Universidade Federal de Goiás
Instituto de Informática

Erika Morais Martins

Comportamento limite da Trajetória Central em Programação Linear

Goiânia
2007

ERIKA MORAIS MARTINS

Comportamento limite da Trajetória Central em Programação Linear

Dissertação apresentada ao Programa de Pós–Graduação do Instituto de Informática da Universidade Federal de Goiás, como requisito parcial para obtenção do título de Mestre em Ciência da Computação.

Área de concentração: Otimização

Orientador: Prof. Dr. Orizon Pereira Ferreira

Goiânia
2007

ERIKA MORAIS MARTINS

Comportamento limite da Trajetória Central em Programação Linear

Dissertação defendida no Programa de Pós–Graduação do Instituto de Informática da Universidade Federal de Goiás como requisito parcial para obtenção do título de Mestre em Ciência da Computação, aprovada em 06 de Fevereiro de 2007, pela Banca Examinadora constituída pelos professores:

Prof. Dr. Orizon Pereira Ferreira
Instituto de Informática – UFG
Presidente da Banca

Prof. Dra. Márcia Helena Costa Fampa
COPPE – UFRJ

Prof. Dr. Luis Román Lucambio Pérez
Instituto de Matemática e Estatística – UFG

Erika Morais Martins

Graduou–se em Ciência da Computação na UCG - Universidade Católica de Goiás. Durante sua graduação, foi monitora no departamento de Computação da UCG e desenvolveu um trabalho de iniciação científica no departamento de Computação. Durante o Mestrado foi bolsista da CAPES.

Dedico este trabalho a minha
mãe Célia e a minha irmã Joicy.

Agradecimentos

À Deus, pelo dom da vida.

Ao professor Orizon, meu orientador, pelos ensinamentos, pela paciência e pelo tempo dedicado à minha orientação.

Ao meu esposo Hebert, pela amizade, companheirismo e pelas contantes ajudas, principalmente durante a preparação para a defesa da dissertação.

Ao professor Humberto, pelas diversas dúvidas esclarecidas em relação ao pacote classe-inf.

À todos os professores, funcionários e amigos do Instituto de Informática da UFG.

À Capes, pela ajuda financeira durante boa parte do curso.

E a todos que não citei, mas que de alguma forma contribuíram para a conclusão de mais uma etapa em minha vida.

Resumo

Martins, Erika Morais. **Comportamento limite da Trajetória Central em Programação Linear**. Goiânia, 2007. Dissertação de Mestrado. Instituto de Informática, Universidade Federal de Goiás.

A Trajetória Central tem importância fundamental no estudo de vários algoritmos para resolução do Problema de Programação Linear (PPL), visto que, os algoritmos que em algum sentido seguem a Trajetória tem comportamento polinomial no pior caso. Como a trajetória converge e é analítica no interior do conjunto viável do PPL, nosso principal objetivo neste trabalho é reescrever a prova, feita em [10], de que a Trajetória Central é analítica no seu ponto limite.

Palavras–chave

Programação Linear e Trajetória Central.

Abstract

Martins, Erika Morais. **Behavior limits of the Central Path in Linear Programming**. Goiânia, 2007. MSc. Dissertation. Instituto de Informática, Universidade Federal de Goiás.

The Central Path has fundamental importance in the study of many algorithms for resolution of the Problem of Linear Programming (PPL), because, the algorithms that follow the Path in some sense have behavior polinomial in the worst case. As the path converges and it is analytical in the interior of the feasible set of PPL, our main objective in this work is rewrite the proof, that has been done in [10], that the Central Path is analytical in its limit point.

Keywords

Linear Programming and Central Path.

Sumário

Notações

Seguem aqui uma lista de notações utilizadas durante o texto.

1. $\mathbb{N}$ é o conjunto dos números naturais.
2. $\mathbb{R}$ é o conjunto dos números reais.
3. $\mathbb{R}_{++}$ é o conjunto dos números reais positivos.
4. $\mathbb{R}^n$ é o conjunto dos vetores reais da forma
$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$
5. $\mathbb{R}^n_{++}$ é o ortante positivo do $\mathbb{R}^n$.
6. $x^T = (x_1, x_2, \ldots, x_n)$ é o transposto do vetor x.
7. x_j é a j-ésima coordenada do vetor x.
8. $||x||$ é a norma euclidiana de $x \in \mathbb{R}^n$.
9. $\langle x, y \rangle$ é o produto interno euclidiano entre $x \in \mathbb{R}^n$ e $y \in \mathbb{R}^n$.
10. $M_{m\times n}(\mathbb{R})$ é o conjunto das matrizes reais de dimensão $m \times n$.
11. A^T é a transposta da matriz $A \in M_{m\times n}(\mathbb{R})$.
12. A^{-1} é a inversa da matriz $A \in M_{n\times n}(\mathbb{R})$, quando existir.
13. $e = (1, \ldots, 1)^T$, isto é, o vetor com todas as coordenadas iguais a um, com dimensão de acordo com o contexto.
14. e_i é um vetor de zeros, com o valor 1 na posição i.
15. $0_{m\times n}$ é a matriz nula de dimensão $m \times n$.

16. I é a matriz identidade com ordem de acordo com o contexto.

17. $X = diag(x)$ é uma matriz diagonal formada pelas coordenadas do vetor x.

Introdução

O Problema de Programação Linear (PPL) é o problema de otimizar uma função linear sujeito à restrições também lineares. Neste trabalho estaremos trabalhando com o PPL no formato padrão, isto é, minimizar uma função linear sujeito a restrições lineares de igualdade e restrições de não negatividade .

Existem vários métodos para a resolução do Problema de Programação Linear. O método simplex, publicado por Dantzig [5] em 1951, foi o primeiro algoritmo efetivo para resolução do PPL. O algoritmo simplex consiste em caminhar pela fronteira do conjunto viável, através de pontos extremos adjacentes, minimizando o valor da função objetivo com relação aos pontos extremos anteriores até atingir uma solução ótima, se existir.

Em 1972, o exemplo de Klee e Minty [17] verificou que, para um certo critério de escolha para entrada na base, o algoritmo simplex tinha comportamento exponencial. Apesar de se comportar bem na prática, o algoritmo simplex tem um comportamento exponencial no pior caso.

A partir de então, começou-se um busca por algoritmos para Programação Linear que tivessem comportamento polinomial. Assim, o primeiro algoritmo polinomial para resolver o PPL foi o algoritmo dos elipsóides, publicado por Khachiyan em 1979 [15] e [16]. Porém este algoritmo não tinha um comportamento eficaz na prática.

Em 1984, Karmarkar [14] publicou um algoritmo de pontos interiores de complexidade polinomial, menor do que os algoritmos dos elipsóides, porém tinha um bom comportamento na prática.

Passou-se então ao estudo dos algoritmos denominados "Algoritmos de Pontos Interiores", na tentativa de baixar a sua complexidade. Algumas classes de algoritmos de pontos interiores geram uma seqüência de pontos no interior do conjunto viável, que convergem para uma solução do PPL, seguindo em algum sentido uma trajetória central, veja em Gonzaga [7]. Os algoritmos de pontos interiores têm comportamento polinomial no pior caso.

Neste sentido, a trajetória central tem importância fundamental no estudo de vários algoritmos para resolução do PPL. Estudando seu comportamento estaremos estimando o melhor comportamento que um algoritmo de ponto interior, que segue a trajetória central, pode obter.

É bem conhecido que a Trajetória Central converge para o centro analítico do conjunto de soluções ótimas do PPL, veja [8], e que a Trajetória Central é analítica no interior do conjunto viável do PPL, veja [27]. Recentemente, em [9] e [10] foi provado que a Trajetória Central é analítica no seu ponto limite.

Desta forma, nosso primeiro objetivo neste trabalho é estudar o comportamento da Trajetória Central, verificando a propriedade analítica no seu ponto limite. E reescrever uma das demonstrações em [10] de uma maneira mais didática.

Nosso segundo objetivo é criar um texto auto-contido, de forma que o leitor não precise de material auxiliar para um bom estudo da Trajetória Central em Programação Linear. Para isso, dividiremos este trabalho em 4 capítulos.

No capítulo 1, revisaremos alguns tópicos de topologia no $\mathbb{R}^n$, estudaremos conceitos básicos de cálculo e de análise convexa de conjuntos e funções, definindo a projeção sobre conjuntos convexos e finalmente enunciar alguns conceitos de algebra linear. Acreditamos obtermos um bom embasamento teórico para a compreensão dos demais capítulos.

No capítulo 2, definiremos o par de PPLs no formato padrão, demonstraremos resultados importantes de dualidade em Programação Linear, obtendo as equações que definem as condições de otimalidade para os pares de PPls.

No capítulo 3, faremos um bom estudo da Trajetória Central. Obteremos as equações que a definem, mostrando que a Trajetória Central é uma curva analítica no interior do conjunto viável do PPL e mostraremos que a Trajetória Central converge para o centro analítico do conjunto de soluções ótimas do PPL.

Por fim, no capítulo 4, estudaremos o comportamento limite da Trajetória Central, mostrando que a mesma é analítica no seu ponto limite.

CAPÍTULO 1

Resultados Básicos

Nosso objetivo neste capítulo é revisar alguns tópicos de Topologia do $\mathbb{R}^n$ e Análise Convexa, incluindo resultados de projeção sobre conjuntos convexos e funções convexas. Revisaremos algumas definições de análise no $\mathbb{R}^n$ e também alguns tópicos de Álgebra Linear. Estes assuntos serão necessários ao desenvolvimento dos capítulos seguintes, dando um embasamento teórico para uma boa compreensão do texto.

1.1 Noções de Topologia do $\mathbb{R}^n$

Nesta seção estaremos interessados em definir alguns conjuntos importantes do espaço euclidiano $\mathbb{R}^n$ e definir continuidade de funções, enunciando o Teorema de Weiertrass sobre funções contínuas. Iniciaremos definindo bolas aberta e fechada.

Sejam dados o ponto $a \in \mathbb{R}^n$ e o número real $r > 0$. A *bola aberta* de centro a e raio r é o conjunto

$$B(a,r) = \{x \in \mathbb{R}^n : ||x - a|| < r\},$$

isto é, o conjunto dos pontos $x \in \mathbb{R}^n$ cuja distância ao ponto a é menor do que r. Análogamente a *bola fechada* de centro a e raio r é o conjunto

$$B[a,r] = \{x \in \mathbb{R}^n : ||x - a|| \leq r\}.$$

Um conjunto $\mathcal{T} \subset \mathbb{R}^n$ é *aberto* se para todo $a \in \mathcal{T}$ existir $r > 0$ tal que $B(a,r) \subset \mathcal{T}$.

Um conjunto $\mathcal{T}$ é *limitado* se existir $r > 0$ tal que $\|x\| \leq r$ para todo $x \in \mathcal{T}$, isto é, $\mathcal{T}$ é limitado se estiver contido em uma bola fechada $B[0,r]$.

Uma *sequência* $\{x^k\} \subset \mathbb{R}^n$ é uma função $x : \mathbb{N} \to \mathbb{R}^n$, que associa a cada número $k \in \mathbb{N}$ um vetor $x^k \in \mathbb{R}^n$. Dizemos que uma sequência $\{x^k\}$ é limitada, se existir $c > 0$ tal que $\|x^k\| \leq c$ para todo $k \in \mathbb{N}$. Diz-se que um ponto $a \in \mathbb{R}^n$ é o limite de uma sequência $\{x^k\}$ quando, para todo $\epsilon > 0$ dado arbitrariamente, é possível obter $k_0 \in \mathbb{N}$ tal que $k > k_0 \Rightarrow \|x^k - a\| < \epsilon$. Escreve-se então

$$\lim_{k\to\infty} x^k = a, \tag{1-1}$$

e dizemos que a sequência converge para a. Se o limite acima existir, dizemos que a sequêcia será *convergente*, caso contrário a sequência será *divergente*.

Um ponto $a \in \mathbb{R}^n$ é dito *aderente* ao conjunto $\mathcal{T} \subset \mathbb{R}^n$ se existir uma sequência de pontos $\{x^k\} \subset \mathcal{T}$ que converge para a. O fecho de um conjunto $\mathcal{T} \subset \mathbb{R}^n$ é o conjunto $\bar{\mathcal{T}}$ de todos os pontos aderentes à $\mathcal{T}$. Assim, um conjunto é dito *fechado* quando $\mathcal{T} = \bar{\mathcal{T}}$.

Um conjunto $\mathcal{T} \subset \mathbb{R}^n$ é *compacto* quando $\mathcal{T}$ for fechado e limitado.

Seja $\mathcal{T} \subset \mathbb{R}^n$, uma função $f : \mathcal{T} \to \mathbb{R}^n$ é *contínua* em $a \in \mathcal{T}$ se

$$\lim_{k\to\infty} x^k = a \quad \Rightarrow \quad \lim_{k\to\infty} f(x^k) = f(a).$$

Se f for contínua em todo ponto $a \in \mathcal{T}$, então f é uma aplicação contínua. Uma função contínua em um conjunto compacto tem uma característica muito importante para nossos estudos. O próximo resultado é o Corolário 1, página 44 de [20].

Teorema 1.1.1 *(Weiertrass) Seja $\mathcal{T} \subset \mathbb{R}^n$ um conjunto compacto e não vazio . Se $f : \mathcal{T} \to \mathbb{R}^n$ é uma função contínua, então o problema*

$$\min\{f(x) : x \in \mathcal{T}\},$$

tem solução.

1.2 Noções de Cálculo

Nesta seção definiremos o conceito de derivada e de diferenciabilidade. Discutiremos a cerca de funções analíticas finalizando com o Teorema da Função Implícita. Inicialmente, definiremos a *derivada* de $f : \mathbb{R} \to \mathbb{R}$ em $x \in \mathbb{R}$, como sendo o limite

$$f'(x) = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}.$$

quando este limite existir.

Uma característica importante de funções de uma variável derivável é que toda função derivável em um ponto é contínua nesse ponto.

Um resultado bem conhecido, sobre funções deriváveis, é o Teorema do Valor Médio, cuja demosntração pode ser encontrada na página 220 em [2].

Teorema 1.2.1 *Seja uma função f definida e contínua num intervalo fechado $[a, b]$ e derivável nos pontos interiores, então, existe pelo menos um ponto c, compreendido entre a e b tal que*

$$f(b) - f(a) = f'(c)(b - a).$$

Agora estaremos interessados nas funções de várias variáveis. Seja $\Omega \subset \mathbb{R}^n$ um conjunto aberto. A *i-ésima derivada parcial* de $f : \Omega \to \mathbb{R}$ no ponto $p \in \Omega$, é definida como sendo o limite

$$\frac{\partial f}{\partial x_i}(p) = \lim_{t \to 0} \frac{f(p + te_i) - f(p)}{t},$$

quando esse limite existe. Definimos o *vetor gradiente* de f no ponto p por

$$\nabla f(p) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(p), & \ldots, & \frac{\partial f}{\partial x_n}(p) \end{pmatrix}^T.$$

Uma função $f : \Omega \to \mathbb{R}$, definida num intervalo aberto $\Omega \subset \mathbb{R}^n$, diz-se de classe $C^1(\Omega)$ quando existem, em cada ponto $x \in \Omega$, as derivadas parciais

$$\frac{\partial f}{\partial x_1}(x), \ldots, \frac{\partial f}{\partial x_n}(x),$$

e são contínuas as n funções assim definidas,

$$\frac{\partial f}{\partial x_i} : \Omega \to \mathbb{R}, \qquad i = 1, \ldots, n.$$

Mais geralmente, diremos que uma função $f : \Omega \to \mathbb{R}$ é de classe $C^k(\Omega)$ quando ela possuir as derivadas parciais em todos os pontos de Ω e as funções

$$\frac{\partial f}{\partial x_1} : \Omega \to \mathbb{R}, \ldots, \frac{\partial f}{\partial x_n} : \Omega \to \mathbb{R},$$

forem de classe $C^{k-1}(\Omega)$. Ainda, diremos que uma função $f : \Omega \to \mathbb{R}$ é de classe $C^0(\Omega)$ quando ela for contínua e diremos que f é de classe $C^\infty(\Omega)$, quando $f \in C^k(\Omega)$ para todo $k \geq 0$.

Uma função $f \in C^\infty(\Omega)$, chama-se *analítica* quando, para cada $p \in \Omega$

e $h \in \mathbb{R}^n$ com $p + h \in \Omega$, existe $\epsilon > 0$ tal que a série

$$\sum_{n=0}^{\infty} \frac{1}{n!} d^n f(p) h^n,$$

converge para $f(p+h)$ desde que $|h| < \epsilon$, onde para $h = (h_1, \ldots, h_n)$ escrevemos a diferencial de ordem zero e de primeira ordem como

$$d^0 f(p) h^0 = f(p), \qquad df(p)h = \sum_{i}^{n} \frac{\partial f}{\partial x_i}(p) h_i,$$

e as demais diferenciais, de maneira análoga, como se seguem,

$$d^2 f(p) h^2 = \sum_{i,j}^{n} \frac{\partial^2 f}{\partial x_j \partial x_i}(p) h_i h_j, \quad d^3 f(p) h^3 = \sum_{i,j,k}^{n} \frac{\partial^3 f}{\partial x_j \partial x_i \partial x_k}(p) h_i h_j h_k, \quad \ldots.$$

Seja $f : \Omega \to \mathbb{R}^m$ dada por $f(x) = (f_1(x), ..., f_m(x))$, onde $\Omega \subset \mathbb{R}^n$. Dizemos que f é *analítica* se suas funções coordenadas, $f_i : \Omega \to \mathbb{R}$ para $i = 1, \ldots, m$, são funções analíticas.

Definição 1.2.2 *Uma curva $w : (0, +\infty) \to \mathbb{R}^n$ é dita analítica em 0, se existe $\epsilon > 0$ e uma curva analítica $\psi : (-\epsilon, \epsilon) \to \mathbb{R}^n$ tal que, $w(t) = \psi(t)$ para todo $t \in [0, \epsilon)$.*

Um resultado de fundamental importância para nosso trabalho é o Teorema de Função Implícita. Este resultado, o Teorema 10.2.1 na página 265 em [6], determina sob quais condições uma relação como $F(x, y) = c$ define y em função de x, carregando a mesma classe de diferenciabilidade da F. Estaremos utilizando a versão analítica deste teorema.

Teorema 1.2.3 *(Teorema da Função Implícita) Sejam $I \times \bar{I} \subset \mathbb{R} \times \mathbb{R}^p$ um conjunto aberto e $(w_0, z_0) \in I \times \bar{I}$, onde $I \subset \mathbb{R}$ e $\bar{I} \subset \mathbb{R}^p$ são conjuntos abertos com $w_0 \in I$ e $z_0 \in \bar{I}$. Suponhamos que $F : I \times \bar{I} \to \mathbb{R}^p$ dada por $F(w, z) = (f_1(w, z), ..., f_m(w, z))$ seja analítica, que $F(w_0, z_0) = c \in \mathbb{R}^p$ e que a matriz*

$$\nabla_w F(w_0, z_0) = \begin{bmatrix} \frac{\partial f_1}{\partial z_1}(w_0, z_0) & \ldots & \frac{\partial f_1}{\partial z_m}(w_0, z_0) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial z_1}(w_0, z_0) & \ldots & \frac{\partial f_m}{\partial z_m}(w_0, z_0) \end{bmatrix} \in M_{m \times m}(\mathbb{R}),$$

seja inversível. Então existe um aberto $J \times U \subset I \times \bar{I}$ contendo (w_0, z_0) tal que $F(w, z) = c$ para $(w, z) \in J \times U$ se, e somente se, $z = \varphi(w)$, onde $\varphi : J \to U$ é a única função satisfazendo $F(w, \varphi(w)) = c, \forall w \in J$. Além disso, φ é analítica.

1.3 Noções de Análise Convexa

Nesta seção, estaremos interessados em estudar conjuntos convexos, cones e funções convexas. Iremos verificar que as funções estritamente convexas possuem uma característica muito importante para problemas de minimização. Iniciaremos com um resultado sobre partição de matrizes, que pode ser encontrado na página 1 da aula 6 em [26].

Teorema 1.3.1 *Seja $A \in M_{m\times n}(\mathbb{R})$ com posto (A) = $m > 0$. Se o conjunto*

$$\mathcal{Q} = \{x \in \mathbb{R}^n : Ax = b,\ x \geq 0\},$$

é não vázio, então existe uma partição da matriz A na forma $A = [A_B : A_N]$, com $A_B \in M_{m\times m}(\mathbb{R})$ inversível, tal que $A_B^{-1}b \geq 0$. Como consequência,

$$x = ((A_B^{-1}b)^T, 0^T)^T \in \mathcal{Q}.$$

Demonstração: Como $\mathcal{Q}$ é não vazio, tome $x \in \mathcal{Q}$. Temos duas posibilidades:

a) $x = 0$;

b) $x \neq 0$,

Primeiro vamos analizar o item a. Suponhamos que x seja o vetor nulo. Como posto$(A) = m > 0$, selecionamos quaisquer m colunas da matriz A de modo que a submatriz $A_B \in M_{m\times m}(\mathbb{R})$ de A formada por estas m colunas seja inversível. Portanto a partição $A = [A_B : A_N]$ satisfaz $A_B^{-1}b = 0$.
Agora vamos analisar o item b. Suponhamos que x seja diferente do vetor nulo. Como posto$(A) = m > 0$, segue-se que b é não nulo. Sem perda de generalidade, x pode ser escrito na forma $x = ((x_{\tilde{B}})^T, 0^T)^T$, onde $x_{\tilde{B}} > 0$. Considere a submatriz $A_{\tilde{B}}$ de A correspondente às coordenadas não nulas do vetor x. Deste modo

$$A_{\tilde{B}} x_{\tilde{B}} = b. \tag{1-2}$$

Seja $\{a_1, ..., a_p\}$ o conjunto formado pelas colunas da matriz $A_{\tilde{B}}$. Temos duas possibilidades:

i) O conjunto $\{a_1, ..., a_p\}$ é linearmente independente;

ii) o conjunto $\{a_1, ..., a_p\}$ é linearmente dependente.

