{sujeito a} && \begin{cases} w \in \text{Im}(W_0 Z) \\ (w - e)^T(w - e) = 1 \end{cases} \end{aligned} \quad (1)$$

uma vez que a esfera; $(w - e)^T(w - e) = 1$ está inteiramente contida no hiperparalelepípedo $0 \leq w \leq \text{diag}(|\bar{z}'|)^{-1}(b^+ - b^-)$, então o problema (1) fornecerá soluções factíveis.

Desde que $w \in \text{Im}(W_0 Z)$ implica que:

$$w = W_0 Z(x) = \bar{Z}^{-1}(Ax - f), \quad x \in P,$$

podemos escrever o problema (1) nas variáveis x , equivalentemente:

$$\begin{cases} \text{minimizar} & c^T x \\ \text{sujeito a} & (x - \bar{x})^T A^T \bar{Z}^{-2} A (x - \bar{x}) = 1 \end{cases} \quad (2)$$

cujas solução, análoga ao problema do capítulo anterior, obtida com a mudança do sistema de coordenadas nas direções dos autovetores ortogonais de $[A^T \bar{Z}^{-2} A]$, é múltipla de:

$$\bar{u} = x - \bar{x} = -[A^T \bar{Z}^{-2} A]^{-1} c.$$

1.4 - Determinação do passo.

Para explicitar os procedimentos do algoritmo para restrições canalizadas numa iteração corrente com $\bar{x}$ e $\bar{z}$ já determinados, determinemos o tamanho do passo a ser dado por $\bar{x}$ e $\bar{z}$ numa direção.

A resolução do problema (2) nos define as direções de busca $d_x \in R^n$ e $d_z \in R^m$ da seguinte forma:

$$d_x = -[A^T \bar{Z}^{-2} A]^{-1} c;$$

$$d_z = A d_x; \quad (\text{pois } \bar{z} + \varepsilon d_z = A(\bar{x} + \varepsilon d_x) \text{ e } \bar{z} = A\bar{x}).$$

Para continuar o processo devemos determinar o tamanho do passo a ser dado numa determinada direção, ou seja, determinar um passo $\varepsilon \in R_+$ a ser dado tal que $\bar{x} + \varepsilon d_x$ seja factível, isto é,

$$b^- \leq A\bar{x} + \varepsilon A d_x \leq b^+,$$

que é equivalente a;

$$b^- \leq \bar{z} + \varepsilon d_z \leq b^+$$

Sejam

$$\varepsilon^1 = \min\{ -(\bar{z}_1 - d_1)/(d_z)_1 \text{ tal que } (d_z)_1 < 0 \};$$

$$\varepsilon^2 = \min\{ (e_1 - \bar{z}_1)/(d_z)_1 \text{ tal que } (d_z)_1 > 0 \}.$$

A seguir, tomamos $\varepsilon = \min\{ \varepsilon^1, \varepsilon^2 \}$, garantindo a factibilidade do problema e observando que, para esta definição de ε , pelo menos uma componente de $z + \varepsilon d_z$ coincide com um dos limites.

Para evitarmos a fronteira, ou seja, manter novos pontos a serem determinados interiores à região de factibilidade do problema, basta que tomemos $\bar{\varepsilon} \in R_+$ tal que:

$$\bar{\varepsilon} = \alpha \varepsilon, \quad 0 < \alpha < 1.$$

Assim, a menos da definição da matriz $\bar{Z}$ e do tamanho do passo a ser dado em uma determinada direção, definido pelas condições de fronteira do problema (já que agora limitantes inferiores e superiores devem ser considerados), todas as expressões deduzidas anteriormente, são idênticas. Como consequência, um algoritmo para o caso de restrições canalizadas pode facilmente ser obtido.

1.5 - Algoritmo para restrições canalizadas.

Encontre x^0 tal que $b^- < Ax^0 < b^+$, $z^0 = Ax^0$, $k = 0$.

Escolha $\alpha \in \mathbb{R}$ tal que $0 < \alpha < 1$.

Enquanto { critério de parada não satisfeito } faça:

Calcule a matriz $\bar{Z}_k = \text{diag} [z_1^k, \dots, z_m^k]$ onde:

$$z_1^k = \begin{cases} z_1^k - b_1^- & \text{se } z_1^k \leq (b_1^- + b_1^+)/2 \\ z_1^k - b_1^+ & \text{se } z_1^k \geq (b_1^- + b_1^+)/2 \end{cases}$$

$$(z^k = Ax^k)$$

Calcule a direção:

$$d_x = - [A^T \bar{Z}_k^{-2} A]^{-1} c$$

$$d_z = A d_x$$

Calcule o passo:

$$\varepsilon^1 = \min \{ - (z_1^k - b_1^-) / (d_z)_1 \text{ tal que } (d_z)_1 < 0 \}.$$

$$\varepsilon^2 = \min \{ (b_1^+ - z_1^k) / (d_z)_1 \text{ tal que } (d_z)_1 > 0 \}.$$

$$\varepsilon = \min \{ \varepsilon^1, \varepsilon^2 \}$$

Atualize a solução:

$$x^{k+1} = x^k + \bar{\varepsilon} d_x ; \quad \bar{\varepsilon} = \alpha \varepsilon ;$$

$$z^{k+1} = z^k + \bar{\varepsilon} d_z \quad \text{ou} \quad z^{k+1} = Ax^{k+1} ;$$

$$k \leftarrow k + 1.$$

1.6 - Critério de parada.

Um critério de parada que pode ser utilizado encontrado em [Gonzaga, (1989)] é:

"Se $(n - m)$ variáveis $\bar{z}_i$ tenderem a zero então pare o processo pois estamos muito próximos do vértice ótimo." Ou ainda, o mesmo critério utilizado em [Adler et al, (1986)], visto no capítulo 2.

2 - Um resultado numérico.

Na presente seção é fornecido um resultado numérico, a fim de ilustrar o comportamento do método computacionalmente.

Para isto utilizamos um micro-computador Itautec, compatível com o padrão IBM PC/AT - 286, (16 bits).

As rotinas foram elaboradas em Pascal, versão 3.0, em precisão simples e são uma modificação da implementação do algoritmo dual afim encontrada em [Orlandi, (1990)].

Exemplo 1:

Seja o seguinte P.P.L de restrições canalizadas;

$$\text{minimizar } z = -x_1 - 2x_2$$

$$\begin{aligned} \text{sujeito a:} \quad & 0 \leq x_1 - x_2 \leq 2 \\ & -4 \leq 2x_1 + x_2 \leq 4 \\ & -1 \leq x_1 \leq 2 \\ & -1 \leq x_2 \leq 2 \end{aligned}$$

Geometricamente, (figura 3), a solução encontrada foi $x_1 = 4/3$ e $x_2 = 4/3$.

figura 3.

Tabela de resultados a partir do ponto $x^0 = (0.25, 0.05)$:

iteração	x_1^k	z_1^k	$z_1'^k$
1	1.3370918922E+0	1.8377567671E-1	2.0000000000E-1
	1.1533162155E+0	3.8227500000E+0	3.4500000000E+0
		1.3370918922E+0	1.2500000000E+0
		1.1533162155E+0	1.0500000000E+0
2	1.3310380839E+0	9.1887838355E-3	1.8377567671E-1
	1.3218493000E+0	3.9839254677E+0	1.7250000000E-1
		1.3310380839E+0	6.6290810777E-1
		1.3218493000E+0	8.4668378447E-1
6	1.3333271520E+0	7.3132884384E-6	1.4626576876E-4
	1.3333198387E+0	3.9999741426E+0	3.2878539059E-5
		1.3333271520E+0	6.6662887092E-1
		1.3333198387E+0	6.6677513669E-1
7	1.3333346851E+0	7.3132884384E-6	7.3132884384E-6
	1.3333293369E+0	3.9999987071E+0	2.5857429137E-5
		1.3333346851E+0	6.6667284804E-1
		1.3333293369E+0	6.6668016134E-1

A solução ótima encontrada computacionalmente foi:

$$x_1 = 1.3333346851E+0 \quad ; \quad x_2 = 1.3333293369E+0$$

e o valor da função objetivo é $z = -3.9999933589E+0$.

Conclusão: Neste capítulo, a interpretação geométrica do algoritmo de Adler e outros, feita no capítulo 2, nos possibilitou estendê-lo para um caso mais geral de problemas de programação linear, o caso de restrições canalizadas.

Resta-nos a curiosidade de saber se este algoritmo, com uma boa implementação computacional trará resultados satisfatórios, já que foi obtido diretamente do algoritmo encontrado em [Adler et al, (1986)], que por sua vez obteve bons resultados computacionais.

CAPÍTULO IV.

PROBLEMAS COM RESTRIÇÕES DO TIPO IGUALDADE.

Os algoritmos que analisaremos neste capítulo são desenvolvidos para problemas com restrições do tipo igualdade e serão denominados por algoritmos "primais".

Faremos inicialmente um desenvolvimento algébrico dos mesmos para depois preocuparmo-nos com uma interpretação geométrica. Para tanto, motivamo-nos em [Gonzaga, (1988), (1989)], em [Chandru e Kochar, (1985)] e em [Karmarkar, (1984)], entre outros.

1 - Algoritmo primal afim.

O primeiro algoritmo a ser analisado, do tipo "primal", será denominado aqui por algoritmo "primal afim", já que outras denominações caracterizam o mesmo. Sabe-se que este já tinha sido analisado em [Dikin, (1967)] e mais recentemente foi reavaliado em [Barnes, (1986)] e [Vanderbei, Meketon e Freedman, (1986)].

1.1 - Preliminares.

Consideremos o problema em formato primal:

$$\begin{aligned} &\text{minimizar } c^T x \\ &\text{sujeito a } \begin{cases} Ax = b \\ x \geq 0; \end{cases} \end{aligned} \quad (1)$$

$$x, c \in R^n, A \in R^{m \times n}; \text{ posto}(A) = m.$$

Sejam $S = \{x \in R^n \text{ tal que } Ax = b, x \geq 0\}$, o conjunto dos pontos factíveis e $\bar{x}$ tal que $A\bar{x} = b$ e $\bar{x} > 0$, ou seja, $\bar{x} \in \text{int}(S)$.

Considere a matriz mudança de escala $\bar{X} = \text{diag}(\bar{x}_1, \dots, \bar{x}_n)$ e a partir dela reescalemos a variável x de acordo com seguinte mudança de variável;

$$y = \bar{X}^{-1}x = \begin{bmatrix} x_1 / \bar{x}_1 \\ \vdots \\ x_n / \bar{x}_n \end{bmatrix};$$

e assim teremos um novo problema equivalente a (1):

$$\begin{aligned} &\text{minimizar } \bar{c}^T y \\ &\text{sujeito a } \begin{cases} \bar{A}y = b \\ y \geq 0; \end{cases} \end{aligned} \quad (2)$$

onde $\bar{A} = A\bar{X}$, $\bar{c}^T = c^T \bar{X}$ e $\bar{y} = \bar{X}^{-1} \bar{x} = e$ é uma solução factível para (2).

O problema (2) será de grande importância para a análise do algoritmo primal afim, principalmente quando fizermos a interpretação geométrica do mesmo e observarmos que a mudança de escala transformando $\bar{x}$ do problema original (1) para o ponto "e", factível ao problema (2) nos facilitará o entendimento dos passos do algoritmo.

1.2 - A direção de busca.

Uma estratégia diferente daquela vista para o algoritmo encontrado em [Adler et al, (1986)], na determinação da direção de busca, será dada aqui, já que a cada iteração, o algoritmo em questão determinará a direção de busca utilizando a estratégia de projetar no espaço das direções factíveis o gradiente da função objetivo escalado pela mudança de variável.

De acordo com a seção 5, capítulo 1, a direção oposta ao gradiente precedida pela mudança de escala, é uma boa direção para possíveis decréscimos da função objetivo se considerarmos um problema de minimização irrestrito. Podemos levar em consideração tal direção, desde que, projetemos tal direção no espaço das direções factíveis de (2), a cada iteração.

