



UNIVERSIDADE FEDERAL DO PIAUÍ
CENTRO DE CIÊNCIAS DA NATUREZA
PÓS-GRADUAÇÃO EM MATEMÁTICA
MESTRADO EM MATEMÁTICA

**Uma Análise Assintótica das Trajetórias em
Programação Linear com Penalidade de
Kullback-Leibler**

Sandoel de Brito Vieira

**Teresina
2016**

Sandoel de Brito Vieira

**Uma Análise Assintótica das Trajetórias em Programação Linear
com Penalidade de Kullback-Leibler**

Dissertação submetida à Coordenação do Curso de Pós-Graduação em Matemática, da Universidade Federal do Piauí, como requisito parcial para obtenção do grau de Mestre em Matemática.

Orientador:

Prof. Dr. João Xavier da Cruz Neto

Teresina

2016

FICHA CATALOGRÁFICA
Serviço de Processamento Técnico da Universidade Federal do Piauí
Biblioteca Setorial do CCN

V657a Vieira, Sandoel de Brito.
Uma análise assintótica das trajetórias em programação
linear com penalidade de Kullback-Leibler / Sandoel de
Brito Vieira. – Teresina, 2016.
61f. il.

Dissertação (Mestrado) – Universidade Federal do Piauí,
Centro de Ciências da Natureza, Pós-Graduação em
Matemática, 2016.

Orientador: Prof. Dr. João Xavier da Cruz Neto.

1. Otimização Matemática. 2. Método do Ponto Proximal.
3. Distâncias de Bregman. I. Título

CDD 519.3



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO PIAUÍ
CENTRO DE CIÊNCIAS DA NATUREZA
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA

Uma Análise Assintótica das Trajetórias em Programação Linear com Penalidade de Kullback-Leibler

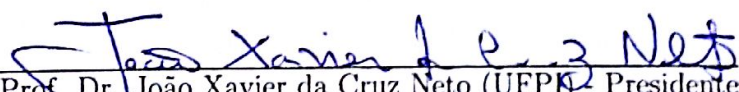
Sandoel de Brito Vieira

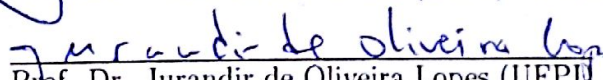
Esta Dissertação foi submetida como parte dos requisitos necessários à obtenção do grau de **Mestre em Matemática**, outorgado pela Universidade Federal do Piauí.

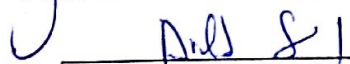
A citação de qualquer trecho deste trabalho é permitida, desde que seja feita em conformidade com as normas da ética científica.

Dissertação aprovada em 01 de Março de 2016.

Banca Examinadora:


Prof. Dr. João Xavier da Cruz Neto (UFPI) Presidente


Prof. Dr. Jurandir de Oliveira Lopes (UFPI)


Prof. Dr. Arnaldo Silva Brito (UESPI)

*Dedico este trabalho aos meus queridos
pais: Francisca (Fransquinha) e Manoel
(Pequeno), sem eles eu nada seria.*

Agradecimentos

Agradeço em primeiro lugar a Deus pela dádiva da vida.

Agradeço aos meus pais Francisca e Manoel pelo imenso amor e carinho compartilhados e por sempre terem sido grandes incentivadores em todas as fases de minha vida, apoiando-me em todos os momentos (Meus heróis!).

Agradeço aos meus irmãos Samara, Augusto e Sylmara pela nossa grande amizade e pelo companheirismo que completa o meu ser (Um por todos e todos por um!).

Agradeço ao professor Antônio Amaral (Amaral) por ter sido o mentor do meus interesses para com a matemática, com o qual aprendi muito, e por ter sido um amigo presente em todas as situações, sobretudo quando precisei de apoio.

Agradeço ao professor João Xavier pelo acompanhamento prestado desde o PIC até a orientação durante toda a minha estadia na UFPI. E pelas conversas que tivemos em sua sala, das quais pude tirar boas lições de vida.

Agradeço ao departamento de matemática da UFPI por ter acolhido-me de modo tão caloroso, em especial aos professores: Jurandir (um grande amigo, desde a época do PIC), Barnabé, Paulo Alexandre, Newton, Juscelino, Marcos Vinícius, Paulo Sérgio, Cícero, Humberto, Alex Marinho, Liane, Benício, Gilvan e Mário.

Aqui gostaria de agradecer a todos as amizades construídas ao longo da minha trajetória, desde os amigos da infância, aos que conheci na universidade, em cursos de verões ou no dia a dia longe de casa. Não me arriscarei a citar nomes para que não seja infortunadamente traído pela minha memória, mas gostaria de salientar que todos foram, da sua maneira, muito importantes na minha formação como pessoa. Embora nascidos de mães diferentes alguns tornaram-se irmãos que levarei para vida toda.

Agradeço ao professores Arnaldo e Jurandir por comporem a banca.

Por fim, agradeço ao CNPQ pelo apoio financeiro.

“O mundo não é um mar de rosas; é um lugar sujo, um lugar cruel, que não quer saber o quanto você é durão. Vai botar você de joelhos e você vai ficar de joelhos para sempre se você deixar. Você, eu, ninguém vai bater tão forte como a vida, mas não se trata de bater forte. Se trata de quanto você aguenta apanhar e seguir em frente, o quanto você é capaz de aguentar e continuar tentando. É assim que se consegue vencer.”

Rocky Balboa

Resumo

Nesta dissertação, consideramos o método do ponto proximal com a divergência de Kullback-Leibler aplicado a problemas de otimização linear, estudamos a sequência proximal primal (SPP) e a sequência proximal dual (SPD). Por meio de resultados obtidos para o comportamento limite das trajetórias primais e duais mostramos que a (SPP) converge para uma solução do problema linear e analisamos seu raio de convergência. Bem como provamos que a (SPD) e a sequência média dual (SMD) convergem para o centróide do conjunto ótimo dual do problema linear. E ainda, estudamos o raio de convergência da (SMD).

Palavras-chave: Método do Ponto Proximal Generalizado, Funções Barreiras, Divergência de Kullback-Leibler, Distâncias de Bregman, Centróide, Trajetórias Dual e Primal, Convergência da Sequência Dual, Raio de Convergência Primal.

Abstract

In this dissertation we consider the Proximal Point Method with the Kullback-Leibler divergence applied to linear optimization problems, study the Primal Proximal Sequence (PPS) and the Dual Proximal Sequence (DPS). With results obtained for the limiting behavior of primal and dual trajectories we show that (PPS) converged to a solution for the linear problem and analyze its radius of convergence. We prove, as well, that the (DPS) and the Dual Average Sequence (DAS) converge to the centroid of the dual optimal set of the linear problem. Also, we study the radius of convergence of the (DAS).

Keywords: Generalized Proximal Point Method, Barrier Functions, Kullback-Leibler Divergence, Bregman Distances, Centroid, Dual and Primal Trajectories, convergence of dual sequence, primal convergence rate.

Sumário

Resumo	iv
Abstract	v
1 Resultados preliminares	3
1.1 Notações e Convenções	3
1.2 Resultados de Álgebra Linear e Análise no \mathbb{R}^n	4
1.3 Otimização e Funções Convexas	7
1.4 Funções e Distâncias Generalizadas de Bregman	13
1.5 Condições de Otimalidade de Karush-Kuhn-Tucker	16
1.6 Dualidade em Programação Linear	20
2 Trajetórias Primal e Dual: Definições e Convergências	24
2.1 Trajetórias Primal e Dual	24
2.2 Suavidade das Trajetórias Primal e Dual	28
2.3 Centróide de S^*	29
2.4 Convergência das Trajetórias Primal e Dual	32
3 Comportamento limite das derivadas das trajetórias	38
3.1 Comportamento Limite de $\Delta s(\mu)$, $\Delta u(\mu)$ e $\dot{s}(\mu)$	38
3.2 Comportamento Limite de $\Delta x(\mu)$	46
4 Método do ponto proximal com distância de Bregman	50
4.1 Sequências Proximais Primal e Dual	50
4.2 Resultados de Convergência	53
4.3 Análise de Convergência	56

Introdução

O algoritmo do ponto proximal com distâncias generalizadas para resolver problemas lineares do tipo:

$$\min \{f(\mathbf{x}) : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq 0\}, \quad (1)$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é uma função convexa diferenciável, \mathbf{A} é uma $m \times n$ matriz real, \mathbf{b} é um m -vetor real e a variável \mathbf{x} é um n -vetor real, gera a *sequência proximal primal* $\{\mathbf{x}^k\}$ de acordo com a interação:

$$\mathbf{x}^{k+1} \equiv \operatorname{argmin}\{f(\mathbf{x}) + \lambda_k D_\varphi(\mathbf{x}, \mathbf{x}^k) : \mathbf{A}\mathbf{x} = \mathbf{b}\}, \quad (2)$$

com $\lambda^0 > 0$ arbitrário, D_φ é uma distância de Bregman e $\{\lambda_k\}$ é uma sequência de escalares positivos satisfazendo $\sum_{k=0}^{\infty} \lambda_k^{-1} = \infty$. As condições de otimalidade para (2) determinam a *sequência proximal dual* $\{\mathbf{s}^k\}$ definida por:

$$\mathbf{s}^k \equiv \lambda_k \nabla D_\varphi(\mathbf{x}^{k+1}, \mathbf{x}^k).$$

Em todo este trabalho usaremos uma distância de Bregman específica, a divergência de Kullback-Leibler:

$$D_\varphi(\mathbf{x}, \tilde{\mathbf{x}}) = \sum_{i=1}^n \left(x_i \log \frac{x_i}{\tilde{x}_i} + \tilde{x}_i - x_i \right), \quad (3)$$

associada a φ , função de entropia, dada por $\varphi(\mathbf{x}) = \sum_{i=1}^n x_i \log x_i$. Os resultados centrais a serem expostos foram baseados no artigo de Cominetti e San Martín [4].

Na literatura, encontra-se um estudo completo sobre a sequência proximal primal, embora pouco tenha-se debatido sobre a sequência proximal dual $\{\mathbf{s}^k\}$. Em geral, ao invés de estudar diretamente esta sequência dual, estuda-se o comportamento da *sequência média dual* $\{\bar{\mathbf{s}}^k\}$, definida a partir de $\{\mathbf{s}^k\}$ como sendo:

$$\bar{\mathbf{s}}^k = \sum_{i=1}^k \mu_k \lambda_i^{-1} \mathbf{s}^i, \quad \text{onde } \mu_k = \left(\sum_{i=1}^k \lambda_i^{-1} \right)^{-1}.$$

Nesta dissertação, para a função barreira particular, será demonstrada a convergência de ambas estas sequências para o centróide (\mathbf{s}^c) do conjunto de folgas duais (\mathbf{S}^*).