Analisemos o caso i. Neste caso temos que $p \leq m$ pois posto(A) = m. Se $p = m$ a matriz $A_{\tilde{B}}$ é inversível. Se $p < m$ podemos adicionar $(m - p)$ colunas de A ao conjunto $\{a_1, ..., a_p\}$ de modo que $\{a_1, ..., a_p, a_{p+1}, ..., a_m\}$ seja ainda um conjunto

linearmente independente. Seja $A_B \in M_{m\times m}$ uma submatriz de A cuja as colunas sejam os vetores $\{a_1, ..., a_p, a_{p+1}, ..., a_m\}$. Desde que $\{a_1, ..., a_p, a_{p+1}, ..., a_m\}$ é um conjunto linearmente independente então A_B é inversível. Considere o vetor $\bar{x}_B = (x_1, ..., x_p, 0, ..., 0)^T = ((x_{\tilde{B}})^T, 0^T)^T \in I\!R^m$ e note que (1-2) é equivalente à $A_B\bar{x}_B = b$, ou seja, $\bar{x}_B = A_B^{-1}b$. Desde que $x = ((\bar{x}_B)^T, 0^T)^T \in \mathcal{Q}$ temos que $A_B^{-1}b = \bar{x}_B \geq 0$.
Analisemos o caso ii. Como o conjunto $\{a_1, ..., a_p\}$ é linearmente dependente, então existe um vetor não nulo $y \in I\!R^p$, tal que $A_{\tilde{B}}y = 0$. Multiplicando esta última igualdade por ε e subtraindo em (1-2) temos que $A_{\tilde{B}}(x_{\tilde{B}} - \varepsilon y) = b$. Tomando $\varepsilon = \min\{x_i/y_i : \;\; x_i > 0, \;\; y_i > 0\}$, teremos $x_{\tilde{B}} - \varepsilon y \geq 0$ e pelo menos uma de suas coordenadas igual a zero. Seja $A_{\hat{B}}$ uma submatriz de $A_{\tilde{B}}$, que por sua vez também é submatriz de A, correspondente às coordenadas não nulas do vetor $x_{\tilde{B}} - \varepsilon y$. Assim

$$A_{\hat{B}}x_{\hat{B}} = b,$$

onde $x_{\hat{B}} > 0$ é o vetor formado pelas coordenadas não nulas do vetor $x_{\tilde{B}} - \varepsilon y$. Se as colunas de $A_{\hat{B}}$ são linearmente independente recaimos no caso i. Caso contrário, observando que b é não nulo e consequentemente $A_{\hat{B}}$ também não é nula, podemos repetir este proceso não mais que $p-1$ vezes para então recair no caso i.
Portanto, em ambos os casos i e ii, existe uma partição da matriz A na forma $A = [A_B : A_N]$, com $A_B \in M_{m\times m}(\mathbb{R})$ inversível, tal que $A_B^{-1}b \geq 0$. Isto conclui a primeira parte.

A prova da segunda parte é imediata pois,

$$A\bar{x} = [A_B : A_N]((A_B^{-1}b)^T, 0^T)^T = b \qquad \text{e} \qquad \bar{x} = ((A_B^{-1}b)^T, 0^T)^T \geq 0.$$

□

A demonstração do próximo resultado pode ser encontrada na página 5 da aula 7 em [26].

Proposição 1.3.2 *Seja $A \in M_{m\times n}(\mathbb{R})$ com posto (A) = $m > 0$. O conjunto*

$$C = \{z \in \mathbb{R}^m : \;\; z = Ax, \;\; x \geq 0\}$$

é fechado.

Demonstração: Seja uma sequência $\{z^k\} \subset C$ tal que $\lim_{k\to\infty} z^k = \bar{z}$. Basta mostrarmos que $\bar{z} \in C$. Como $\{z^k\} \subset C$ existe uma sequência $\{x^k\}$ tal que

$Ax^k = z^k$ e $x^k \geq 0$ para todo k. Desta forma, para cada k o conjunto

$$\mathcal{Q}_k = \left\{x \in \mathbb{R}^n : Ax = z^k,\ x \geq 0\right\},$$

é não vazio. Então pelo Teorema 1.3.1, existe uma partição da matriz A na forma $A = [A_B : A_N]$ com $A_B \in M_{m\times m}(\mathbb{R})$ inversível tal que $A_B^{-1}z^k \geq 0$. Agora, defina $\tilde{x}^k = \left((A_B^{-1}z^k)^T, 0^T\right)^T$. Note que $\tilde{x}^k \geq 0$ converge para o ponto $\bar{x} = \left((A_B^{-1}\bar{z})^T, 0^T\right)^T \geq 0$. Além disso $A\bar{x} = \bar{z}$, de fato,

$$A\bar{x} = [A_B : A_N]\begin{pmatrix} \bar{x}_B \\ 0 \end{pmatrix} = [A_B : A_N]\begin{pmatrix} A_B^{-1}\bar{z} \\ 0 \end{pmatrix} = \bar{z}.$$

Como $\bar{x} \geq 0$ a última igualdade implica que $\bar{z} \in C$. Portanto C é fechado. □

Um subconjunto C de um espaço vetorial euclidiano é dito *cone* se para todo $x \in C$ e $\alpha \geq 0$ tem-se

$$\alpha x \in C.$$

Um conjunto $C \subset \mathbb{R}^n$ é dito *convexo* se, dados dois pontos $x_1, x_2 \in C$, tivermos

$$x_1 + (1-\lambda)x_2 \in C, \qquad \forall\, \lambda \in [0,1].$$

Geometricamente significa que, um conjunto é convexo se dados dois pontos quaisquer do conjunto, então o segmento de reta que une os dois pontos pertence ao conjunto.

Proposição 1.3.3 *Seja $A \in M_{m\times n}(\mathbb{R})$ com posto (A) = $m > 0$. Então o conjunto*

$$C = \{z \in \mathbb{R}^m : \ \ z = Ax, \ \ x \geq 0\},$$

é um cone não vazio, convexo e fechado.

Demonstração: Da definição de C segue imediatamente que $0 \in C$. Agora vamos mostrar que C é um cone. Seja $\alpha > 0$. Tome $z \in C$, e sejam $x \geq 0$ tal que $Ax = z$. Deste modo, como A é linear temos,

$$\alpha z = \alpha A x = A(\alpha x).$$

Agora sendo $\alpha > 0$ e $x \geq 0$ temos que $\alpha x \geq 0$ o que juntamente com a última igualdade implica que $\alpha z \in C$. Portanto C é um cone. Agora verifiquemos que C é convexo. Sejam $z^1,\ z^2 \in C$ e $\lambda \in [0,1]$. Tome $x^1, x^2 \geq 0$ tais que $Ax^1 = z^1$ e

$Ax^2 = z^2$. Novamente usando a linearidade de A temos,

$$\lambda z^1 + (1-\lambda)z^2 = \lambda A x^1 + (1-\lambda)Ax^2 = A(\lambda x^1 + (1-\lambda)x^2).$$

Desde que $\lambda \in [0,1]$ e $x^1, x^2 \geq 0$ temos que $\lambda x^1 + (1-\lambda)x^2 \geq 0$, assim a igualdade acima implica que a combinação convexa $\lambda z^1 + (1-\lambda)z^2 \in C$. Logo C é convexo. Finalmente para verificarmos que C é fechado basta utilizarmos a Proposição 1.3.2. Portanto C é um cone não vazio, convexo e fechado. $\square$

Um ponto $x \in \mathcal{Q} = \{x \in \mathbb{R}^n : Ax = b,\ x \geq 0\}$ é dito *ponto extremo* ou *vértice*, se x não puder ser escrito como combinação linear convexa de dois outros pontos distintos de $\mathcal{Q}$. Note que a cada equação ou restrição que define o conjunto $\mathcal{Q}$ podemos associar um hiperplano de $\mathbb{R}^n$. Geometricamente podemos dizer que estes hiperplanos definem o conjunto $\mathcal{Q}$. Existem várias maneiras equivalentes de definir pontos extremos. Na próxima proposição daremos uma caracterização geométrica para pontos extremos de $\mathcal{Q}$, cuja demonstração pode ser encontrada na página 63 em [3].

Proposição 1.3.4 *Sejam $A \in M_{m\times n}(\mathbb{R})$ com posto (A) = $m > 0$ e*

$$\mathcal{Q} = \{x \in \mathbb{R}^n : Ax = b,\ x \geq 0\} \neq \emptyset.$$

Um ponto $\bar{x} \in \mathcal{Q}$ é ponto extremo de $\mathcal{Q}$ se , e somente se, $\bar{x}$ se encontrar em n hiperplanos linearmente independentes que definem $\mathcal{Q}$.

Demonstração: Suponha que o número máximo de hiperplanos linearmente independentes que contém $\bar{x} \in \mathcal{Q}$ seja $r < n$. Considere $G \in M_{r\times n}(\mathbb{R})$ a matriz formada pelos r vetores que definem estes hiperplanos e seja $g = G\bar{x}$. Como $r < n$ existe $d \neq 0$ tal que, $Gd = 0$. Então para todo λ temos que $G(\bar{x} + \lambda d) = g$. Desta forma, tome $\epsilon > 0$ de modo que

$$x' = \bar{x} + \epsilon d \in \mathcal{Q}, \qquad x'' = \bar{x} - \epsilon d \in \mathcal{Q}.$$

Consequentemente $\bar{x}$ pode ser escrito como

$$\bar{x} = \frac{1}{2}x' + \frac{1}{2}x'',$$

onde $x', x'' \in \mathcal{Q}$. Desta forma, $\bar{x}$ não é um ponto extremo. Portanto, concluímos que se $\bar{x} \in \mathcal{Q}$ é ponto extremo de $\mathcal{Q}$, então $\bar{x}$ se encontrar em n hiperplanos linearmente independentes que definem $\mathcal{Q}$.

Agora mostraremos a recíproca. Suponha que $\bar{x} \in \mathcal{Q}$ se encontra em n hiperplanos linearmente independentes que definem $\mathcal{Q}$. Note que o conjunto $\mathcal{Q}$ pode ser escrito da seguinte maneira equivalente,

$$\mathcal{Q} = \left\{x \in \mathbb{R}^n : \tilde{A}x \geq \tilde{b},\ Ax \leq b\right\}, \qquad \tilde{A} = \begin{pmatrix} A \\ I \end{pmatrix}, \quad \tilde{b} = \begin{pmatrix} b \\ 0 \end{pmatrix}.$$

Agora seja $G \in M_{n\times n}(\mathbb{R})$ a submatriz inversível de $\tilde{A}$ correspondente aos n hiperplanos linearmente independentes que contém $\bar{x}$ e seja g o subvetor de $\tilde{b}$ definido por $g := G\bar{x}$. Agora suponha que $\bar{x}$ pode ser escrito como

$$\bar{x} = \lambda\tilde{x} + (1-\lambda)\bar{\bar{x}},$$

onde $\lambda \in (0,1)$ e $\tilde{x}, \bar{\bar{x}} \in \mathcal{Q}$. Note que, como $\tilde{x}, \bar{\bar{x}} \in \mathcal{Q}$, pela definição de G temos

$$G\tilde{x} \geq g, \qquad G\bar{\bar{x}} \geq g. \tag{1-3}$$

Da definição de g e da combinação linear acima obtemos com simples manipulações algébricas que

$$0 = G\bar{x} - g = \lambda(G\tilde{x} - g) + (1-\lambda)(G\bar{\bar{x}} - g).$$

Como $\lambda > 0$ e $1 - \lambda > 0$ obtemos de (1-3) e da última igualdade que

$$G\tilde{x} = g, \qquad G\bar{\bar{x}} = g.$$

Como a matriz G é inversível, as igualdades acima implicam que $\tilde{x} = \bar{\bar{x}}$. Portanto $\bar{x}$ é ponto extremo de $\mathcal{Q}$. Finalizando assim a prova do teorema. □

Agora, estamos interessados em verificar que o número de pontos extremos do conjunto $\mathcal{Q}$ é finito. Veja Teorema 2.1 na página 69 em [3].

Proposição 1.3.5 *Sejam $A \in M_{m\times n}(\mathbb{R})$ com posto (A) = $m > 0$ e*

$$\mathcal{Q} = \{x \in \mathbb{R}^n : Ax = b,\ x \geq 0\} \neq \emptyset.$$

Então o conjunto de pontos extremos de $\mathcal{Q}$ é não vazio e finito.

Demonstração: Primeiramente, mostraremos que o conjunto de pontos extremos de $\mathcal{Q}$ é não vazio. Para isso, considere $\bar{x} \in \mathcal{Q}$. Se $\bar{x}$ é um ponto extremo então a primeira parte do resultado está concluída. Senão, considere o sistema

$$Gx = g,$$

onde G é a matriz formada pelos vetores que definem os hiperplanos que contém $\bar{x}$. Seja r o posto da matriz G. Como estamos assumindo que $\bar{x}$ não é ponto extremo, pela Proposição 1.3.4, temos $0 \leq r < n$. Desde que Posto$(G) = r < n$, existe $d \neq 0$ tal que

$$Gd = 0.$$

Agora, movendo a partir de $\bar{x}$ ao longo de d e $-d$, temos que estas duas direções permitem passos de tamanhos positivos mas, ambas não permitem tamanhos de passos infinitos pois, $\mathcal{Q} \subseteq \{x \in \mathbb{R}^n : x \geq 0\}$. Desta forma, sem perda de generalidade, podemos assumir que

$$y^1 = \bar{x} - \bar{\gamma}d \in \mathcal{Q}, \qquad 0 < \bar{\gamma} = \max\{\gamma : \bar{x} - \gamma d \in \mathcal{Q}\} < \infty.$$

Note que, como $Gd = 0$ temos que $Gy^1 = g$ e, como $0 < \bar{\gamma} < \infty$ segue-se que y^1 está contido em um hiperplano que não contém $\bar{x}$. Seja v o vetor que define este novo hiperplano que contém y^1 e note que

$$\langle v, d \rangle \neq 0,$$

pois caso contrário, $\bar{x}$ se encontraria neste novo hiperplano. Assim

$$\text{Posto}\left(\begin{bmatrix} G \\ v^T \end{bmatrix}\right) = r + 1,$$

pois senão, existiria $u \in \mathbb{R}^n$ tal que $v = G^T u$ e assim obteríamos a seguinte contradição,

$$0 \neq \langle v, d \rangle = \left\langle G^T u, d \right\rangle = u^T G d = 0.$$

Se $r + 1 = n$, então y^1 é um ponto extremo. Caso contrário, podemos repetir este processo até obter n hiperplanos linearmente independentes e assim o conjunto de pontos extremos de $\mathcal{Q}$ é não vazio. Agora para verificar que o conjunto de pontos extremos de $\mathcal{Q}$ é finito, basta notar que o número de maneiras de se escolher n hiperplanos linearmente independentes dentre $(m + n)$ hiperplanos, é um número finito. $\square$

Gostaríamos de observar que uma prova analítica da Proposição 1.3.5 pode ser feita usando o Teorema 1.3.1 acima.

Podemos definir direções em conjuntos convexos. Sejam $A \in M_{m\times n}(\mathbb{R})$ com posto (A) = $m > 0$ e

$$\mathcal{Q} = \{x \in \mathbb{R}^n : Ax = b,\ x \geq 0\} \neq \emptyset.$$

O conjunto das *direções de recessão* de $\mathcal{Q}$ é o conjunto

$$\mathcal{D} = \left\{d \in \mathbb{R}^n : Ad = 0,\ e^T d = 1,\ d \geq 0\right\}.$$

Agora, estamos aptos à demonstrar que qualquer ponto no conjunto $\mathcal{Q}$ pode ser escrito como combinação convexa de pontos extremos e direções de recessão em $\mathcal{Q}$. Veja Teorema 2.1 na página 69 em [3].

Teorema 1.3.6 *Sejam* $A \in M_{m\times n}(\mathbb{R})$ *com posto (A)* = $m > 0$ *e*

$$\mathcal{Q} = \{x \in \mathbb{R}^n : Ax = b,\ x \geq 0\} \neq \emptyset.$$

Todo ponto $\bar{x} \in \mathcal{Q}$ *pode ser escrito como*

$$\bar{x} = \sum_{j=1}^{k} \lambda_j x^j + \sum_{j=1}^{l} u_j d^j, \qquad \sum_{j=1}^{k} \lambda_j = 1, \tag{1-4}$$

onde $x^1, \ldots, x^k$ *são pontos extremos de* $\mathcal{Q}$, $d^1, \ldots, d^l$ *são as direções de recessão em* $\mathcal{Q}$, $\lambda_j \geq 0$ *para* $j = 1, \cdots, k$ *e* $u_j \geq 0$ *para* $j = 1, \ldots, l$.

Demonstração: Considere $\bar{x} \in \mathcal{Q}$. Defina o conjunto

$$\bar{\mathcal{Q}} = \mathcal{Q} \cap \{x : e^T x \leq M\},$$

onde M é grande o suficiente para que $e^T x^j \leq M$ para $j = 1, \ldots, k$ e $e^T \bar{x} \leq M$. Note que $\bar{\mathcal{Q}}$ é limitado e que os pontos extremos de $\mathcal{Q}$ estão contidos no conjunto $\bar{\mathcal{Q}}$. Seja $\bar{V} = \{x^1, \ldots, x^k, \ldots, x^{k+\ell}\}$ o conjunto dos pontos extremos de $\bar{\mathcal{Q}}$, onde $0 \leq \ell < \infty$. Em príncipio, iremos mostrar que $\bar{x}$ pode ser escrito como combinação convexa dos pontos em $\bar{V}$. Se $\bar{x} \in \bar{V}$ então este pode ser escrito como combinação convexa dos pontos de $\bar{V}$. Senão, considere o sistema $Gx = g$, que representa o sistema constituído pelos hiperplanos onde $\bar{x}$ se encontra e note que, Posto$(G) \leq (n-1)$. Como Posto$(G) \leq (n-1)$ encontre uma solução $d \neq 0$ para o sistema $Gd = 0$ e calcule

$$\bar{\gamma}_1 = \max\{\gamma :\ \bar{x} + \gamma d \in \bar{\mathcal{Q}}\},$$

onde $0 < \bar{\gamma}_1 < \infty$ pois o conjunto $\bar{\mathcal{Q}}$ é limitado. Denote $y^1 = \bar{x} + \bar{\gamma}_1 d \in \bar{\mathcal{Q}}$. Note que, como $Gd = 0$ temos que $Gy^1 = g$ e, como $0 < \bar{\gamma} < \infty$ segue-se que y^1 está contido em um hiperplano que não contém $\bar{x}$. Seja v o vetor que define este novo hiperplano que contém y^1 e note que

$$\langle v, d\rangle \neq 0,$$

pois caso contrário, $\bar{x}$ se encontraria neste novo hiperplano. Assim

$$\text{Posto}\left(\begin{bmatrix} G \\ v^T \end{bmatrix}\right) = r + 1,$$

pois senão, existiria $u \in \mathbb{R}^n$ tal que $v = G^T u$ e assim obteríamos a seguinte contradição,

$$0 \neq \langle v, d\rangle = \left\langle G^T u, d\right\rangle = u^T G d = 0.$$

Se $r + 1 = n$ então, pela Proposição 1.3.4, $\bar{x}$ é ponto extremo. Caso contrário, podemos repetir este processo não mais que $(n - \text{Posto}(G))$, obtendo um ponto extremo $\bar{x}^1 \in \bar{\mathcal{Q}}$ que satisfaz $G\bar{x}^1 = g$. Agora defina

$$\bar{\gamma}_2 = \max\{\gamma : \ \bar{x} + \gamma(\bar{x} - \bar{x}^1) \in \bar{\mathcal{Q}}\},$$

e denote

$$y^2 = \bar{x} + \bar{\gamma}_2(\bar{x} - \bar{x}^1) \in \bar{\mathcal{Q}}.$$

Como $\bar{\mathcal{Q}}$ é limitado, temos $\bar{\gamma}_2 < \infty$. Agora, desde que $G[\bar{x} + \gamma(\bar{x} - y^1)] = g$ obtemos que y^2 está contido em alguns hiperplanos que não contém $\bar{x}$, e assim temos também que $\bar{\gamma}_2 > 0$. Em particular, $Gy^2 = g$ e novamente um hiperplano linearmente independente adicional contém y^2. Além disso, $\bar{x}$ é uma combinação convexa de $\bar{x}^1$ e y^2,

$$\bar{x} = \delta\bar{x}^1 + (1 - \delta)y^2,$$

onde $\delta = \bar{\gamma}_2/(1+\bar{\gamma}_2)$. Como $\bar{x}^1$ é ponto extremo de $\bar{\mathcal{Q}}$ temos que, se y^2 também for ponto extremo, então $\bar{x}$ é escrito como combinação convexa de pontos extremos do conjunto $\bar{\mathcal{Q}}$. Caso contrário, y^2 pode ser escrito como combinação convexa de $\bar{x}^2$ e y^3, onde $\bar{x}^2$ é um ponto extremo de $\bar{\mathcal{Q}}$ e y^3 esta contido em pelo menos um hiperplano linearmente independente adicional. Continuando com este processo, podemos escrever $\bar{x}$ como combinação convexa de, não mais que, $n-\text{Posto}(G)+1$ pontos extremos de $\bar{\mathcal{Q}}$. Consideremos então, esta representação de $\bar{x}$ dada por

$$\bar{x} = \sum_{j=1}^{k+\ell} \delta_j x^j, \qquad \text{onde } \sum_{j=1}^{k+\ell} \delta_j = 1, \ \text{ e } \ \delta \geq 0. \tag{1-5}$$

Se $\delta_j = 0$, para $j > k$, então a equação acima esta na forma (1-4). Senão considere algum x^r com $r > k$ e $\delta_r > 0$. Note que x^r é um ponto extremo gerado pela restrição $e^T x \leq M$, isto é, $e^T x = M$ é um dos hiperplanos linearmente independentes que contém $x^r \in \bar{\mathcal{Q}}$. Os outros $(n-1)$ hiperplanos linearmente independentes estão associados à algumas equações que definem $\mathcal{Q}$. Seja $x^{i(r)} \in \mathcal{Q}$, com $1 \leq i(r) \leq k$, um ponto extremo contido nestes $n-1$ hiperplanos linearmente independentes. Note que $x^{i(r)}$ se encontra nos mesmos hiperplanos de x^r mais um hiperplano, que define o conjunto $\mathcal{Q}$. Desta forma, como $x^{i(r)}$ se encontra em um hiperplano a mais que x^r, este tem uma coordenada nula a mais do que o ponto x^r. Assim $(x^r - x^{i(r)}) \geq 0$ define uma direção de recessão em $\mathcal{Q}$ da seguinte forma

$$\bar{d} = \frac{(x^r - x^{i(r)})}{\theta_r},$$

onde $\theta_r = e^T(x^r - x^{i(r)}) > 0$. Os $(n-1)$ hiperplanos linearmente independentes do sistema $Ad = 0, \ d \geq 0$ que correspondem aos $(n-1)$ hiperplanos linearmente independentes de $\mathcal{Q}$ contendo x^r, também contém $\bar{d}$. Também, estes $(n-1)$ hiperplanos linearmente independentes contendo x^r, mais o hiperplano $e^t d = 1$, produzem n hiperplanos linearmente independentes de D que contém $\bar{d}$. Assim, $\bar{d}$ deve ser um ponto extremo $d^{j(r)}$ de D. Consequentemente, temos que

$$x^r = x^{i(r)} + \theta_v d^{j(r)}.$$

Substituindo a equação acima na equação (1-5) para cada v e colocando $i(r) = j(r) = 1$ se $\delta_r = 0$, obtemos que

$$\bar{x} = \sum_{j=1}^{k} \delta_j x^j + \sum_{v=k+1}^{k+\ell} \delta_v x^{i(r)} + \sum_{v=k+1}^{k+\ell} \delta_v \theta_v d^{j(r)},$$

que é da forma (1-4). E isto completa a demonstração. $\square$

1.3.1 Projeção em Conjuntos Convexos

Nesta seção, estaremos interessados em propriedades da projeção sobre conjuntos convexos. Para isso começamos definindo distância de um ponto x à um conjunto convexo. A demonstração do próximo resultado pode ser encontrada na página 116 em [12].