Consideremos o gradiente da função objetivo do problema (2), $\bar{c}$ e sua decomposição em:

$$\bar{c} = \bar{c}_p + \bar{c}_p^- \quad \text{onde} \quad \bar{c}_p \in N(\bar{A}), \quad \bar{c}_p^- \in I(\bar{A}^T), \quad c_p = P_A \bar{c} \quad \text{e}$$

$$P_A = I - \bar{A}^T (\bar{A} \bar{A}^T)^{-1} \bar{A}.$$

Como consequência direta da expressão anterior, definiremos a direção de busca utilizada pelo algoritmo a cada iteração.

A partir de uma solução factível $\bar{x}$ em (1), consideremos $\bar{y} = e$; se "perturbarmos" $\bar{y}$ na seguinte direção, $h \in \mathbb{R}^n$ tal que:

$$h = -\bar{c}_p \quad (3)$$

então h é uma direção factível para $\bar{y}$ pois $h \in N(\bar{A})$.

Resta-nos mostrar que h é uma direção de descida.

Proposição 1: Da maneira como foi definida acima, h é direção de descida, ou seja, $\bar{c}^T h < 0$.

Prova:

Basta notar que, de $\bar{c} = \bar{c}_p + \bar{c}_p^-$; $\bar{c}_p \in N(\bar{A})$ e $\bar{c}_p^- \in I(\bar{A}^T)$ temos:

$$0 < \bar{c}_p^T \bar{c}_p = \bar{c}_p^T (\bar{c} - \bar{c}_p^-) = \bar{c}_p^T \bar{c} - \bar{c}_p^T \bar{c}_p^- = \bar{c}_p^T \bar{c}, \text{ pois } \bar{c}_p^T \text{ e } \bar{c}_p^-$$

são ortogonais.

Logo, $\bar{c}^T h < 0$. Portanto, $h = -\bar{c}_p$ é uma direção de descida. $\square$

1.3 - Determinação do passo

Se fizermos uma perturbação no vetor $\bar{y} \in \mathbb{R}^n$, factível à (2) de tal forma que geramos o novo ponto $\bar{y} + \epsilon h$, $\epsilon \geq 0$, então a seguinte condição deve ser respeitada (condição de fronteira):

$$\bar{y} + \epsilon h \geq 0$$

Para isto, determinemos o valor de ϵ .

$$\bar{y} + \epsilon h \geq 0 \Rightarrow 1 + \epsilon h_i \geq 0; \quad (i = 1 \dots n)$$

$$\Rightarrow \epsilon \leq -1/h_i \text{ se } h_i < 0. \quad (i = 1 \dots n)$$

Assim, temos que :

$$\epsilon = \min_{i=1 \dots n} \{ -1/h_i \text{ tal que } h_i < 0 \}.$$

Se $h \geq 0$, então temos um raio de descida, acusando assim que o problema não tem solução ótima.

Uma heurística é adotada para a escolha do tamanho do passo, de tal forma que os novos pontos sejam interiores a (2):

$$\bar{\epsilon} = \alpha \epsilon, \text{ onde } 0 < \alpha < 1.$$

1.4 - Algoritmo.

Após as considerações anteriores podemos enunciar o seguinte algoritmo:

Encontre x^0 tal que $x^0 > 0$ e $Ax^0 = b$, $k = 0$ e

$$0 < \alpha < 1.$$

Parada := falso.

Enquanto {parada := falso} faça:

início;

Sejam

$$X_k = \text{diag}(x_1^k, \dots, x_n^k),$$

$$A_k = AX_k; \quad c^k = c^T X_k.$$

Determine a direção:

$$P_{A_k} = I - A_k^T (A_k A_k^T)^{-1} A_k.$$

$$h = -P_{A_k} c^k.$$

Calcule o passo:

$$\varepsilon = \min_{i=1..n} \{-1/h_i \text{ tal que } h_i < 0\}.$$

$$\bar{\varepsilon} = \alpha \varepsilon.$$

Se $h \geq 0$, então parada := verdadeiro.

Atualize a solução:

$$y^{k+1} = e + \bar{\varepsilon} h.$$

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

$$k = k + 1.$$

Se $n - m$ componentes $x_j \leq \text{tol}$, onde tol é uma pequena tolerância positiva, então parada := verdadeiro, pois estamos próximos do vértice ótimo.

1.5 - Obtenção de um ponto inicial.

Um procedimento comum para obtenção de um ponto inicial resulta em um problema de dimensão $n + 1$, dado a seguir:

$$\begin{aligned} &\text{minimizar } c^T x + Mw \\ &\text{sujeito a } \begin{cases} Ax + w(b - Ae) = b \\ x \geq 0, w \geq 0; \end{cases} \end{aligned} \quad (4)$$

onde M é escolhido grande de forma que em qualquer solução ótima $w = 0$, assim " e " é uma solução inicial de (4).

2 - Uma interpretação geométrica.

Nesta seção pretendemos deduzir o algoritmo afim escala primal com uma motivação geométrica do problema.

Para isto, consideremos $\bar{x} \in \text{int}(S)$ e seja $V(\bar{x}) \subset S$ tal que seja fácil resolver o seguinte problema:

$$\begin{aligned} &\text{minimizar } c^T(\bar{x} + d) \\ &\text{sujeito a } \begin{cases} A(\bar{x} + d) = b \\ \bar{x} + d \in V(\bar{x}) \end{cases} \end{aligned} \quad (5)$$

É válida a mesma observação do capítulo 2, seção 2, de que utilizar $V(\bar{x})$ como uma esfera conduz a tomar d como a direção oposta ao gradiente projetado no núcleo de A , o que não é uma boa escolha, pois esta é super influenciada pela fronteira e uma vez atingida a fronteira não será mais possível caminhar na busca de novos pontos, mesmo que o ponto encontrado não seja ótimo (ver figura 2, capítulo 2), pois para alguma componente i , $\bar{x}_i \approx 0$ então $d \approx 0$, mesmo que outras componentes possam sofrer maiores alterações.

Já vimos que uma escolha aparentemente mais razoável para $V(\bar{x})$ é uma elipse, que se ajusta melhor às condições de fronteira do problema, permite "diminuir" $V(\bar{x})$ numa determinada direção sem prejudicar outras direções e mantém o problema (5) fácil de ser resolvido.

O elipsóide mais simples, centrado em $\bar{x}$, que "pune" apenas as direções dos eixos coordenados para os quais as componentes de $\bar{x}$ se aproximam de zero e que está contido em R_+^n é:

$$E_x = V(x) = \{x \in R_+^n \text{ tal que } (x - \bar{x})^T \bar{X}^{-2}(x - \bar{x}) = 1\} \quad (6)$$

Definindo-se $d = x - \bar{x}$, teremos $d^T \bar{X}^{-2}d = 1$ e o problema (5) se torna:

$$\begin{aligned} &\text{minimizar} && c^T d \\ &\text{sujeito a} && \begin{cases} Ad = 0 \\ d^T \bar{X}^{-2}d = 1 \end{cases} \end{aligned} \quad (7)$$

A estratégia utilizada para solucionar (7) consiste em definir a seguinte transformação mudança de escala:

$$Y: S \rightarrow R_+^n \text{ tal que } y = Y(x) = \bar{X}^{-1}x.$$

Proposição 2: Y transforma $E_x \subset S \subset R_+^n$ em uma esfera $E_y \subset \text{Im}(Y) \subset R_+^n$ dada por:

$$E_y = \{y \in R^n \text{ tal que } (y - e)^T(y - e) = 1\}; (\bar{y} = e).$$

Prova:

$\forall x \in S$, $\exists! y \in R_+^n$ tal que $y = Y(x) = \bar{X}^{-1}x$ pois Y é bijetiva. Assim;

$$\begin{aligned} x \in E_x &\Leftrightarrow (x - \bar{x})^T \bar{X}^{-2}(x - \bar{x}) = 1 && \Leftrightarrow \\ (\bar{X}y - \bar{X}e)^T \bar{X}^{-2}(\bar{X}y - \bar{X}e) &= 1 \text{ (pois } x = \bar{X}y) && \Leftrightarrow \\ (y - e)^T(y - e) &= 1. \end{aligned}$$

Logo, pontos do elipsóide $E_x \subset S$ são mapeados para a esfera $E_y \subset \text{Im}(Y)$ de centro "e" pela transformação Y .

Pelo proposição 2 temos que $x - \bar{x} = \bar{X}(y - e)$.

Se considerarmos $h \in R^n$ tal que $h = y - e$; teremos que $d = \bar{X}h$, assim (7) é equivalente ao seguinte problema transformado:

$$\begin{aligned} &\text{minimizar} && \bar{c}^T h \\ &\text{sujeito a} && \begin{cases} \bar{A}h = 0 \\ h^T h = 1 \end{cases} \end{aligned}$$

cuja solução é;

$$h = -\frac{\bar{c}}{p}.$$

A direção h é a direção (3), já definida neste capítulo, usada pelo algoritmo primal afim.

Note que, poderíamos voltar em (7) com d determinado por:

$$d = -(\bar{X}^2 - \bar{X}^2 A^T [A \bar{X}^2 A^T]^{-1} A \bar{X}^2) c; \quad (8)$$

e definir um algoritmo equivalente ao afim escala primal, trabalhando diretamente com o espaço das direções factíveis do problema (1), com a direção de busca (3) sendo obtida por resolver (5) com $V(\bar{x})$ definida por (6).

Como já foi visto, isto é feito implicitamente pois, a estratégia do algoritmo primal afim não é esta, já que ele trabalha diretamente no espaço transformado, para depois devolver um ponto factível ao espaço inicial.

Um algoritmo equivalente ao algoritmo "primal afim" (algoritmo de Dikin), que trabalha diretamente no espaço das direções factíveis do problema (1), utilizando a direção de busca (8) na determinação de novos pontos será visto a seguir.

3 - Algoritmo dos quadrados mínimos ponderados (Chandru/Kochar).

Baseado em [Chandru and Kochar, (1985)], daremos o próximo algoritmo primal, que possibilitou a demonstração de convergência aos algoritmos primais, para passos curtos, como já havia demonstrado Dikin em 1967.

Além disso, uma extensão deste algoritmo para variáveis canalizadas será objeto do próximo capítulo.

3.1 - Preliminares.

Consideremos o problema (1) e seu problema dual:

$$\begin{aligned} &\text{maximizar } b^T w \\ &\text{sujeito a } A^T w \leq c; \end{aligned} \quad (9)$$

dados x primal factível e w dual factível, respectivamente para (1) e (9), a medida do desvio de complementação é dada por $\| \Gamma_{xw} \|$

onde;

$$\Gamma_{xw} = X (c - A^T w) ; X = \text{diag}(x_1, \dots, x_n).$$

Pela condição iii) do teorema das folgas complementares, (capítulo 1, seção 2), se tivermos $\| \Gamma_{xw} \| = 0$, x e w são complementares e soluções ótimas para (1) e (9), respectivamente.

Vejam os como a partir de um ponto $\bar{x}$ dado, podemos calcular um ponto $\bar{w}$, objetivando $\| \Gamma_{xw} \| = 0$.

3.2 - Determinação de $\bar{w}$.

Chandru e Kochar propuseram a seguinte estratégia:

dado $\bar{x}$ primal factível e interior, determine $\bar{w}$ que;

$$\underset{w}{\text{minimize}} \| \Gamma_{xw}^- \|^2,$$

na tentativa de satisfazer a complementação das folgas. Caso isto não ocorra, $\bar{w}$ será utilizada para definir uma direção de descida para $\bar{x}$.

Ao resolvermos:

$$\underset{w}{\text{minimizar}} \| \Gamma_{xw}^- \|^2 = \underset{w}{\text{minimizar}} \| \bar{X}c - \bar{X}A^T w \|^2, \quad (10)$$

obtemos a seguinte solução :

$$\bar{w} = (A\bar{X}^2 A^T)^{-1} A\bar{X}^2 c;$$

que é a solução quadrados mínimos para o problema (10).