Com o intuito de analisar a convergência e a velocidade com que convergem as sequências proximais primal, média dual e dual será feito um estudo sobre a *trajetória primal* associada ao problema linear (1) perturbado, de modo que para cada parâmetro $\mu > 0$ será definida como sendo:

$$\mathbf{x}(\mu) = \operatorname{argmin} \{ \mathbf{c}^T \mathbf{x} + \mu D_\varphi^1(\mathbf{x}) : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq 0 \}, \quad (4)$$

onde $D_\varphi^1(\mathbf{x}) = D_\varphi(\mathbf{x}, \mathbf{x}^1)$, com $\mathbf{x}^1 > 0$, $\mathbf{A}\mathbf{x}^1 = \mathbf{b}$ e D_φ dada por (3). Com o mesmo objetivo será estudado a *trajetória dual*, cuja definição é motivada pelas condições de otimalidade de (4) e dada por:

$$\mathbf{s}(\mu) \equiv -\mu(\nabla\varphi(\mathbf{x}(\mu)) - \nabla\varphi(\mathbf{x}^1)) = -\mu(\log \mathbf{x}(\mu) - \log \mathbf{x}^1).$$

Como mostrado no trabalho de Cruz Neto, Iusem e Svaiter [9], as sequências proximal primal e média dual estão contidas, respectivamente, nas trajetória primal e dual de uma maneira conveniente. Assim sendo, estudaremos o comportamento limite das trajetórias primal e dual, de suas derivadas e também dos quocientes $\Delta\mathbf{s}(\mu)$, $\Delta\mathbf{u}(\mu)$ e $\Delta\mathbf{x}(\mu)$ (definidos *a posteriori*), para depois utilizar estes resultados no estudo das sequências obtidas pelo método do ponto proximal generalizado em questão.

Este trabalho está dividido em quatro capítulos. O primeiro deles conterá definições e resultados sobre álgebra linear, análise no \mathbb{R}^n , convexidade e otimização que darão suporte às demonstrações ao longo do texto. No segundo Capítulo, introduziremos os conceitos de trajetórias primal e dual, bem como mostraremos que estas convergem para pontos particulares. No terceiro Capítulo, estudaremos o comportamento limite das derivadas destas trajetórias e de alguns quocientes importantes para o último capítulo. No quarto Capítulo, fazendo uso dos resultados obtidos nos Capítulos 2 e 3 e da boa conexão estabelecida em Cruz Neto, Iusem e Svaiter [9], já comentada acima, mostraremos a convergência das sequências proximais primal, média dual e dual, além disso analisaremos as velocidades de convergência de algumas delas.

Capítulo 1

Resultados preliminares

Neste capítulo apresentaremos algumas definições e resultados preliminares que darão suporte as demonstrações no transcorrer desta dissertação. Para alguns resultados apresentaremos provas e para outros indicaremos apenas as referências usadas, principalmente para aqueles cujas provas são muito técnicas e fogem do escopo do trabalho.

1.1 Notações e Convenções

As seguintes notações estabelecidas serão usadas ao longo de todo o texto. A partir de agora \mathbb{R}^n será o espaço Euclidiano n -dimensional, onde a i -ésima componente do vetor $\mathbf{x} \in \mathbb{R}^n$ será o número real denotado por x_i . Seja $\mathbb{R}_+^n \equiv \{\mathbf{x} \in \mathbb{R}^n : x_i \geq 0, \forall i = 1, \dots, n\}$ e $\mathbb{R}_{++}^n \equiv \{\mathbf{x} \in \mathbb{R}^n : x_i > 0, \forall i = 1, \dots, n\}$. A menos de menção contrária, o \mathbb{R}^n sempre será munido com o produto interno canônico, a saber, para $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ temos $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y} \equiv x_1 y_1 + \dots + x_n y_n$. E com a norma Euclidiana, definida por $\|\mathbf{x}\| \equiv \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$. Dado $S \subset \mathbb{R}^n$ indicaremos por \bar{S} , o fecho de S . Dado um conjunto de índices $J = \{j_1, \dots, j_r\} \subset \{1, \dots, n\}$, indicaremos a cardinalidade de J por $|J| = r$ e $\bar{J} = \{1, \dots, n\} \setminus J$. Mais ainda, dado $\mathbf{x} \in \mathbb{R}^n$ definimos \mathbf{x}_J como sendo o subvetor $([x_i]_{i \in J}) = (x_{j_1}, \dots, x_{j_r}) \in \mathbb{R}^{|J|}$.

No que diz respeito a notação matricial, o conjunto das matrizes reais de ordem $m \times n$ será denotado por $\mathbb{R}^{m \times n}$, cujos elementos serão escritos da forma $\mathbf{A} = (\mathbf{a}_{ij})_{i,j} \in \mathbb{R}^{m \times n}$ (aqui faremos a identificação $\mathbb{R}^n \simeq \mathbb{R}^{n \times 1}$). A matriz transposta de $\mathbf{A} \in \mathbb{R}^{m \times n}$ será denotado por $\mathbf{A}^T \in \mathbb{R}^{n \times m}$. E quando existir, a matriz inversa de $\mathbf{B} \in \mathbb{R}^{n \times n}$ será representada por $\mathbf{B}^{-1} \in \mathbb{R}^{n \times n}$. Dado $\mathbf{A} \in \mathbb{R}^{m \times n}$ e $J \subset \{1, \dots, n\}$ denotaremos por \mathbf{A}_J a matriz em $\mathbb{R}^{m \times |J|}$ formada pelas colunas de \mathbf{A} indexadas por J . Daí, podemos escrever

$$A = [A_j \ A_{\bar{j}}].$$

Por fim, sendo $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ e $\mathbf{u} \in \mathbb{R}_{++}^n$ usaremos as seguintes razoáveis notações:

- $\mathbf{xy} \equiv (x_1 y_1, \dots, x_n y_n)^T \in \mathbb{R}^n$;
- $\mathbf{x} \geq \mathbf{y} \Leftrightarrow x_i \geq y_i, \forall i = 1, \dots, n$;
- $\mathbf{e} \equiv (1, \dots, 1)^T \in \mathbb{R}^n$;
- $\log \mathbf{u} \equiv (\log u_1, \dots, \log u_n)^T, \mathbf{e}^{\mathbf{x}} \equiv (e^{x_1}, \dots, e^{x_n})^T \in \mathbb{R}^n$;
- $\mathbf{X} = \text{Diag}(x_1, \dots, x_n) \in \mathbb{R}^{n \times n}$.

1.2 Resultados de Álgebra Linear e Análise no \mathbb{R}^n

Neste trabalho assumiremos uma certa familiaridade com os conceitos básicos de análise no \mathbb{R}^n e álgebra linear, os quais em sua maioria podem ser encontrados em [14], [15] e [16]. Sendo assim, as ferramentas que faremos menção explícita nesta parte são apenas aquelas que serão muito utilizadas posteriormente.

Proposição 1.1. *Todo subespaço vetorial $E \subset \mathbb{R}^n$ é fechado.*

Demonstração. Tome uma base ortonormal $\mathbf{G} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$ de \mathbb{R}^n , de modo que $\mathbf{G}' = \{\mathbf{w}_1, \dots, \mathbf{w}_r\}$, com $r \leq n$, seja uma base de E (o que é possível pelo processo de ortogonalização de Gram-Schmidt, conforme seção 10 de [16]). Dada uma sequência (\mathbf{x}^k) dada por $\mathbf{x}^k = \sum_{i=1}^r \alpha_i^k \mathbf{w}_i$, de modo que $\mathbf{x}^k \rightarrow \mathbf{x} \in \mathbb{R}^n$ com $\mathbf{x} \equiv \sum_{i=1}^n \alpha_i \mathbf{w}_i$. Daí,

$$\begin{aligned} 0 &= \lim_{k \rightarrow \infty} \|\mathbf{x}^k - \mathbf{x}\|^2 = \lim_{k \rightarrow \infty} \left\langle \sum_{i=1}^r \alpha_i^k \mathbf{w}_i - \sum_{i=1}^n \alpha_i \mathbf{w}_i, \sum_{i=1}^r \alpha_i^k \mathbf{w}_i - \sum_{i=1}^n \alpha_i \mathbf{w}_i \right\rangle \\ &= \lim_{k \rightarrow \infty} \sum_{i=1}^r (\alpha_i^k - \alpha_i)^2 + \alpha_{r+1}^2 + \dots + \alpha_n^2. \end{aligned}$$

Logo, $\alpha_{r+1} = \dots = \alpha_n = 0$ e então $\mathbf{x} = \left(\sum_{i=1}^r \alpha_i \mathbf{w}_i \right) \in E$. Portanto, E é fechado. \square

Corolário 1.1. *Seja E um subespaço vetorial do \mathbb{R}^n . Se $\Gamma : (a, b) \subset \mathbb{R} \rightarrow \mathbb{R}^n$ é um caminho diferenciável que tem imagem contida em $\mathbf{v} + E$, para algum $\mathbf{v} \in \mathbb{R}^n$, então $\Gamma'(\mu) \in E, \forall \mu \in (a, b)$.*

Demonstração. Dado $\mu \in (a, b)$ e $h > 0$ tal que $\mu + h \in (a, b)$, como E é subespaço vetorial temos

$$\left(\frac{\Gamma(\mu + h) - \Gamma(\mu)}{h} \right) \in E.$$

Fazendo $h \rightarrow 0$, obtemos pelo teorema anterior que $\Gamma'(\mu) \in E$. \square

Definição 1.1. *Seja $S \subset \mathbb{R}^n$ não vazio. O complementar ortogonal de S , denotado por S^\perp , é definido como sendo o conjunto*

$$S^\perp = \{x \in \mathbb{R}^n : \langle x, y \rangle = 0, \forall y \in S\}.$$

Definição 1.2. *Seja $A \in \mathbb{R}^{m \times n}$, define-se:*

- i) o núcleo de A , denotado por $\text{Null } A$, como sendo $\text{Null } A = \{x \in \mathbb{R}^n : Ax = 0\}$;*
- ii) a imagem de A , denotada por $\text{Im } A$, como sendo $\text{Im } A = \{Ax \in \mathbb{R}^m : x \in \mathbb{R}^n\}$;*
- iii) o posto de A , denotado por $\text{posto}(A)$, como sendo o número máximo de colunas de A linearmente independentes(LI)(que é igual ao número de linhas LI, conforme Teorema 8.2 de [16]).*

Pela definição acima, $\text{Null } A$ e $\text{Im } A^T$ são subespaços vetoriais de \mathbb{R}^n . E mais ainda, estes estão bem relacionados, como mostra a proposição abaixo.