Proposição 1.3.7 *Seja C um conjunto convexo, fechado e não vazio de $\mathbb{R}^n$. Para $x \in \mathbb{R}^n$ fixado, o problema*

$$\inf \left\{ ||y - x||^2 : \ y \in C \right\}, \tag{1-6}$$

tem solução.

Demonstração: Seja $c \in C$. Considere o conjunto de subnível

$$S := \{y \in \mathbb{R}^n : \ ||y - x|| \leq ||c - x||\}.$$

Então o problema (1-6) é equivalente a

$$\inf\{||y - x|| : \ y \in C \cap S\}.$$

Agora como a aplicação $y \mapsto ||y - x||$ é contínua, segue-se que o conjunto S é compacto e consequentemente o conjunto $C \cap S$ também é compacto. Portanto, pelo Teorema de Weiertrass 1.1.1 o último problema tem solução o que implica que o problema (1-6) tem solução. □

Pelo resultado acima, deduzimos a existência de um ponto em C que minimiza a distância a x e desta forma o problema (1-6) é de fato um mínimo. Veremos que este mínimo é único. Veja Teorema 3.1.1 na página 117 em [12].

Teorema 1.3.8 *Seja $x \in \mathbb{R}^n$. Se para algum $y_x \in C$ tem-se $||y_x - x|| \leq ||y - x||$, para todo $y \in C$, então*

$$(x - y_x)^T (y - y_x) \leq 0, \tag{1-7}$$

para todo $y \in C$.

Demonstração: Tome $y \in C$ arbitrário de forma que $y_x + \alpha(y - y_x) \in C$ para qualquer $\alpha \in (0,1)$. Então podemos escrever

$$\begin{aligned} ||y_x - x||^2 &\leq ||y_x + \alpha(y - y_x) - x||^2 \\ &= ||(y_x - x) + \alpha(y - y_x)||^2 \\ &= ||y_x - x||^2 + 2\alpha(y_x - x)^T(y - y_x) + \alpha^2||y - y_x||^2. \end{aligned}$$

Desenvolvendo esta última equação obtemos

$$0 \leq \alpha(y_x - x)^T(y - y_x) + \frac{1}{2}\alpha^2||y - y_x||^2.$$

Dividindo ambos os lados por $\alpha > 0$ e fazendo α tender a zero concluimos que

$$(x - y_x)^T(y - y_x) \leq 0, \ \ \forall\, y \in C.$$

□

Para o próximo resultado veja página 116 em [12].

Teorema 1.3.9 *Seja $x \in \mathbb{R}^n$. Existe um único ponto $y_x \in C$ tal que*

$$||y_x - x|| \leq ||y - y_x||, \qquad \forall\, y \in C.$$

Demonstração: Primeiramente iremos mostrar a existência do ponto y_x. Para isso, suponha que $y_x \in C$ satisfaça $(x - y_x)^T(y - y_x) \leq 0$, para todo $y \in C$. Se $x \in C$, então o resultado vale. Senão

$$\begin{aligned} 0 &\geq (x - y_x)^T(y - y_x) \\ &= (x - y_x)^T(y - x + x - y_x) \\ &\geq ||x - y_x||^2 - ||y_x - x||||y - x||, \end{aligned}$$

onde a desigualdade de Cauchy-Schwarz foi utilizada. Dividindo por $||y_x - x|| > 0$, obtemos $||y_x - x|| \leq ||y - x||$. Agora, para mostrar unicidade suponha que existam $y_x \in C$ e $y_x^{'} \in C$ tais que

$$||y_x - x|| \leq ||y - x|| \qquad \text{e} \qquad ||y_x^{'} - x|| \leq ||y - x||,$$

para todo $y \in C$. Em particular, valem as desigualdades

$$(x - y_x)^T(y_x^{'} - y_x) \leq 0 \qquad \text{e} \qquad (x - y_x^{'})^T(y_x - y_x^{'}) \leq 0,$$

e somando as desigualdades obtemos que

$$(y_x^{'} - y_x)^T (y_x^{'} - y_x) \leq 0.$$

Logo, $y_x^{'} = y_x$. Portanto y_x é único. $\square$

Seja o conjunto $C \subset \mathbb{R}^n$, defina a função distância à C como sendo

$$\begin{aligned} d(.,C) : \mathbb{R}^n &\to \mathbb{R} \\ x &\mapsto d(x,C) = \inf\{||y - x|| : y \in C\}. \end{aligned}$$

A *projeção ortogonal* sobre C é a função $P_C : \mathbb{R}^n \to C$ definida como sendo o único ponto $P_C(x) \in C$, que existe Pelo Teorema 1.3.9, tal que

$$d(x,C) = ||P_C(x) - x||.$$

Teorema 1.3.10 *Seja $C \subset \mathbb{R}^n$ um conjunto não vazio, convexo e fechado. Então para todo $x \notin C$ a projeção ortogonal $P_C(x)$ de x sobre C está bem definida e satisfaz a seguinte desigualdade*

$$(x - P_C(x))^T (y - P_C(x)) \leq 0, \qquad \forall\, y \in C. \tag{1-8}$$

Demonstração: Segue imediatamente do Teorema 1.3.9 que a projeção está bem definida. E o Teorema 1.3.9 implica a desigualdade (1-8). $\square$

Toda nossa discussão sobre projeção em conjuntos convexos, nos possibilitou obter ferramentas para demonstração do Teorema da Separação, cuja demonstração pode ser encontrada no Teorema 2.4.4 na página 45 em [4].

Teorema 1.3.11 *(Teorema da Separação) Seja C um conjunto convexo, fechado e não vazio em $\mathbb{R}^n$ e considere $b \notin C$. Então existe um vetor $a \in \mathbb{R}^n$, $a \neq 0$, e um escalar $\delta \in \mathbb{R}$ tais que, para todo $x \in C$*

$$a^T x \geq \delta > a^T b.$$

Como consequência, $a^T b < \inf\{a^T x : x \in C\}$.

Demonstração: Sejam C um conjunto convexo, fechado e não vazio e $b \notin C$. Pelo Teorema 1.3.10 a projeção $P_C(b) \in C$ está bem definida e satisfaz

$$(b - P_C(b))^T (x - P_C(b)) \leq 0, \tag{1-9}$$

para cada $x \in C$. Defina

$$a := -(b - P_C(b)) \neq 0, \qquad \delta := a^T P_C(b).$$

A definição de a juntamente com (1-9) resulta que $-a^T(x - P_C(b)) \leq 0$ ou equivalentemente, $a^T x \geq a^T P_C(b)$. Esta última equação juntamente com a definição de δ implica em $a^T x \geq \delta$, que é a primeira desigualdade desejada. Para verificar a segunda desigualdade note que pelas definições de a e δ temos que

$$a^T b - \delta = (-b + P_C(b))^T b + a^T P_C(b) = -||b - P_C(b)||^2 < 0,$$

onde a desigualdade estrita segue do fato que $b \notin C$. Assim, $a^T b < \delta$ o que conclui a primeira parte. A segunda parte é imediata. $\square$

O Teorema da Separação afirma que, o ponto $b \notin C$ está separado de C pelo hiperplano $a^T x = \delta$.

1.3.2 Funções convexas

Nesta seção vamos tratar de alguns conceitos básicos a respeito de funções convexas. Nosso objetivo principal é obter as seguintes propriedades importantes:

- o conjunto dos subníveis de uma função convexa é um conjunto convexo;
- funções estritamente convexas possuem no máximo um ponto de mínimo.

Iniciaremos esta seção, tratando de funções convexas de uma variável.

Seja $I \subset \mathbb{R}$ um intervalo. Uma função $f : I \to \mathbb{R}$ é dita *convexa* quando, para todo $x, y \in I$ e todo $\lambda \in [0, 1]$

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y), \tag{1-10}$$

e é estritamente convexa se (1-10) é estrita para todo $x, y \in I$ com $x \neq y$ e todo $\lambda \in (0, 1)$.

O próximo resultado fornece uma caracterização de funções convexas, cuja demonstração pode ser encontrada no Teorema 4 e Corolário 2, respectivamente nas páginas 106 e 108 em [18].

Proposição 1.3.12 *Sejam $f : I \to \mathbb{R}$ uma função diferenciável no intervalo aberto I. As seguintes afirmações valem*

i) *A função f é convexa se, e somente se,*

$$f'(x)(y-x)+f(x)\leq f(y), \qquad \forall\, x,y\in I. \tag{1-11}$$

Se (1-11) *é estrita para todo $x,y\in I$, então f é estritamente convexa.*

ii) *A função f é convexa se, e somente se,*

$$\big(f'(x)-f'(y)\big)(x-y)\geq 0, \qquad \forall\, x,y\in I. \tag{1-12}$$

Se (1-12) *é estrita para todo $x,y\in I$, então f é estritamente convexa.*

iii) *A função f é convexa se, e somente se*

$$f''(x)\geq 0, \qquad \forall\, x\in I. \tag{1-13}$$

Se (1-13) *é estrita para todo $x\in I$, então f é estritamente convexa.*

Demonstração: Dados $x,y\in I$, temos por hipótese que

$$f\big(\lambda y+(1-\lambda)x\big)\leq \lambda f(y)+(1-\lambda)f(x).$$

Após algumas manipulações algébricas segue-se que

$$\frac{f\big(x+\lambda(y-x)\big)+f(x)}{\lambda}\leq f(y)-f(x),$$

para todo $\lambda\in[0,1]$. Fazendo $\lambda\to 0$ na última desigualdade temos

$$f'(x)(y-x)+f(x)\leq f(y),$$

que é o resultado desejado. Reciprocamente, dados $x,y\in I$ e $\lambda\in[0,1]$ tome $z=\lambda x+(1-\lambda)y$. Então da equação (1-11) segue-se que

$$f'\big(\lambda x+(1-\lambda)y\big)(1-\lambda)(x-y)+f\big(\lambda x+(1-\lambda)y\big)\leq f(x) \tag{1-14}$$

e

$$f'\big(\lambda x+(1-\lambda)y\big)t(y-x)+f\big(\lambda x+(1-\lambda)y\big)\leq f(y). \tag{1-15}$$

Multiplicando (1-14) por λ e (1-15) por $(1-\lambda)$ e adicionando o resultado obtemos $f\big(\lambda x+(1-\lambda)y\big)\leq \lambda f(x)+(1-\lambda)f(y)$. Portanto f é convexa. Agora, se (1-11) é estrita para todo $x,y\in I$, então (1-14) e (1-15) valem para desigualdade estrita e analogamente concluímos que f é estritamente convexa. Isto prova o item i.

Para provar o item ii, tome $x, y \in I$, como f é convexa em I temos pela Proposição 1.3.12 item i que

$$f'(x)(y-x) + f(x) \leq f(y) \quad \text{e} \quad f'(y)(x-y) + f(y) \leq f(x), \tag{1-16}$$

para todo x, y em I. Então, segue-se de (1-16) que

$$f(x) + f'(x)(y-x) \leq f(y) \leq f(x) + f'(y)(y-x),$$

isto implica que $\big(f'(x) - f'(y)\big)(x-y) \geq 0$, para todo x, y em I. Isto prova a primeira parte. Reciprocamente, é claro que o resultado é válido para $x = y$. Agora, dados $x, y \in I$ com $x \neq y$ e $\lambda \in [0,1]$, temos pelo Teorema do Valor Médio Teorema 1.2.1 que, para todo $a \in (x, y)$, existem c' e c'' com $c' \in (x, a)$ e $c'' \in (a, y)$ tais que

$$f(a) - f(x) = f'(c')(a-x) \quad \text{e} \quad f(y) - f(a) = f'(c'')(y-a).$$

Multiplicando a primeira igualdade por $-\lambda$, a segunda por $1-\lambda$, tomando $a = \lambda x + (1-\lambda)y$ e adicionando o resultado obtemos a seguinte igualdade

$$\lambda f(x) + (1-\lambda)f(y) - f\big(\lambda x + (1-\lambda)y\big) = f'(c'')(1-\lambda)\lambda(y-x) - f'(c')\lambda(1-\lambda)(y-x). \tag{1-17}$$

Note que existe $k > 0$ tal que $c'' - c' = k(y-x)$ ou seja, $y - x = (c'' - c')/k$. Assim a igualdade anterior se reduz a

$$\lambda f(x) + (1-\lambda)f(y) - f\big(\lambda x + (1-\lambda)y\big) = \frac{(1-\lambda)\lambda}{k}\big(f'(c'') - f'(c')\big)(c'' - c'). \tag{1-18}$$

Como $k > 0$ e $\lambda \in [0,1]$ segue da igualdade anterior e desigualdade (1-12) que

$$f\big(\lambda x + (1-\lambda)y\big) \leq \lambda f(x) + (1-\lambda)f(y),$$

e portanto f é convexa. Agora, se (1-12) é estrita para todo $x, y \in I$ e como $k > 0$ e $\lambda \in (0,1)$ na equação (1-17), então (1-18) vale para desigualdade estrita e f é estritamente convexa. Isto prova o item ii.

Prova do item iii. Se fé convexa, segue-se do item ii que

$$\frac{\big(f'(x+h) - f'(x)\big)h}{h^2} \geq 0, \tag{1-19}$$

para $x \in I$ e $h \in \mathbb{R}$ tal que $x + h \in I$. Fazendo $h \to 0$ na equação (1-19) obtemos $f''(x) \geq 0$, e isto prova a primeira parte. Reciprocamente, é claro que o resultado é válido para $x = y$. Agora, dados $x, y \in I$ com $x \neq y$ temos, pelo Teorema do Valor Médio 1.2.1, que existe a entre x e y tal que

$$f'(y) - f'(x) = f''(a)(y - x). \tag{1-20}$$

Como $f''(a) \geq 0$, segue-se desta igualdade que

$$\big(f'(y) - f'(x)\big)(y - x) \geq 0,$$

para todo $x, y \in I$. Portanto, pelo item ii, temos que f é convexa. Agora, se (1-13) é estrita para todo $x, y \in I$, então (1-14) vale para desigualdade estrita e pelo item ii segue que f é estritamente convexa. $\square$

Agora iremos tratar sobre funções convexas definidas em subconjuntos convexos de $\mathbb{R}^n$. Seja C um conjunto não vazio convexo em $\mathbb{R}^n$. Uma função $f : C \to \mathbb{R}$ é dita convexa quando, para todo $x, y \in C$ e todo $\lambda \in [0, 1]$ vale

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y), \tag{1-21}$$

e é estritamente convexa se a desigualdade (1-21) é estrita para todo $x, y \in C$ com $x \neq y$ e todo $\lambda \in (0, 1)$.

A próxima proposição afirma que os conjuntos dos subníveis de uma função convexa são convexos. Veja Teorema 3.4.1 página 133 em [13].

Proposição 1.3.13 *Sejam $C \subset \mathbb{R}^n$ um conjunto convexo e $f : C \to \mathbb{R}$ uma função convexa. Então para todo $r \in \mathbb{R}$ o conjunto subnível*

$$\{x \in C : f(x) \leq r\},$$

é um conjunto convexo.

Demonstração: Tome $r \in \mathbb{R}$ arbitrário. Se o conjunto $L = \{x \in C : f(x) \leq r\}$ for vazio, o resultado segue. Caso contrário tomemos $x \in L$ e $y \in L$. Assim, temos que

$$f(x) \leq c \quad \text{e} \quad f(y) \leq c. \tag{1-22}$$

Como C é convexo $\alpha x + (1 - \alpha)y \in C$. Agora como a f é convexa obtemos que

$$f(x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y).$$

Da desigualdade acima e de (1-22) obtemos que $f(x + (1-\alpha)y) \leq c$. Portanto $x + (1-\alpha)y \in L$, concluindo a demonstração. $\square$

Agora, estaremos interessados em verificar que uma função convexa sobre um conjunto convexo tem no máximo um ponto de mínimo. Para isso precisamos definir ponto de mínimo local e global. Sejam Ω um subconjunto de $\mathbb{R}^n$ e uma função $f : \Omega \to \mathbb{R}$. Um ponto $x^* \in \Omega$ é um mínimo local de f se existir $r > 0$ tal que

$$f(x^*) \leq f(x), \qquad \forall\, x \in B(x^*, r) \cap \Omega,$$

e um mínimo global de f se $f(x^*) \leq f(x)$, para todo $x \in \Omega$.

É fácil ver que a bola aberta é convexa e que a interseção de dois conjuntos convexos é convexo. Vamos usar este fato no seguinte lema, cuja demonstração pode ser encontrada no Teorema 3.1.5 na página 69 em [13].

Lema 1.3.14 *Sejam C um subconjunto convexo de $\mathbb{R}^n$ e $f : C \to \mathbb{R}$ uma função convexa. As seguintes alternativas valem:*

i) *Se x^* é um mínimo local de f em C. Então x^* é mínimo global de f.*

ii) *Se $f : C \to \mathbb{R}$ é estritamente convexa, então f possui no máximo um ponto de mínimo.*

Demonstração: Suponhamos que f admita um mínimo local x^* que não seja um mínimo global. Sendo x^* um mínimo local, existe $B(x^*, r)$ tal que $f(x^*) \leq f(x)$ para todo $x \in B(x^*, r) \cap C$. Como x^* não é mínimo global, existe $z \in \mathbb{R}^n$ tal que $f(z) < f(x^*)$. Sendo $B(x^*, r) \cap C$ um conjunto convexo, $\lambda x^* + (1-\lambda)z \in B(x^*, r) \cap C$, para todo $\lambda \in [0, 1]$. Agora escolhendo λ suficientemente próximo da unidade, pela convexidade de f, temos

$$f(\lambda x^* + (1-\lambda)z) \leq \lambda f(x^*) + (1-\lambda) f(z) < f(x^*), \tag{1-23}$$

pois $f(z) < f(x^*)$. Mas pela construção, $\lambda x^* + (1-\lambda)z$ está em $B(x^*, r) \cap C$, assim $f(x^*) \leq f(\lambda x^* + (1-\lambda)z)$, chegando assim a uma contradição com (1-23). Portanto, x^* é um mínimo global de f em $\mathbb{R}^n$. Isto prova o item i.

Prova do item ii. Suponhamos que $x, y \in C$ sejam mínimos de f. Suponhamos por absurdo que $x \neq y$. Desde que f é estritamente convexa temos

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y), \qquad \forall\, \lambda \in (0, 1).$$

Pelo item i temos que x e y são mínimos globais, assim $f(x) = f(y)$. Deste modo a equação anterior se reduz a

$$f(\lambda x + (1-\lambda)y) < f(x) = f(y), \qquad \forall\, \lambda \in (0,1).$$

Isto é um absurdo, pois x e y são mínimos globais. □

A demonstração do próximo resultado pode ser encontrada no Teorema 8 na página 76 em [19].