3.3 - A direção de busca.

O algoritmo utiliza a seguinte direção de busca a cada iteração:

$d \in R^n$ tal que $d = -\bar{X}^2 \tilde{c}$, onde $\tilde{c} = c - A^T \bar{w}$.

Observamos que:

$$\begin{aligned} d &= -\bar{X}^2 \tilde{c} = -\bar{X}^2 (c - A^T [A\bar{X}^2 A^T]^{-1} A\bar{X}^2 c) = \\ &= -[\bar{X}^2 - \bar{X}^2 A^T (A\bar{X}^2 A^T)^{-1} A\bar{X}^2] c = \bar{X}h, \text{ por (8),} \end{aligned}$$

ou seja, a direção definida por Chandru e Kochar é equivalente à direção (8) definida na seção 2. Portanto d é múltipla da solução do problema:

$$\begin{aligned} &\underset{d}{\text{minimizar}} c^T d \\ &\text{sujeito a } \begin{cases} Ad = 0 \\ d^T d = 1 \end{cases} \end{aligned}$$

3.4 - Algoritmo.

- 1- Dado x^0 tal que $Ax^0 = b$ e $x^0 > 0$; faça $k = 0$;
- 2- Calcule $w^k = (A\bar{X}_k^2 A^T)^{-1} A\bar{X}_k^2 c$;
com $\bar{X}_k = \text{diag}(x_1^k, \dots, x_n^k)$;
- 3- Calcule $\tilde{c} = c - A^T w^k$ e pare se $\tilde{c} \geq 0$ e
 $\| \Gamma_{x^k w^k} \| \cong 0$, caso contrário vá para 4.
- 4- Calcule
 $c^k = X_k^2 \tilde{c}$ e faça
 $d = -c^k$.
- 5- Faça
 $x^{k+1} = x^k + \bar{\epsilon} d$ onde
 $\bar{\epsilon} = \alpha \epsilon$; $0 < \alpha < 1$;
com $\epsilon = \min\{-x_j^k / d_j \mid \text{tal que } d_j < 0; j=1 \dots n\}$
e volte em 2 com $k = k + 1$.

4 - Convergência dos algoritmos com restrição de igualdade.

Partindo da hipótese da não degeneração do P.P.L. primal, ou seja, o conjunto solução de $\{x \in \mathbb{R}^n \mid Ax = b; x \geq 0\}$, é não degenerado, Chandru e Kochar demonstraram o seguinte teorema para passos curtos (considerando $z = c^T x$):

Teorema 1: A sequência $\{c^T x^k\}$ converge para o valor ótimo z^* do problema (1), se o problema admite solução ótima finita.

Para a demonstração deste teorema, algumas proposições foram utilizadas e demonstradas por Chandru e Kochar, vistas a seguir.

Proposição 3: Se tomarmos $\epsilon = (r_k)^{-1/2}$, onde $r_k = \Gamma_{x^k w^k}^T \Gamma_{x^k w^k}$, então, ϵ é um passo factível a cada iteração k do algoritmo.

Prova:

$$\epsilon_{\max} = -x_j^k / d_j = x_j^k / c_j^k; \text{ com } j, \text{ índice bloqueio de fronteira.}$$

Como $c_j^k = (x_j^k)^2 \tilde{c}_j$, temos;

$$\epsilon_{\max} = 1/(x_j^k \tilde{c}_j) \geq 1/\|X_k \tilde{c}\| = (r_k)^{-1/2}. \quad \square$$

Proposição 4: A diferença entre os valores da função objetivo na iteração k e na iteração $k+1$ é limitado inferiormente por $(r_k)^{1/2}$, ou seja;

$$c^T x^k - c^T x^{k+1} \geq (r_k)^{1/2}$$

Prova:

$$\begin{aligned} r_k &= \Gamma_{xw}^{-T} \Gamma_{xw}^{-} = \tilde{c}^T \bar{X}^{-2} \tilde{c} = (c^T - \bar{w}^T A) \bar{X}^{-2} (c - A^T \bar{w}) = \\ &= c^T \bar{X}^{-2} (c - A^T \bar{w}) - \bar{w} A \bar{X}^{-2} (c - A^T \bar{w}) = \\ &= c^T c^k + \bar{w} A d = c^T c^k, \quad \text{pois } d \in N(A). \end{aligned}$$

assim, $c^T x^k - c^T x^{k+1} = \epsilon c^T c^k \geq (r_k)^{-1/2} c^T c^k = (r_k)^{-1/2} r_k = (r_k)^{1/2}$.

Portanto, $c^T x^k - c^T x^{k+1} \geq (r_k)^{1/2}$.

Antes de enunciarmos a próxima proposição, consideremos a seguinte técnica de perturbação para uma iteração k do algoritmo 3.4:

Calcule $\bar{X}'_k = \text{diag}(x_j'^k; j = 1, \dots, n)$, onde:

$$x_j'^k = \begin{cases} x_j^k & \text{se } x_j^k \geq \sigma \text{ ou se } \tilde{c}_j > -2\delta \\ \sigma & \text{caso contrário} \end{cases}$$

onde δ e σ são pequenas tolerâncias positivas definidas a "priori".

Faça $\tilde{c}' = (I - A^T (\bar{A} \bar{X}'_k{}^2 A^T)^{-1} \bar{A} \bar{X}'_k{}^2) c$ e

$x^{k+1} = x^k - \epsilon c^k$, onde $c^k = \bar{X}'_k{}^2 \tilde{c}'$.

Para esta técnica de perturbação é válida a proposição demonstrada em [Chandru e Kochar, (1985)]:

Proposição 5: Admitindo-se a não degeneração primal de (1), para todo $\delta = 2^{-O(L)}$ existe $\sigma = 2^{-O(L)}$ tal que $\tilde{c}'_j \leq -\delta$ quando $\tilde{c}_j \leq -2\sigma$.

L : é o comprimento da representação binária da entrada dos dados do P.P.L. (1).

O uso desta técnica de perturbação será necessário para obtermos um limite inferior no decréscimo da função objetivo ao analisarmos $c^T x^k - c^T x^{k+1}$, quando $\tilde{c}_j < -2\sigma$, ou seja, w^k é infactível para o problema (9).

Proposição 6: Se $\tilde{c}_j > -2\delta$, com $\delta = 2^{-O(L)}$; então,

$$c^T x^k - c^T x^{k+1} \geq 2\delta^2.$$

Prova:

Mostremos que $c^T x^k - c^T x^{k+1} \geq 2\delta^2$.

Desde que $-X_k^2 \tilde{c}$ é uma direção de descida para $c^T x^k$ e:

$$c^T c^k = c^T X_k^2 \tilde{c} = \Gamma_{x^k w^k}^T \Gamma_{x^k w^k} = r_k = \|X_k \tilde{c}\|^2;$$

então, para um passo $\epsilon \geq 1/\|X_k \tilde{c}\|$ temos que:

$$\begin{aligned} c^T (x^{k+1} - x^k) &= -\epsilon c^T X_k^2 \tilde{c} = -\epsilon \|X_k \tilde{c}\|^2 \leq \\ &\leq (-\|X_k \tilde{c}\|^2)/\|X_k \tilde{c}\| = -\|X_k \tilde{c}\| \leq -\sigma 2\delta \leq -2\delta^2, \end{aligned}$$

o que conclui a prova. $\square$

Agora, estamos prontos para demonstrar o principal resultado, o teorema da convergência. Para isto, dois casos foram analisados por Chandru e Kochar.

Caso 1: $(\tilde{c}_j \geq -2\delta ; \forall j = 1, \dots, n)$.

Neste caso $\tilde{c}' = \tilde{c}$ e mostra-se através de manipulações algébricas e uso das proposições anteriores que:

$$\left\{ \frac{c^T x^{k+1} - z^*}{c^T x^k - z^*} \right\} \leq 1 - 2/n.$$

Prova:

Considere o par de problema duais:

$$\text{minimizar } \{c^T x: Ax = b, 0 \leq x \leq u\} \quad (11)$$

$$\text{maximizar } \{b^T w - u^T v: A^T w + Iy - Iv = c\} \quad (12)$$

Sem perda de generalidade, assumimos $u_j = 2^I$, $j = 1, \dots, n$; onde I : é um número inteiro mais próximo de $O(L)$ tal que $I < O(L)$.

Dado x^k factível para (11) e y^k estimador dual do algoritmo, pode-se construir uma solução (y^k, v^k, w^k) factível em (12), da seguinte forma:

$$\begin{cases} w_j^k = \max(0, \tilde{c}_j) \\ v_j^k = \max(0, -\tilde{c}_j); \end{cases} \quad j = 1, \dots, n$$

então a folga para a otimalidade pode ser limitada como segue:

$$c^T x^k - z^* \leq c^T x^k - (b^T y^k - u^T v^k) = (v^k)^T (u - x^k) + \\ + (w^k)^T x^k + (y^k)^T (Ax^k - b) = \sum_{\tilde{c}_j < 0} (-\tilde{c}_j) (u_j - x_j^k) + \sum_{\tilde{c}_j > 0} \tilde{c}_j x_j^k.$$

Desde que:

$$i) \sum_{\tilde{c}_j < 0} (-\tilde{c}_j) (u_j - x_j^k) \leq \sum_{\tilde{c}_j > 0} 2\delta 2^I = 2n\delta 2^I < 2n 2^{-I} 2^I = 2n.$$

$$ii) \sum_{\tilde{c}_j > 0} x_j^k \tilde{c}_j = \left(\sum_{\tilde{c}_j > 0} (x_j^k \tilde{c}_j)^2 \right)^{1/2} \leq \left(\sum_{\tilde{c}_j > 0} r_k \right)^{1/2} = (nr_k)^{1/2} \leq \\ \leq n(r_k)^{1/2};$$

$$\text{obtemos que } c^T x^k - z^* \leq 2n + n(r_k)^{1/2}.$$

Além disso, $(r_k)^{1/2} \geq \delta^2$, pois:

$$(r_k)^{1/2} = \|X_k \tilde{c}\| \geq \max_j |x_j^k \tilde{c}_j| \geq \sigma\delta = \delta^2.$$

Logo:

$$\frac{c^T x^{k+1} - z^*}{c^T x^k - z^*} = \frac{c^T x^{k+1} - c^T x^k}{c^T x^k - z^*} + 1 \leq 1 - \frac{2\delta^2}{n + n(r_k)^{1/2}} \leq \\ \leq 1 - 2\delta^2/n - 2\delta^2/n\delta^2 \leq 1 - 2/n.$$

Como x^k é ponto interior e $c^T x^k \geq z^*$, para $k \rightarrow \infty$ então:

$$c^T x^k > z^*.$$

Portanto, $\{c^T x^k\} \rightarrow z^*$, para $k \rightarrow \infty$, como queríamos provar.

Caso 2: $(\tilde{c}_j \leq -2\delta$ para pelo menos um $j = 1, \dots, n)$

Observarmos que, este é o pior caso, já que $\tilde{c}_j \geq 0$ indica a factibilidade dual do algoritmo e a infactibilidade de w_j^k pode ser preocupante.

Neste caso, pode-se somente afirmar que o decréscimo do valor da função objetivo é limitado inferiormente por δ^2 , pois:

$$c^T x^k - c^T x^{k+1} \geq (r_k)^{1/2} \geq \max_j |x_j^k \tilde{c}_j| \geq \sigma\delta = \delta^2, \text{ (proposição 5).}$$

Isto determina, um decréscimo muito pequeno da diferença entre $c^T x^k$ e $c^T x^{k+1}$, o que acarretará um aumento no número de iterações exponencial em L para a convergência do algoritmo.