Proposição 1.2. *Seja $A \in \mathbb{R}^{m \times n}$, então $\text{Null } A = (\text{Im } A^T)^\perp$ e $(\text{Null } A)^\perp = \text{Im } A^T$.*

Demonstração. Dado $x \in (\text{Im } A^T)^\perp$, temos $(Ax)^T y = x^T (A^T y) = 0$, para todo $y \in \mathbb{R}^m$. Logo, $Ax = 0$ e daí $x \in \text{Null } A$. Reciprocamente, se $x \in \text{Null } A$ e dado $y \in \mathbb{R}^m$, vale que $x^T (A^T y) = (Ax)^T y = 0$. Então, $x \in (\text{Im } A^T)^\perp$. Portanto, $\text{Null } A = (\text{Im } A^T)^\perp$. A outra igualdade segue da primeira juntamente com o fato que $(E^\perp)^\perp = E$, sempre que E for subespaço vetorial do \mathbb{R}^n (ver [16], Corolário 1 da seção 11). \square

Os dois lemas seguintes garantem a invertibilidade de certas matrizes, e serão utilizados como ferramentas técnicas nos próximos capítulos.

Lema 1.1. *Seja $A \in \mathbb{R}^{m \times n}$, com $m \leq n$ e $\text{posto}(A) = m$. Então, $AA^T \in \mathbb{R}^{m \times m}$ é inversível.*

Demonstração. Pelo Teorema do Núcleo e da Imagem (Ver [16], Teorema 6.6) é suficiente mostrarmos que $\text{Null}(AA^T) = 0$, pois neste caso $\text{Im}(AA^T) = \mathbb{R}^m$ e então AA^T é invertível. Dado $u \in \text{Null}(AA^T)$, temos que:

$$\|A^T u\| = (A^T u)^T (A^T u) = (u^T A)(A^T u) = u^T (AA^T u) = 0.$$

Portanto, $A^T \mathbf{u} = 0$. Como $\text{posto}(A^T) = \text{posto}(A) = m$ segue que A^T é injetiva. Daí, $\mathbf{u} = 0$ e segue o resultado. \square

Lema 1.2. *Sejam $D = \text{diag}(d_1, \dots, d_n)$ e $A \in \mathbb{R}^{m \times n}$, onde $\mathbf{d} \in \mathbb{R}_{++}^n$. Se $m \geq n$ e $\text{posto}(A) = \text{posto}(A^T) = m$, então é inversível a seguinte matriz*

$$M \equiv \begin{pmatrix} A^T & D \\ 0 & A \end{pmatrix} \in \mathbb{R}^{(m+n) \times (m+n)}.$$

Demonstração. Pelo Teorema do Núcleo e da Imagem é suficiente mostrarmos que $\text{Null } M = 0$, pois neste caso $\text{Im } M = \mathbb{R}^{m+n}$ e então, M é invertível. Dado $(\mathbf{u}, \mathbf{v})^T \in \text{Null } M \subset \mathbb{R}^m \times \mathbb{R}^n$, temos que:

$$\begin{pmatrix} A^T & D \\ 0 & A \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Daí,

$$\begin{cases} A^T \mathbf{u} + D\mathbf{v} = 0 \\ A\mathbf{v} = 0 \end{cases}.$$

Logo, $\mathbf{v} = -D^{-1}A^T \mathbf{u}$ e daí, $0 = A\mathbf{v} = -AD^{-1}A^T \mathbf{u}$. Sendo assim,

$$\|D^{-1/2}A^T \mathbf{u}\|^2 = (D^{-1/2}A^T \mathbf{u})^T (D^{-1/2}A^T \mathbf{u}) = \mathbf{u}^T (AD^{-1}A^T \mathbf{u}) = 0.$$

Portanto, $D^{-1/2}A^T \mathbf{u} = 0$ e multiplicando a esquerda por $D^{1/2}$ temos $A^T \mathbf{u} = 0$. Como $\text{posto}(A^T) = m$, temos que $\text{Null } A^T = 0$, então $\mathbf{u} = 0$ e $\mathbf{v} = -D^{-1/2}A^T \mathbf{u} = 0$. Desta maneira, concluímos que $(\mathbf{u}, \mathbf{v})^T = (0, 0)^T \in \mathbb{R}^m \times \mathbb{R}^n$ e segue o resultado. \square

Os únicos teoremas de análise que aqui destacaremos são o *Teorema do Valor Médio de Cauchy*, *Teorema de Bolzano-Weierstrass* e o *Teorema da Função Implícita*, os quais serão expostos apenas com referências para suas demonstrações.

Teorema 1.1. (Teorema do Valor Médio de Cauchy)

Sejam $f, g : [a, b] \rightarrow \mathbb{R}$ contínuas em $[a, b]$ e deriváveis em (a, b) . Então, existe $c \in (a, b)$ tal que:

$$f'(c)[g(b) - g(a)] = g'(c)[f(b) - f(a)].$$

Demonstração. Vide Teorema 19.8 de [1] \square

Teorema 1.2. (Teorema de Bolzano-Weierstrass)

Toda sequência limitada em \mathbb{R}^n possui uma subsequência convergente.

Demonstração. Ver [14], mais precisamente o Teorema 3 do Capítulo 1. \square

Teorema 1.3. (Teorema da Função Implícita)

Seja $f = (f_1, \dots, f_N) : U \rightarrow \mathbb{R}^N$ de classe C^k ($k \geq 1$) no aberto $U \subset \mathbb{R}^{M+N}$. Suponhamos que no ponto $p = (a, b) \in U$, com $f(p) = c$, a matriz de ordem $N \times N$ dada por:

$$J_2 f(p) \equiv \left[\frac{\partial f_i}{\partial y_j}(p) \right] \quad (i, j = 1, \dots, N)$$

seja inversível. Então existem $Z \subset U$ um aberto contendo p , $V \subset \mathbb{R}^M$ um aberto contendo a e $\xi : V \rightarrow \mathbb{R}^N$ de classe C^k , com $\xi(a) = b$ e valendo a seguinte propriedade:

$$f^{-1}(c) \cap Z = \{(x, \xi(x)) : x \in V\}.$$

Demonstração. Ver [14], mais precisamente o Teorema 6 do capítulo 6. \square

1.3 Otimização e Funções Convexas

Apresentaremos agora algumas definições e resultados sobre otimização, com maior destaque àqueles obtidos para funções convexas, para as quais será feito um estudo mais detalhado. As referências aqui usadas foram Solodov e Izmailov[10] e Ribeiro e Karas[18].

Seja $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ uma função, um problema é dito ser de minimização quando é da forma:

$$\min\{f(x) : x \in D\}. \tag{1.1}$$

O conjunto D é chamado de *conjunto viável* do problema, os pontos de D são chamados de *pontos viáveis* e f é chamada de *objetivo*. Mais ainda, $\bar{v} \in [-\infty, \infty)$ definido por $\bar{v} = \inf\{f(x) : x \in D\}$ recebe o nome de *valor ótimo* do problema (1.1).

Definição 1.3. Dizemos que um ponto $\bar{x} \in D$ é:

- i) *minimizador global* de (1.1) se, e só se, $f(\bar{x}) \leq f(x), \forall x \in D$;
- ii) *minimizador local* de (1.1) se, e só se, existe $\varepsilon > 0$ tal que

$$f(\bar{x}) \leq f(x), \forall x \in D \cap B(\bar{x}, \varepsilon).$$

O conjunto $D^* = \{\bar{x} \in D : f(\bar{x}) \leq f(x), \forall x \in D\}$ é chamado de *conjunto solução* do problema (1.1). Em alguns casos, mesmo quando \bar{v} é finito, pode não haver um minimizador global para o problema, isto é, $D^* = \emptyset$. Considere, por exemplo, o problema: $\min\{e^x : x \in \mathbb{R}\}$. A seguir, serão expostos critérios que garantem a existência de solução global.

Teorema 1.4. (Teorema de Weierstrass)

Sejam $D \subset \mathbb{R}^n$ um conjunto compacto e $f : D \rightarrow \mathbb{R}$ uma função contínua. Então f assume máximo e mínimo em D . Em particular, o problema (1.1) possui solução global.

Demonstração. Ver Teorema 1.2.1 em [10]. □

Definição 1.4. O conjunto de nível da função $f : D \rightarrow \mathbb{R}$ associado a $\beta \in \mathbb{R}$ é o conjunto dado por

$$L_{f,D}(\beta) = \{x \in D : f(x) \leq \beta\}.$$

Corolário 1.2. Sejam $D \subset \mathbb{R}^n$ um conjunto e $f : D \rightarrow \mathbb{R}$ uma função contínua. Supondo que existe β real tal que o conjunto $L_{f,D}(\beta)$ seja não-vazio e compacto. Então, o problema (1.1) possui solução global.

Demonstração. Usando o Teorema 1.4, o problema

$$\min\{f(x) : x \in L_{f,D}(\beta)\}$$

tem solução global \bar{x} , isto é, $f(\bar{x}) \leq f(x)$, para todo $x \in L_{f,D}(\beta)$. E dado $x \in D/L_{f,D}(\beta)$, temos que $f(x) > \beta \geq f(\bar{x})$. Portanto, \bar{x} é minimizador global de f em D . □

O seguinte lema será crucial em várias demonstrações do Capítulo 3.

Lema 1.3. (Hoffman) Sejam E um subespaço de \mathbb{R}^n , J um conjunto de índices contido em $I_n = \{1, \dots, n\}$ e $\bar{J} = I_n/J$. Então, existe uma constante $M = M(E, J)$ com a seguinte propriedade: para cada u_J e $u_{\bar{J}}$ tais que $(u_J, u_{\bar{J}})^T \in E$, existe $\tilde{u}_{\bar{J}}$ tal que $(u_J, \tilde{u}_{\bar{J}})^T \in E$ e $\|\tilde{u}_{\bar{J}}\| \leq M\|u_J\|$.