Proposição 1.3.15 *Sejam $C \subset \mathbb{R}^n$ um conjunto convexo e uma função $f : C \to \mathbb{R}$. A função f é convexa se, e somente se, a restrição de f a todo segmento de reta, que esta contido em C, é uma função convexa. Mais precisamente, a função f é convexa se, e somente se, para todo $x \in C$ e $v \in \mathbb{R}^n$ a função $h : I \to \mathbb{R}$ definida por*

$$h(t) = f(x + tv), \tag{1-24}$$

é convexa, onde $I = \{t \in \mathbb{R} : x + tv \in C\}$.

Demonstração: Fixe $x \in C$ e $v \in \mathbb{R}^n$ e suponha que f seja convexa. Dados $t_1, t_2 \in I$ e $\lambda \in [0,1]$, pela definição de h e convexidade de f temos

$$\begin{aligned} h\big(\lambda t_1 + (1-\lambda)t_2\big) &= f\big(\lambda(x + t_1 v) + (1-\lambda)(x + t_2 v)\big) \\ &\leq \lambda f(x + t_1 v) + (1-\lambda) f(x + t_2 v) \\ &= \lambda h(t_1) + (1-\lambda) h(t_2). \end{aligned}$$

e isto implica que h é convexa em I. Reciprocamente, dados $x, y \in C$ e $\lambda \in [0,1]$. Seja $I = \{t \in \mathbb{R} : x + t(y-x) \in C\}$. Note que, sendo C convexo, $x + \lambda(y-x) = \lambda y + (1-\lambda)x \in C$ para $\lambda \in [0,1]$ e assim $\lambda \in I$. Defina $h : I \to \mathbb{R}$ por $h(t) = f(x + t(y-x))$. Como, por hipótese, h é convexa segue-se que

$$\begin{aligned} f(\lambda y + (1-\lambda)x) &= f(x + t(y-x)) = h(\lambda) = h(\lambda 1 + (1-\lambda)0) \\ &\leq \lambda h(1) + (1-\lambda)h(0) = \lambda f(y) + (1-\lambda) f(x), \end{aligned}$$

e isto implica que f é convexa. □

É imediato do resultado acima que a restrição de uma função convexa em um conjunto convexo é convexa.

Veja no Teorema 3.4.7 na página 146 em [13] que, se uma função f é duas vezes diferenciável ela pode ser caracterizada da seguinte forma.

Proposição 1.3.16 *Sejam C um subconjunto aberto e convexo de $\mathbb{R}^n$ e $f : C \to \mathbb{R}$ uma função duas vezes diferenciável. A função f é convexa se, e somente se, para todo $x \in C$, $v \in \mathbb{R}^n$*

$$\langle f''(x)v, v\rangle \geq 0. \tag{1-25}$$

Se a desigualdade (1-25) *é estrita , então f é estritamente convexa.*

Demonstração: Dados $x \in C$ e $v \in \mathbb{R}^n$, defina h como na Proposição 1.3.15. Note que I é aberto, pois C é aberto e que $0 \in I$. Sabendo-se que f é convexa, pela Proposição 1.3.15, segue-se que h é convexa. Então, pela Proposição 1.3.12, temos que

$$h''(0) \geq 0. \tag{1-26}$$

Pela definição de h, e regra da cadeia, temos que $h''(t) = \langle f''(x+tv)v, v\rangle$, assim (1-26) é equivalente a

$$\Big\langle f''(x)v, v\Big\rangle \geq 0,$$

o que prova a primeira parte. Reciprocamente, dados $x \in C$ e $v \in \mathbb{R}^n$. Defina h como na Proposição 1.3.15. Note que o intervalo I é aberto e que

$$h''(t) = \langle f''(x+tv)v, v\rangle.$$

Então pela definição de h e equação (1-25), temos que para todo $t \in I$

$$h''(t) = \langle f''(x+tv)v, v\rangle \geq 0.$$

Portanto, pela Proposição 1.3.12 segue-se que h é convexa e isto implica, pela Proposição 1.3.15, que f é convexa e a proposição está demonstrada. □

Exemplo 1.3.17 *Dados $c \in \mathbb{R}^n$ e $\mu > 0$, a função $\varphi : \mathbb{R}^n_{++} \to \mathbb{R}$ definida por*

$$\varphi(x) = c^T x - \mu \sum_{i=1}^{n} \log x_i,$$

é estritamente convexa. Como consequência, dado $r \in \mathbb{R}$, o conjunto

$$\Omega = \left\{ x \in \mathbb{R}^n : \quad Ax = b, \quad x > 0, \quad c^T x - \mu \sum_{i=1}^{n} \log x_i \leq r \right\},$$

é convexo.

De fato, a derivada segunda de φ é dada por

$$\varphi''(x) = \mu X^{-2}.$$

Utilizando propriedades do produto interno e definição de norma de vetores temos que,

$$\langle \varphi''(x)v, v\rangle = \left\langle \mu X^{-2}v, v\right\rangle = \mu \left\langle X^{-1}v, X^{-1}v\right\rangle = \mu||X^{-1}v||.$$

Como $x \in \mathbb{R}^n_{++}$ temos que para todo $v \neq 0$, vale $||X^{-1}v|| > 0$. Este fato juntamente com $\mu > 0$ e a equação acima implica que

$$\langle \varphi''(x)v, v\rangle > 0, \qquad \forall\, v \neq 0.$$

Portanto, pela Proposição 1.3.16, φ é estritamente convexa.

Agora, da Proposição 1.3.13 podemos concluir que o conjunto subnível

$$S = \left\{ x \in \mathbb{R}^n_{++} : \quad c^T x - \mu \sum_{i=1}^{n} \log x_i \leq r \right\},$$

é convexo. Por outro lado, o conjunto $L = \{x \in \mathbb{R}^n : \quad Ax = b\}$ é convexo. Como interseção de conjuntos convexos é convexo e $\Omega = S \cap L$ segue-se que Ω é convexo.

1.4 Resultados Básicos de Álgebra Linear

Matrizes têm fundamental importância no estudo de álgebra linear. Uma matriz quadrada A é dita *não-singular* se a única solução do sistema $Ax = 0$ é $x = 0$.

Lema 1.4.1 *Sejam $u \in \mathbb{R}^n_{++}$, $v \in \mathbb{R}^n_{++}$ e $W \in M_{m\times n}(\mathbb{R})$ com Posto(W) = $m < n$. Se $U = diag(u)$ e $V = diag(v)$, então a matriz*

$$M = \begin{pmatrix} W & 0_{m\times m} & 0_{m\times n} \\ 0_{n\times n} & W^T & I \\ V & 0_{n\times m} & U \end{pmatrix},$$

é não-singular, onde I é a matriz identidade $n \times n$.

Demonstração: Defina o vetor $z = (z^1, z^2, z^3)^T \in \mathbb{R}^{2n+m}$ onde $z^1, z^3 \in \mathbb{R}^n$ e $z^2 \in \mathbb{R}^m$. Basta mostrar que a única solução do sistema $Mz = 0$ é o vetor

$0 \in \mathbb{R}^{2n+m}$. A igualdade $Mz = 0$ é equivalente ao seguinte sistema:

$$Wz^1 = 0, \tag{1-27}$$

$$W^T z^2 + z^3 = 0, \tag{1-28}$$

$$Vz^1 + Uz^3 = 0. \tag{1-29}$$

Isolando o vetor z^3 na expressão (1-29) e substituindo na expressão (1-28) temos que

$$W^T z^2 - U^{-1}Vz^1 = 0 \Rightarrow W^T z^2 = U^{-1}Vz^1 \Rightarrow z^1 = V^{-1}UW^T z^2.$$

Substituindo o vetor z^1, obtido na última implicação, na expressão (1-27) obtemos que

$$WV^{-1}UW^T z^2 = 0. \tag{1-30}$$

Agora note que $V^{-1} = V^{-1/2}V^{-1/2}$ e $U = U^{1/2}U^{1/2}$ desta forma, a expressão (1-30) pode ser reescrita como

$$WV^{-1/2}V^{-1/2}U^{1/2}U^{1/2}W^T z^2 = 0. \tag{1-31}$$

Com algumas manipulações algébricas, e observando que V e U são matrizes diagonais, a expressão (1-31) pode ser reformulada da seguinte forma

$$WV^{-1/2}U^{1/2}V^{-1/2}U^{1/2}W^T z^2 = 0 \Rightarrow WU^{1/2}V^{-1/2}V^{-1/2}U^{1/2}W^T z^2 = 0. \tag{1-32}$$

Denotando $B = WU^{1/2}V^{-1/2}$, a expressão (1-32) é reescrita da seguinte forma $BB^T z^2 = 0$. Como Posto(W) = $m < n$, segue das definições de B, U, V e última igualdade que $z^2 = 0$.
Com $z^2 = 0$ temos que na expressão (1-28), $z^3 = 0$ e como V é inversível temos que na expressão (1-29) $Vz^1 = 0$ implica que $z^1 = 0$. Desta forma a única solução do sistema $Mz = 0$ é o vetor nulo. Portanto a matriz M é não-singular. $\square$

Uma matriz A pode representar diversas estruturas em Álgebra Linear. Associados à matriz $A \in M_{m\times n}(\mathbb{R})$ podemos destacar dois subespaços do $\mathbb{R}^n$, a saber, o espaço nulo de A definido por

$$\textbf{Null}(A) = \{x \in \mathbb{R}^n : Ax = 0\},$$

e o espaço imagem de A^T (ou espaço linha de A), definido por

$$\text{Im}(A^T) = \left\{x \in \mathbb{R}^n : x = A^T z,\ z \in \mathbb{R}^m\right\}.$$

A seguir vamos obter uma relação entre estes subespaços. A demonstração da próxima proposição pode ser encontrada em [25].

Proposição 1.4.2 *Considere o $\mathbb{R}^n$. O subespaço nulo de A é ortogonal ao subespaço linha de A.*

Demonstração: Considere o espaço $\mathbb{R}^n$. Suponha que $x \in \text{Null}(A)$ então por definição temos que

$$Ax = 0. \tag{1-33}$$

E suponha também que $v \in \text{Im}(A^T)$ então por definição

$$v = A^T z, \ \text{ para algum } z \in \mathbb{R}^m. \tag{1-34}$$

Agora usando (1-33) e (1-34) e realizando algumas manipulações matriciais obtemos que

$$v^T x = (A^T z)x = z^T Ax = 0.$$

E por definição, o resultado segue. $\square$

Matrizes também estão associadas à sistemas de equações lineares. Existem diversos métodos de resolução de sistemas de equações lineares. Seja a matriz $A \in M_{m\times n}(\mathbb{R})$ e considere o sistema $Ax = b$. O método de Eliminação de Gauss utiliza sistemas equivalentes para resolver sistemas de equações lineares. O método de Gauss transforma um sistema de equações lineares $Ax = b$, através de operações elementares, em um sistema equivalente $QAx = Qb$, onde a matriz QA fica em uma forma triangular superior, porém os pivôs nem sempre estarão na diagonal principal e Q é a matriz das operações elementares. O importante neste processo, é que a matriz QA possui o mesmo posto da matriz A. Veja Teorema 2D em [25].

Teorema 1.4.3 *Seja $A \in M_{m\times n}(\mathbb{R})$ tal que sua primeira coluna seja não nula. Sejam também os conjuntos de índices $B = \{1, 2, \ldots, k\}$ e $N = \{k+1, k+2, \ldots, n\}$. Se Posto$(A) = m$, então existe uma matriz $Q \in M_{m\times m}(\mathbb{R})$ inversível tal que*

$$QA = \begin{bmatrix} \tilde{A}_{l\times B} & \tilde{A}_{l\times N} \\ 0_{(m-l)\times B} & \tilde{A}_{(m-l)\times N} \end{bmatrix},$$

onde $\tilde{A}_{l\times B} \in M_{l\times B}(\mathbb{R})$ *e* $\tilde{A}_{m-l\times N} \in M_{m-l\times N}(\mathbb{R})$. *Além disso, vale a igualdade* ***Posto***$(\tilde{A}_{l\times B}) +$ ***Posto***$(\tilde{A}_{(m-l)\times N}) = m =$ ***Posto***(A).

No próximo capítulo, estaremos interessados em apresentar os pares de problemas de Programação Linear e obter as equações que definem as condições de otimalidade para estes pares de problemas.

CAPÍTULO 2

Condições de Otimalidade para o Problema de Programação Linear

Nosso objetivo neste capítulo é definir os problemas de programação linear primal e dual no formato padrão, mostrar algumas relações entre o par de problemas através dos resultados de dualidade e, como consequência, obter as equações que definem as condições de otimalidade para o Problema de Programação Linear.

2.1 Os problemas de Programação Linear Primal e Dual

O *Problema de Programação Linear* (PPL) primal no formato padrão é o seguinte problema de Otimização:

$$\begin{array}{lll} \textbf{(P)} & minimizar & c^T x \\ & sujeito\ a: & Ax = b \\ & & x \geq 0, \end{array}$$

onde $A \in M_{m \times n}(\mathbb{R})$, com $m \leq n$, $b \in \mathbb{R}$ e $c \in \mathbb{R}^n$, são os dados do problema, e os pontos $x \in \mathbb{R}^n$ são as variáveis primais.

A seguir faremos algumas notações associadas ao problema (P). O conjunto

$$\mathcal{F}(P) = \{x \in \mathbb{R}^n : Ax = b, \ \ x \geq 0\},$$

é chamado conjunto viável e um ponto $x \in \mathcal{F}(P)$ é denominado ponto viável. O conjunto

$$\mathcal{F}^0(P) = \{x \in \mathbb{R}^n : Ax = b, \ \ x > 0\},$$

é chamado conjunto de pontos interiores viáveis e um ponto deste conjunto é denominado ponto interior viável. O conjunto

$$P^* = \{x^* \in \mathcal{F}(\mathcal{P}) : c^T x^* \leq c^T x, \text{ para todo } x \in \mathcal{F}(P)\},$$

é chamado conjunto de soluções ótimas e um ponto $x^* \in P^*$ é denominado solução ótima.

Resolver o problema (P) consiste em encontrar um ponto $x^* \in P^*$ ou mostrar que $\mathcal{F}(P)$ é vazio, isto é que o problema (P) é inviável, ou ainda, mostrar que a função objetivo é ilimitada no conjunto viável ou seja, (P) é um problema ilimitado.

O problema dual correspondente ao problema primal (P), é o seguinte problema de Otimização:

$$\begin{array}{lll} \text{(D)} & maximizar & b^T y \\ & sujeito \ a: & A^T y + s = c \\ & & s \geq 0, \end{array}$$

onde $y \in \mathbb{R}^m$ e $s \in \mathbb{R}^n$ são as variáveis duais.

A seguir, faremos algumas definições associadas ao problema (D). O conjunto

$$\mathcal{F}(D) = \{(y, s) \in \mathbb{R}^m \times \mathbb{R}^n : A^T y + s = c, \ \ s \geq 0\},$$

é chamado conjunto viável para o problema dual e um par $(y, s) \in \mathcal{F}(D)$ é denominado ponto viável para o problema dual. O conjunto

$$\mathcal{F}^0(D) = \{(y, s) \in \mathbb{R}^m \times \mathbb{R}^n : A^T y + s = c, \ \ s > 0\},$$

é chamado conjunto de pontos interiores viáveis duais e um par neste conjunto é denominado ponto interior viável dual. O conjunto

$$D^* = \{(y^*, s^*) \in \mathcal{F}(D) : b^T y^* \geq b^T y, \text{ para todo } (y, s) \in \mathcal{F}(D)\},$$

é chamado conjunto de soluções ótimas para o problema dual e um par $(y^*, s^*) \in D^*$ é denominado solução ótima para o problema dual.

De agora em diante assumiremos a seguinte hipótese:

(H1) Posto$(A) = m$.

Esta hipótese de fato não é necessária, ela apenas garante uma correspondência unívoca entre as variáveis duais y e s, e isto simplifica nossa análise.

2.2 Condições de Otimalidade

Nesta seção estaremos interessados em apresentar as relações entre o par de problemas Primal e Dual, obtendo as equações que definem as condições de otimalidade para estes problemas. Iniciaremos com o teorema de Dualidade Fraco cuja a demonstração pode ser encontrada no Teorema 3.2.1 na página 70 em [24].

Teorema 2.2.1 *(Dualidade Fraco) Dados $x \in \mathcal{F}(P)$ e $(y, s) \in \mathcal{F}(D)$. Então vale a igualdade*

$$x^T s = c^T x - b^T y.$$

Como consequência $c^T x \geq b^T y$.

Demonstração: Por hipótese $Ax = b$ e $A^T y + s = c$, assim

$$c^T x - b^T y = (A^T y + s)^T x - (Ax)^T y = (A^T y)^T x + x^T s - (Ax)^T y = x^T s.$$

Desde que $x \geq 0$ e $s \geq 0$ a última igualdade implica que $c^T x \geq b^T y$. □

O problema (P) chama-se ilimitado, quando existe uma sequência $\{x^k\} \subset \mathcal{F}(P)$ tal que $c^T x^k \to -\infty$, quando $k \to \infty$. Caso contrário, o problema (P) é limitado. De maneira análoga dizemos que o problema (D) é ilimitado, quando existe uma sequência $\{y^k\} \subset \mathcal{F}(D)$ tal que $b^T y^k \to \infty$, quando $k \to \infty$. Caso contrário, o problema (D) é limitado.

O próximo resultado pode ser encontrado no Corolário 3.2.2 na página 71 em [24].

Corolário 2.2.2 *As seguintes afirmações valem:*

i) *Se o problema (P) é ilimitado, então o problema (D) é um problema inviável;*

ii) *se o problema (D) é ilimitado, então o problema (P) é um problema inviável.*

Demonstração: Suponha que (P) seja ilimitado, então existe uma sequência $\{x^k\}$ em $\mathcal{F}(P)$ tal que $\lim_{k\to\infty} c^T x^k = -\infty$. Então pelo Teorema 2.2.1 segue imediatamente que $\mathcal{F}(D) = \emptyset$, isto prova o item **i**. A prova do item **ii** é feita de modo análogo. □

Agora vamos demonstrar um teorema de dualidade que, entre outras coisas, apresenta relações entre as soluções dos problemas primal e dual, a demonstração pode ser encontrada no Teorema de Dualidade na página 89 em [21].

Teorema 2.2.3 *Uma, e somente uma, das seguintes afirmações acontece:*

a) *O problema (P) tem uma solução ótima se, e somente se, o problema (D) também tem solução ótima. Além disso, os valores das funções objetivo de ambos são iguais.*

b) *Se (P) é um problema ilimitado, então (D) é um problema inviável.*

c) *Se (D) é um problema ilimitado, então (P) é um problema inviável.*

Demonstração: Para o item **a**, iremos mostrar que se o problema (P) tem uma solução ótima, então o problema (D) também tem uma solução ótima e os valores das funções objetivos de ambos são iguais. Assim suponha que o problema (P) tem uma solução ótima com valor ótimo p^*. Defina o conjunto

$$C = \left\{ (w^T, r)^T \in \mathbb{R}^m \times \mathbb{R};\;\; B(x^T, t)^T = (w^T, r)^T;\;\; x \geq 0,\;\; t \geq 0 \right\},$$

onde

$$B = \begin{bmatrix} -A & b \\ -c^T & p^* \end{bmatrix} \in M_{(m+1)\times(n+1)}(\mathbb{R}).$$

Agora verifiquemos que $(0,1) \notin C$, onde $0 \in R^m$. Suponhamos que $(0,1) \in C$, então existem $\bar{x} \geq 0$ e $\bar{t} \geq$ tais que

$$\bar{t}b - A\bar{x} = 0, \tag{2-1}$$
$$\bar{t}p^* - c^T\bar{x} = 1. \tag{2-2}$$

Temos duas possibilidades $\bar{t} > 0$ e $\bar{t} = 0$. Se $\bar{t} > 0$, como $\bar{x} \geq 0$ temos de (2-1) que $\bar{x}/\bar{t}$ é viável para (P). Já que $\bar{x}/\bar{t}$ é viável para (P) e p^* é o valor ótimo de (P) temos $p^* \leq c^T(\bar{x}/\bar{t})$. Como estamos assumindo que $\bar{t} > 0$ a última desigualdade é equivalente à $\bar{t}p^* - c^T\bar{x} \leq 0$, o que contradiz (2-2). Agora vamos supor que $\bar{t} = 0$. Então o sistema acima reduz-se à

$$A\bar{x} = 0,$$
$$c^T\bar{x} = -1.$$

Note que, se x é viável para (P) então desde que $\bar{x} \geq 0$ temos que $x + \alpha\bar{x} \geq 0$, para todo $\alpha \geq 0$. Além disso, a primeira igualdade no último sistema implica que $x+\alpha\bar{x}$ também é viável para (P). Como p^* é o valor ótimo e, para todo $\alpha \geq 0$, o ponto $x+\alpha\bar{x}$ é viável para (P), temos da segunda igualdade no sistema acima que

$$p^* \leq c^T(x + \alpha\bar{x}) = c^T x - \alpha.$$

Como a igualdade acima vale para todo $\alpha \geq 0$ temos um absurdo, visto que podemos fazer α tender para mais infinito, o que resulta no lado direito desta igualdade tender a menos infinito. Assim concluímos a prova que $(0,1) \notin C$.