Sobre a hipótese da não degenerescência primal, em [Chandru and Kochar, (1985)] encontram-se demonstrados os seguintes teoremas:

Teorema 3: $\lim_{k \rightarrow \infty} \{ r_k \} = 0$;

ou seja, as folgas complementares serão satisfeitas com o processo iterativo.

Teorema 4: $\lim_{k \rightarrow \infty} -\inf \{ \tilde{c}(x^k) \} \geq 0$;

ou seja, a factibilidade dual é alcançada com o processo iterativo.

Proposição 7: existe uma subsequência k de estimadores duais tal que $\lim_{k \rightarrow \infty} \{ w^k \} = w^*$, solução ótima de (10).

Isto quer dizer que, se o primal alcança a solução ótima no processo iterativo para (1), também a solução ótima de (10) é encontrada.

Observação: Recentemente, Gonzaga (1990), publicou a demonstração de convergência dos algoritmos primais para passos longos, admitindo a não degeneração primal do problema e utilizando a prova feita por Dikin (1967).

5 - A complexidade dos algoritmos com restrição de igualdade.

No trabalho que Karmarkar publicou em 1984, um resultado importante foi que seu algoritmo tinha complexidade de $O(n^{3.5}L^2)$ obtida pelos seguintes resultados:

i) $O(n^{2.5}L)$: tempo de execução gasto pelo algoritmo na resolução do problema dos quadrados mínimos, que envolve "inversões" de matrizes.

ii) $O(nL)$: n^0 máximo de iterações até se obter a solução ótima; onde $L = L + n + 1$.

O resultado i) já era conhecido e a demonstração de ii) feita por Karmarkar é de difícil entendimento. Preferimos utilizar, aquela encontrada em [Gonzaga, (1988), (1989)], como segue abaixo.

Se considerarmos o problema:

$$\begin{aligned} &\text{minimizar } z(x) = c^T x + \mu p(x); \mu: \text{ parâmetro fornecido.} \\ &\text{sujeito a } Ax = b, \end{aligned}$$

com $p(x) = -\sum_{i=1}^n \log(x_i)$, recaímos no clássico método de barreiras,

conhecido como método das barreiras logarítmica, analisado por Frisch em 1955.

O fato de $p(x)$ só ser definida se $x \in R_{++}^n$ nos possibilita abandonar a restrição de não negatividade do problema (1), em oposição aos métodos anteriores onde regiões de confiança eram necessárias. Isto faz com que se trabalhe iterativamente com soluções interiores ao conjunto viável de (1). Este clássico método pode ser utilizado então para evitarmos a fronteira do conjunto viável e a heurística utilizada nos algoritmos anteriores pode ser abandonada.

Observe que, a função z , controlada pelo parâmetro μ , tende a $+\infty$ quando x se aproxima da solução ótima, a qual está na fronteira.

Para $x \in R_{++}^n$, temos que:

$$\nabla p(x) = (x_1^{-1}, \dots, x_n^{-1}),$$

$$\nabla^2 p(x) = \text{diag}[x_i^{-2}];$$

no ponto "e", $\nabla p(e) = -e$, $\nabla^2 p(e) = I$, o que motiva a fazermos uma mudança de escala que leve x^k em "e", já que a função barreira é muito simples neste ponto.

Assim, se tomarmos $x = X_k y$, obtemos:

$$p(y) = p(X_k^{-1} x) = -\sum_{i=1}^n \log(x_i/x_i^k) = -\sum_{i=1}^n \log(x_i) + \sum_{i=1}^n \log(x_i^k), \text{ ou}$$

seja, $p(y) = p(x) + \text{cte.}$

Estudar a função próxima de "e" não muda as propriedades diferenciais da função, pois, $p(y)$ nada mais é que $p(x)$ deslocada por uma constante.

Isto baseia-se no seguinte teorema demonstrado por Gonzaga:

Teorema 5: Para $h \in R^n$ tal que $\|h\| \leq 0.5$; $e + h \in R_{++}^n$, a função barreira pode ser escrita como:

$$p(e + h) = -\nabla p(e)^T h + O(h) = -e^T h + O(h), \quad \text{com} \\ 0 \leq O(h) \leq \|h\|^2.$$

prova:

- i) $O(h) \geq 0$ segue diretamente de $p(\cdot)$ ser convexa.
 ii) Se considerarmos a função logarítmica, desenvolvida em série de Taylor nas proximidades do ponto $(1 + \lambda) \in \mathbb{R}$, temos:

$$\log(1 + \lambda) = \lambda + O_1(\lambda); \quad O_1(\lambda) = -\lambda^2/2 + \lambda^3/3 - \dots$$

Se $|\lambda| < 1$, é fácil mostrar que $|O_1(\lambda)| \leq \lambda^2/2(1 - |\lambda|)$; se utilizarmos este resultado para a função barreira, em torno de "e" para o vetor $e + h \in \mathbb{R}^n$ tal que $\|h\|_\infty < 1$, temos,

$$p(e + h) = -e^T h + O(h) \quad \text{onde} \quad 0 \leq O(h) \leq \|h\|^2 / 2 * (1 - \|h\|_\infty),$$

$$\text{pois } p(e + h) = -\sum_{i=1}^n \log(1 + h_i) = -\sum_{i=1}^n h_i + \sum_{i=1}^n O_1(h_i).$$

$$|O(h)| \leq \sum_{i=1}^n |O_1(h_i)| \leq \sum_{i=1}^n h_i^2 / 2 * (1 - \|h\|_\infty) =$$

$$\|h\|^2 / 2 * (1 - \|h\|_\infty).$$

Se considerarmos $\|h\|_\infty \leq 0.5$, obtemos então;

$p(e + h) = -e^T h + O(h)$ e $0 \leq O(h) \leq \|h\|^2$, que é o resultado desejado.

5.1 - A função potencial e o algoritmo afim polinomial.

Se considerarmos a seguinte hipótese forte:

$c^T x \geq 0$, $\forall x$ factível e que na otimalidade $c^T x^* = 0$, podemos definir a seguinte função;

$$f(x) = q \log(c^T x) - \sum_{i=1}^n \log(x_i), \quad q \in \mathbb{Z}, \quad q \geq n;$$

de tal forma que, na otimalidade $f(x^k) \rightarrow +\infty$, pois, $c^T x^k \rightarrow c^T x^*$.

Esta função foi chamada por Karmarkar como "função potencial" e foi utilizada na demonstração de complexidade do seu algoritmo, com $q = n$.

Vale observar que, pode-se remover a hipótese forte acima e trabalhar com limitantes inferiores $v^k = b^T w^k$, w^k : estimador dual, de tal forma que, $c^T x^k - v^k \geq 0$ e na otimalidade $v^* = c^T x^*$. Isto é encontrado em [Todd e Burrell, (1985)]

A função potencial obtida após a mudança de escala será:

$$g(y) = q \log((c^k)^T y) + p(y).$$

Para esta função, a direção de busca para a determinação do novo ponto $e + \lambda h$, $h \in \mathbb{R}^n$, $\lambda > 0$ (o passo λ é agora determinado por algum método de busca unidimensional), é dada por:

$$h = -P_A^{-1} \nabla g(e) = - (q c_p) / (c^k)^T e - e_p; \text{ onde}$$

$$c_p = P_A^{-1} c^k, \quad e_p = P_A^{-1} e.$$

Com esta nova estratégia na determinação da direção de busca, controlada pela utilização da "função potencial", este algoritmo recebeu o nome de algoritmo "afim polinomial" e possibilitou a demonstração de complexidade para os algoritmos de pontos interiores para problemas com restrição de igualdade, de uma maneira menos complicada que a demonstração feita por Karmarkar.

Para o novo ponto $e + \lambda h$ é válido o seguinte teorema que mostra o bom comportamento de $g(\cdot)$ nas proximidades de $e + \lambda h$.

Teorema 6: Se $\lambda \|h\|_\infty \leq 0.5$ então $g(e + \lambda h) \leq g(e) + \nabla g(e)^T h + \lambda^2 \|h\|^2$.

Prova:

Desde que $g(y) = g_1(y) + p(y)$, com $g_1(y) = q \log(c^T y)$; este resultado segue diretamente do teorema 5 e do fato de função $g_1(\cdot)$ ser côncava, ou seja, $g_1(e + \lambda h) \leq g_1(e) + \lambda \nabla g_1(e)^T h$. $\square$

Antes de enunciarmos o próximo teorema, vejamos a seguinte proposição.

Proposição 8: Considere a direção de busca $h = -P_A^{-1} \nabla g(e)$.

Se $q \geq n + \sqrt{n}$, então $\|h\| \geq 1$.

Antes de obtermos o resultado principal, façamos as seguintes hipóteses sobre o conjunto de restrições, (considerando também a hipótese forte $c^T x^* = 0$):

- i) Todos os dados de entrada são inteiros;
- ii) O conjunto viável é limitado, portanto, compacto e possui pontos interiores. Isto garante que existe pelo menos uma solução ótima e pelo teorema fundamental da programação linear, existe um vértice ótimo;
- iii) O conjunto de soluções é não degenerado;

Consideremos o teorema 8:

Teorema 8: se x é um vértice do conjunto viável, então:

- i) para $i = 1, \dots, n$; $x_i = 0$ ou $x_i > 2^{-L}$
- ii) $c^T x = 0$ ou $c^T x > 2^{-L}$.
- iii) Para qualquer ponto viável x , $c^T x \leq 2^L$; $e^T x \leq 2^L$.

prova:

O número L , que relaciona as grandezas do problema com o seu tamanho em "bits", é dado de uma forma explícita por:

$$L = \sum_{i=1}^m \sum_{j=1}^n \log_2(|a_{ij}| + 1) + \sum_{i=1}^m \log(|b_i| + 1) + \sum_{j=1}^n \log_2(|c_j| + 1)$$

Obviamente, 2^L é um número grande, ao passo que 2^{-L} é um número muito pequeno.

Dada uma solução básica factível $x = [x_B, x_N]$ onde $x_N = 0$ e $A = [B, N]$ tal que $\exists B^{-1}$, então:

- i) $x_B = B^{-1}b = ([\text{cof}(B)]^T b) / \det(B) \Rightarrow x_{B_j} > 2^{-L}$; pois a base é não degenerada e $\det(B) < 2^L$; ($x_{B_j} = 0$ se o numerador é nulo).

- ii) se $x = 0 \Rightarrow c^T x = 0$, caso contrário, desde que $c^T x \geq 0$ e c_j é inteiro, para $j = 1, \dots, n$, é válido:

$$c^T x = \|c^T x\| \geq \|c^T x\|_{\infty} = \max_{j=1..n} |c_j x_j| = |c_p| x_p \geq 1 * 2^{-L} = 2^{-L}, \text{ pois } x_p \geq 2^{-L} \text{ por i).}$$

$$c^T x = c_B^T x_B = c_B^T B^{-1} b = (c_B^T (\text{cof } B)^T b) / \det B < 2^L;$$

$$e^T x \leq n 2^L \leq 2^L$$

Prova:

Seja $y^* = e + \lambda^* h$ tal que $c^T y^* = 0$, ou seja, y^* é solução ótima do problema escalado.

Como $y^* - e \in N(A)$, temos:

$$h^T(y^* - e) = -P_A^- \nabla g(e)^T(y^* - e) = -\nabla g(e)^T(y^* - e).$$

Por outro lado, sabemos que:

$$-\nabla g(e)^T(y^* - e) = [(-q \ c_p / c^T e) + e]^T(y^* - e).$$

Comparando os dois resultados obtemos;

$$h^T(y^* - e) = q - n + e^T y^* \text{ já que } c^T y^* = 0 \text{ e } e^T e = n.$$

$$e^T y^* = \|y^*\|_1 \geq \|y^*\|, \text{ pois } y^* \geq 0.$$

$$\text{Daí, } h^T(y^* - e) \geq q - n + \|y^*\| \Rightarrow \|h\| \|y^* - e\| \geq q - n +$$

$$+ \|y^*\| \Rightarrow \|h\| (\|y^*\| + \|e\|) \geq q - n + \|y^*\| \Rightarrow$$

$$\|h\| \geq (q - n + \|y^*\|) / (\sqrt{n} + \|y^*\|);$$

Para $q \geq n + \sqrt{n}$, temos que $\|h\| \geq 1$. $\square$

Já que $g(.) \rightarrow -\infty$ é interessante para atingir a otimalidade, o teorema abaixo mostra que, iterativamente, o decréscimo de $g(.)$ é razoavelmente bom, se considerarmos $q \geq n + \sqrt{n}$ e a direção de busca $h = -P_A^- \nabla g(e)$.