Demonstração. Reordenando as coordenadas, sem perda de generalidade, vamos supor $J = \{1, \dots, m\}$ com $m < n$. A priori, encontraremos a projeção de E sobre o \mathbb{R}^m (aqui estamos considerando $\mathbb{R}^n = \mathbb{R}^m \oplus \mathbb{R}^{n-m}$). Sendo $P_J : \mathbb{R}^n \rightarrow \mathbb{R}^m$ definida por $P_J(x) = P_J(x_J, x_{\bar{J}}) = x_J$, note que $P_J(E)$ é um subespaço vetorial de \mathbb{R}^m , logo é fechado. Portanto, $E_p \equiv P_J(E) \cup B[0, 1]$ é compacto. Seja $u \in E_p$, então $F_u = (\{u\} \times \mathbb{R}^{n-m}) \cap E$ é não-vazio e fechado. E definindo $h_u : F_u \rightarrow \mathbb{R}$ por $h_u(u, v) = \|v\|$. Pelo Corolário 1.2, a função $g : E_p \rightarrow \mathbb{R}$ dada por $g(u) = \min\{h_u(u, v) : (u, v)^T \in F_u\}$ é bem-definida. Mais ainda, existe $(u, v_u)^T \in F_u$ tal que $g(u) = \|v_u\|$. Agora, note que F_u varia continuamente com u , já que a interseção de $\{u\} \times \mathbb{R}^{n-m}$ com E varia continuamente com u e daí, g é contínua em E_p . Então, pelo Teorema 1.4 existe $M = \max\{g(u) : u \in E_p\} \geq 0$. Afirmamos que M resolve o lema. De fato, dado $u = (u_J, u_{\bar{J}})^T \in E$:

- Se $\mathbf{u}_J = 0$, então podemos tomar $\tilde{\mathbf{u}}_J = 0$, pois $(0, 0)^T \in E$ e $\|0\| \leq M\|0\|$.
- Se $\mathbf{u}_J \neq 0$, considere $\mathbf{u} = \frac{\mathbf{u}_J}{\|\mathbf{u}_J\|} \in E_P$. Pela construção acima, existe \mathbf{v}_u tal que $(\mathbf{u}, \mathbf{v}_u)^T \in E$ e $\|\mathbf{v}_u\| \leq M$. Portanto, sendo $\tilde{\mathbf{u}}_J = \|\mathbf{u}_J\|\mathbf{v}_u$, temos:

$$(\mathbf{u}_J, \tilde{\mathbf{u}}_J)^T = \|\mathbf{u}_J\|(\mathbf{u}, \mathbf{v}_u)^T = (\mathbf{u}_J, \|\mathbf{u}_J\|\mathbf{v}_u)^T \in E$$

é o vetor procurado, pois $\|\tilde{\mathbf{u}}_J\| = \|\|\mathbf{u}_J\|\mathbf{v}_u\| = \|\mathbf{u}_J\|\|\mathbf{v}_u\| \leq M\|\mathbf{u}_J\|$.

□

Observação 1.1. No lema acima a constante M pode ser tomada independente da escolha de J , a saber, é suficiente tomar $M = M(E) = \max_J M(E, J)$. Esta constante M é bem definida pois existem finitas possibilidades para a escolha de $J \subset \{1, \dots, n\}$.

A seguir apresentaremos um resultado que, sob certas hipóteses, dá condições (as chamadas *condições necessárias de otimalidade*) a serem satisfeitas quando \bar{x} é minimizador local do problema (1.1).

Teorema 1.5. Seja a função $f : D \rightarrow \mathbb{R}$, onde $D \subset \mathbb{R}^n$ é um conjunto aberto. Se \bar{x} é um minimizador local de f em D e f é diferenciável em \bar{x} , então $\nabla f(\bar{x}) = 0$.

Demonstração. Vide Teorema 1.3.1 de [10].

□

Agora, apresentaremos a definição de conjuntos convexos. A saber, um conjunto convexo se caracteriza por conter todos os segmentos cujos extremos pertencem ao mesmo. Mais precisamente:

Definição 1.5. Um conjunto $D \subset \mathbb{R}^n$ é dito convexo se, e só se, para quaisquer $\mathbf{x}, \mathbf{y} \in D$ e $\alpha \in [0, 1]$, temos $\alpha\mathbf{x} + (1 - \alpha)\mathbf{y} \in D$.

Exemplo 1.1. Sendo dados $A \in \mathbb{R}^{m \times n}$, $\mathbf{a} \in \mathbb{R}^n$ e $\mathbf{b} \in \mathbb{R}^m$ os conjuntos \mathbb{R}^n , \mathbb{R}_+^n , $B(\mathbf{a}, r) = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{a}\| \leq r\}$ e $L = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{b}\}$ são convexos. Como exemplo de conjunto não-convexo, temos o conjunto da direita na Figura 1.1.

Proposição 1.3. Sejam $D_j \subset \mathbb{R}^n$ conjuntos convexos, para todo $j \in J$. Então, $D = \bigcap_{j \in J} D_j$ é um conjunto convexo.

Demonstração. Sejam \mathbf{x} e \mathbf{y} em D , logo $\mathbf{x} \in D_j$ e $\mathbf{y} \in D_j$, para todo $j \in J$. Dado $\alpha \in [0, 1]$, como cada D_j é convexo, temos $\alpha\mathbf{x} + (1 - \alpha)\mathbf{y} \in D_j$, para todo $j \in J$. Logo, $\alpha\mathbf{x} + (1 - \alpha)\mathbf{y} \in D$. Portanto, D é convexo.

□

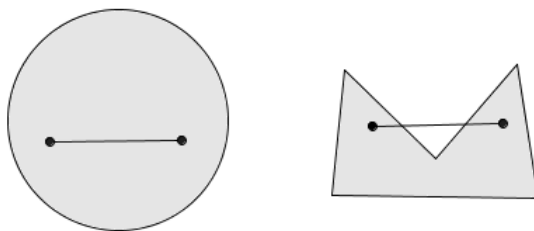


Figura 1.1: Conjuntos convexo(à esquerda) e não convexo(à direita)

Como aplicação, vemos que $D = \{x \in \mathbb{R}^n : Ax = b, x \geq 0\} = \mathbb{R}_+^n \cap \{x \in \mathbb{R}^n : Ax = b\}$ é convexo.

A seguir, introduziremos o conceito de funções convexas. Esta classe de funções tem grande importância no contexto de otimização, por exemplo, pelo fato de que condições necessárias de otimalidade tornarem-se suficientes.

Definição 1.6. *Seja $D \subset \mathbb{R}^n$ um conjunto convexo. Dizemos que uma função $f : D \rightarrow \mathbb{R}$ é convexa em D quando*

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y), \tag{1.2}$$

para todos $x, y \in D$ e $\alpha \in [0, 1]$.

A função f é chamada estritamente convexa quando a desigualdade (1.2) for estrita para todos $x \neq y$ e $\alpha \in (0, 1)$.

Veja na Figura 1.2 uma função convexa (à esquerda) e outra não convexa (à direita).

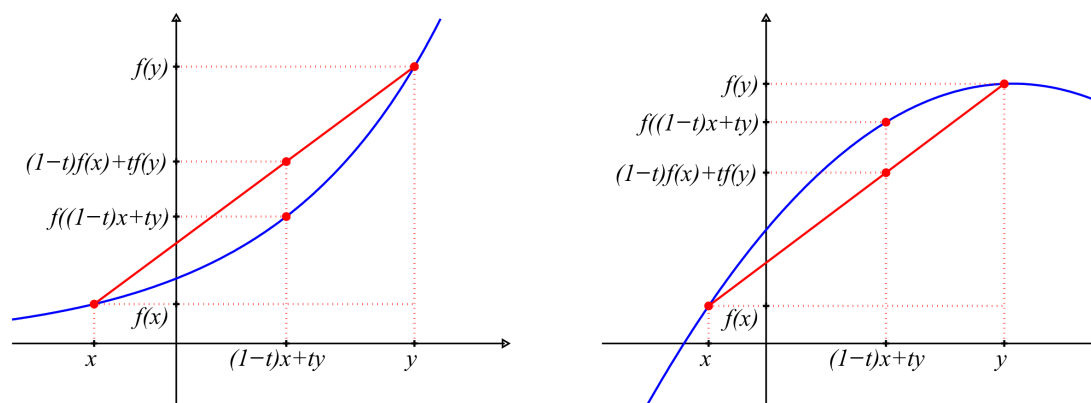


Figura 1.2: Funções convexas e não convexas

Exemplo 1.2. A função $f : \mathbb{R} \rightarrow \mathbb{R}$, dada por $f(x) = x$ é convexa, mas não é estritamente convexa. Enquanto a função $g : \mathbb{R} \rightarrow \mathbb{R}$, definida por $g(x) = x^2$ é estritamente convexa. De fato, dados $x \neq y$ e $\alpha \in (0, 1)$, temos:

$$\begin{aligned} g(\alpha x + (1 - \alpha)y) &= (y + \alpha(x - y))^2 = y^2 + 2y\alpha(x - y) + \alpha^2(x - y)^2 \\ &< y^2 + 2y\alpha(x - y) + \alpha(x - y)^2 = y^2 + \alpha(x^2 - y^2) \\ &= \alpha g(x) + (1 - \alpha)g(y). \end{aligned}$$

Em todo o resto desta seção o conjunto $D \subset \mathbb{R}^n$ será convexo. O teorema seguinte é uma justificativa para o fato de funções convexas serem muito bem vistas em otimização.

Teorema 1.6. (Teorema de minimização convexa)

Seja $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa em D . Então, todo minimizador local do problema $\min\{f(x) : x \in D\}$ é global. Além disso, o conjunto dos minimizadores é convexo. E se f é estritamente convexa, não pode haver mais de um minimizador.

Demonstração. Ver Teorema 3.1.5 em [10]. □

Geralmente, a tarefa de verificar, por definição, que funções são convexas é muito árdua. Para facilitar esta missão, a seguir serão expostos resultados que caracterizam este tipo de função.

Proposição 1.4. (i) Sejam $f_i : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$, $i = 1, \dots, p$, funções convexas em D . Então, para quaisquer reais $\mu_i > 0$, $i = 1, \dots, p$, a função $f : D \rightarrow \mathbb{R}$ definida por $f(x) = \sum_{i=1}^p \mu_i f_i(x)$ é convexa em D . Mais ainda, se alguma f_i for estritamente convexa, então f é estritamente convexa.