Agora, da Proposição 1.3.3, temos que C é convexo, fechado e não vazio. Desde que $(0,1) \notin C$ temos, pelo Teorema da Separação 1.3.11, que existe um hiperplano que separa $(0,1)$ de C, isto é, existe um vetor não zero $(v^T, \lambda)^T \in \mathbb{R}^m \times \mathbb{R}$ e uma constante ℓ tal que

$$\lambda < \ell := inf\{\lambda r + v^T w : \quad (w,r) \in C\}. \tag{2-3}$$

Temos duas possibilidades $\ell < 0$ e $\ell \geq 0$. Se $\ell < 0$, então existe $(\bar{w}, \bar{r}) \in C$ tal que $\lambda\bar{r} + v^T\bar{w} < 0$, pois ℓ é o ínfimo de C. Como C é cone $\alpha(\bar{r}, \bar{w})$ pertence a C para todo $\alpha > 0$, logo

$$\ell \leq \lambda(\alpha\bar{r}) + v^T(\alpha\bar{w}) = \alpha(\lambda\bar{r} + v^T\bar{w}).$$

Fazendo α tender ao infinito na última desigualdade temos um absurdo, pois $\lambda\bar{r} + v^T\bar{w} < 0$. Portanto $\ell \geq 0$, mas como $(0,0) \in C$ segue-se que $\ell = 0$. Como consequência da equação (2-3) temos que $\lambda < 0$ e $\lambda r + v^T w \geq 0$, para todo $(w^T, r) \in C$, ou equivalentemente $-r + \left(-v^T/\lambda\right) w \geq 0$. Agora, denotando $-v/\lambda = y$ a última equação é equivalente à

$$-r + y^T w \geq 0, \quad \forall \ (r,w) \in C. \tag{2-4}$$

Definindo $s = -A^T y + c$ resta mostrar que (y,s) é solução ótima de (D). Primeiro mostremos que (y,s) é viável para (D). Pela definição de s resta mostrar que $s \geq 0$. Segue da definição de C que $(w^T, r)^T = B(x^T, t)^T$ ou equivalentemente $w = -Ax + tb$ e $r = -c^T x + tp^*$, para $x \geq 0$ e $t \geq 0$. Substituindo estas igualdades em (2-4) temos

$$0 \leq c^T x - tp^* + y^T(-Ax + tb) = (c^T - y^T A)x - tp^* + ty^T b = s^T x - tp^* + ty^T b,$$

para todo $x \geq 0$ e $t \geq 0$. Já que a equação acima vale para todo $x \geq 0$ e $t \geq 0$, fazendo $t = 0$, temos que $s^T x \geq 0$, para todo $x \geq 0$, o que implica que $s \geq 0$. Logo (y,s) é viável. Ainda, fazendo $t = 1$ e $x = 0$ na equação acima temos que $p^* \leq y^T b$. Mas pelo Teorema de Dualidade Fraco $p^* \geq b^T y$ deste modo $p^* = b^T y$. Portanto (y,s) é solução ótima para (D). A demonstração de que (D)

tem solução ótima implica que (P) tem solução ótima é análoga.
Para os itens b e c use o Corolário 2.2.2. □

Note que a volta do item b do teorema acima nem sempre é válida. O que ocorre de fato, é que se um dos problemas (P) ou (D) é inviável o seu dual pode ser ilimitado ou inviável. Para verificarmos isto analisemos o seguinte exemplo.

Exemplo 2.2.4 *Considere os seguintes pares de problemas primal e dual no formato padrão.*

$$
\begin{array}{lll}
\textit{(P)} & \textit{minimizar} & -x_1 - x_2 \\
& \textit{sujeito a}: & x_1 - x_2 - x_3 = 2 \\
& & -x_1 + x_2 - x_4 = 2 \\
& & x_1, x_2, x_3, x_4 \geq 0
\end{array}
$$

$$
\begin{array}{lll}
\textit{(D)} & \textit{maximizar} & 2y_1 + 2y_2 \\
& \textit{sujeito a}: & y_1 - y_2 + s_1 = -1 \\
& & -y_1 + y_2 + s_2 = -1 \\
& & -y_1 + s_3 = 0 \\
& & -y_2 + s_4 = 0 \\
& & s_1, s_2, s_3, s_4 \geq 0
\end{array}
$$

Iremos verificar que $\mathcal{F}(P) = \mathcal{F}(D) = \emptyset$. Realizando operações elementares na matriz aumentada $[A \mid b]$ do problema (P) acima, obtemos o seguinte sistema equivalente

$$
\begin{aligned}
x_1 - x_2 - x_3 &= 2 \\
-x_3 - x_4 &= 4.
\end{aligned}
$$

Da segunda equação do sistema acima obtemos que $x_3 = -4 - x_4$. Como as variáveis do problema (P) são todas não negativas, a última equação implica $\mathcal{F}(P) = \emptyset$. Agora, analisemos o problema (D). Das duas últimas equações do sistema, temos que $y_1 = s_3$ e $y_2 = s_4$. Substituindo estas igualdades nas primeiras equações do sistema temos

$$
\begin{aligned}
s_3 - s_4 + s_1 &= -1 \\
-s_3 + s_4 + s_2 &= -1.
\end{aligned}
$$

Somando as equações acima obtemos que $s_1 = -2 - s_2$. Como as variáveis de folga dual são todas não negativas a última equação implica $\mathcal{F}(D) = \emptyset$.

Portanto a volta do item b *do Teorema 2.2.3 nem sempre é verdadeira pois, exibimos um problema primal inviável cujo seu dual é também inviável.*

Para o próximo resultado, veja página 82 em [21].

Proposição 2.2.5 *Se o problema (P) é viável e limitado então (P) tem solução ótima, isto é* $P^* \neq \emptyset$.

Demonstração: Temos duas possibilidades para o conjunto viável, $\mathcal{F}(P)$ limitado ou ilimitado. Primeiro suponhamos que $\mathcal{F}(P)$ seja limitado. Como $\mathcal{F}(P)$ é fechado temos que ele é compacto. Então, pelo teorema de Weierstrass e pela continuidade da função objetivo, segue-se que (P) tem solução ótima. Agora suponhamos $\mathcal{F}(P)$ ilimitado, pelo Teorema 1.3.6 podemos escrever

$$\mathcal{F}(P) = \{v + d : v \in V, \;\; d \in K\}, \tag{2-5}$$

onde V é o conjunto de todas as combinações convexas dos vértices de $\mathcal{F}(P)$ e K é o conjunto de todas as combinações convexas das direções de recessão de $\mathcal{F}(P)$. Como $\mathcal{F}(P)$ é convexo segue-se que $V \subset \mathcal{F}(P)$. Assim, pela definição de K temos que $v+\alpha d \in \mathcal{F}(P)$, para todo $\alpha \geq 0$. Como por hipótese, o problema (P) é limitado existe ℓ^* tal que $\ell^* \leq c^T(v + \alpha d) = c^T v + \alpha c^T d$, ou equivalentemente, $\ell^*/\alpha \leq c^T v/\alpha + c^T d$. Já que a última equação vale para todo $\alpha > 0$, fazendo α tender ao infinito obtemos que $c^T d \geq 0$. Deste modo,

$$c^T(v + d) \geq c^T v, \qquad \forall\, v \in V, \quad \forall\, d \in K.$$

Portanto da última equação e de (2-5), segue-se que o problema (P) é equivalente à

$$\begin{array}{ll} minimizar & c^T v \\ sujeito\ a: & v \in V \end{array}$$

Pela Proposição 1.3.5 o número de vértices em $\mathcal{F}(P)$ é finito. Assim, dado $v \in V$ temos que

$$v = \sum_{i=1}^{\ell} \lambda_i v^i,$$

onde $\sum_{i=1}^{\ell} \lambda_i = 1$, $\lambda_i \geq 0$ e $v^1, ..., v^\ell$ são vértice de $\mathcal{F}(P)$. Portanto,

$$c^T v = \sum_{i=1}^{\ell} \lambda_i c^T v^i \geq \sum_{i=1}^{\ell} \lambda_i (\min\{c^T v^i : i = 1, ..., \ell\}) = \min\{c^T v^i : i = 1, ..., \ell\}.$$

Deste modo concluímos que o problema acima admite ponto de mínimo em um vértice de $\mathcal{F}(P)$. Portanto em ambos os casos, quando $\mathcal{F}(P)$ é limitado ou ilimitado, obtemos que o problema (P) tem solução ótima. $\square$

Proposição 2.2.6 *Se o problema (D) é viável e limitado então (D) tem solução ótima, isto é $D^* \neq \emptyset$.*

Demonstração: Com demonstração análoga ao Teorema 1.3.6 podemos escrever

$$\mathcal{F}(D) = \{v + d : v \in V, \ \ d \in K\},$$

onde V é o conjunto de todas as combinações convexas dos vértices de $\mathcal{F}(D)$ e K é o conjunto de todas as combinações convexas das direções de recessão de $\mathcal{F}(D)$. Então com argumento análogo utilizado na Proposição 2.2.5 podemos verificar o resultado. $\square$

O Teorema de Dualidade Forte apresenta uma relação entre as soluções ótimas dos problemas (P) e (D), veja Teorema 3.2.3 na página 71 em [24].

Teorema 2.2.7 *(Teorema de Dualidade Forte) Se $\mathcal{F}(P) \neq \emptyset$ e $\mathcal{F}(D) \neq \emptyset$ então:*

- ***a)*** *P^* é não vazio;*
- ***b)*** *D^* é não vazio;*
- ***c)*** *Para todo $x^* \in P^*$ e $(y^*, s^*) \in D^*$ tem-se $c^T x^* = b^T y^*$.*

Demonstração: Verifiquemos o item a. Seja $(y, s) \in \mathcal{F}(D)$ então para todo $x \in \mathcal{F}(P)$ temos, pelo Teorema de Dualidade Fraco , que $c^T x \geq b^T y$ que implica que o problema (P) é viável e limitado. Então pela Proposição 2.2.5 o conjunto P^* é não vazio. Para verificar os itens b e c basta utilizar a hipótese do item a, isto é, P^* é não vazio. Então pelo Teorema de Dualidade 2.2.3 item a o problema (D) tem solução ótima e os valores das funções objetivo de ambos são iguais, ou seja, D^* é não vazio e para todo $x^* \in P^*$ e $(y^*, s^*) \in D^*$ tem-se $c^T x^* = b^T y^*$. $\square$

Como consequência dos teoremas de dualidade, obtemos as condições de otimalidade para o problema de Programação Linear.

Teorema 2.2.8 *(Condições de Otimalidade) Considere os problemas (P) e (D) e o sistema de equações*

$$Ax = b, \quad x \geq 0 \tag{2-6}$$

$$A^T y + s = c, \quad s \geq 0 \tag{2-7}$$

$$x^T s = 0. \tag{2-8}$$

Então as seguintes afirmações valem:

i) *Um ponto $x^* \in \mathbb{R}^n$ é uma solução ótima de (P) se, e somente se, existe um par $(y^*, s^*) \in \mathbb{R}^m \times \mathbb{R}^n$ solução ótima de (D), tal que (x^*, y^*, s^*) seja solução do sistema* (2-6)-(2-8).

ii) *Um ponto $(y^*, s^*) \in \mathbb{R}^m \times \mathbb{R}^n$ solução ótima de (D) se, e somente se, existe $x^* \in \mathbb{R}^n$ uma solução ótima de (P), tal que (x^*, y^*, s^*) seja solução do sistema* (2-6)-(2-8).

Demonstração: Prova do item i. Seja x^* solução ótima de (P), então x^* satisfaz (2-6). Além disso, pelo item a do Teorema 2.2.3 existe (y^*, s^*) solução de (D), e em particular, este ponto satisfaz a equação (2-7). Desde que x^* é viável para (P) e (y^*, s^*) é viável para (D) temos pelo Teorema 2.2.1 que

$$c^T x^* - b^T y^* = x^{*T} s^*.$$

Ainda pelo item a do Teorema 2.2.3 temos que $c^T x^* = b^T y^*$, que juntamente com a última igualdade implica que x^* e s^* satisfaz (2-8). Portanto (x^*, y^*, s^*) é solução de (2-6)-(2-8).

Agora, suponhamos que (x^*, y^*, s^*) satisfaça o sistema (2-6) - (2-8). Pela equação (2-6) temos que x^* é um ponto viável para (P) e pela equação (2-7) temos que (y^*, s^*) é um par viável para (D). Pelo Teorema 2.2.1 temos que $c^T x^* - b^T y^* = x^{*T} s^*$, o que juntamente com a equação (2-8) implica $c^T x^* = b^T y^*$. Esta igualdade juntamente com o Teorema 2.2.1 implica que para todo x viável para (P) vale

$$c^T x^* = b^T y^* \leq c^T x.$$

Portanto x^* é solução de (P).

A Prova do item ii é feita de modo análoga a prova do item i. □

2.3 Resultados com viabilidade estrita

Na prova do Teorema da Dualidade Forte supomos apenas que os conjuntos viáveis primal e dual são não vazios, isto é, $\mathcal{F}(P) \neq \emptyset$ e $\mathcal{F}(D) \neq \emptyset$. Nesta seção, veremos que exigindo um pouco mais, isto é, $\mathcal{F}^0(D) \neq \emptyset$, a prova deste resultado pode ser bastante simplificada.

Proposição 2.3.1 *Seja $\bar{x} \in \mathcal{F}(P)$. Suponha que $\mathcal{F}^0(D) \neq \emptyset$ então o conjunto*

$$\mathcal{L} = \{x \in \mathcal{F}(P) : \quad c^T x \leq c^T \bar{x}\},$$

é compacto.

Demonstração: Primeiro mostremos que $\mathcal{L}$ é fechado. Seja a sequência $\{x^k\} \subset \mathcal{L}$ tal que $\lim_{k\to\infty} x^k = \tilde{x} \geq 0$. Desde que $Ax^k = b$ e $x^k \geq 0$ segue imediatamente que $\tilde{x} \in \mathcal{F}(P)$. Como $c^T x^k \leq c^T \bar{x}$, para todo k, segue-se que

$$\lim_{k\to\infty} c^T x^k = c^T \lim_{k\to\infty} x^k = c^T \tilde{x} \leq c^T \bar{x}.$$

Isto implica que $\tilde{x} \in \mathcal{L}$, assim $\mathcal{L}$ é fechado.

Agora verifiquemos que $\mathcal{L}$ é limitado. Por hipotése $\mathcal{F}^0(D) \neq \emptyset$. Sejam $(\bar{y}, \bar{s}) \in \mathcal{F}^0(D)$ e $x \in \mathcal{L}$. Como $x \in \mathcal{L} \subset \mathcal{F}(P)$, temos pelo Teorema de Dualidade Fraco que

$$\sum_{i=1}^{n} x_i \bar{s}_i = x^T \bar{s} = c^T x - b^T \bar{y} \leq c^T \bar{x} - b^T \bar{y} = \bar{x}^T \bar{s}.$$

Como $x \geq 0$ e $\bar{s} > 0$ concluimos da última equação que

$$x_i \leq \frac{1}{\bar{s}_i} \bar{x}^T \bar{s} \leq \left(\max_{i=1,\cdots,n} \left\{ \frac{1}{\bar{s}_i} \right\} \right) \bar{x}^T \bar{s}, \qquad i = 1, \cdots, n.$$

Como tomamos $x \in \mathcal{L}$ arbitrário na última equação concluímos imediatamente que $\mathcal{L}$ é limitado. Portanto $\mathcal{L}$ é compacto. □

Proposição 2.3.2 *Se $\mathcal{F}(P) \neq \emptyset$ e $\mathcal{F}^0(D) \neq \emptyset$ então o problema primal (P) tem conjunto solução P^* não vazio e compacto.*

Demonstração: Sejam $\tilde{x} \in \mathcal{F}(P)$ e $(\tilde{y}, \tilde{s}) \in \mathcal{F}^0(D)$. Defina o seguinte problema:

$$\begin{array}{lll} \text{(GP)} & \textit{minimizar} & c^T x \\ & \textit{sujeito a}: & c^T x \leq c^T \bar{x} \\ & & Ax = b \\ & & x \geq 0. \end{array}$$

Note que o conjunto viável de (GP) é dado por $\mathcal{L} = \{x \in \mathcal{F}(P) : \ c^T x \leq c^T \bar{x}\}$, e pela Proposição 2.3.1 é compacto. Como a função objetivo de (GP) é contínua, então pelo Teorema de Weiertrass o problema (GP) tem solução. Logo P^* é não vazio.

Mostremos então que P^* é compacto. Note que $P^* \subset \mathcal{L}$, então P^* é limitado. Agora, resta mostrar que P^* é fechado. Para isso, tomemos uma sequência $\{x^k\} \in P^*$ tal que $\lim_{k\to\infty}(x^k)^* = \tilde{x}$. Então

$$p^* = \lim_{k\to\infty} c^T (x^k)^* = c^T \lim_{k\to\infty} (x^k)^* = c^T \tilde{x},$$

onde p^* é o valor ótimo de (P). Assim $\tilde{x} \in P^*$ e P^* é fechado, portanto P^* é compacto. □

Teorema 2.3.3 *(Teorema de Dualidade Forte) Se $\mathcal{F}(P) \neq \emptyset$ e $\mathcal{F}^0(D) \neq \emptyset$ então:*

a) *P^* é não vazio e compacto.*

b) *D^* é não vazio.*

c) *Para todo $x^* \in P^*$ e $(y^*, s^*) \in D^*$ tem-se $c^T x^* = b^T y^*$.*

Demonstração: Para o item a basta usar a Proposição 2.3.2. Para os itens b e c basta utilizar o item a, e o resultado segue diretamente pelo Teorema 2.2.3 item a. □

2.4 Teorema da Complementaridade Estrita

Considere $x^* \in P^*$ e $(y^*, s^*) \in D^*$ soluções ótimas para os problemas primal e dual respectivamente. A desigualdade $x^* + s^* > 0$ é chamada condição de folga complementar estrita. O Teorema de Complementaridade Estrita afirma que qualquer PPL com solução ótima, possui uma solução ótima que satisfaz a condição de folga complementar estrita. Para sua demonstração, Teorema 3.3.5 na página 77 em [24], precisaremos definir os seguintes conjuntos.

Para cada $x^* \in P^*$ defina o conjunto $B(x^*) \subset \{1, \cdots, n\}$ como sendo

$$B(x^*) = \{i : x_i^* > 0\}.$$

Assim, defina o conjunto B como sendo a união dos conjuntos $B(x^*)$, isto é,

$$B = \bigcup_{x^* \in P^*} B(x^*). \tag{2-9}$$

De maneira análoga vamos definir para cada $(y^*, s^*) \in D^*$ o seguinte conjunto $N(s^*) \subset \{1, \cdots, n\}$ como sendo

$$N(s^*) = \{i : s_i^* > 0\}.$$

Desta maneira, defina o conjunto N como sendo a união dos conjuntos $N(s^*)$, ou seja,

$$N = \bigcup_{(y^*, s^*) \in D^*} N(s^*). \tag{2-10}$$

Teorema 2.4.1 *Suponha que os problemas (P) e (D) admitam soluções ótimas. Então*

$$B \cup N = \{1, 2, \ldots, n\}.$$

Como consequência, existem soluções $x^* \in P^*$ e $(y^*, s^*) \in D^*$ satisfazendo

$$x^* + s^* > 0.$$

Demonstração: Sejam B e N como definidos em (2-9) e (2-10) e defina o conjunto $\tilde{N}$, como sendo o complementar do conjunto B em relação ao conjunto $\{1, 2, \ldots, n\}$, isto é,

$$\tilde{N} = \{1, 2, \ldots, n\}/B.$$

Nosso objetivo é mostrar que $\tilde{N} = N$. Primeiro note que, se $s_N^* > 0$ para alguma solução ótima dual s^*, então pelo Teorema 2.2.8, temos que $x_N^* = 0$ para toda solução ótima x^* do problema primal ou seja, $N \subset \tilde{N}$.