Teorema 7: Considere a direção de busca $h = -P_A^- \nabla g(e)$.

Se $q \geq n + \sqrt{n}$ então, $g(e + \bar{\lambda}h) - g(e) \leq -0.2$; onde $\bar{\lambda} = \min_{\lambda} g(e + \lambda h)$.

Prova:

$$\text{Se } \bar{\lambda} = \min_{\lambda} g(e + \lambda h) \text{ e } \bar{y} = e + \bar{\lambda}h \Rightarrow g(\bar{y}) \leq g(e + \lambda h).$$

Em particular, para $\lambda \in [0; 0.5/\|h\|]$, temos $\lambda \|h\|_{\infty} \leq 0.5$.

$$g(\bar{y}) \leq g(e + \lambda h) \leq g(e) + \lambda \nabla g(e)^T h + \lambda^2 \|h\|^2, \text{ pelo teorema 6.}$$

$$\text{Como } h \in N(A), \nabla g(e)^T h = -\|h\|^2 e.$$

$$g(\bar{y}) - g(e) \leq -\lambda \|h\|^2 (1 - \lambda) \text{ se } \lambda = 0.3/\|h\|.$$

$$\text{Se } q \geq n + \sqrt{n} \Rightarrow \|h\| \geq 1 \Rightarrow g(\bar{y}) - g(e) \leq -0.2. \quad \square$$

pois todos os denominadores são inteiros e os numeradores tem elementos menores ou iguais a 2^L . $\square$

O teorema assegura que se encontrarmos um vértice com custo menor que 2^{-L} , ele é uma solução ótima do problema.

Finalmente, chega-se ao resultado principal, dado a seguir, que mostra que o ponto x^* é alcançado em $O(qL)$ iterações.

Teorema 9: Dado um ponto inicial x^0 tal que $x_1^0 > 2^{-L}$, então o algoritmo determina a solução ótima em $O(qL)$ iterações.

prova:

$$p(x^0) = \sum_{i=1}^n -\log(x_i^0) \leq \sum_{i=1}^n L \log 2 < nL < nL.$$

Pelo teorema 7, podemos afirmar que: $f(x^k) \leq f(x^{k-1}) - 0.2$,
então, $f(x^k) \leq f(x^0) - 0.2k$.

Pelo teorema 8 é válido que: $x_1 \leq 2^L$ e então $\sum_{i=1}^n \log(x_i^k) \leq nL$;

$$\text{Além disso, } f(x^0) = q \log(c^T x^0) - \sum_{i=1}^n \log(x_i^0) < q \log(c^T x^0) + nL$$

$$< (q + n)L \text{ e } q \geq n.$$

Logo,

$$f(x^k) = q \log(c^T x^k) - \sum_{i=1}^n \log(x_i^k) < f(x^0) - 0.2k \text{ e então}$$

$$q \log(c^T x^k) < f(x^0) + \sum_{i=1}^n \log(x_i^k) - 0.2k < (q + n)L + nL - 0.2k \\ < 3qL - 0.2k.$$

Se tomarmos $k = 20 qL$, ou seja, $k = O(qL)$ obtemos:

$$\log(c^T x^k) < -L \text{ e então } c^T x^k < e^{-L} < 2^{-L}.$$

Portanto, o problema é resolvido em $O(qL)$ iterações. $\square$

6 - O algoritmo de Karmarkar.

Como este algoritmo já vem sendo comentado desde o início do trabalho, não poderíamos deixar de mencioná-lo aqui. Isto será feito de maneira simples, sem entrarmos em maiores detalhes.

A denominação original deste algoritmo, como já foi escrito, é algoritmo "das transformações projetivas".

6.1 - Preliminares.

Desde que façamos algumas manipulações algébricas no problema (1), como a descrita em [Dodani e Babu, (1987)], este pode ser transformado no seguinte problema - "problema na forma homogênea":

$$\begin{array}{ll} \text{minimizar} & c^T x \\ \text{sujeito a} & \begin{cases} Ax = 0 \\ e^T x = 1 \\ x \geq 0 \end{cases} \end{array} \quad (13)$$

Se considerarmos $S = \{x \in \mathbb{R}_+^n \text{ tal que } Ax = 0\}$;

$$\Delta = \{x \in \mathbb{R}_+^n \text{ tal que } e^T x = 1\},$$

obviamente, se $x \in \mathbb{R}_+^n$ é uma solução factível de (13) então $x \in S \cap \Delta$, (ver figura 1).

figura 1.

Pela figura vemos que, se um ponto $x' \in S$, ou seja, $Ax' = 0$ e $x' \geq 0$, este ponto será factível para (13) desde que façamos sua projeção no subspaço $S \cap \Delta$, que chamaremos de "projeção cônica".

Isto é simples de se fazer, basta que:

$$x = P(x') = x' / e^T x'.$$

6.2 - A mudança de escala.

Karmarkar definiu uma transformação mudança de escala diferente daquela utilizada pelos métodos afins:

$$T: \mathbb{R}^n \rightarrow \mathbb{R}^n \text{ tal que } y = T(x) = \bar{X}^{-1}x / e^T \bar{X}^{-1}x, \\ \text{com } \bar{X} = \text{diag}(\bar{x}_1, \dots, \bar{x}_n).$$

Analogamente à transformação $P(\cdot)$ definida anteriormente, $T(\cdot)$ é uma transformação projetiva com as seguintes propriedades:

P_1) $T(\cdot)$ é bijetiva e sua inversa é dada por:

$$x = T^{-1}(y) = \bar{X}y / e^T \bar{X}y$$

e (13) é transformado no seguinte problema equivalente:

$$\begin{aligned} &\text{minimizar } c^T \bar{X}y / e^T \bar{X}y \\ &\text{sujeito a } \begin{cases} A\bar{X}y / e^T \bar{X}y = 0 \\ e^T \bar{X}y / e^T \bar{X}y = 1 \\ y \geq 0; \end{cases} \end{aligned} \quad (14)$$

além disso, $\bar{y} = T(\bar{x}) = e/n$, que é o centro geométrico de Δ' , onde

$$\Delta' = \{y \in \mathbb{R}^n \text{ tal que } e^T \bar{X}y / e^T \bar{X}y = 1\}.$$

P_2) Se considerarmos o elipsóide:

$$E_x = \{x \in \mathbb{R}_+^n \text{ tal que } (x - \bar{x})^T \bar{X}^{-2} (x - \bar{x}) = 1\} \subset S,$$

temos que:

$$E_x \cap \Delta = \{x \in \mathbb{R}_+^n \text{ tal que } (x/e^T \bar{X}^{-1}x - x/e^T \bar{X}^{-1}\bar{x})^T \bar{X}^{-2} \\ (x/e^T \bar{X}^{-1}x - x/e^T \bar{X}^{-1}\bar{x})\} \subset S \cap \Delta.$$

A aplicação direta de T em $E_x \cap \Delta$ transforma este elipsóide na esfera:

$$E_y = \{y \in \mathbb{R}_+^n \text{ tal que } (y - e/n)^T (y - e/n) = 1\} \subset S', \text{ onde}$$

$$S' = \{y \in \mathbb{R}_+^n \text{ tal que } A\bar{X}y/e^T \bar{X}y = 0 \text{ e } e^T \bar{X}y/e^T \bar{X}y = 1\}.$$

Se assumirmos que:

- i) $Ae = 0$ tal que $\bar{x} = e$ é uma solução factível inicial;
- ii) $c^T e > 0$ pois se $c^T e = 0$ então e é solução ótima;
- iii) Na otimalidade $c^T x^* = 0$, (que é uma condição que pode ser relaxada, como já comentamos na seção anterior); então o problema pode ser reescrito como:

$$\begin{aligned} & \text{minimizar } c^T \bar{X}y \\ & \text{sujeito a } \begin{cases} A\bar{X}y = 0 \\ e^T \bar{X}y = 1 \\ y \geq 0; \end{cases} \end{aligned} \quad (15)$$

pois $e^T \bar{X}y \neq 0$ e na otimalidade $c^T y^* = 0$.

6.3 - A função potencial.

É idêntica àquela utilizada por Gonzaga, apenas fazendo-se a substituição de q por n :

$$f(x) = n \log(c^T x) - \sum_{i=1}^n \log(x_i);$$

que após a mudança de escala se escreve:

$$g(y) = n \log(\bar{c}^T y) - \sum_{i=1}^n \log(y_i), \text{ onde } \bar{c} = \bar{X}c.$$

Para esta função é válida a seguinte proposição:

Proposição 9: A função $f(\cdot)$ é invariante sobre S ou $S \cap \Delta$.

Prova:

Seja $x \in S$ e $\lambda x \in S \cap \Delta$, $\lambda \in \mathbb{R}$.

$$\begin{aligned} f(\lambda x) &= \log(\bar{c}^T \lambda x / \prod_{i=1}^n \lambda x_i)^n = \log(\lambda \bar{c}^T x / \lambda \prod_{i=1}^n x_i)^n = \\ &= \log(\bar{c}^T x / \prod_{i=1}^n x_i) = f(x). \end{aligned}$$

Este fato é importante pois permite definir o algoritmo ignorando temporariamente a restrição $e^T x = 1$, até atingirmos a solução ótima x'^* em S , para depois fazer a projeção cônica deste ponto para $S \cap \Delta$:

$$x^* = P(x'^*) = x'^* / e^T x'^*.$$

6.4 - A direção de busca.

Se considerarmos o problema (15), podemos definir a seguinte direção de busca:

$$d = -c_p = -P_B \bar{X}c = -[I - B^T(BB^T)^{-1}B] \bar{X}c, \text{ onde;}$$

$$B = [A\bar{X}, e^T \bar{X}]^T, \text{ isto é direto de } A\bar{X}d = 0 \text{ e } e^T \bar{X}d = 1.$$

Podemos também definir $d = -P_B \nabla g(e)$, pois em [Gonzaga, (1988)] está demonstrado que com uma ou outra direção obtém-se os mesmos resultados.

6.5 - Algoritmo.

- 1- Dados $x^0 \in S$ e $k = 0$.
- 2- Enquanto um critério de parada não for satisfeito, faça:

$$\text{Calcule } c^k = X_k c; \quad A_k = A X_k; \quad e^k = X_k e;$$

$$\text{com } X_k = \text{diag}(x_1^k, \dots, x_n^k)$$

$$\text{Calcule } B = \begin{bmatrix} A_k \\ (e^k)^T \end{bmatrix};$$

Determine $d = -nc_p/c_p^T e + e$ (não utilizamos e_p pois e pertence ao núcleo de A);

$$\text{calcule } \bar{\lambda} = \min\{g(e + \lambda d) \text{ tal que } \lambda \geq 0 \text{ e } e + \lambda d > 0\}$$

(busca unidimensional)

$$\text{atualize a solução, } y = e + \bar{\lambda} d.$$

- 3- Determine a projeção cônica de y :

$$\bar{y} = X_k y / e^T X_k y.$$

No algoritmo observamos que, o passo 2 pode ser alterado da seguinte forma:

$$\text{Determine } d = -c_p;$$

$$\text{atualize a solução, } y = e/n + \alpha r d;$$

onde $\alpha \in (0,1)$ e $r = 1/\sqrt{n(n-1)}$ é o raio da maior esfera inscrita em Δ .

Determine a projeção cônica de y :

$$\bar{y} = X_k y / e^T X_k y,$$

antes de atualizar a solução.