(ii) Seja $f : \mathbb{R}^m \rightarrow \mathbb{R}$ convexa e $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$, dada por $g(x) = Ax - b$ com $A \in \mathbb{R}^{m \times n}$ e $b \in \mathbb{R}^m$. Então, $f \circ g : \mathbb{R}^n \rightarrow \mathbb{R}$ é convexa.

Demonstração. Segue diretamente da definição de função convexa. □

A caracterização seguinte é muito útil para determinar quando funções diferenciáveis são convexas ou estritamente convexas.

Teorema 1.7. (Caracterização de funções convexas diferenciáveis)

Seja $D \subset \mathbb{R}^n$ um conjunto convexo aberto e $f : D \rightarrow \mathbb{R}$ uma função diferenciável em D . Então, as propriedades seguintes são equivalentes:

(i) A função f é convexa em D ;

(ii) Para todo $\mathbf{x} \in D$ e todo $\mathbf{y} \in D$,

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle;$$

(iii) Para todo $\mathbf{x} \in D$ e todo $\mathbf{y} \in D$,

$$\langle \nabla f(\mathbf{y}) - \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \geq 0.$$

Mais ainda, sendo f estritamente convexa estas caracterizações seguem com as desigualdades estritas.

Demonstração. Vide Teorema 3.4.7 em [10]. □

Corolário 1.3. *Seja $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ uma função diferenciável no conjunto D aberto. Se $\nabla f(\mathbf{x}) = 0$, então \mathbf{x} é minimizador global de f em D .*

Demonstração. Segue do Teorema 1.7(ii). □

Teorema 1.8. *Seja $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ uma função duas vezes diferenciável em D aberto. A função f é convexa se, e somente se, $\text{Hess}(f) \equiv f''(\mathbf{x}) \geq 0$, isto é,*

$$\langle f''(\mathbf{x})\mathbf{d}, \mathbf{d} \rangle \geq 0, \quad \forall \mathbf{x} \in D \text{ e } \forall \mathbf{d} \in \mathbb{R}^n. \quad (1.3)$$

Se a desigualdade (1.3) é estrita ($\text{Hess}(f) \equiv f''(\mathbf{x}) > 0$), então f é estritamente convexa.

Demonstração. Ver Teorema 3.15 em [18]. □

Exemplo 1.3. *As funções $f, g : \mathbb{R}^n \rightarrow \mathbb{R}$, dadas por $f(\mathbf{x}) = \|\mathbf{x}\|^2$ e $g(\mathbf{x}) = e^{\mathbf{x}}$ são estritamente convexas, já que $\text{Hess}(f) = 2I_n > 0$ e $\text{Hess}(g) = \text{Diag}(e^{x_1}, \dots, e^{x_n}) > 0$.*

Exemplo 1.4. *Seja $f : \mathbb{R}_+ \rightarrow \mathbb{R}$ definida por $f(x) = x \log x$, com a convenção de que $0 \cdot \log 0 = 0$. Como $f''(x) = 1/x > 0$, para todo $x > 0$, temos f estritamente convexa em \mathbb{R}_{++} . E sendo $t \in (0, 1)$ e $x > 0$, vale:*

$$f((1-t)0 + tx) = (tx)(\log(tx)) < (1-t)(0 \cdot \log 0) + t(x \log x).$$

Na desigualdade acima usamos que $tx < x$ e o fato da função $x \mapsto \log x$ ser crescente. Portanto, segue o resultado.

A seguir, estudaremos as propriedades de funções convexas não necessariamente diferenciáveis. Começaremos com a seguinte definição que sob um certo sentido funciona como uma derivada generalizada.

Definição 1.7. *Seja $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ convexa. Dizemos que $s \in \mathbb{R}^n$ é um subgradiente de f no ponto $x \in \mathbb{R}^n$, se*

$$f(y) \geq f(x) + \langle s, y - x \rangle, \quad \forall y \in D.$$

O conjunto de todos os subgradientes de f em x é chamado de subdiferencial de f em x e denotado por $\partial f(x)$.

Exemplo 1.5. *Seja $f : \mathbb{R} \rightarrow \mathbb{R}$, dada por $f(x) = |x|$, é possível verificar que:*

$$\partial f(x) = \begin{cases} -1, & \text{se } x < 0, \\ [-1, 1], & \text{se } x = 0, \\ 1, & \text{se } x > 0. \end{cases}$$

Neste exemplo, o ponto $\bar{x} = 0$ é ponto de mínimo global. E ainda $0 \in \partial f(0)$, mas note que $0 \notin \partial f(x)$, para todo $x \neq 0$. Mais geralmente vale a seguinte proposição:

Proposição 1.5. *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ um função convexa. O ponto $x \in \mathbb{R}^n$ é mínimo global de f se, e somente se, $0 \in \partial f(x)$.*

Demonstração. Segue da definição de subgradiente. □

1.4 Funções e Distâncias Generalizadas de Bregman

Nesta seção apresentaremos a definição e algumas propriedades de distância generalizada de Bregman, introduzidas pioneiramente por Bregman em [2]. A relevância destes conceitos para o recorrente trabalho deve-se ao fato de estas distâncias generalizadas serem utilizadas como funções barreiras para penalização de problemas lineares. A referência usada na construção desta seção foi [7].

Definição 1.8. *Seja $S \subset \mathbb{R}^n$ aberto e convexo. Considere uma função convexa $\varphi : \bar{S} \rightarrow \mathbb{R}$ e seja $D_\varphi : \bar{S} \times S \rightarrow \mathbb{R}$ dada por*

$$D_\varphi(x, y) = \varphi(x) - \varphi(y) - \langle \nabla \varphi(y), x - y \rangle. \quad (1.4)$$

Dizemos que φ é uma função de Bregman e D_φ uma distância generalizada de Bregman induzida por φ , com zona S , se valem as seguintes condições:

B1) φ é continuamente diferenciável em S ;

B2) φ é estritamente convexa e contínua em \bar{S} ;

B3) Para todo $\delta \in \mathbb{R}$ os conjuntos de níveis parciais $\Gamma_1(\mathbf{y}, \delta) \equiv \{\mathbf{x} \in \bar{S} : D_\varphi(\mathbf{x}, \mathbf{y}) \leq \delta\}$ e $\Gamma_2(\mathbf{x}, \delta) \equiv \{\mathbf{y} \in S : D_\varphi(\mathbf{x}, \mathbf{y}) \leq \delta\}$ são limitados para todo $\mathbf{y} \in S$ e para todo $\mathbf{x} \in \bar{S}$, respectivamente;

B4) Se $\{\mathbf{y}_k\} \subset S$ converge para \mathbf{y}^* , então $D_\varphi(\mathbf{y}^*, \mathbf{y}^k)$ converge a 0;

B5) Se $\{\mathbf{x}_k\} \subset \bar{S}$ e $\{\mathbf{y}_k\} \subset S$ são sequências tais que $\{\mathbf{x}_k\}$ é limitada, $\lim_{k \rightarrow \infty} \mathbf{y}^k = \mathbf{y}^*$ e $\lim_{k \rightarrow \infty} D_\varphi(\mathbf{x}^k, \mathbf{y}^k) = 0$, então $\lim_{k \rightarrow \infty} \mathbf{x}^k = \mathbf{y}^*$.

Como φ é estritamente convexa, dados $\mathbf{x} \in \bar{S}$ e $\mathbf{y} \in S$, pelo Teorema 1.7 segue que $D_\varphi(\mathbf{x}, \mathbf{y}) \geq 0$. Além disso, $D_\varphi(\mathbf{x}, \mathbf{y}) = 0$ se, e somente se, $\mathbf{x} = \mathbf{y}$. Por esta razão D_φ é chamada de distância generalizada.

Em seguida, definiremos duas importantes subclasses de funções de Bregman: fronteira coerciva e zona coerciva.

Uma função de Bregman φ é dita ser *fronteira coerciva* se:

B6) Dado $\{\mathbf{y}^k\} \subset S$ tal que $\lim_{k \rightarrow \infty} \mathbf{y}^k = \mathbf{y} \in \partial S$ (fronteira de S), então para todo $\mathbf{x} \in S$, temos $\lim_{k \rightarrow \infty} \langle \nabla \varphi(\mathbf{y}^k), \mathbf{x} - \mathbf{y}^k \rangle = -\infty$.

Uma função de Bregman φ é dita ser *zona coerciva* se:

B7) Para todo $\mathbf{y} \in \mathbb{R}^n$ existe $\mathbf{x} \in S$ tal que $\nabla \varphi(\mathbf{x}) = \mathbf{y}$.

Observação 1.2. Observe que se φ satisfaz **B6)**, temos $\lim_{k \rightarrow \infty} D_\varphi(\mathbf{x}, \mathbf{y}^k) = +\infty$, para todo $\mathbf{x} \in S$. Em outras palavras, $D_\varphi(\mathbf{x}, \mathbf{y})$ se aproxima do infinito quando \mathbf{y} se aproxima da fronteira ∂S de S , sempre que $\mathbf{x} \in S$. Este é o conceito chave da utilização de distâncias generalizadas de Bregman como barreiras em problemas de minimização convexa, o que discutiremos mais adiante.

Exemplo 1.6. Sendo $S = \mathbb{R}^n$ e $\varphi(\mathbf{x}) = \langle \mathbf{x}, \mathbf{x} \rangle$ estritamente convexa, pelo Exemplo 1.3, temos $D_\varphi(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|^2$ e as condições **B1)** a **B5)** são verificadas. Portanto, D_φ é uma distância generalizada de Bregman.

Exemplo 1.7. Seja $S = \mathbb{R}_{++}^n$ e $\varphi(\mathbf{x}) = \sum_{j=1}^n x_j \log x_j$, que é contínua em \mathbb{R}_{++}^n com a convenção de que $0 \cdot \log 0 = 0$. A convexidade estrita de φ em \bar{S} segue da definição e do

Exemplo 1.4. Em S , temos $\nabla\varphi(\mathbf{x}) = (\log \mathbf{x} + \mathbf{e})$, onde $\mathbf{e} = (1, \dots, 1)$. Daí,

$$D_\varphi(\mathbf{x}, \mathbf{y}) = \sum_{j=1}^n \left(x_j \log \frac{x_j}{y_j} + y_j - x_j \right). \quad (1.5)$$

A saber, (1.5) é chamada na literatura de *divergência de Kullback-Leibler* ou *entropia relativa* e é muito usada em *Probabilidade e Estatística*, como pode ser visto no trabalho de Eckstein [5]. Como valem as condições **B1)** a **B5)**, D_φ é uma distância generalizada de Bregman. Mais ainda, sem grandes esforços verifica-se valem **B6)** e **B7)**. Portanto, φ é zona e fronteira coerciva.