Agora vamos mostrar a recíproca. Seja $x^* = ((x_B^*)^T, 0^T)^T \in P^*$, é suficiente mostrar que existe uma solução $(y^*, s^*) \in D^*$ tal que $s_{\tilde{N}}^* > 0$. Para isso, considere p^* o valor ótimo primal, d^* o valor ótimo dual e, pelo Teorema 2.2.7, $p^* = d^* =: \kappa$. Seja o seguinte problema

$$\begin{aligned} \text{(1)}\ \ & minimizar - \tilde{e}^T \tilde{x} \\ & sujeito\ a : \tilde{A}\tilde{x} = \tilde{b} \\ & \qquad\qquad \tilde{x} \geq 0, \end{aligned}$$

onde a variável $\tilde{x} \in \mathbb{R}^{n+1}$ e os dados do problema, $\tilde{A} \in M_{m+1 \times n+1}(\mathbb{R})$, $\tilde{b} \in \mathbb{R}^{m+1}$

e $\tilde{e} \in \mathbb{R}^{n+1}$, são respectivamente iguais a

$$\tilde{x} = \begin{pmatrix} x \\ t \end{pmatrix}, \quad \tilde{A} = \begin{bmatrix} A & 0 \\ -c^T & -1 \end{bmatrix}, \quad \tilde{b} = \begin{pmatrix} b \\ -\kappa \end{pmatrix}, \quad \tilde{e}_j = \begin{cases} 0, & \textbf{para} \quad j \in B \\ 1, & \textbf{para} \quad j \in \tilde{N} \\ 0, & \textbf{para} \quad j = n+1. \end{cases}$$

É fácil ver que o conjunto viável do problema (1) é igual ao conjunto de soluções ótimas de (P) denotado por P^*. Assim, segue da definição do conjunto B e do vetor $\tilde{e}$, que o valor ótimo do problema (1) é zero. O problema dual de (1) é o seguinte problema de otimização,

$$\begin{aligned} (2) \quad & maximizar\ \tilde{b}^T \tilde{y} \\ & sujeito\ a: \tilde{A}^T \tilde{y} + \tilde{s} = \tilde{e} \\ & \qquad \tilde{s} \geq 0, \end{aligned}$$

onde as variáveis $\tilde{y} \in \mathbb{R}^{m+1}$ e $\tilde{s} \in \mathbb{R}^{n+1}$ são dadas por

$$\tilde{y} = \begin{pmatrix} \bar{y} \\ \bar{y}_{m+1} \end{pmatrix}, \qquad \tilde{s} = \begin{pmatrix} \bar{s} \\ \bar{s}_{n+1} \end{pmatrix}.$$

Pelo Teorema 2.2.7, o problema (2) também admite solução ótima $(\tilde{y}^*, \tilde{s}^*)$ com valor ótimo

$$\tilde{b}^T \tilde{y}^* = b^T \bar{y}^* - \kappa \bar{y}^*_{m+1} = 0. \tag{2-11}$$

Agora, considere uma solução $(\bar{\bar{y}}^*, \bar{\bar{s}}^*) \in D^*$. Desde que $(\tilde{y}^*, \tilde{s}^*)$ é solução ótima para o problema (2), obtemos que

$$\begin{aligned} A^T \bar{\bar{y}}^* + \bar{\bar{s}}^* = c, \quad & \bar{\bar{s}}^* \geq 0, \\ A^T \bar{y}^* - \bar{y}^*_{m+1} c + \bar{s}^* = \bar{e}, \quad & \bar{s}^* \geq 0, \\ \bar{y}^*_{m+1} = \bar{s}^*_{n+1} \geq 0, & \end{aligned}$$

onde $\tilde{e} = ((\bar{e})^T, 0)^T$. Usando as equações acima obtemos, após algumas manipulações algébricas, que

$$A^T \left(\frac{\bar{\bar{y}}^* + \bar{y}^*}{1 + \bar{s}^*_{n+1}} \right) + \frac{\bar{\bar{s}}^* + \bar{s}^* + \bar{e}}{1 + \bar{s}^*_{n+1}} = c. \tag{2-12}$$

Agora, defina os seguintes vetores

$$y^* := \frac{\bar{\bar{y}}^* + \bar{y}^*}{1 + \bar{s}^*_{n+1}} \qquad s^* := \frac{\bar{\bar{s}}^* + \bar{s}^* + \bar{e}}{1 + \bar{s}^*_{n+1}} \geq 0, \tag{2-13}$$

onde a última desigualdade é válida pois, $\bar{\bar{s}}^* \geq 0$, $\bar{s}^* \geq 0$, $\bar{e} \geq 0$ e $1 + \bar{s}^*_{n+1} > 0$. Segue diretamente de (2-12) e (2-13) que o par $(y^*, s^*) \in \mathcal{F}(D)$, resta verificar que este par é solução ótima. Para isso, vamos combinar as equações (2-11) e (2-13) obtendo

$$b^T y^* = b^T \left(\frac{\bar{\bar{y}}^* + \bar{y}^*}{1 + \bar{s}^*_{n+1}} \right) = \left(b^T \bar{\bar{y}}^* + b^T \bar{y}^* \right) \frac{1}{1 + \bar{s}^*_{n+1}} = \left(b^T \bar{\bar{y}}^* + \kappa \bar{y}^*_{m+1} \right) \frac{1}{1 + \bar{s}^*_{n+1}}.$$

Agora, como $(\bar{\bar{y}}^*, \bar{\bar{s}}^*) \in D^*$ temos que $b^T \bar{\bar{y}}^* = \kappa$. Substituindo esta igualdade na equação acima e observando que $\bar{y}^*_{m+1} = \bar{s}^*_{n+1}$ temos

$$b^T y^* = \left(\kappa + \kappa \bar{y}^*_{m+1} \right) \frac{1}{1 + \bar{s}^*_{n+1}} = \kappa \frac{1 + \bar{y}^*_{m+1}}{1 + \bar{s}^*_{n+1}} = \kappa.$$

O que prova que $(y^*, s^*) \in D^*$. Finalmente, note que pela definição de $\bar{e}$ e sendo $\bar{\bar{s}}^* \geq 0$ e $\bar{s}^* \geq 0$, obtemos

$$s^*_{\tilde{N}} = \frac{\bar{\bar{s}}^*_{\tilde{N}} + \bar{s}^*_{\tilde{N}} + \bar{e}_{\tilde{N}}}{1 + \bar{s}^*_{n+1}} > 0.$$

E isto implica que $\tilde{N} \subset N$. Portanto $N = \tilde{N}$. Pela definição de $\tilde{N}$, segue-se que $B \cup N = \{1, 2, \ldots, n\}$. Logo, pela definição de B e N, existem soluções $x^* \in P^*$ e $(y^*, s^*) \in D^*$ com $x^* + s^* > 0$. Com isso, finalizamos a prova. $\square$

CAPÍTULO 3

A Trajetória Central em Programação Linear

A Trajetória Central tem importância fundamental no estudo de vários algoritmos para resolução de Problemas de Programação Linear, visto que os algoritmos que, em algum sentido, seguem a Trajetória Central tem comportamento polinomial. Neste capítulo iremos mostrar que, o interior dos conjuntos viáveis primal e dual não vazios asseguram a existência de uma curva analítica de $\mathbb{R}_{++}$ em $\mathbb{R}^n_{++} \times \mathbb{R}^m \times \mathbb{R}^n_{++}$ que converge para o centro analítico da face ótima. Esta curva é denominada Trajetória Central. Neste capítulo vamos assumir

$$\mathcal{F}^0(P) \neq \emptyset, \qquad \mathcal{F}^0(D) \neq \emptyset,$$

sem mencionar este fato explicitamente nos enunciados dos resultados.

3.1 As equações que definem a Trajetória Central

Nesta seção, estamos interessados em obter o sistema de equações que define a Trajetória Central.

Proposição 3.1.1 *Sejam $\hat{x} > 0$ e $\hat{s} > 0$. Então o conjunto*

$$\Omega = \left\{ x \in \mathbb{R}^n : \;\; Ax = b, \;\; x > 0, \;\; c^T x - \mu \sum_{i=1}^{n} \log x_i \leq c^T \hat{x} - \mu \sum_{i=1}^{n} \log \hat{x}_i \right\},$$

é compacto.

Demonstração: Primeiro mostremos que Ω é limitado. Para isso afirmamos que o conjunto Ω pode ser reescrito da seguinte forma

$$\Omega = \left\{ x \in \mathbb{R}^n : \quad Ax = b, \quad x > 0, \quad x^T\hat{s} - \mu \sum_{i=1}^{n} \log x_i \leq \hat{x}^T\hat{s} - \mu \sum_{i=1}^{n} \log \hat{x}_i \right\}.$$

De fato, seja $\hat{y} \in \mathbb{R}^m$ tal que $A^T\hat{y} + \hat{s} = c$. Assim pelo Teorema 2.2.1 temos

$$c^Tx - c^T\hat{x} = x^T\hat{s} + b^T\hat{y} - (\hat{x}^T\hat{s} + b^T\hat{y}) = x^T\hat{s} - \hat{x}^T\hat{s},$$

e esta igualdade juntamente com a definiçao de Ω prova a afirmação. Agora suponhamos por absurdo, que Ω seja ilimitado. Então existe uma sequência de pontos $\{x^k\} \subset \Omega$ tal que $\lim_{k\to\infty} ||x^k|| = +\infty$. Note que, tomando $0 < \hat{s}^*_\ell = \min\{\hat{s}_i : i = 1, ..., n\}$ temos que

$$(x^k)^T\hat{s} = \sum_{i=1}^{n} x_i^k \hat{s}_i \geq \sum_{i=1}^{n} x_i^k \hat{s}^*_\ell.$$

Como $\{x^k\} \subset \Omega$, segue da desigualdade acima que

$$\hat{x}^T\hat{s} - \mu \sum_{i=1}^{n} \log \hat{x}_i \geq (x^k)^T\hat{s} - \mu \sum_{i=1}^{n} \log x_i^k \geq \sum_{i=1}^{n} x_i^k \hat{s}^*_\ell - \mu \sum_{i=1}^{n} \log x_i^k. \tag{3-1}$$

Com algumas manipulações algébricas é fácil ver que

$$\sum_{i=1}^{n} x_i^k \hat{s}^*_\ell - \mu \sum_{i=1}^{n} \log x_i^k = \sum_{i=1}^{n} x_i^k \left(\hat{s}^*_\ell - \frac{\mu \log x_i^k}{x_i^k} \right). \tag{3-2}$$

Nestes termos dividiremos os índices em dois conjuntos: $I = \{i : \lim x_i^k = +\infty\}$ e $\bar{I} = \{i : \{x_i^k\}$ é limitada $\}$. Como $\lim ||x^k|| = +\infty$ e $x^k > 0$ então, $\lim x_i^k = +\infty$ para pelo menos um índice i, logo $I \neq \emptyset$. Analisemos os dois casos. Para $i \in I$ temos que

$$\lim_{k\to\infty} x_i^k \left(\hat{s}^*_\ell - \mu \frac{\log x_i^k}{x_i^k} \right) = +\infty, \tag{3-3}$$

pois $\lim_{t\to\infty} \log t/t = 0$, $\hat{s}^*_\ell > 0$ e $\lim x_i^k = +\infty$. Agora para $i \in \bar{I}$ temos, na expressão acima, que $\hat{s}^*_\ell - (\mu \log x_j^k / x_j^k)$ é limitado inferiormente. Assim, concluímos a partir de (3-2) que

$$\lim_{k\to\infty} \left(\sum_{i=1}^{n} x_i^k \hat{s}^*_\ell - \mu \sum_{i=1}^{n} \log x_i^k \right) = +\infty,$$

o que contradiz (3-1). Logo Ω é limitado.
Mostremos que Ω é fechado. Seja uma sequência $\{x^k\} \subset \Omega$ tal que $\lim x^k = \bar{x}$. Note que $\bar{x} > 0$ pois, caso contrário,

$$\lim_{k\to\infty} \left[(x^k)^T \hat{s} - \mu \sum_{i=1}^{n} \log x_i^k \right] = +\infty,$$

que contradiz o fato de Ω ser limitado. Agora como $\{x^k\} \subset \Omega$, por continuidade temos que

$$\bar{x}^T \hat{s} - \mu \sum_{i=1}^{n} \log \bar{x}_i = \lim \left((x^k)^T \hat{s} - \mu \sum_{i=1}^{n} \log x_i^k \right) \leq \hat{x}^T \hat{s} - \mu \sum_{i=1}^{n} \log \hat{x}_i.$$

Logo $\bar{x} \in \Omega$ e temos que Ω é fechado. Portanto o conjunto Ω é compacto. $\square$

Os próximos resultados podem ser encontrados em [22].

Proposição 3.1.2 *Para cada $\mu > 0$ o seguinte problema*

$$\begin{array}{lll} (P_\mu) & \textit{minimizar} & c^T x - \mu \sum_{i=1}^{n} \log x_i \\ & \textit{sujeito a}: & Ax = b \\ & & x > 0, \end{array}$$

tem solução única $x(\mu) > 0$.

Demonstração: Sejam $\hat{x} \in \mathcal{F}^0(P)$ e $(\hat{y}, \hat{s}) \in \mathcal{F}^0(D)$. Considere o seguinte problema

$$\begin{array}{lll} (P_{\mu s}) & \textit{minimizar} & c^T x - \mu \sum_{i=1}^{n} \log x_i \\ & \textit{sujeito a}: & Ax = b \\ & & c^T x - \mu \sum_{i=1}^{n} \log x_i \leq c^T \hat{x} - \mu \sum_{i=1}^{n} \log \hat{x}_i \\ & & x > 0, \end{array}$$

O problema $(P_{\mu s})$ tem conjunto viável Ω definido como na Proposição 3.1.1. Como Ω é um conjunto de nível da função objetivo do problema (P_μ) podemos concluir que $x(\mu)$ é solução do problema (P_μ) se, e somente se, $x(\mu)$ é solução do problema $(P_{\mu s})$. Pela Proposição 3.1.1 Ω é compacto. Logo pelo Teorema de Weiertrass o problema (P_μ) tem solução. Note que pelo Exemplo 1.3.17, a função objetivo de $(P_{\mu s})$ é estritamente convexa, assim Ω é convexo o que juntamente com o Lema 1.3.14 implica que o problema $(P_{\mu s})$ tem uma única solução. Portanto (P_μ) tem única solução. $\square$

Também podemos mostrar que, para cada $\mu > 0$, o problema

$$\begin{array}{lll} (D_\mu) & maximizar & b^T y + \mu \sum_{i=1}^n \log s_i \\ & sujeito\ \ a: & A^T y + s = c \\ & & s > 0, \end{array}$$

tem uma única solução, digamos $(y(\mu), s(\mu))$, com $s(\mu) > 0$.

Proposição 3.1.3 *Para cada $\mu > 0$ o seguinte sistema*

$$Ax = b, \qquad x > 0 \tag{3-4}$$

$$A^T y + s = c, \qquad s > 0 \tag{3-5}$$

$$Xs = \mu e, \quad \mu > 0, \tag{3-6}$$

tem solução única, digamos $(x(\mu), y(\mu), s(\mu))$.

Demonstração: A função de Lagrange associada ao problema (P_μ) é dada por

$$L_P(x, u) = c^T x - \langle u, Ax - b\rangle - \mu \sum_{i=1}^n \log x_i,$$

onde $u \in \mathbb{R}^m$ é o vetor de multiplicadores de Lagrange. Por definição de $x(\mu)$ existe $u(\mu)$ tal que o par $(x(\mu), u(\mu))$ é o único mínino de L_P, que ocorre quando seu gradiente é nulo, isto é, quando

$$\nabla_x L_P(x(\mu), u(\mu)) := c - A^T u(\mu) - \mu x^{-1}(\mu) = 0, \tag{3-7}$$

$$\nabla_u L_P(x(\mu), u(\mu)) := Ax(\mu) - b = 0. \tag{3-8}$$

Como Posto$(A) = m$ existe um único $u(\mu) \in \mathbb{R}^m$ satisfazendo a equação (3-7). Agora defina o seguinte vetor $s(\mu) := \mu x^{-1}(\mu) > 0$ e note que esta igualdade é equivalente à

$$X(\mu)s(\mu) = \mu e. \tag{3-9}$$

Portanto $s(\mu) > 0$ e $x(\mu) > 0$ satisfazem a equação (3-6). Agora, fazendo $y(\mu) = u(\mu)$ e como $s(\mu) = \mu x^{-1}(\mu) > 0$ temos que a equação (3-7) pode ser reescrita da seguinte forma

$$A^T y(\mu) + s(\mu) = c. \tag{3-10}$$

Portanto segue das equações (3-8), (3-9) e (3-10) que a terna $(x(\mu), y(\mu), s(\mu))$ é a única solução do sistema (3-4)-(3-6). □

Também, de modo análogo, podemos mostrar que as condições de otimalidade do problema (D_μ) são as equações (3-4) - (3-6), calculando o gradiente da função de Lagrange associada ao problema (D_μ) definida por

$$L_D(z,y,s) = b^T y - \left\langle z, A^T y + s - c \right\rangle - \mu \sum_{i=1}^{n} \log s_i.$$

De agora em diante $\{(x(\mu), y(\mu), s(\mu)) \,:\, \mu > 0\}$ denota a solução do sistema de equações (3-4) - (3-6) e será chamado de *Trajetória Central*.

3.2 A Trajetória Central é uma curva analítica

Nesta seção estamos interessados em mostrar que as equações (3-4), (3-5) e (3-6) definem uma curva analítica no conjunto $\mathbb{R}^n_{++} \times \mathbb{R}^m \times \mathbb{R}^n_{++}$. O próximo resultado pode ser encontrado em [1].

Teorema 3.2.1 *As equações* (3-4)*,* (3-5) *e* (3-6) *definem uma curva analítica no conjunto* $\mathbb{R}^n_{++} \times \mathbb{R}^m \times \mathbb{R}^n_{++}$.

Demonstração: Primeiro, defina a seguinte função

$$\begin{aligned} F : \mathbb{R} \times \mathbb{R}^n_{++} \times \mathbb{R}^m \times \mathbb{R}^n_{++} &\to \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^n \\ (\mu, x, y, s) &\mapsto (Ax - b, A^T y + s - c, Xs - \mu e). \end{aligned}$$

Note que a matriz jacobiana de F com relação a (x,y,s) é dada por

$$\nabla_{(x,y,s)} F(\mu,x,y,s) := \begin{bmatrix} A & 0_{m\times m} & 0_{m\times n} \\ 0_{n\times n} & A^T & I_n \\ S & 0_{n\times m} & X \end{bmatrix}.$$

Note que o sistema (3-4) - (3-6) é equivalente ao seguinte sistema

$$F(\mu,x,y,s) = (0_m, 0_n, 0_n) \qquad \mu > 0, \quad x > 0, \quad s > 0. \tag{3-11}$$

Dado $\bar\mu$ temos de (3-4)-(3-6) que o ponto $(\bar\mu, x(\bar\mu), y(\bar\mu), s(\bar\mu))$ é solução do sistema acima. Então, como $x(\bar\mu) > 0$ e $s(\bar\mu) > 0$ temos pelo Lema 1.4.1 que a matriz jacobiana de F no ponto $(\bar\mu, x(\bar\mu), y(\bar\mu),\ s(\bar\mu))$ é não-singular. Então segue-se do Teorema da Função Implícita 1.2.3 que existe um aberto

$$(\bar\mu - \epsilon, \bar\mu + \epsilon) \times U \subset \mathbb{R} \times \mathbb{R}^n_{++} \times \mathbb{R}^m \times \mathbb{R}^n_{++},$$

e uma única curva analítica $\varphi : (\bar{\mu} - \epsilon, \bar{\mu} + \epsilon) \to U$, satisfazendo a seguinte igualdade $\varphi(\bar{\mu}) = (x(\bar{\mu}), y(\bar{\mu}), s(\bar{\mu}))$ e além disso

$$F(\mu, \varphi(\mu)) = (0_m, 0_n, 0_n), \qquad \mu \in (\bar{\mu} - \epsilon,\ \bar{\mu} + \epsilon). \tag{3-12}$$

Pela Proposição 3.1.3, para cada $\mu > 0$, o sistema (3-4) - (3-6) tem uma única solução $(x(\mu), y(\mu), s(\mu)) \in \mathbb{R}^n_{++} \times \mathbb{R}^m \times \mathbb{R}^n_{++}$. Assim, como $\varphi(\mu) \in \mathbb{R}^n_{++} \times \mathbb{R}^m \times \mathbb{R}^n_{++}$ segue-se de (3-12), (3-11) e do sistema (3-4) - (3-6) que

$$\varphi(\mu) = (x(\mu), y(\mu), s(\mu)), \qquad \mu \in (\bar{\mu} - \epsilon,\ \bar{\mu} + \epsilon).$$

Como $\bar{\mu}$ é qualquer ponto no intervalo aberto $(0, +\infty)$ segue-se que a trajetória $\{(x(\mu), y(\mu), s(\mu)) : \mu > 0\}$ é analítica. $\square$

Seja $\{(x(\mu), y(\mu), s(\mu)) : \mu > 0\}$ a Trajetória Central. A curva analítica $\phi : (0, +\infty) \to \mathbb{R}$ definida por

$$\phi(\mu) = (x(\mu), y(\mu), s(\mu)),$$

tem como imagem a Trajetória Central e como é usual também será chamada de Trajetória Central.

3.3 Convergência da Trajetória Central

Agora estamos interessados em mostrar que a Trajetória Central $\{(x(\mu), y(\mu), s(\mu)) : \mu > 0\}$ converge para um ponto específico da face ótima, em outras palavras, a função $\varphi : (0, +\infty) \to \mathbb{R}$ definida por

$$\phi(\mu) = (x(\mu), y(\mu), s(\mu)),$$

pode ser extendida continuamente no ponto 0. Para os próximos resultados veja Teorema 5.10.3 na página 225 em [24].