6.6 - Critério de parada e complexidade.

Para demonstrar a complexidade do algoritmo, ou seja, que o algoritmo determina uma solução ótima em $O(nL)$ iterações, (resultado obtido por Karmarkar), Gonzaga demonstrou o seguinte teorema:

Teorema 10: Em qualquer iteração, c_p tem uma componente não positiva.

Isto garante que $\|d\| \geq 1$ pois $d = e - nc_p/c_p^T e$ tem pelo menos uma componente $d_i \geq 1$.

Com isto são válidos os teoremas abaixo, com demonstrações análogas àsquelas feitas na seção 5:

Teorema 11: Para uma iteração k do algoritmo,

$$f(x^{k+1}) \geq f(x^k) - 0.2.$$

Teorema 12: Desde que $q = n$ e o teorema 11 é satisfeito, então o algoritmo determina uma solução ótima em $O(nL)$ iterações.

Isto sugere o seguinte critério de parada:

"se em uma iteração k , $f(x^k) < -nL$, então pare a execução pois uma solução ótima foi encontrada".

Conclusão: Neste capítulo dedicado a problemas com restrições de igualdade, analisamos alguns algoritmos variantes do algoritmo de Karmarkar que possibilitaram-nos interessantes interpretações geométricas, análise de convergência e análise de complexidade para problemas em "formato primal".

Estes resultados podem ser estendidos aos algoritmos para problemas em "formato dual" devido à inter-relação existente entre os mesmos.

Apesar de teoricamente os algoritmos para restrições de igualdade possibilitarem a obtenção de excelentes resultados algébricos, tais algoritmos não trouxeram bons resultados computacionais na prática. Erros acumulativos na operação de projeção, a cada iteração podem gerar pontos infactíveis ao problema.

CAPÍTULO 5

PROBLEMAS COM VARIÁVEIS CANALIZADAS.

Neste capítulo, algoritmos para variáveis canalizadas são desenvolvidos e uma atenção especial é dada para uma extensão do algoritmo dos quadrados mínimos ponderados, de Chandru e Kochar, visto na seção 2, do capítulo 4, para problemas com restrições de igualdade e variáveis canalizadas. Uma interpretação geométrica define elipsóides a cada iteração, os quais podem não estar totalmente contidos na região viável, o que não ocorria com os algoritmos vistos até aqui.

1 - Algoritmo de Chandru/Kochar para variáveis canalizadas.

1.1 - Preliminares.

Consideremos o seguinte problema com variáveis canalizadas:

$$\begin{aligned} &\text{minimizar } c^T x \\ &\text{sujeito a } \begin{cases} Ax = b \\ l \leq x \leq u \end{cases} ; \end{aligned} \quad (1)$$

Sejam:

$$S = \{x \in R^n \text{ tal que } Ax = b \text{ e } l \leq x \leq u\};$$

$$H = \{x \in R^n \text{ tal que } l \leq x \leq u\}.$$

$$\text{int}(S) = \{x \in R^n \text{ tal que } Ax = b \text{ e } l < x < u\}.$$

Pode-se fazer uma extensão do algoritmo definido na seção 2, capítulo 2, desde que leve-se em consideração o hiperparalelepípedo definido por H , (no caso anterior exigia-se apenas $x \geq 0$), redefinindo-se a matriz mudança de escala $\bar{X}$, bem como o tamanho do passo a ser dado a cada iteração.

1.2 - A mudança de escala.

A definição que daremos para $\bar{X}$ a seguir considerará a cada iteração o elipsóide E_x totalmente contido no hiperparalelepípedo $l \leq x \leq u$, ou seja $E_x \subset H$, onde E_x será definida a seguir em analogia ao que está proposto e definido no capítulo 3.

Sejam $\bar{x} \in \text{int}(S)$ e $\bar{X} = \text{diag}(\bar{x}'_1, \dots, \bar{x}'_n)$ tal que

$$\bar{x}'_j = \begin{cases} \bar{x}_j - l_j & \text{se } \bar{x}_j \leq (u_j + l_j)/2 \\ \bar{x}_j - \bar{u}_j & \text{se } \bar{x}_j \geq (u_j + l_j)/2 \end{cases}$$

Esta definição de $\bar{x}'_j$ toma o limite inferior ou superior de $\bar{x}_j$, mais próximo de $\bar{x}_j$.

Se definirmos $f \in \mathbb{R}^n$ tal que:

$$f_j = \begin{cases} l_j & \text{se } \bar{x}_j \leq (u_j + l_j)/2 \\ u_j & \text{se } \bar{x}_j \geq (u_j + l_j)/2 \end{cases}$$

então podemos escrever $\bar{X} = \text{diag}(\bar{x}'_1 - f_1)$ e f será o vértice de H mais próximo de $\bar{x}$, (ver figura 1).

Consideremos o elipsóide E_x :

$$E_x = \{x \in \mathbb{R}^n \text{ tal que } (x - \bar{x})^T \bar{X}^{-2}(x - \bar{x}) = 1\}.$$

Para esta definição de $\bar{x}'_j$, o elipsóide $E_x \subset H$.

Observe que, os mesmos resultados seriam obtidos se considerássemos:

$$\bar{x}'_j = \min\{(\bar{x}_j - l_j); (u_j - \bar{x}_j)\};$$

sendo agora $\bar{x}'_j > 0$.

figura 1

Consideremos agora a seguinte transformação afim;

$$Y: \mathbb{R}^n \rightarrow \mathbb{R}^n \quad \text{tal que} \quad y = Y(x) = \bar{X}^{-1}(x - f).$$

Para esta definição de Y , como já foi visto no capítulo 3, podemos afirmar que, o novo sistema de variáveis y_1 , além de promover uma mudança de escala em x , tem como origem a imagem do vértice de f e pode também inverter a orientação do eixo coordenado uma vez que podemos ter $\bar{x}'_j < 0$.

Desde que $x = Y^{-1}(y) = \bar{X}y + f$, o problema (1) é equivalente a:

$$\begin{aligned} &\text{minimizar} \quad \bar{c}^T y + \{c^T f\} \\ &\text{sujeito a} \quad \begin{cases} \bar{A}y = b - Af \\ 0 \leq y \leq \text{diag}(|\bar{x}'_j|^{-1})(u - 1). \end{cases} \end{aligned} \quad (3)$$

Analogamente ao capítulo 3, são válidas também as seguintes proposições:

Proposição 1: O elipsóide E_x é transformado na esfera de centro "e" e raio unitário pela transformação Y .

Proposição 2: O hiperparalelepípedo H corresponde nas variáveis y_1 ao seguinte hiperparalelepípedo:

$$0 \leq y \leq \text{diag}(|\bar{x}'_j|^{-1})(u - 1).$$

Como estamos considerando o elipsóide E_x totalmente contido em H , então a esfera $(y - e)^T(y - e) = 1$ está totalmente contida em H^1 ; $H^1 = \{y \in \mathbb{R}^n \text{ tal que } 0 \leq y \leq \text{diag}(|\bar{x}'_j|^{-1})(u - 1)\}$.

1.3 - A direção de busca.

Se analisarmos o seguinte problema:

$$\begin{aligned} &\text{minimizar} \quad c^T d \\ &\text{sujeito a} \quad \begin{cases} Ad = 0 \\ d^T \bar{X}^{-2} d = 1; \end{cases} \end{aligned} \quad (4)$$

onde $d \in \mathbb{R}^n$ é a direção de busca para um ponto $\bar{x}$ factível em (1), $d = x - \bar{x}$, então podemos considerar a mudança de escala, $h = \bar{X}^{-1}d$ e obtermos o seguinte problema equivalente:

$$\text{minimizar } \bar{c}^T h$$

$$\text{sujeito a } \begin{cases} \bar{A}h = 0 \\ h^T h = 1, \end{cases}$$

já analisado no capítulo anterior e que nos dá a solução de (4):

$$d = -(\bar{X}^2 - \bar{X}^2 A^T (\bar{A} \bar{X}^2 A^T)^{-1} \bar{A} \bar{X}^2) c \in N(A), \quad (5)$$

Na próxima seção, analisaremos uma mudança de escala proposta em [Chandru e Kochar, (1986)], a qual é interessante pois possibilitará que trabalhem com elipsóides E_x que podem não estar totalmente contidos na região de factibilidade.

1.4 - A mudança de escala de Chandru\Kochar.

Em [Chandru e Kochar, (1986)], está definida a seguinte matriz mudança de escala:

$$\bar{X} = \text{diag} (\bar{x}'_1, \dots, \bar{x}'_n) \text{ onde}$$

$$\bar{x}'_j = \begin{cases} \bar{x}_j - l_j & \text{se } \tilde{c}_j > \delta \\ u_j - \bar{x}_j & \text{se } \tilde{c}_j < -\delta \\ [(\bar{x}_j - l_j)(u_j - \bar{x}_j)] / (u_j - l_j) & \text{se } |\tilde{c}_j| \leq \delta; \end{cases} \quad (6)$$

onde $\delta > 0$ (pequeno), $\tilde{c} = \bar{X}(c - A^T w^K)$ e w^K é uma solução aproximada do seguinte problema:

$$\begin{aligned} &\text{maximizar } b^T w \\ &\text{sujeito a } A^T w \leq c, \end{aligned}$$

que não é o dual do problema (1), e sim do seguinte problema:

$$\text{minimizar } \{x \in \mathbb{R}^n \text{ tal que } Ax = b \text{ e } x \geq 0\}.$$

Para esta definição de $\tilde{c}$, a direção de busca utilizada por Chandru e Kochar, $d = -\bar{X}^2 \tilde{c}$ é equivalente à direção (5) definida na seção anterior, de acordo com o capítulo 4, seção 3. Assim, a direção d é factível para (1), pois $d \in N(A)$.

Observe que, se ocorrer:

$$\bar{x}_j \geq (u_j + l_j)/2, \quad \tilde{c}_j > \delta,$$

desde que $d = -\bar{X}^2 \tilde{c}$ (o que mostra que d_j é influenciada diretamente por $\tilde{c}_j$), vemos claramente que $\bar{x}_j$ está mais próxima de u_j e no entanto a definição de $\bar{x}_j$ leva em consideração o limite inferior de $\bar{x}_j$, não se preocupando com a maior proximidade de $\bar{x}_j$ com u_j , uma vez que a componente $x_j = \bar{x}_j + \varepsilon d$, (ε é o tamanho do passo a ser visto na próxima seção), irá decrescer pois $d_j = -x_j^2 \tilde{c}_j < 0$.

Assim, se definirmos o elipsóide E_x , com $\bar{X}$ definido em (6), este pode não estar totalmente contido na região viável, pois $\bar{x}_j$ define o tamanho do eixo j para E_x e este não leva em conta a menor proximidade de $\bar{x}_j$ com u_j ou l_j , (ver figura 2).

figura 2

Outro fato que podemos observar é que, não é óbvio definir uma mudança de escala tal que uma relação entre as coordenadas escaladas e as coordenadas x_j levem o ponto $\bar{x}$ no ponto "e".

Esta pode ser esboçada da seguinte forma; $Y: S \rightarrow R^n$ tal que:

$$y_j = \begin{cases} \bar{x}_j^{-1} (x_j - l_j) & \text{se } \tilde{c}_j > \delta \\ \bar{x}_j^{-1} (u_j - x_j) & \text{se } \tilde{c}_j < -\delta \\ \bar{x}_j^{-1} t & \text{se } |\tilde{c}_j| \leq \delta \end{cases}$$

onde $t = [(u_j - x_j)(\bar{x}_j - l_j)] / (u_j - l_j)$.

Obviamente $\bar{y}_j = 1$ e $Y(\bar{x}) = e$.

1.5 - Condições de fronteira.