A seguinte proposição fornece propriedades que serão muito úteis no restante deste trabalho.

Proposição 1.6. *Se φ é uma função de Bregman com zona S , então:*

- (i) $D_\varphi(\mathbf{x}, \mathbf{y}) - D_\varphi(\mathbf{x}, \mathbf{z}) - D_\varphi(\mathbf{z}, \mathbf{y}) = \langle \nabla\varphi(\mathbf{y}) - \nabla\varphi(\mathbf{z}), \mathbf{z} - \mathbf{x} \rangle$, para todos $\mathbf{x} \in \bar{S}$ e $\mathbf{y}, \mathbf{z} \in S$;
- (ii) $\nabla_{\mathbf{x}} D_\varphi(\mathbf{x}, \mathbf{y}) = \nabla\varphi(\mathbf{x}) - \nabla\varphi(\mathbf{y})$, para todos $\mathbf{x}, \mathbf{y} \in S$;
- (iii) $D_\varphi(\cdot, \mathbf{y})$ é estritamente convexa, para todo $\mathbf{y} \in S$.

Demonstração. (i) Da definição de D_φ , temos:

$$D_\varphi(\mathbf{x}, \mathbf{y}) = \varphi(\mathbf{x}) - \varphi(\mathbf{y}) - \langle \nabla\varphi(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle, \quad (1.6)$$

$$D_\varphi(\mathbf{x}, \mathbf{z}) = \varphi(\mathbf{x}) - \varphi(\mathbf{z}) - \langle \nabla\varphi(\mathbf{z}), \mathbf{x} - \mathbf{z} \rangle \text{ e} \quad (1.7)$$

$$D_\varphi(\mathbf{z}, \mathbf{y}) = \varphi(\mathbf{z}) - \varphi(\mathbf{y}) - \langle \nabla\varphi(\mathbf{y}), \mathbf{z} - \mathbf{y} \rangle. \quad (1.8)$$

Subtraindo (1.7) e (1.8) de (1.6), obtemos o resultado.

(ii) Por **B1)**, o resultado segue derivando (1.4), com relação a primeira variável.

(iii) Seja $\mathbf{y} \in S$ fixo, $\mathbf{r}, \mathbf{s} \in \bar{S}$ diferentes e $t \in (0, 1)$, então por **B2)** temos:

$$\begin{aligned} D_\varphi((1-t)\mathbf{r} + t\mathbf{s}, \mathbf{y}) &= \varphi((1-t)\mathbf{r} + t\mathbf{s}) - \varphi(\mathbf{y}) - \langle \nabla\varphi(\mathbf{y}), (1-t)\mathbf{r} + t\mathbf{s} - \mathbf{y} \rangle \\ &< (1-t)\varphi(\mathbf{r}) + t\varphi(\mathbf{s}) - [(1-t)\varphi(\mathbf{y}) + t\varphi(\mathbf{y})] - \\ &\quad - (1-t)\langle \nabla\varphi(\mathbf{y}), \mathbf{r} - \mathbf{y} \rangle - t\langle \nabla\varphi(\mathbf{y}), \mathbf{s} - \mathbf{y} \rangle \\ &= (1-t)D_\varphi(\mathbf{r}, \mathbf{y}) + tD_\varphi(\mathbf{s}, \mathbf{y}). \end{aligned}$$

Portanto, $D_\varphi((1-t)\mathbf{r} + t\mathbf{s}, \mathbf{y}) < (1-t)D_\varphi(\mathbf{r}, \mathbf{y}) + tD_\varphi(\mathbf{s}, \mathbf{y})$ e segue o resultado. \square

1.5 Condições de Otimalidade de Karush-Kuhn-Tucker

Nesta seção analisaremos uma importante ferramenta utilizada em problemas de otimização, as chamadas *condições necessárias de otimalidades de Karush-Kuhn-Tucker (KKT)*. A bibliografia aqui usada como referência foi basicamente [6], [10] e [13].

Sejam $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ e $g : \mathbb{R}^n \rightarrow \mathbb{R}^l$, consideraremos o seguinte problema com restrições mistas de igualdade e desigualdade

$$\min\{f(x) : x \in D\}, \quad \text{onde} \quad (1.9)$$

$$D = \{x \in \mathbb{R}^n : h(x) = 0, g(x) \leq 0\}. \quad (1.10)$$

Para que as tais condições de otimalidade, a serem discutidas a *posteriori*, possam ser aplicadas o problema acima deve satisfazer certas condições de regularidades, conhecidas como *condições de qualificação das restrições*. Algumas condições de qualificações das restrições são:

(C1)(Ponto Regular)

Um ponto $\bar{x} \in \mathbb{R}^n$ é um *regular* quando o conjunto

$$\{\nabla h_i(\bar{x}); i = 1, \dots, m\} \cup \{\nabla g_i(\bar{x}); i \in I(\bar{x})\}$$

é linearmente independente, onde $I(\bar{x}) = \{i \in \{1, \dots, l\} : g_i(\bar{x}) = 0\}$;

(C2)(Mangasarian-Fromovitz)

Um ponto $\bar{x} \in \mathbb{R}^n$ satisfaz a *condição de regularidade de Mangasarian-Fromovitz*, quando $\{\nabla h_i(\bar{x}); i = 1, \dots, m\}$ é um conjunto linearmente independente e ainda existe $\bar{d} \in \text{Null}(h'(\bar{x}))$ tal que $\langle \nabla g_i(\bar{x}), \bar{d} \rangle < 0$ para todo $i \in I(\bar{x})$;

(C3)(Slater)

A *condição de Slater* é satisfeita quando h é uma função afim ($h(x) = Ax - b$, onde $A \in \mathbb{R}^{m \times n}$), g é uma função convexa, existe \tilde{x} tal que $h(\tilde{x}) = 0$ e $g(\tilde{x}) < 0$;

(C4)(Linearidade das Restrições)

Neste caso, h e g são funções afins, isto é, $h(x) = Ax - b$ e $g(x) = Cx - d$, onde $A \in \mathbb{R}^{m \times n}$, $C \in \mathbb{R}^{l \times n}$, $b \in \mathbb{R}^m$ e $d \in \mathbb{R}^l$.

Teorema 1.9. (Condições de Karush-Kuhn-Tucker/KKT)

Sejam $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $g : \mathbb{R}^n \rightarrow \mathbb{R}^l$ funções diferenciáveis no ponto $\bar{x} \in \mathbb{R}^n$ e $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ uma função diferenciável numa vizinhança do ponto \bar{x} , com derivada contínua neste ponto.

Se \bar{x} é um minimizador local do problema (1.9)-(1.10) e vale qualquer uma das condições de regularidade (C1) – (C4), então existem $\bar{\lambda} \in \mathbb{R}^m$ e $\bar{\eta} \in \mathbb{R}^l$ tais que:

$$\nabla f(\bar{x}) + \sum_{i=1}^m \bar{\lambda}_i \nabla h_i(\bar{x}) + \sum_{j=1}^l \bar{\eta}_j \nabla g_j(\bar{x}) = 0, \quad (1.11)$$

$$\bar{\eta}_j \geq 0, \quad j = 1, \dots, l \quad e \quad (1.12)$$

$$\bar{\eta}_j g_j(\bar{x}) = 0, \quad j = 1, \dots, l. \quad (1.13)$$

A saber, o conjunto dos pontos $(\bar{\lambda}, \bar{\eta}) \in \mathbb{R}^m \times \mathbb{R}^l$ que satisfazem (1.11), (1.12) e (1.13) é chamado o Conjunto dos Multiplicadores de Lagrange no ponto \bar{x} , e denotado por $M(\bar{x})$.

Demonstração. Ver Teorema 4.2.1 em [10]. □

Observação 1.3. Como $g(\bar{x}) \leq 0$ e $\bar{\eta} \geq 0$, temos que (1.13) é satisfeita se, e somente se, $\langle g(\bar{x}), \bar{\eta} \rangle = 0$.

Observação 1.4. A expressão (1.11) pode ser reescrita, de maneira compacta, como abaixo:

$$\nabla f(\bar{x}) + (h'(\bar{x}))^T \bar{\lambda} + (g'(\bar{x}))^T \bar{\eta} = 0.$$

No caso em que $l = 0$, o problema (1.9)-(1.10) torna-se um problema com restrições apenas de igualdade. E se vale uma das condições de regularidades (C1) ou (C4), então (1.11), (1.12) e (1.13) resumem-se a:

$$\nabla f(\bar{x}) + \sum_{i=1}^m \bar{\lambda}_i \nabla h_i(\bar{x}) = \nabla f(\bar{x}) + (h'(\bar{x}))^T \bar{\lambda} = 0, \quad \text{para algum } \bar{\lambda} \in \mathbb{R}^m.$$

Neste caso as condições dadas pelo último teorema são chamadas na literatura de *Condições de Otimalidade de Lagrange*.

Em um problema de programação convexo, as condições de otimalidade de KKT, além de necessárias, tornam-se suficientes. Como mostra a proposição abaixo.

Proposição 1.7. No problema (1.9)-(1.10), sejam f , g_i ($i = 1, \dots, l$) funções convexas e $h(x) = Ax - b$, onde $A \in \mathbb{R}^{m \times n}$ e $b \in \mathbb{R}^m$. Se $\bar{x} \in D$ é tal que $M(\bar{x}) \neq \emptyset$, isto é, se as condições de otimalidade (1.11), (1.12) e (1.13) são satisfeitas, então \bar{x} é um minimizador do problema.