Proposição 3.3.1 *Seja $\{(x(\mu), y(\mu), s(\mu)) : \mu > 0\}$ a trajetória central. Então,*

i) *$c^T x(\mu)$ é estritamente crescente em μ;*

ii) *$b^T y(\mu)$ é estritamente decrescente em μ.*

Demonstração: Para a prova do item i, considere $0 < \mu < \bar{\mu}$. Para cada $\mu > 0$ temos que $x(\mu)$ é a única solução de (P_μ) assim,

$$c^T x(\mu) - \mu \sum_{i=1}^{n} \log x_i(\mu) < c^T x(\bar{\mu}) - \mu \sum_{i=1}^{n} \log x_i(\bar{\mu}).$$

Utilizando propriedades da função $\log$ obtemos que

$$c^T x(\mu) - c^T x(\bar{\mu}) < \mu \sum_{i=1}^{n} \log \frac{x_i(\mu)}{x_i(\bar{\mu})}. \tag{3-13}$$

Analogamente, para $\bar{\mu} > 0$ temos que $x(\bar{\mu})$ é a única solução de $(P_{\bar{\mu}})$ neste caso,

$$c^T x(\bar{\mu}) - \bar{\mu} \sum_{i=1}^{n} \log x_i(\bar{\mu}) < c^T x(\mu) - \bar{\mu} \sum_{i=1}^{n} \log x_i(\mu).$$

Manipulando convenientemente esta última inequação e usando propriedades da função $\log$ obtemos que

$$\bar{\mu} \sum_{i=1}^{n} \log \frac{x_i(\mu)}{x_i(\bar{\mu})} < c^T x(\mu) - c^T x(\bar{\mu}). \tag{3-14}$$

Adicionando as inequações (3-13) e (3-14) temos que

$$\bar{\mu} \sum_{i=1}^{n} \log \frac{x_i(\mu)}{x_i(\bar{\mu})} < \mu \sum_{i=1}^{n} \log \frac{x_i(\mu)}{x_i(\bar{\mu})}.$$

Como $0 < \mu < \bar{\mu}$, temos pela equação acima que $\sum_{i=1}^{n} \log(x_i(\mu)/x_i(\bar{\mu})) < 0$, então de (3-13) obtemos

$$c^T x(\mu) - c^T x(\bar{\mu}) < 0.$$

Como $0 < \mu < \bar{\mu}$ a última desigualdade prova que $c^T x(\mu)$ é estritamente crescente em μ. O que finaliza a prova do item i.
A prova do item ii é análoga à demonstração do item i. $\square$

Proposição 3.3.2 *Para cada $\bar{\mu} > 0$, o conjunto $\{(x(\mu), y(\mu), s(\mu)) : \bar{\mu} > \mu > 0\}$ é limitado.*

Demonstração: Seja $\bar{\mu} > 0$. Pela Proposição 2.3.1 o conjunto

$$\Gamma = \{x \in \mathbb{R}^n : \; Ax = b, \;\; c^T x \leq c^T x(\bar{\mu}), \;\; x > 0\},$$

é compacto, portanto limitado. Pela Proposição 3.1.2, segue-se que $Ax(\mu) = b$ e $x(\mu) > 0$. Agora para todo $0 < \mu < \bar{\mu}$ temos, pela Proposição 3.3.1 que $c^T x(\mu) < c^T x(\bar{\mu})$. Logo $x(\mu) \in \Gamma$, o que implica que $\{x(\mu) : \bar{\mu} > \mu > 0\}$ é limitado. Com demonstração análoga à Proposição 3.1.2, podemos verificar que o conjunto

$$\Psi = \{(y, s) \in \mathbb{R}^m \times \mathbb{R}^n : \; A^T y + s = c, \;\; b^T y \leq b^T y(\bar{\mu}), \;\; s > 0\},$$

é compacto. Então, com argumento análogo utilizado na primeira parte, podemos verificar que o conjunto $\{(y(\mu), s(\mu)) : \bar{\mu} > \mu > 0\}$ é limitado. Portanto $\{(x(\mu), y(\mu), s(\mu)) : \bar{\mu} > \mu > 0\}$ é limitado. □

A partir de agora, vamos definir as seguintes trajetórias: a Trajetória Central primal é o conjunto dos pontos solução do problema (P_μ) para cada $\mu > 0$, isto é, o conjunto

$$\{x(\mu) : \mu > 0\},$$

a Trajetória Central dual é definida como sendo o conjunto dos pontos solução do problema (D_μ) para cada $\mu > 0$, isto é

$$\{(y(\mu), s(\mu)) : \mu > 0\}.$$

Corolário 3.3.3 *Se $\bar{x}$ é um ponto de acumulação da Trajetória Central primal $\{x(\mu) : \mu > 0\}$ e $(\bar{y}, \bar{s})$ é um ponto de acumulação da Trajetória Central dual $\{(y(\mu), s(\mu)) : \mu > 0\}$ quando μ vai a zero. Então $(\bar{x}, \bar{y}, \bar{s})$ satisfaz as equações* (2-6) - (2-8)*, isto é, $\bar{x} \in P^*$ e $(\bar{y}, \bar{s}) \in D^*$.*

Demonstração: Tome a sequência $\{\mu_k\}$ tal que $\mu_k \to 0$ e $\lim_{k\to\infty} x(\mu_k) = \bar{x}$, $\lim_{k\to\infty} y(\mu_k) = \bar{y}$ e $\lim_{k\to\infty} s(\mu_k) = \bar{s}$. Como $(x(\mu), y(\mu), s(\mu))$ satisfaz (3-4) - (3-6) para cada $\mu > 0$, o resultado segue passando o limite nestas equações. □

Lema 3.3.4 *Suponha que $x^* \in P^*$ e $(y^*, s^*) \in D^*$ e seja a trajetória central $\{(x(\mu), y(\mu), s(\mu)) : \mu > 0\}$. Então*

$$\langle x(\mu), s^* \rangle + \langle x^*, s(\mu) \rangle = n\mu.$$

Demonstração: Note que, para todo $x \in \mathcal{F}(P)$ temos que $x(\mu) - x \in \text{Null}(A)$. De fato,

$$A(x(\mu) - x) = Ax(\mu) - Ax = b - b = 0.$$

De forma análoga, para todo par $(y, s) \in \mathcal{F}(D)$ temos que $s(\mu) - s \in \text{Im}(A^T)$. De fato,

$$A^T(y - y(\mu)) + (s - s(\mu)) = (A^T y + s) - (A^T y(\mu) + s(\mu)) = c - c = 0,$$

o que implica que $A^T(y(\mu) - y) = s(\mu) - s$. Pela Proposição 1.4.2 segue-se que $x(\mu) - x$ e $s(\mu) - s$ são ortogonais para todo $x \in \mathcal{F}(P)$ e $(y, s) \in \mathcal{F}(D)$ e $\mu > 0$. Em particular,

$$\langle x(\mu) - x^*, s(\mu) - s^* \rangle = 0, \qquad \mu > 0.$$

Simples manipulações algébricas na última igualdade implicam que

$$\langle x(\mu), s^* \rangle + \langle x^*, s(\mu) \rangle = \langle x(\mu), s(\mu) \rangle + \langle x^*, s^* \rangle . \tag{3-15}$$

Por hipótese $x^* \in P^*$ e $(y^*, s^*) \in D^*$, então por (2-8)

$$\langle x^*, s^* \rangle = (x^*)^T s^* = 0. \tag{3-16}$$

Por outro lado, segue de (3-6) que $X(\mu)s(\mu) = \mu e$, ou equivalentemente, $x_i(\mu)s_i(\mu) = \mu$, para $i = 1, 2, ..., n$, desta forma

$$\langle x(\mu), s(\mu) \rangle = n\mu.$$

Portanto, combinando a última equação com (3-15) e (3-16) obtemos a igualdade desejada. □

Proposição 3.3.5 *Seja $\bar{x}$ um ponto de acumulação da Trajetória Central primal $\{x(\mu) : \mu > 0\}$. Se $\bar{x}_i = 0$ então $x_i^* = 0$ para todo $x^* \in P^*$.*

Demonstração: Seja $x^* \in P^*$ e $(y^*, s^*) \in D^*$, pelo Lema 3.3.4 temos que

$$\langle x(\mu), s^* \rangle + \langle x^*, s(\mu) \rangle = n\mu.$$

Como $x(\mu) > 0$ e $s^* \geq 0$ a última equação implica em $\langle x^*, s(\mu) \rangle \leq n\mu$. Também, como $s(\mu) > 0$ e $x^* \geq 0$ temos

$$x_i^* s_i(\mu) \leq n\mu \qquad \forall \;\; i = 1, 2, \cdots, n. \tag{3-17}$$

Por outro lado, segue de (3-6) que $x_i(\mu)s_i(\mu) = \mu$ para todo $i = 1, 2, \cdots, n$ o que implica

$$s_i(\mu) = x_i^{-1}(\mu)\mu \qquad \forall \;\; i = 1, 2, \cdots, n. \tag{3-18}$$

Substituindo (3-18) em (3-17) obtemos que $x_i^* x_i^{-1}(\mu)\mu \leq n\mu$ que pode ser reescrito como

$$0 \leq x_i^* \leq n x_i(\mu) \qquad \forall \ i = 1, 2, \cdots, n. \tag{3-19}$$

Agora, por hipótese $\bar{x}$ é ponto de acumulação, desta forma, tome a sequência $\{\mu_k\}$ tal que $\mu_k \to 0$ e $\lim_{k\to\infty} x(\mu_k) = \bar{x}$. Note que a expressão (3-19) é válida para todo μ e, em particular, vale para μ_k assim,

$$0 \leq x_i^* \leq n x_i(\mu_k) \qquad \forall \ i = 1, 2, \cdots, n.$$

Agora, fazendo $k \to \infty$ na última expressão, temos que

$$0 \leq x_i^* \leq n \bar{x}_i \qquad \forall \ i = 1, 2, \cdots, n.$$

Portanto, da última desigualdade segue o resultado desejado. □

Proposição 3.3.6 *Seja* $(\bar{y}, \bar{s})$ *um ponto de acumulação da Trajetória Central dual* $\{(y(\mu), s(\mu)) : \mu > 0\}$. *Se* $\bar{s}_i = 0$ *então* $s_i^* = 0$ *para todo* $(y^*, s^*) \in D^*$.

Demonstração: Análoga à demonstração da Proposição 3.3.5. □

A partir das definições dos conjuntos B e N, na seção 2.4 do Capítulo 2, podemos obter os seguintes resultados.

Corolário 3.3.7 *Se* $\bar{x}$ *é um ponto de acumulação da Trajetória Central primal* $\{x(\mu) : \mu > 0\}$. *Então* $B(\bar{x}) = B$, *onde* $B(\bar{x}) = \{i : \bar{x}_i > 0\}$.

Demonstração: Segue imediatamente da definição de B e da Proposição 3.3.5. □

Corolário 3.3.8 *Se* $(\bar{y}, \bar{s})$ *é um ponto de acumulação da Trajetória Central dual* $\{(y(\mu), s(\mu)) : \mu > 0\}$. *Então* $N(\bar{s}) = N$, *onde* $N(\bar{s}) = \{i : \bar{s}_i > 0\}$.

Demonstração: Segue imediatamente da definição de N e da Proposição 3.3.6. □

Corolário 3.3.9 *Seja* $(\bar{x}, \bar{y}, \bar{s})$ *um ponto de acumulação da Trajetória Central* $\{(x(\mu), y(\mu), s(\mu)) : \mu > 0\}$ *quando* $\mu \to 0$. *Então*

$$\bar{x} + \bar{s} > 0.$$

Demonstração: Segue imediatamente do Teorema 2.4.1 e dos Corolário 3.3.7 e Corolário 3.3.8. □

Para os próximos resultados, veja Teorema 5.10.3 na página 225 em [24].

Teorema 3.3.10 *A Trajetória Central primal* $\{x(\mu) : \mu > 0\}$ *converge para o centro analítico da face ótima quando* $\mu \to 0$*, isto é, para a única solução do problema*

$$minimizar - \sum_{i \in B} \log x_i \tag{3-20}$$

$$sujeito \;\; a : x \in P^*. \tag{3-21}$$

Demonstração: Seja $x^* \in P^*$, $(y^*, s^*) \in D^*$ e $\{(x(\mu), y(\mu), s(\mu)) : \mu > 0\}$ a trajetória central. Então pelo Lema 3.3.4 temos que

$$\left\langle \frac{x(\mu)}{\mu}, s^* \right\rangle + \left\langle x^*, \frac{s(\mu)}{\mu} \right\rangle = n. \tag{3-22}$$

Por outro lado $X(\mu)s(\mu) = \mu e$, esta igualdade implica que

$$\frac{s(\mu)}{\mu} = x^{-1}(\mu), \qquad \frac{x(\mu)}{\mu} = s^{-1}(\mu).$$

Substituindo as duas últimas igualdades em (3-22) temos que

$$\left\langle s^{-1}(\mu), s^* \right\rangle + \left\langle x^*, x^{-1}(\mu) \right\rangle = n. \tag{3-23}$$

Pela definição dos conjuntos B e N, a última equação pode ser reescrita da seguinte forma

$$\sum_{i \in N} s_i^{-1}(\mu) s_i^* + \sum_{i \in B} x_i^{-1}(\mu) x_i^* = n > 0. \tag{3-24}$$

Tome a sequência $\{\mu_k\}$ tal que $\mu_k \to 0$ e $\lim_{k\to\infty} x(\mu_k) = \bar{x}$ e $\lim_{k\to\infty} s(\mu_k) = \bar{s}$. Pelo Corolário 3.3.7 temos que $\bar{x}_i > 0$ para $i \in B$ e pelo Corolário 3.3.8 que $\bar{s}_i > 0$ para $i \in N$. Assim pela equação (3-24) obtemos que

$$\frac{\sum_{i \in N} \bar{s}_i^{-1} s_i^* + \sum_{i \in B} \bar{x}_i^{-1} x_i^*}{n} = 1.$$

Pelo Teorema da complementaridade estrita $B \cup N = \{1, \dots n\}$, assim utilizando a propriedade de que a média geométrica é menor ou igual do que média aritmética , obtemos da última equação que

$$\prod_{i\in N} \bar{s}_i^{-1}s_i^* \prod_{i\in B} \bar{x}_i^{-1}x_i^* \leq 1.$$

Note que a fórmula acima vale para qualquer solução $(y^*, s^*) \in D^*$. Pelo Corolário 3.3.3 $(\bar{y}, \bar{s}) \in D^*$ assim fazendo $s_i^* = \bar{s}_i$ para $i \in N$, a última desigualdade se reduz à

$$\prod_{i\in B} \bar{x}_i^{-1}x_i^* \leq 1. \tag{3-25}$$

Após algumas manipulações algébricas simples nesta desigualdade e utilizando propriedades da função $\log$, obtemos que

$$\sum_{i\in B} \log x_i^* \leq \sum_{i\in B} \log \bar{x}_i.$$

Portanto o ponto $\bar{x}$ é uma solução do problema 3-20 e 3-21. Como este problema tem solução única, o teorema está provado. □

Teorema 3.3.11 *A Trajetória Central dual* $\{(y(\mu), s(\mu)) : \mu > 0\}$ *converge para o centro analítico da face ótima quando* $\mu \to 0$*, isto é, para a única solução do problema*

$$minimizar - \sum_{i\in N} \log s_i \tag{3-26}$$

$$sujeito\ \ a : (y, s) \in D^*. \tag{3-27}$$

Demonstração: Análoga à demosntração do Teorema 3.3.10. □

Sejam $x^* \in P^*$ e $(y^*, s^*) \in D^*$. De acordo com a definição dos conjuntos B e N podemos assumir, sem perda de generalidade, que

$$x^* = ((x_B^*)^T, 0^T)^T, \qquad s^* = (0^T, (s_N^*)^T)^T.$$

De fato, existe uma matriz P tal que Px^* e Ps^* tem a forma acima. E assim, fazendo a mudança de variável $\tilde{x} = Px$ e $\tilde{s} = Ps$ nos problemas (P) e (D), obtemos problemas equivalentes cuja a solução está no formato acima. Portanto podemos enunciar o seguinte resultado.

Corolário 3.3.12 *Seja a Trajetória Central* $\{(x(\mu), y(\mu), s(\mu)) : \mu > 0\}$*. Então valem as seguintes afirmações:*

i) $\lim_{\mu \to 0} x(\mu) = x^* = ((x_B^*)^T, 0^T)^T$ ***e*** $x_B^* > 0$*;*

ii) $\lim_{\mu \to 0} s(\mu) = s^* = (0^T, (s_N^*)^T)^T$ ***e*** $s_N^* > 0$*;*

iii) $x^* + s^* > 0$*;*

iv) $\lim_{\mu \to 0} \frac{x_N(\mu)}{\mu} = (s_N^*)^{-1} > 0$*;*

v) $\lim_{\mu \to 0} \frac{s_B(\mu)}{\mu} = (x_B^*)^{-1} > 0$*.*

Demonstração: Os itens **i** e **ii** seguem diretamente dos Teoremas 3.3.10 e 3.3.11. O item **iii** segue imediatamente do Corolário 3.3.3 e do Corolário 3.3.9. Agora, considere novamente os Teoremas 3.3.10 e 3.3.11 e note que valem as seguintes expressões

$$\lim_{\mu \to 0} x_B(\mu) = x_B^* > 0, \qquad \lim_{\mu \to 0} s_N(\mu) = s_N^* > 0. \tag{3-28}$$

Desde que $X(\mu)s(\mu) = \mu e$, temos que

$$\frac{s_B(\mu)}{\mu} = x_B^{-1}(\mu), \qquad \frac{x_N(\mu)}{\mu} = s_N^{-1}(\mu).$$

Os itens **iv** e **v** seguem tomando o limite na última expressão e usando (3-28).
□

CAPÍTULO 4

Comportamento limite da Trajetória Central

No capítulo anterior definimos e descrevemos as equações que definem a Trajetória Central

$$\{(x(\mu), y(\mu), s(\mu)) : \mu > 0\}.$$

Verificamos que as equações da Trajetória Central (3-4) - (3-6) definem uma curva analítica

$$\mu > 0 \mapsto \phi(\mu) = (x(\mu), y(\mu), s(\mu)),$$

contida no conjunto $\mathcal{F}^0(P) \times \mathcal{F}^0(D) \subset I\!\!R^n_{++} \times (I\!\!R^m \times I\!\!R^n_{++})$ e que converge para um ponto na fronteira deste conjunto, isto é,

$$\lim_{\mu \to 0} \phi(\mu) = \lim_{\mu \to 0} (x(\mu), y(\mu), s(\mu)) = (x^*, y^*, s^*) \in P^* \times D^* \subset \mathcal{F}^0(P) \times \mathcal{F}^0(D).$$

Neste sentido podemos dizer que ϕ está definida em $\mu = 0$, ou seja,

$$\phi(0) = (x^*, y^*, s^*).$$

Neste capítulo estaremos interessados em mostrar que a Trajetória Central é analítica em $\mu = 0$.

Uma primeira prova para este fato, foi feita em [9], porém esta prova era muito extensa e difícil, exigindo um bom conhecimento matemático para o seu entendimento. Ja em [10] foi feita uma prova mais simples, utilizando o teorema da função implícita em um certo sistema equivalente. Durante a demonstração, a autora utilizou uma matriz não-singular sem mostrar sua existência e também não foi entendido algumas passagens durante a demonstração.

Desta forma, resolvemos fazer uma nova prova, utilizando sistemas equivalentes, baseando-se na prova feita para Programação Semidefinida em [23].

4.1 Mudança de Variáveis

A Trajetória Central é a única solução do sistema (3-4) - (3-6). Visto que a matriz que define este sistema é não-singular sempre que $x > 0$, $s > 0$ e Posto(A) é completo. Aplicamos o Teorema da Função Implícita para concluir que a Trajetória Central é analítica para $\mu > 0$, veja Teorema 3.2.1. Agora para $\mu = 0$ o ponto limite da Trajetória Central (x^*, y^*, s^*) é solução do sistema (2-6) - (2-8) que pode admitir infinitas soluções. Neste caso, sendo $x^* \geq 0$ e $s^* \geq 0$, não podemos garantir que a matriz que define esse sistema seja não-singular. E consequentemente, não podemos aplicar o Teorema da Função Implícita.

Nosso objetivo nesta sessão é realizar uma mudança de variáveis conveniente no sistema (3-4) - (3-6), de forma que podemos obter um sistema equivalente cuja a matriz que o define seja não-singular, podendo assim aplicar o Teorema da Função Implícita e como consequência concluir que a Trajetória Central é analítica neste ponto.

A partir de agora, iremos assumir a hipótese de que a primeira coluna da matriz A é não nula sem mencionar este fato nos enunciados dos resultados. Esta hipótese não é restritiva pois, se tivermos a primeira coluna da matriz A nula, podemos eliminar a variável do problema (P) associada à esta coluna obtendo uma nova matriz cuja primeira coluna seja não nula.

Lema 4.1.1 *Sejam os conjuntos de índices B e N definidos em* (2-9) *e* (2-10) *, então existe uma matriz $Q \in M_{m\times m}(\mathbb{R})$ inversível tal que*

$$QA = \begin{bmatrix} \tilde{A}_{l\times B} & \tilde{A}_{l\times N} \\ 0_{m-l\times B} & \tilde{A}_{(m-l)\times N} \end{bmatrix},$$

onde $\tilde{A}_{l\times B} \in M_{l\times B}(\mathbb{R})$ e $\tilde{A}_{m-l\times N} \in M_{m-l\times N}(\mathbb{R})$, possivelmente com $l = m$ e Posto$(\tilde{A}_{l\times B})$ + Posto$(\tilde{A}_{(m-l)\times N}) = m =$ Posto(A).

Demonstração: Como a primeira coluna da matriz A é não nula, o resultado segue diretamente do Teorema 1.4.3. □

Note que, se as primeiras m colunas da matriz A são linearmente independentes e a cardinalidade do conjunto B é maior que m então neste caso teremos $l = m$, isto é, $QA = [\tilde{A}_{m\times B} \;\; \tilde{A}_{m\times N}]$.