Para garantir que o novo ponto $\bar{x} + \varepsilon d$ seja factível é preciso que determinemos $\varepsilon \in \mathbb{R}_+$ tal que:

$$\begin{cases} A(\bar{x} + \varepsilon d) = b \\ l \leq \bar{x} + \varepsilon d \leq u. \end{cases}$$

Sejam:

$$\varepsilon^1 = \min\{-(\bar{x}_1 - l_1)/d_1 \mid \text{tal que } d_1 < 0\}$$

$$\varepsilon^2 = \min\{(u_1 - \bar{x}_1)/d_1 \mid \text{tal que } d_1 > 0\};$$

basta que tomemos $\varepsilon = \min\{\varepsilon^1, \varepsilon^2\}$ para garantirmos a factibilidade de $\bar{x} + \varepsilon d$.

A definição de ε leva pelo menos uma componente de $\bar{x} + \varepsilon d$ na fronteira do problema. Para manter o ponto interior, fazemos:

$$\bar{\varepsilon} = \alpha \varepsilon, \text{ com } 0 < \alpha < 1.$$

1.6 - Algoritmo para variáveis canalizadas.

1- Dado x^0 tal que $Ax^0 = b$ e $l \leq x^0 < u$; faça $k = 0$.

2- Calcule $w^k = (A\bar{X}_k^2 A^T)^{-1} A\bar{X}_k^2 c$ e

$$\tilde{c} = c - A^T w^k;$$

onde $X_k = \text{diag}(x_1^k, \dots, x_n^k)$; com

$$x_j^k = \begin{cases} x_j^k - l_j & \text{se } \tilde{c}_j > \delta \\ u_j - x_j^k & \text{se } \tilde{c}_j < -\delta \\ (x_j^k - l_j)(u_j - x_j^k)/(u_j - l_j) & \text{se } |\tilde{c}_j| \leq \delta \end{cases}$$

ou

$$x_j^k = \begin{cases} x_j^k - l_j & \text{se } x_j^k \leq (u_j + l_j)/2 \\ x_j^k - u_j & \text{se } x_j^k \geq (u_j + l_j)/2 \end{cases}$$

3- Se $\| \Gamma_{x^k w^k} \| \leq \delta$ e $\tilde{c}_j > -\delta$, onde

$$\Gamma_{x^k w^k} = X_k (c - A^T w^k), \text{ pare o processo, senão}$$

Calcule

$$c^k = X_k^2 \tilde{c} \text{ e faça}$$

$$d = -c^k.$$

4- Atualize a solução:

$$x^{k+1} = x^k + \bar{\epsilon} d \quad \text{onde}$$

$$\bar{\epsilon} = \alpha \epsilon \quad ; \quad 0 < \alpha < 1 \quad \text{e}$$

$$\epsilon = \min\{\epsilon^1, \epsilon^2\} \quad \text{com}$$

$$\epsilon^1 = \min\{-(x_j^k - l_j) / d_j \mid \text{tal que } d_j < 0; j=1 \dots n\} \text{ e}$$

$$\epsilon^2 = \min\{(u_j - x_j^k) / d_j \mid \text{tal que } d_j > 0; j=1 \dots n\}$$

5- Volte em 2 com $k \leftarrow k + 1$.

2 - Algoritmos para outros problemas com variáveis canalizadas.

O problema de variáveis canalizadas e restrições do tipo $\leq$:

$$\text{minimizar } c^T x$$

$$\text{sujeito a } \begin{cases} Ax \leq b \\ l \leq x \leq u; \end{cases} \quad (7)$$

e também o seguinte problema;

$$\text{minimizar } c^T x$$

$$\text{sujeito a } \begin{cases} Ax \leq b \\ x \geq 0 \end{cases} \quad (8)$$

abordado em [Cavalier e Schall, (1987)], podem ser tratados como casos particulares do algoritmo para restrições canalizadas visto no capítulo 3.

O problema (7) pode ser redefinido, como:

$$\begin{aligned} & \text{minimizar } c^T x \\ & \text{sujeito a } \begin{cases} b^- \leq Ax \leq b^+ ; & b_1^- = -\infty ; & b^+ = b; \\ l \leq Ix \leq u \end{cases} \end{aligned}$$

que por sua vez é equivalente a:

$$\begin{aligned} & \text{minimizar } c^T x \\ & \text{sujeito a } \tilde{b}^- \leq \tilde{A}x \leq \tilde{b}^+ \end{aligned}$$

$$\text{com } \tilde{A} = \begin{bmatrix} A \\ I \end{bmatrix} ; \quad \tilde{b}^- = \begin{bmatrix} b^- \\ l \end{bmatrix} ; \quad \tilde{b}^+ = \begin{bmatrix} b^+ \\ u \end{bmatrix}.$$

Agora, se redefinirmos:

$$\tilde{Z} = \begin{bmatrix} Z & 0 \\ 0 & X \end{bmatrix};$$

onde $Z = \text{diag}(z'_i)$; com

$$\tilde{z}_i = \begin{cases} z_i - b_i^- & \text{se } z_i \leq (b_i^+ + b_i^-)/2 \\ z_i - b_i^+ & \text{se } z_i \geq (b_i^+ + b_i^-)/2 \end{cases}$$

($z = Ax$);

e $X = \text{diag}(\tilde{x}_i)$, com

$$\tilde{x}_i = \begin{cases} x_i - l_i & \text{se } x_i \leq (u_i + l_i)/2 \\ x_i - u_i & \text{se } x_i \geq (u_i + l_i)/2 \end{cases}$$

No algoritmo para restrições canalizadas ficamos:

$$\tilde{A}^T \tilde{Z}^{-2} \tilde{A} = [A^T \ I] \begin{bmatrix} Z^{-2} & 0 \\ 0 & X^{-2} \end{bmatrix} \begin{bmatrix} A \\ I \end{bmatrix} = A^T Z^{-2} A + X^{-2};$$

ou seja, nada se altera, com exceção de que na diagonal de $A^T Z^{-2} A$ é somado $\tilde{x}_i^{-2}$; e o passo a ser dado é semelhante ao estudado para restrições canalizadas.

Para o caso (8), basta que modifiquemos apenas a definição de X , voltando à definição original, $X = \text{diag}(x_i)$, $x_i > 0$, $i = 1 \dots n$.

Logo, para ambos os casos, o algoritmo para restrições canalizadas pode ser readaptado.

O problema para variáveis canalizadas e restrições de igualdade:

$$\begin{aligned} &\text{minimizar } c^T x \\ &\text{sujeito a } \begin{cases} Ax = b \\ 0 \leq x \leq u; \end{cases} \end{aligned} \quad (9)$$

é um caso mais difícil para adaptação do algoritmo para restrições canalizadas, visto no capítulo 3.

Em [Martins, (1990)] e [Oliveira, (1989)] , este problema é abordado e obtém-se uma técnica de resolução com o uso direto do algoritmo dual afim de Adler e outros, visto no capítulo 2.

O problema (9), pode ser escrito como:

$$\begin{aligned} &\text{minimizar } c^T x \\ &\text{sujeito a } \begin{cases} Ax = b \\ x + y = u \\ x, y \geq 0; \end{cases} \end{aligned}$$

cujo dual é:

$$\begin{aligned} &\text{maximizar } b^T v + u^T w \\ &\text{sujeito a } \begin{cases} A^T v + w \leq c \\ w \leq 0 \\ v, w : \text{irrestritos.} \end{cases} \end{aligned}$$

Se fizermos:

$$\tilde{A} = \begin{bmatrix} A & 0 \\ I & I \end{bmatrix}; \quad \tilde{b} = \begin{bmatrix} b \\ u \end{bmatrix}; \quad \tilde{c} = \begin{bmatrix} c \\ 0 \end{bmatrix}; \quad \tilde{v} = \begin{bmatrix} v \\ w \end{bmatrix};$$

ficamos com o seguinte problema:

$$\begin{aligned} &\text{minimizar } \tilde{c}^T \tilde{v} \\ &\text{sujeito a } \tilde{A} \tilde{v} \leq \tilde{c}; \end{aligned}$$

que pode ser resolvido pela aplicação do algoritmo afim dual, visto no capítulo 2, fazendo - se algumas manipulações algébricas no sistema de restrições.

A seguir daremos os passos do algoritmo desenvolvido em [Martins, (1990)] e [Oliveira, (1989)]:

1- Sejam v^0 , w^0 e γ tais que,

$$\begin{bmatrix} A^T & I \\ 0 & I \end{bmatrix} \begin{bmatrix} v^0 \\ w^0 \end{bmatrix} < \begin{bmatrix} c \\ 0 \end{bmatrix}, \quad 0 < \gamma < 1 \quad \text{e} \quad k = 0$$

2- Enquanto "critério de parada não satisfeito" faça:

$$y_1^k = c - A^T v^k - w^k,$$

$$y_2^k = -w^k,$$

$$D_{y_1}^k = \text{diag} [1/(y_1^k)_1, \dots, 1/(y_1^k)_n],$$

$$D_{y_2}^k = \text{diag} [1/(y_2^k)_1, \dots, 1/(y_2^k)_n].$$

$$A \bar{D} A^T d_v = b - A \bar{D} \text{diag}[(y_2^k)_1^2] u,$$

$$d_w = \bar{D} \text{diag}[(y_2^k)_1^2] \{ \text{diag}[(y_1^k)_1^2] u - A^T d_v \},$$

$$\text{onde } \bar{D} = \begin{bmatrix} (D_{y_1}^k)^2 & (D_{y_2}^k)^2 \\ (D_{y_2}^k)^2 & (D_{y_1}^k)^2 + (D_{y_2}^k)^2 \end{bmatrix}^{-1},$$

$$\begin{bmatrix} d_{y_1} \\ d_{y_2} \end{bmatrix} = \begin{bmatrix} -A^T & -I \\ 0 & -I \end{bmatrix} \begin{bmatrix} d_v \\ d_w \end{bmatrix} = \begin{bmatrix} -A^T d_v - d_w \\ -d_w \end{bmatrix}, \quad (i = 1, \dots, n).$$

Calcule $x_i = (d_{y_1})_i / (y_1^k)_i^2$, $(i = 1, \dots, n)$.

Determine $\varepsilon = \gamma \min \{ -(y_1^k)_i / (d_{y_1})_i; -(y_2^k)_i / (d_{y_2})_i \}$ para

$(d_{y_1})_i$ e $(d_{y_2})_i$ negativos, $(i = 1, \dots, n)$.

Atualize:

$$\begin{bmatrix} v^{k+1} \\ w^{k+1} \end{bmatrix} = \begin{bmatrix} v^k \\ w^k \end{bmatrix} + \varepsilon \begin{bmatrix} d_v \\ d_w \end{bmatrix}.$$

Faça $k = k + 1$.

Conclusão: Todos os métodos estudados até aqui, com exceção do método de Chandru e Kochar para variáveis canalizadas, trabalham a cada iteração considerando os elipsóides totalmente contidos na região viável do problema.

O método de Chandru e Kochar não se preocupa com este fato, já que os elipsóides podem não estar totalmente contidos na região viável do problema, pois, no momento de definir a mudança de escala, já é sabido se uma variável irá sofrer um acréscimo ou um decréscimo dependendo do sinal da coordenada correspondente de $\tilde{c}$.

Assim, a fronteira a ser considerada é aquela em que a direção aponta, (aquela "adiante"), pois esta depende diretamente de $\tilde{c}$, e não a mais próxima.

APÊNDICE A.

O Método simplex.

Este clássico método é muito utilizado na programação matemática e foi desenvolvido por Dantzig na década de 40.

A1 - Preliminares.

Seja o problema:

$$\begin{array}{ll} \text{minimizar} & c^T x \\ \text{sujeito a} & \begin{cases} Ax = b \\ x \geq 0; \end{cases} \end{array} \quad (1)$$

onde $A \in \mathbb{R}^{m \times n}$ e $\text{posto}(A) = m$.

Considere uma partição nas colunas de A:

$$A = [B, N],$$

com $B \in \mathbb{R}^{m \times m}$, não singular, $N \in \mathbb{R}^{m \times (n-m)}$ e

$I_B = \{j \text{ tal que } j \text{ é um índice coluna relacionado à base } B\}$,

$I_N = \{j \text{ tal que } j \text{ é um índice coluna relacionado à } N\}$.