Demonstração. Seja $(\bar{\lambda}, \bar{\eta}) \in M(\bar{x}) \subset \mathbb{R}^m \times \mathbb{R}_+^l$. Para todo $x \in D$, como $h(x) = 0$ e $g(x) \leq 0$, temos:

$$f(x) \geq f(\bar{x}) + \langle \bar{\lambda}, h(x) \rangle + \sum_{i=1}^l \bar{\eta}_i g_i(x). \quad (1.14)$$

Pela convexidade de f e g_i ($i = 1, \dots, l$), usando o Teorema 1.7, obtemos:

$$f(\mathbf{x}) \geq f(\bar{\mathbf{x}}) + \langle \nabla f(\bar{\mathbf{x}}), \mathbf{x} - \bar{\mathbf{x}} \rangle \quad \text{e} \quad (1.15)$$

$$g_i(\mathbf{x}) \geq g_i(\bar{\mathbf{x}}) + \langle \nabla g_i(\bar{\mathbf{x}}), \mathbf{x} - \bar{\mathbf{x}} \rangle, \quad i = 1, \dots, l. \quad (1.16)$$

E ainda,

$$\mathbf{h}(\mathbf{x}) = \mathbf{h}(\bar{\mathbf{x}}) + \mathbf{A}(\mathbf{x} - \bar{\mathbf{x}}). \quad (1.17)$$

De (1.15), (1.16) e (1.17) em (1.14), usando (1.11) e (1.13) temos:

$$\begin{aligned} f(\mathbf{x}) &\geq f(\bar{\mathbf{x}}) + \langle \nabla f(\bar{\mathbf{x}}), \mathbf{x} - \bar{\mathbf{x}} \rangle + \langle \bar{\boldsymbol{\lambda}}, \mathbf{h}(\bar{\mathbf{x}}) + \mathbf{A}(\mathbf{x} - \bar{\mathbf{x}}) \rangle \\ &+ \sum_{i=1}^l \bar{\eta}_i g_i(\bar{\mathbf{x}}) + \sum_{i=1}^l \langle \bar{\eta}_i \nabla g_i(\bar{\mathbf{x}}), \mathbf{x} - \bar{\mathbf{x}} \rangle \\ &= f(\bar{\mathbf{x}}) + \langle \nabla f(\bar{\mathbf{x}}) + \mathbf{A}^T \bar{\boldsymbol{\lambda}} + (\mathbf{g}'(\bar{\mathbf{x}}))^T \bar{\boldsymbol{\eta}}, \mathbf{x} - \bar{\mathbf{x}} \rangle \\ &= f(\bar{\mathbf{x}}). \end{aligned}$$

O resultado segue da desigualdade acima. \square

Corolário 1.4. *Sejam $B \subset \{1, \dots, n\}$, $N = \{1, \dots, n\}/B$, $\mathbf{A} = [\mathbf{A}_B \ \mathbf{A}_N] \in \mathbb{R}^{m \times n}$ e $f: \mathbb{R}^{|B|} \rightarrow \mathbb{R}$ convexa. Considere o problema*

$$\min \left\{ f(\mathbf{x}) : \begin{pmatrix} \mathbf{x} \\ \mathbf{a} \end{pmatrix} \in \text{Null } \mathbf{A} \right\}, \quad (1.18)$$

onde $\mathbf{a} \in \mathbb{R}^{|N|}$ é fixado. Se $\bar{\mathbf{x}}$ viável ao problema (1.18), então $\bar{\mathbf{x}}$ é um minimizador se, e somente se, $\nabla f(\bar{\mathbf{x}}) \in \text{Im } \mathbf{A}_B^T$.

Demonstração. Sendo $D = \{\mathbf{x} \in \mathbb{R}^{|B|} : \mathbf{A}_B \mathbf{x} + \mathbf{A}_N \mathbf{a} = 0\}$, o problema (1.18) pode ser reescrito da forma:

$$\min \{f(\mathbf{x}) : \mathbf{x} \in D\}. \quad (1.19)$$

Pelo Teorema 1.9 e pela Proposição 1.7 temos que $\bar{\mathbf{x}} \in D$ é minimizador de (1.19) se, e somente se, existe $\mathbf{y} \in \mathbb{R}^m$ tal que:

$$\nabla f(\bar{\mathbf{x}}) + \mathbf{A}_B^T \mathbf{y} = 0.$$

Portanto, segue o resultado. \square

Corolário 1.5. *Sejam $J \subset \{1, \dots, n\}$, $\mathbf{B} = \begin{pmatrix} \mathbf{B}_{\bar{J}} \\ \mathbf{B}_J \end{pmatrix} \in \mathbb{R}^{n \times m}$ e $f: \mathbb{R}^{|J|} \rightarrow \mathbb{R}$ convexa. Considere o problema*

$$\min \left\{ f(\mathbf{x}) : \begin{pmatrix} \mathbf{a} \\ \mathbf{x} \end{pmatrix} \in \text{Im} \begin{pmatrix} \mathbf{B}_{\bar{J}} \\ \mathbf{B}_J \end{pmatrix} \right\}, \quad (1.20)$$

onde $\bar{J} = \{1, \dots, n\}/J$ e $\mathbf{a} \in \mathbb{R}^{|J|}$ fixado. Se $\bar{\mathbf{x}}$ viável ao problema (1.20), então $\bar{\mathbf{x}}$ é um minimizador se, e somente se, $\mathbf{B}_{\bar{J}}^T (\nabla f(\bar{\mathbf{x}})) \in \text{Im } \mathbf{B}_J^T$.

Demonstração. Sendo $D = \{\lambda \in \mathbb{R}^m : B_J \lambda - \alpha = 0\}$ e $g : \mathbb{R}^m \rightarrow \mathbb{R}$ dada por $g(\lambda) = f(B_J \lambda)$, o problema (1.20) equivale a

$$\min \{g(\lambda) : \lambda \in D\}, \quad (1.21)$$

no sentido que $\bar{\lambda} \in D$ é solução de (1.21) se, e somente se, $\bar{x} = B_J \bar{\lambda}$ for solução de (1.20). Agora, usando o Teorema 1.9 e a Proposição 1.7 temos que $\bar{\lambda} \in D$ é minimizador de (1.21) se, e somente se, existe $y \in \mathbb{R}^{|J|}$ tal que:

$$\nabla g(\bar{\lambda}) + B_J^T y = B_J^T (\nabla f(B_J \bar{\lambda})) + B_J^T y = 0.$$

Daí, $\bar{x} = B_J \bar{\lambda}$ é minimizador de (1.20) se, e somente se, $B_J^T (\nabla f(\bar{x})) = B_J^T (-y)$, para algum $y \in \mathbb{R}^{|J|}$, e segue o resultado. \square

O resultado abaixo estabelece condições de otimalidade para o problema de programação linear, as quais serão muito usado em todo o texto.

Corolário 1.6. *Considere o problema*

$$\min \{c^T x : Ax = b, x \geq 0\}, \quad (1.22)$$

onde $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ e $c \in \mathbb{R}^n$. Um ponto $\bar{x} \in \mathbb{R}^n$ é solução de (1.22) se, e somente se, \bar{x} é viável e $M(\bar{x}) \neq \emptyset$. Em outras palavras, $\bar{x} \in \mathbb{R}^n$ é solução do problema acima se, e somente se, existe $(\bar{\lambda}, \bar{s}) \in \mathbb{R}^m \times \mathbb{R}^n$ tal que $(\bar{x}, \bar{\lambda}, \bar{s})$ seja solução do sistema:

$$\left\{ \begin{array}{l} Ax = b \\ x \geq 0 \\ s + A^T \lambda = c \\ s \geq 0 \\ x^T s = 0 \end{array} \right.$$

Demonstração. Consequência do Teorema 1.9 e da Proposição 1.7. \square

Os próximos resultados técnicos darão propriedades do conjunto dos multiplicadores de Lagrange, para problemas do tipo (1.9)-(1.10) sobre certas restrições, que serão utilizadas nos próximos capítulos.

Proposição 1.8. *Seja \bar{x} um minimizador do problema (1.9)-(1.10) satisfazendo (C3), com $\text{posto}(A) = m \leq n$. Então $M(\bar{x})$ é não-vazio, convexo e compacto.*

Demonstração. Ver [13], precisamente o Teorema 2.3.2 do Capítulo VII. \square

Proposição 1.9. *Sejam \bar{x} e x' soluções do problema (1.9)-(1.10) com f convexa diferenciável e sob restrições lineares, isto é, satisfazendo (C4). Então, $M(\bar{x}) = M(x')$.*

Demonstração. Seja $(\lambda, \eta) \in M(\bar{x}) \subset \mathbb{R}^m \times \mathbb{R}^n$, por definição:

$$\nabla f(\bar{x}) + A^T \lambda + C^T \eta = 0, \quad (1.23)$$

$$\eta \geq 0 \text{ e } \langle \eta, C\bar{x} - d \rangle = 0. \quad (1.24)$$

Considere a função convexa $H_{(\lambda, \eta)} : \mathbb{R}^n \rightarrow \mathbb{R}$, definida por:

$$H_{(\lambda, \eta)}(x) = f(x) + \langle \lambda, Ax - b \rangle + \langle \eta, Cx - d \rangle.$$

Por (1.23) obtemos que $\nabla H_{(\lambda, \eta)}(\bar{x}) = 0$. E pelo Corolário 1.3 temos que \bar{x} é minimizador global de $H_{(\lambda, \eta)}$. Daí, $H_{(\lambda, \eta)}(x') \geq H_{(\lambda, \eta)}(\bar{x})$ e por (1.24) segue:

$$f(x') + \langle \eta, Cx' - d \rangle \geq f(\bar{x}) = f(x').$$

Como $\eta \geq 0$ e $Cx' - d \leq 0$, concluímos da expressão acima que $\langle \eta, Cx' - d \rangle = 0$. Assim,

$$H_{(\lambda, \eta)}(x') = f(x') = f(\bar{x}) = H_{(\lambda, \eta)}(\bar{x}).$$

Assim, x' é também um minimizador global de $H_{(\lambda, \eta)}$, e então:

$$\nabla H_{(\lambda, \eta)}(x') = \nabla f(x') + A^T \lambda + C^T \eta = 0.$$

Portanto, $(\lambda, \eta) \in M(x')$. Desta maneira, mostramos que $M(\bar{x}) \subset M(x')$ e a outra inclusão segue de maneira análoga. \square

1.6 Dualidade em Programação Linear

A chave da teoria de dualidade é a associação feita entre o problema original (primal) a um outro problema (dual) a ser definido. Este último, sob algumas condições tem resolução mais simples e conveniente. As relações entre as soluções destes problemas são muito úteis em questões teóricas e computacionais. Um estudo mais aprofundado sobre esta teoria pode ser encontrado em [10], [17] e [19] que nos serviram como referências neste ponto.

Seja o seguinte problema de programação linear que chamaremos de *problema primal (PP)*:

$$\min\{c^T x : Ax = b, x \geq 0\}, \quad (1.25)$$

onde $\mathbf{c} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m$, com $m \leq n$ e $\mathbf{A} \in \mathbb{R}^{m \times n}$ tem posto máximo igual a m .