Proposição 4.1.2 *Sejam $x^* \in P^*$, $(y^*, s^*) \in D^*$ e $\{(x(\mu), y(\mu), s(\mu)) : \mu > 0)\}$ a trajetória central. Considere os conjuntos B e N definidos em* (2-9) *e* (2-10) *e* *as matrizes* $\tilde{A}_{l\times B} \in M_{l\times B}(\mathbb{R})$ *e* $\tilde{A}_{m-l\times N} \in M_{m-l\times N}(\mathbb{R})$ *definidas no Lema 4.1.1. Então existe uma curva* $\{\tilde{y}(\mu) : \mu > 0\}$ *tal que a curva* $\{(\tilde{x}(\mu), \tilde{y}(\mu)), \tilde{s}(\mu) : \mu > 0\}$*, onde*

$$\tilde{x}(\mu) := \begin{pmatrix} x_B(\mu) \\ \frac{x_N(\mu)}{\mu} \end{pmatrix}, \quad \tilde{s}(\mu) := \begin{pmatrix} \frac{s_B(\mu)}{\mu} \\ s_N(\mu) \end{pmatrix},$$

é a única solução do sistema

$$A_\mu(\tilde{x} - x^*) = 0, \quad \tilde{x} > 0 \tag{4-1}$$

$$A_\mu^T \tilde{y} + (\tilde{s} - s^*) = 0, \quad \tilde{s} > 0 \tag{4-2}$$

$$\tilde{X}\tilde{s} = e, \quad \mu > 0, \tag{4-3}$$

onde a matriz A_μ *tem posto completo e é dada por*

$$A_\mu = \begin{bmatrix} \tilde{A}_{l\times B} & \mu\tilde{A}_{l\times N} \\ 0_{m-l\times B} & \tilde{A}_{(m-l)\times N} \end{bmatrix}.$$

Demonstração: Sejam $x^* \in P^*$, $(y^*, s^*) \in D^*$ e $\{(x(\mu), y(\mu), s(\mu)) : \mu > 0)\}$ a trajetória central. Desta forma, para cada $\mu > 0$, a terna $(x(\mu), y(\mu), s(\mu))$ é a única solução do sistema (3-4) - (3-6). Como $x^* \in P^*$, $(y^*, s^*) \in D^*$ o sistema (3-4) - (3-6) é equivalente ao seguinte sistema

$$\begin{aligned} A(x - x^*) = 0, &\quad x > 0 \\ A^T(y - y^*) + (s - s^*) = 0, &\quad s > 0 \\ Xs = \mu e, &\quad \mu > 0. \end{aligned}$$

Logo, para cada $\mu > 0$, a terna $(x(\mu), y(\mu), s(\mu))$ é a única solução do sistema acima. Agora, considere a seguinte mudança de variáveis

$$\tilde{x} = P_\mu x, \qquad \tilde{s} = D_\mu s, \qquad \mu > 0, \tag{4-4}$$

onde P_μ e D_μ são matrizes diagonais inversíveis definidas da seguinte forma

$$P_\mu = \begin{bmatrix} I_B & 0_{B\times N} \\ 0_{N\times B} & \frac{1}{\mu}I_N \end{bmatrix}, \qquad D_\mu = \begin{bmatrix} \frac{1}{\mu}I_B & 0_{B\times N} \\ 0_{N\times B} & I_N \end{bmatrix}.$$

Note que pela definição de P_μ e D_μ e Corolário 3.3.12, itens **i** e **ii**, temos que

$$P_\mu x^* = x^*, \qquad D_\mu s^* = s^*.$$

Assim, utilizando a mudança de variáveis (4-4) no sistema acima, obtemos um novo sistema equivalente

$$A(P_\mu^{-1}\tilde{x} - P_\mu^{-1}x^*) = 0, \quad \tilde{x} > 0 \tag{4-5}$$
$$A^T(y - y^*) + (D_\mu^{-1}\tilde{s} - D_\mu^{-1}s^*) = 0, \quad \tilde{s} > 0 \tag{4-6}$$
$$P_\mu^{-1}\tilde{X}D_\mu^{-1}\tilde{s} = \mu e, \quad \mu > 0. \tag{4-7}$$

A partir da mundança de variáveis (4-4), defina as seguintes trajetórias,

$$\tilde{x}(\mu) = P_\mu x(\mu) = \begin{pmatrix} x_B(\mu) \\ \frac{x_N(\mu)}{\mu} \end{pmatrix}, \qquad \tilde{s}(\mu) = D_\mu s(\mu) = \begin{pmatrix} \frac{s_B(\mu)}{\mu} \\ s_N(\mu) \end{pmatrix}.$$

Deste modo, a trajetória $\{\tilde{x}(\mu), y(\mu), \tilde{s}(\mu) : \mu > 0\}$ é a única solução do sistema (4-5) - (4-7). Agora, veja que

$$P_\mu^{-1}\tilde{X}D_\mu^{-1} = \mu\tilde{X},$$

o que implica que a última equação do sistema acima é equivalente à $\tilde{X}\tilde{s} = e$. Desta forma este sistema é equivalente à

$$\begin{aligned} AP_\mu^{-1}(\tilde{x} - x^*) = 0, &\quad \tilde{x} > 0 \\ A^T(y - y^*) + D_\mu^{-1}(\tilde{s} - s^*) = 0, &\quad \tilde{s} > 0 \\ \tilde{X}\tilde{s} = e, &\quad \mu > 0. \end{aligned}$$

Considere os conjuntos B e N definidos em (2-9) e (2-10). Logo, pelo Lema 4.1.1 existe uma matriz $Q \in M_{m\times m}(\mathbb{R})$ inversível tal que

$$QA = \begin{bmatrix} \tilde{A}_{l\times B} & \tilde{A}_{l\times N} \\ 0_{m-l\times B} & \tilde{A}_{(m-l)\times N} \end{bmatrix},$$

onde $\tilde{A}_{l\times B} \in M_{l\times B}(\mathbb{R})$, $\tilde{A}_{m-l\times N} \in M_{m-l\times N}(\mathbb{R})$ e Posto$(\tilde{A}_{l\times B})$ + Posto$(\tilde{A}_{(m-l)\times N})$ = Posto(A). Assim, segue-se que o último sistema é equivalente ao seguinte sistema

$$\begin{aligned} QAP_\mu^{-1}(\tilde{x} - x^*) = 0, &\quad \tilde{x} > 0 \\ (QA)^T(Q^T)^{-1}(y - y^*) + D_\mu^{-1}(\tilde{s} - s^*) = 0, &\quad \tilde{s} > 0 \\ \tilde{X}\tilde{s} = e, &\quad \mu > 0. \end{aligned}$$

É imediato verificar que a segunda equação deste sistema pode ser reescrita da seguinte forma

$$D_\mu(QA)^T(Q^T)^{-1}(y - y^*) + (\tilde{s} - s^*) = 0.$$

Usando esta igualdade, o sistema acima é equivalente ao seguinte sistema,

$$\begin{aligned} QAP_\mu^{-1}(\tilde{x} - x^*) &= 0, \quad \tilde{x} > 0 \\ D_\mu(QA)^T(Q^T)^{-1}(y - y^*) + (\tilde{s} - s^*) &= 0, \quad \tilde{s} > 0 \\ \tilde{X}\tilde{s} &= e, \quad \mu > 0. \end{aligned}$$

Agora defina as seguintes matrizes diagonais inversíveis

$$\bar{P}_\mu = \begin{bmatrix} I_l & 0_{l\times m-l} \\ 0_{m-l\times l} & \frac{1}{\mu}I_{m-l} \end{bmatrix}, \qquad \bar{D}_\mu = \begin{bmatrix} \frac{1}{\mu}I_l & 0_{l\times m-l} \\ 0_{m-l\times l} & I_{m-l} \end{bmatrix}.$$

A partir da definição acima, o último sistema pode ser reescrito da seguinte forma.

$$\begin{aligned} \bar{P}_\mu QAP_\mu^{-1}(\tilde{x} - x^*) &= 0 \\ D_\mu(QA)^T(\bar{D}_\mu)^{-1}\left[\bar{D}_\mu(Q^T)^{-1}(y - y^*)\right] + (\tilde{s} - s^*) &= 0 \\ \tilde{X}\tilde{s} &= e. \end{aligned}$$

Note que as matrizes A_μ e A_μ^T podem ser escritas da seguinte forma,

$$A_\mu = \bar{P}_\mu QAP_\mu^{-1}, \qquad A_\mu^T = D_\mu(QA)^T(\bar{D}_\mu)^{-1} = \begin{bmatrix} \tilde{A}_{l\times B}^T & 0_{l\times N} \\ \mu\tilde{A}_{l\times N}^T & \tilde{A}_{(m-l)\times N}^T \end{bmatrix}.$$

Finalmente, substituindo A_μ e A_μ^T no sistema acima, obtemos o sistema equivalente,

$$A_\mu(\tilde{x} - x^*) = 0, \quad \tilde{x} > 0 \tag{4-8}$$

$$A_\mu^T\tilde{y} + (\tilde{s} - s^*) = 0, \quad \tilde{s} > 0 \tag{4-9}$$

$$\tilde{X}\tilde{s} = e, \quad \mu > 0. \tag{4-10}$$

onde a variável $\tilde{y}$ é dada por $\tilde{y} := [\bar{D}_\mu(Q^T)^{-1}(y - y^*)]$. Como o sistema (4-5) - (4-7) é equivalente ao sitema (4-8) - (4-10) podemos concluir que a trajetória $\{(\tilde{x}(\mu), \tilde{y}(\mu), \tilde{s}(\mu)) : \mu > 0\}$ é a única curva solução do sistema acima, onde a trajetória $\{\tilde{y}(\mu) : \mu > 0\}$ é dada por $\tilde{y}(\mu) := [\bar{D}_\mu(Q^T)^{-1}(y(\mu) - y^*)]$. Isto conclui a demonstração. $\square$

4.2 Analiticidade no ponto limite

Pelo Teorema 3.2.1 e da definição da trajetória $\{(\tilde{x}(\mu), \tilde{y}(\mu), \tilde{s}(\mu)) : \mu > 0\}$ na Proposição 4.1.2 concluímos que esta Trajetória é analítica no intervalo $(0, \infty)$. Pelo Corolário 3.3.12 podemos definir,

$$\tilde{x}^* := \begin{pmatrix} x_B^* \\ (s_N^*)^{-1} \end{pmatrix} = \lim_{\mu \to 0} \tilde{x}(\mu), \quad \tilde{y}^* = \lim_{\mu \to 0} \tilde{y}(\mu), \quad \tilde{s}^* := \begin{pmatrix} (x_B^*)^{-1} \\ s_N^* \end{pmatrix} = \lim_{\mu \to 0} \tilde{s}(\mu),$$

e assim, a Trajetória está bem definida para $\mu = 0$, isto é,

$$\{(\tilde{x}(\mu), \tilde{y}(\mu), \tilde{s}(\mu)) : \mu \geq 0\}, \qquad (\tilde{x}(0), \tilde{y}(0), \tilde{s}(0)) = (\tilde{x}^*, \tilde{y}^*, \tilde{s}^*).$$

Nosso primeiro objetivo nesta seção é mostrar analiticidade desta trajetória em $\mu = 0$. Agora, pelos Teoremas 3.3.10 e 3.3.11 temos que

$$x^* = \lim_{\mu \to 0} x(\mu), \qquad y^* = \lim_{\mu \to 0} y(\mu), \qquad s^* = \lim_{\mu \to 0} s(\mu),$$

e assim, a Trajetória Central $\{(x(\mu), y(\mu), s(\mu)) : \mu > 0\}$ está definida para $\mu = 0$, isto é,

$$\{(x(\mu), y(\mu), s(\mu)) : \mu \geq 0\}, \qquad (x(0), y(0), s(0)) = (x^*, y^*, s^*).$$

E desta forma, mostraremos a analiticidade da Trajetória Central em $\mu = 0$.

Teorema 4.2.1 *A Trajetória* $\{\tilde{x}(\mu), \tilde{y}(\mu), \tilde{s}(\mu) : \mu \geq 0\}$ *é uma curva analítica.*

Demonstração: Primeiro, defina a seguinte função

$$\begin{aligned} F : \mathbb{R} \times \mathbb{R}^n_{++} \times \mathbb{R}^m \times \mathbb{R}^n_{++} &\to \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^n \\ (\mu, \tilde{x}, \tilde{y}, \tilde{s}) &\mapsto (A_\mu(\tilde{x} - x^*), A_\mu^T \tilde{y} + (\tilde{s} - s^*), \tilde{X}\tilde{s} - e). \end{aligned}$$

Note que a matriz jacobiana de F com relação à $(\tilde{x}, \tilde{y}, \tilde{s})$ é dada por

$$\nabla_{(\tilde{x}, \tilde{y}, \tilde{s})} F(\mu, \tilde{x}, \tilde{y}, \tilde{s}) = \begin{bmatrix} A_\mu & 0_{m \times m} & 0_{m \times n} \\ 0_{n \times n} & A_\mu^T & I_n \\ \tilde{S} & 0_{n \times m} & \tilde{X} \end{bmatrix}.$$

Agora, considere o seguinte sistema

$$F(0,\tilde{x},\tilde{y},\tilde{s}) = (0,0,0), \qquad \tilde{x} > 0, \quad \tilde{s} > 0,$$

e note que, pelo Corolário 3.3.12 e definição de $\{\tilde{x}(\mu),\tilde{y}(\mu),\tilde{s}(\mu) : \mu \geq 0\}$, o ponto $(0,\tilde{x}^*,\tilde{y}^*,\tilde{s}^*)$ é solução do sistema acima. Por outro lado, temos que a matriz

$$A_0 = \begin{bmatrix} \tilde{A}_{l\times B} & 0_{l\times N} \\ 0_{m-l\times B} & \tilde{A}_{(m-l)\times N} \end{bmatrix},$$

tem posto completo pois, Posto$(\tilde{A}_{l\times B})$ + Posto$(\tilde{A}_{(m-l)\times N}) = m$, e como $\tilde{x}^* > 0$ e $\tilde{s}^* > 0$ obtemos, pelo Lema 1.4.1, que a matriz

$$\nabla_{(\tilde{x},\tilde{y},\tilde{s})}F(0,\tilde{x}^*,\tilde{y}^*,\tilde{s}^*) = \begin{bmatrix} A_0 & 0_{m\times m} & 0_{m\times n} \\ 0_{n\times n} & A_0^T & I_n \\ \tilde{S}^* & 0_{n\times m} & \tilde{X}^* \end{bmatrix}.$$

é não-singular. Então segue-se do Teorema da Função Implícita 1.2.3 que existe um aberto $(-\epsilon,\epsilon)\times U \subset \mathbb{R}\times\mathbb{R}^n_{++}\times\mathbb{R}^m\times\mathbb{R}^n_{++}$ e uma única curva analítica, $\varphi : (-\epsilon,\epsilon) \to U$, satisfazendo $\varphi(0) = (\tilde{x}^*,\tilde{y}^*,\tilde{s}^*)$, além disso,

$$F(\mu,\varphi(\mu)) = (0_m,0_n,0_n), \qquad \mu \in (-\epsilon,\epsilon). \tag{4-11}$$

Finalmente, note que o sistema (4-1) - (4-3) é equivalente ao seguinte sistema

$$F(\mu,\tilde{x},\tilde{y},\tilde{s}) = (0_m,0_n,0_n) \qquad \mu > 0, \quad \tilde{x} > 0, \quad \tilde{s} > 0. \tag{4-12}$$

Pela Proposição 4.1.2, para cada $\mu > 0$, o sistema (4-1) - (4-3) tem uma única solução $(\tilde{x}(\mu),\tilde{y}(\mu),\tilde{s}(\mu))$ e $(\tilde{x}(0),\tilde{y}(0),\tilde{s}(0)) = (\tilde{x}^*,\tilde{y}^*,\tilde{s}^*)$. Assim, como $\varphi(\mu) \subset \mathbb{R}^n_{++}\times\mathbb{R}^m\times\mathbb{R}^n_{++}$ segue-se de (4-11) e (4-12) que

$$\varphi(\mu) = (\tilde{x}(\mu),\tilde{y}(\mu),\tilde{s}(\mu)), \qquad \mu \in [0,\epsilon).$$

E assim, pela Definição 1.2.2, a trajetória $\{(\tilde{x}(\mu),\tilde{y}(\mu),\tilde{s}(\mu)) : \mu \geq 0\}$ é analítica em $\mu = 0$. $\square$

Teorema 4.2.2 *A Trajetória Central* $\{(x(\mu),y(\mu),s(\mu)) : \mu \geq 0\}$ *é uma curva analítica em* $\mu = 0$*. Em particular, todas as derivadas da trajetória central convergem quando* $\mu \to 0$*.*

Demonstração: Pela Proposição 4.1.2 temos que para $\mu > 0$,

$$x_B(\mu) = \tilde{x}_B(\mu), \quad x_N(\mu) = \mu\tilde{x}_N(\mu), \qquad s_B(\mu) = \mu\tilde{s}_B(\mu), \quad s_N(\mu) = \tilde{s}_N(\mu).$$

Como a trajetória $\{(\tilde{x}(\mu), \tilde{y}(\mu), \tilde{s}(\mu)) : \mu \geq 0\}$ é analítica em $\mu = 0$, segue imediatamente da igualdade acima que a trajetória $\{(x(\mu), s(\mu)) : \mu \geq 0\}$ é analítica em $\mu = 0$. Agora, note que

$$y(\mu) = (AA^T)^{-1}A(c - s(\mu)).$$

Desta forma concluímos que a Trajetória Central $\{(x(\mu), y(\mu), s(\mu)) : \mu \geq 0\}$ é analítica em $\mu = 0$. A última parte do teorema segue imediatamente da definição de analiticidade. $\square$

Considerações Finais

Neste trabalho estudamos e demonstramos resultados acerca da Trajetória Central usual, isto é, a Trajetória Central obtida alterando a função objetivo do PPL através da função barreira logarítma.

No capítulo 4 foi demonstrada a analiticidade na fronteira da Trajetória Central (3-4) - (3-6). Veja que com simples adaptações da prova apresentada neste capítulo, podemos mostrar que os resultados são válidos para a w-weighted Central Path, isto é, substituindo a equação (3-6) pela equação $Xs = \mu w$, para qualquer $w \in \mathbb{R}^n$, no sistema que define a Trajetória Central.

O comportamento limite da Trajetória Central foi demonstrado nas referências [9] e [10]. A demonstração feita em [9] é trabalhosa e exige uma boa bagagem matemática para o seu bom entendimento. Enquanto que em [10] a prova é simples e utiliza o Teorema da Função Implícita para concluir o resultado. Porém, nesta última, é utilizado alguns argumentos sem demonstrá-los.

Nossa contribuição foi provar que a Trajetória Central é analítica no seu ponto limite, com uma abordagem um pouco diferente daquela feita em [10]. O que tornou a prova mais didática e acessível inclusive para alunos dos primeiros anos de graduação.

Foi feito um estudo completo da Trajetória Central onde, quase todos os resultados necessários para obtenção do comportamento limite foram demonstrados. Tentamos desta forma, deixar o texto auto-contido de maneira que o leitor interessado não necessite de leitura auxiliar para o bom entendimento do texto.

Por fim, como proposta de trabalho futuro sugerimos uma demonstração análoga para o Problema de Programação Semi-Definida. Visto que, vários resultados para Programação Semi-definida seguem demonstrações análogas às feitas em Programação Linear. Esta proposta se baseia no fato que [11] fez uma prova deste fato para Programação Semi-definida.

Referências Bibliográficas

[1] I. Adler e R. C. D. Monteiro. *Limiting Behavior of the affine scaling continuous trajectories for linear programming problems.* Contemporary Mathematics, Vol. 114, 1990.

[2] G. Ávila. *Cálculo I: Funções de uma variável*. Livros Técnicos e Científicos Editora, 6ª edição, 1994.

[3] M. S. Bazaraa, J. J. Jarvis e H. D. Sherali. *Linear Programming and Network Flows*. John Wiley & Sons, 2ª edição, 1990.

[4] M. S. Bazaraa, H. D. Sherali e C. M. Shetty. *Nonlinear Programming: Theory and algorithms*. John Wiley & Sons, 2ª edição, 1993.

[5] G. B. Dantzig. *Maximization of a Linear Function of Variables Subject to Linear Inequalities*. Activity Analysis of Production and Allocation, T. C. Koopmans, Ed., John Wiley and Sons, New York, pp. 339-347, 1951.

[6] J. Dieudonné. *Foundations of Modern Analisys*. Academic Press, New York and London, 1960.

[7] C. C. Gonzaga. *Path-following methods for linear programming*. SIAM Review 34 (1992) 167-224.

[8] O. Güler, C. Roos, T. Terlaky, J. -Ph. Vial. *A survey of the implications of the behavior of the central path for the duality theory of linear programming*. Management Science 41 (1995) 1922-1934.

[9] M. Halická. *Analytical properties of the central path at boundary point in linear programming*. Mathematical Programming 84 (1999) 335-355.

[10] M. Halická. *Two simple Proofs for analyticity of the central path in linear programming*. Operations Research Letters 28 (2001) 9-19.

[11] M. Halická. *Analyticity of the central path at the boundary point in semidefinite programming*. European Journal of Operational Research 143 (2002) 311-324.

[12] J.-B. Hiriart-Urruty, C. Lemaréchal. *Convex Analysis and Minimization Algorithms I*. Springer-Verlag, New York, 1993.

[13] A. Izmailov e M. Solodov. *Otimização - Volume 1: Condições de Otimalidade, Elementos de Análise Convexa e de Dualidade*. IMPA, 2005.

[14] N. Karmarkar. *A new polynomial time algorithm for linear programming.* Combinatorica 4 (1984) 373-395.

[15] L. G. Khachiyan. *A polynomial algorithm in linear programming*. Soviet Mathematics Doklady 20 (1979) 191-194.

[16] L. G. Khachiyan. *Polynomial algorithms in linear programming.* USSR Computational Mathematics and Mathematical Physics 20 (1980) 53-72.

[17] V. Klee e G. Minty. *How good is the simplex algorithm?*. Inequalities III, O. Sisha, ed., Academic Press, New York, NY, 1972.

[18] E. L. Lima. *Análise Real: Funções de Uma Variável - Volume 1*. IMPA, 8ª edição, 2006.

[19] E. L. Lima. *Análise Real - Volume 2*. IMPA, 2004.

[20] E. L. Lima. *Curso de Análise - Volume 2*. IMPA, 3ª edição, 1981.

[21] D. G. Luenberger. *Linear and Nonlinear Programming*. Addison-Wesley Publishing Company, 2ª edição.

[22] R. D. C. Monteiro e I. Adler. *Interior path following primal-dual algorithms. Part I: linear programming*. Mathematical Programming 44 (1989) 27-41.

[23] J. X. da C. Neto, O. P. Ferreira e R. D. C. Monteiro. *Assymptotic behavior of the central path for a special class of degenerate SDP problems*. Math. Programming Ser. A 103 (2005) 487-514.

[24] R. Saigal. *Linear Programming: A Modern Integrated Analysis*. Kluwer Academic Publishers, 2ª edição, 1997.

[25] G. Strang. *Linear Algebra and Its Applications*. Harcourt Brace and International Edition, 3ª edição, 1988.

[26] M. J. Todd. *Mathematical Programming: Lecture 6*. Notas de Aula acessada em: http://legacy.orie.cornell.edu/ miketodd/or630/or630.html, 2005.

[27] G. Zhao, J. Zhu. *Analytical properties of the central trajectory in interior points methods*. Advances in Optimization and Approximation, Kluwer Academic Publishers, Dordrecht (1994) 362-375.