Para esta partição podemos associar a solução básica:

$$\bar{x} = \begin{bmatrix} \bar{x}_B \\ \bar{x}_N \end{bmatrix}, \quad \text{com } \bar{x}_B = B^{-1}b \text{ e } \bar{x}_N = 0,$$

e esta representa um vértice do politopo $Ax = b$ se $x_B \geq 0$.

Se $\bar{x}_B \geq 0$, a solução é chamada básica factível.

Pelo teorema da programação linear é válido que:

"se o problema (1) tem solução ótima então existe uma solução básica ótima". Levando-se em consideração este fato, o método simplex percorre soluções básicas factíveis até alcançar uma solução básica ótima, estratégia diferente dos métodos de pontos interiores.

A2 - Direção de busca.

Já foi visto que, se $d \in \mathbb{R}^n$ é uma direção factível então $d \in N(A)$.

Se particionarmos d em:

$$d = \begin{bmatrix} d_B \\ d_N \end{bmatrix},$$

$$\text{então } Ad = 0 \iff Bd_B + Nd_N = 0 \iff d_B = -B^{-1}Nd_N.$$

Uma escolha particular para d_N é :

$$d_N = e_j = (0, \dots, 1, \dots, 0)^T \in R^{n-m},$$

com 1 na posição j , que fornece,

$$d_B = -B^{-1}(a_N)_j,$$

onde $j \in I_N$ e $(a_N)_j$ é uma coluna da submatriz N .

Portanto,

$$d = \begin{bmatrix} -B^{-1}(a_N)_j \\ e^j \end{bmatrix} \in N(A).$$

A3 - Determinação do passo.

Considerando-se em uma iteração corrente a solução básica e factível $\bar{x} = (\bar{x}_B, \bar{x}_N)^T$, então a factibilidade será garantida se:

$$\bar{x} + \varepsilon d \geq 0 \iff \bar{x}_B + \varepsilon d_B \geq 0 \iff (\bar{x}_B)_i + \varepsilon (d_B)_i \geq 0, i \in I_B.$$

Isto ocorre, conforme a seção 4, capítulo 1, se:

$$\varepsilon = \min_{i \in I_B} \{ -(\bar{x}_B)_i / (d_B)_i \text{ tal que } (d_B)_i < 0 \}.$$

Suponha que

$$\varepsilon = -(\bar{x}_B)_p / (d_B)_p, p \in I_B,$$

então $(\bar{x}_B)_p + \varepsilon (d_B)_p = 0$, se torna não básica e é substituída por $(\bar{x}_N)_j = \varepsilon, j \in I_N$.

Se $(d_B)_i > 0, \forall i \in I_B$, então o conjunto de soluções factíveis é ilimitado. Além disso, se duas ou mais componentes de $\bar{x}_B$ se anulam para o valor de ε , temos o caso de degeneração da solução.

A3 - Critério de mudança da base.

Para a nova solução factível $\bar{x} + \varepsilon d$, temos:

$$\begin{aligned} c^T(\bar{x} + \varepsilon d) &= c^T \bar{x} + \varepsilon c_B^T d_B + \varepsilon c_N^T d_N = c^T \bar{x} + \varepsilon (c_B^T d_B + c_N^T d_N) = \\ &= c^T \bar{x} + \varepsilon (c_j - c_B^T B^{-1}(a_N)_j) = c^T + \varepsilon r_j, \end{aligned}$$

onde $r_j = c_j - c_B^T B^{-1}(a_N)_j$ é o chamado custo relativo da componente $j \in I_N$.

Se $r_j \leq 0$, então $c^T(\bar{x} + \varepsilon d) \leq c^T \bar{x}$, logo, d é uma direção de descida. Assim:

- Se $\exists j \in I_N$ tal que $r_j \leq 0$, isto é utilizado como um critério de mudança de base.
- Se $r_j \geq 0, \forall j \in I_N$, então a otimalidade é atingida pois não conseguimos mais decréscimos para a função objetivo e isto é utilizado como critério de parada do método.

Após as considerações das seções anteriores, podemos enunciar um algoritmo que segue os seguintes passos:

A4 - Algoritmo simplex, [Luenberger, (1984)].

1- Dada uma solução corrente $x_B^0 = B_0^{-1}b$, onde B_0 é uma base inicial, faça $k = 0$.

2- Calcule o vetor custo relativo $r_N^k = c_N^T - c_B^T B_k^{-1} N_k$.

Se $r_N^k \geq 0$, pare, a solução encontrada é ótima.

3- Determine

$$(r_N^k)_q = \min_{j \in I_N} \{(r_N^k)_j \text{ tal que } (r_N^k)_j < 0\},$$

para determinar qual vetor de N_k irá entrar na base.

4- Calcule $d_B^k = -B_k^{-1}(a_N)_q$.

5- Se não existe $(d_B^k)_i < 0$, pare, o problema não tem solução, caso contrário determine:

$$-(x_B^k)_p / (d_B^k)_p = \min_{i=1 \dots m} \{-(x_B^k)_i / (d_B^k)_i \text{ tal que } (d_B^k)_i < 0\},$$

para determinar qual vetor de B_k irá sair da base.

6- Atualize:

$$(B^{k+1})_p \Leftarrow (N^k)_q ;$$

$$(N^{k+1})_q \Leftarrow (B^k)_p ;$$

$$k \Leftarrow k + 1.$$

BIBLIOGRAFIA.

- [Adler et al, (1986)]- Adler, I. & Karmarkar, N. & Resende, M. & Veiga, G.: *An implementation of Karmarkar's algorithm for linear programming*. Report ORC86-8, Operations Research, University of California, Berkeley, 1986, (revised in (1989)).
- [Bayer e Lagarias, (1986)]- Bayer, K. & Lagarias, J.C. : *The non-linear geometry of linear programming: I- affine and projective scaling trajectories, II- Legendre transform Coordinates, III- Central trajectories*. Preprints, AT & Bell Laboratories, Murray Hill, N.J. , 1986.
- [Barnes, (1986)]- Barnes, E. R.: *A variation on Karmarkar's algorithm for solving linear programming problems*. Mathematical Programming, 36:174-182, 1986.
- [Bazaraa e Shetty, (1979)]- Bazaraa, M.S. & Shetty, M.: *Nonlinear programming. Theory and Algorithms*. Canada, by John-Wiley & Sons Inc., pp 80-358, 1979.
- [Cavalier e Soyster, (1985)]- Cavalier, T. & Soyster, A.: *Some Computational Experience and a Modification of the Karmarkar algorithm*. Working Paper 85-105, Dp. of Industrial Management System, Pennsylvania State Univ., february 1985.

- [Cavalier e Schall, (1987)]- Cavalier, T & Schall, K.C. : *Implementing an affine scaling algorithm for linear programming*. Comput Opns. Res. vol 14, No. 5, pp. 341-347, 1987.
- [Chandru e Kochar, (1985)]- Chandru, V. & Kochar, B.: *A class of algorithms for linear programming*. Research Memorandum No. 86-10, School of Industrial Engineering, Purdue University, W. Lafayette, Indiana, 1985.
- [Chandru e Kochar, (1986)]- Chandru, V. & Kochar, B.: *Exploiting special structures using a variant of Karmarkar's algorithm*. Research Memorandum No. 86-10, School of Industrial Engineering, Purdue University, W. Lafayette, Indiana, 1986.
- [Dantzig, (1963)]- Dantzig, G.B.: *Linear Programming and extensions*. Princeton University Press, Princeton , N.J., 1963.
- [Dikin, (1967)]- Dikin, I.: *Iterative solution of problems of linear and quadratic programming*. Soviet Mathematics Doklady, 8:674-675, 1967.
- [Dodani e Babu, (1987)]- Dodani, M.H. & Babu, J.G.: *Karmarkar's projective method for linear programming: A computational survey*. Computers Ind. Engineering , vol. 13, Nos. 1-4, Printed in Great Britain, pp. 285-289, 1987.

- [Frisch, (1955)]- Frisch, K.R.: *The Logarithm Potential Method of Convex Programming*. Memorandum, University Institute of Economics, Oslo, Norway, 1955.
- [Golub e Van Loan, (1983)]- Golub, G.H. & Van Loan, C.F.: *Matrix computations*. The Johns Hopkins Press, Baltimore, 1983.
- [Gonzaga, (1988)]- Gonzaga, C.: *Polinomial affine algorithm for linear programming*. Internal report, programa de Eng. de Sist. e computação, COPPE/UFRJ, Rio de Janeiro, January, 1988.
- [Gonzaga, (1989)]- Gonzaga, C.: *Algoritmos de pontos interiores em programação linear*. Notas da I-EBO, Rio de Janeiro, Janeiro, 1989.
- [Gonzaga, (1989)b]- Gonzaga, C.: *Path following methods for linear programming*. Programa de Eng. Sist. e Computação, COPPE/UFRJ, Rio de Janeiro, January, 1989
- [Gonzaga, (1990)]- Gonzaga, C.: *Convergence of the large steps primal affine scaling algorithm for primal non degenerate linear programs*. Programa de Eng. de Sist. e Computação, COPPE/UFRJ, ES-230/90, Rio de Janeiro, setembro, 1990.
- [Hooker, (1986)]- Hooker, J.N.: *Karmarkar's linear programming algorithm*. Interfaces, 16:4, pp. 75-90, July-August 1986.
- [Karmarkar, (1984)]- Karmarkar, N.: *A new polynomial-time algorithm for linear programming*. Combinatorica, vol.4, No.4, pp. 373-395, 1984.
- [Luenberger, (1984)]- Luenberger, D.G.: *Linear and Nonlinear Programming*. USA, Second Edition, by Adison Wesley Publishing Company, 1984.

- [Martins, (1990)]- Martins, P.J.F.: *Otimização de um sistema de mecanização agrícola através de um algoritmo de ponto interior num processo de branch and bound*. Tese de mestrado, F.E.E. - UNICAMP, Campinas-S.P.- 1990.
- [Murthag e Saunders, (1977)]- Murthag, B.A. & Saunders, M.A.: *MINO'S User's guide*. Technical Report SOL77-9, Department of Operations Research, Stanford University, Stanford, C.A., 1977.
- [Oliveira, (1989)]- Oliveira, A.L.: *Métodos de ponto interior em Programação Linear - Estudo e implantação*. Tese de mestrado, F.E.E. - UNICAMP, Campinas-S.P.- 1989.
- [Orlandi, (1990)]- Orlandi, M.A.: *Métodos de pontos interiores para Programação Linear*. Projeto de iniciação científica-FAPESP, ICM-USP-São Carlos, 1990.
- [Renegar, (1988)]- Renegar, J.: *A polynomial-time algorithm based on Newton's method for linear programming*. Mathematical Programming, 40:59-94, 1988.
- [Todd e Burrel, (1985)]- Burrel, B.P.: & Todd, M.J.: *The ellipsoid method generates dual variables*. Mathematics of Operations Research, 10: 688-700, 1985.
- [Vaydia, (1987)]- Vaydia, P.M.: *An algorithm for linear programming which requires $O((m+n)n^2 + (m+n)^{1.5}n)L$ Arithmetics Operations*. Preprint, AT & T Bell Laboratories, Murray Hill, N.J., 1987.

- [Vanderbei et al, (1986)]- Vanderbei, R.J. & Meketon, M.J. & Freedman, B.A.: *A modification of Karmarkar's linear programming algorithm*. Algorithmica, 1:395-407, 1986.
- [Ye, (1991)]- Ye, Y.: *An $O(n^3L)$ potential reduction algorithm for linear programming*. Mathematical Programming, 50:239-258, North-Holland, 1991.
- [Wilkinson, (1965)]- Wilkinson, J.H.: *The algebraic eigenvalue problem*. Oxford, University Press, 1965.