O problema dual (PD) associado ao problema primal acima é dado por:

$$\max\{\mathbf{b}^T \boldsymbol{\lambda} : \mathbf{A}^T \boldsymbol{\lambda} + \mathbf{s} = \mathbf{c}, \mathbf{s} \geq 0\}. \quad (1.26)$$

A saber, na literatura a variável \mathbf{s} é chamada de *folga dual* (*dual slack*).

Para facilitar a escrita estabeleceremos as seguintes notações para os conjuntos viáveis do problemas primal e dual como sendo, respectivamente:

$$\mathbf{D} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq 0\} \text{ e}$$

$$\Delta = \{(\boldsymbol{\lambda}, \mathbf{s}) \in \mathbb{R}^m \times \mathbb{R}^n : \mathbf{A}^T \boldsymbol{\lambda} + \mathbf{s} = \mathbf{c}, \mathbf{s} \geq 0\}.$$

Assim denotaremos os conjuntos soluções para os problemas primal e dual como sendo, respectivamente:

$$\mathbf{X}^* = \{\bar{\mathbf{x}} \in \mathbf{D} : \mathbf{c}^T \bar{\mathbf{x}} \leq \mathbf{c}^T \mathbf{x}, \forall \mathbf{x} \in \mathbf{D}\} \text{ e}$$

$$\mathbf{Y}^* = \{(\bar{\boldsymbol{\lambda}}, \bar{\mathbf{s}}) \in \Delta : \mathbf{b}^T \bar{\boldsymbol{\lambda}} \geq \mathbf{b}^T \boldsymbol{\lambda}, \forall (\boldsymbol{\lambda}, \mathbf{s}) \in \Delta\}.$$

Observamos inicialmente que dados quaisquer $\mathbf{x} \in \mathbf{D}$ e $(\boldsymbol{\lambda}, \mathbf{s}) \in \Delta$, temos:

$$\mathbf{c}^T \mathbf{x} = (\mathbf{A}^T \boldsymbol{\lambda} + \mathbf{s})^T \mathbf{x} = (\mathbf{A}^T \boldsymbol{\lambda})^T \mathbf{x} + \mathbf{s}^T \mathbf{x} = \boldsymbol{\lambda}^T (\mathbf{A}\mathbf{x}) + \mathbf{s}^T \mathbf{x} \geq \mathbf{b}^T \boldsymbol{\lambda}. \quad (1.27)$$

A diferença $\mathbf{c}^T \mathbf{x} - \mathbf{b}^T \boldsymbol{\lambda} \geq 0$ é conhecida como *gap* ou folga de dualidade.

A seguir, o primeiro resultado que apresentaremos estabelece uma relação entre as soluções dos problemas (1.25) e (1.26).

Teorema 1.10. (Teorema de Dualidade Fraca)

Um ponto $\bar{\mathbf{x}} \in \mathbf{D}$ é solução do problema primal (1.25) se, e somente se, existe $(\bar{\boldsymbol{\lambda}}, \bar{\mathbf{s}}) \in \Delta$ tal que $\mathbf{c}^T \bar{\mathbf{x}} = \mathbf{b}^T \bar{\boldsymbol{\lambda}}$. Mais ainda, esta última afirmação equivale a dizer que $(\bar{\boldsymbol{\lambda}}, \bar{\mathbf{s}})$ é solução do problema dual (1.26), ou que $(\bar{\boldsymbol{\lambda}}, \bar{\mathbf{s}})$ é um Multiplicador de Lagrange associado a solução $\bar{\mathbf{x}}$ do problema primal (1.25), isto é, $(\bar{\boldsymbol{\lambda}}, \bar{\mathbf{s}}) \in \mathbf{M}(\bar{\mathbf{x}})$.

Demonstração. Pelo Corolário 1.6, temos que $\bar{\mathbf{x}}$ viável ao problema (1.25) é solução se, e somente se, existe $(\bar{\boldsymbol{\lambda}}, \bar{\mathbf{s}}) \in \mathbf{M}(\bar{\mathbf{x}})$, isto é,

$$\bar{\mathbf{s}} + \mathbf{A}^T \bar{\boldsymbol{\lambda}} = \mathbf{c}, \quad \bar{\mathbf{s}} \geq 0 \text{ e } \bar{\mathbf{x}}^T \bar{\mathbf{s}} = 0.$$

Daí, concluímos que $(\bar{\boldsymbol{\lambda}}, \bar{\mathbf{s}}) \in \Delta$. Note que:

$$\bar{\mathbf{x}}^T \bar{\mathbf{s}} = \bar{\mathbf{c}}^T (\mathbf{c} - \mathbf{A}^T \bar{\boldsymbol{\lambda}}) = \bar{\mathbf{x}}^T \mathbf{c} - (\mathbf{A}\bar{\mathbf{x}})^T \bar{\boldsymbol{\lambda}} = \bar{\mathbf{x}}^T \mathbf{c} - \mathbf{b}^T \bar{\boldsymbol{\lambda}}.$$

Portanto,

$$\bar{\mathbf{x}}^T \bar{\mathbf{s}} = 0 \quad \Leftrightarrow \quad \bar{\mathbf{x}}^T \mathbf{c} = \mathbf{b}^T \bar{\boldsymbol{\lambda}}.$$

Finalmente, (1.27) implica que $(\bar{\boldsymbol{\lambda}}, \bar{\mathbf{s}})$ é solução do problema dual (1.26). \square

Sejam dadas $\bar{\mathbf{x}} \in X^*$ e $(\bar{\boldsymbol{\lambda}}, \bar{\mathbf{s}}) \in Y^*$ soluções dos problemas primal e dual, nesta ordem. Pela demonstração do Teorema 1.10, temos que $\bar{\mathbf{x}}^T \bar{\mathbf{s}} = 0$. E sendo $\bar{\mathbf{s}}, \bar{\mathbf{x}} \in \mathbb{R}_+^n$ temos que:

$$\bar{s}_i \bar{x}_i = 0, \quad \text{para todo } i = 1, \dots, n.$$

Desta maneira, dizemos que estas soluções satisfazem a *condição de complementaridade de folga* e vale que:

$$\bar{x}_i > 0 \Rightarrow \bar{s}_i = 0, \quad \forall i = 1, \dots, n;$$

$$\bar{s}_i > 0 \Rightarrow \bar{x}_i = 0, \quad \forall i = 1, \dots, n.$$

Dizemos que um par de soluções, como acima, satisfaz a *condição de complementaridade de folga estrita*, quando:

$$\bar{x}_i = 0 \Rightarrow \bar{s}_i > 0, \quad \forall i = 1, \dots, n;$$

$$\bar{s}_i = 0 \Rightarrow \bar{x}_i > 0, \quad \forall i = 1, \dots, n.$$

O próximo teorema trata sobre a existência de soluções para os problemas primais e duais satisfazendo este último tipo de condição.

Teorema 1.11. *Suponha que os problemas (1.25) e (1.26) admitam soluções. Então, existem $\mathbf{x}^* \in X^*$ e $(\boldsymbol{\lambda}^*, \mathbf{s}^*) \in Y^*$ tais que $\mathbf{x}^* + \mathbf{s}^* > 0$.*

Demonstração. Analisaremos o problema em cada coordenada e depois usaremos a convexidades de X^* e Y^* para concluir o resultado. Mostraremos que $x_j^* = 0$ para todo $\mathbf{x}^* \in X^*$ se, e somente se, $s_j^* > 0$ para algum $(\boldsymbol{\lambda}^*, \mathbf{s}^*) \in Y^*$. Se existe $(\boldsymbol{\lambda}^*, \mathbf{s}^*) \in Y^*$ com $s_j^* > 0$, então pela condição de complementaridade de folga temos $x_j^* = 0$, para toda solução do problema primal. Considere $\gamma \in \mathbb{R}$ o valor ótimo para os problemas primal e dual, já que estes são iguais pelo Teorema 1.10. Reciprocamente, suponha que $x_j^* = 0$ para todo $\mathbf{x}^* \in X^*$, o que é equivalente ao problema:

$$\min\{-\mathbf{e}_j^T \mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{b}, -\mathbf{c}^T \mathbf{x} - t = -\gamma, \mathbf{x} \geq 0, t \geq 0\}, \quad (1.28)$$

onde \mathbf{e}_j é o j -ésimo vetor da base canônica do \mathbb{R}^n , possuir solução ótima, digamos $(\tilde{\mathbf{x}}, \tilde{\mathbf{t}})$, com valor ótimo $-\mathbf{e}_j^T \tilde{\mathbf{x}} = 0$. O problema dual de (1.28) é o seguinte problema:

$$\max\{\mathbf{b}^T \mathbf{y} - \gamma \mathbf{y}_{n+1} : \mathbf{A}^T \mathbf{y} - \mathbf{c} \mathbf{y}_{n+1} + \mathbf{s} = -\mathbf{e}_j, -\mathbf{y}_{n+1} + \mathbf{s}_{n+1} = 0, \mathbf{s} \geq 0, \mathbf{s}_{n+1} \geq 0\}. \quad (1.29)$$

Pelo Teorema 1.10, o problema (1.29) admite solução ótima $((\tilde{\mathbf{y}}, \tilde{\mathbf{y}}_{n+1}), (\tilde{\mathbf{s}}, \tilde{\mathbf{s}}_{n+1}))$ com valor ótimo $\mathbf{b}^T \tilde{\mathbf{y}} - \gamma \tilde{\mathbf{y}}_{n+1} = -\mathbf{e}_j^T \tilde{\mathbf{x}} = 0$. Agora, temos duas possibilidades:

(i) Se $\tilde{\mathbf{y}}_{n+1} = 0$, então $\mathbf{b}^T \tilde{\mathbf{y}} = 0$ e $\mathbf{A}^T \tilde{\mathbf{y}} + \tilde{\mathbf{s}} = -\mathbf{e}_j$. Sendo $(\hat{\lambda}, \hat{\mathbf{s}}) \in \mathbf{Y}^*$ solução do problema dual (1.26), dado $\alpha > 0$ temos:

$$\mathbf{A}^T (\hat{\mathbf{y}} + \alpha \tilde{\mathbf{y}}) + (\alpha \mathbf{e}_j + \hat{\mathbf{s}} + \alpha \tilde{\mathbf{s}}) = (\mathbf{A}^T \hat{\mathbf{y}} + \hat{\mathbf{s}}) + \alpha (\mathbf{A}^T \tilde{\mathbf{y}} + \tilde{\mathbf{s}} + \mathbf{e}_j) = \mathbf{c},$$

com $(\alpha \mathbf{e}_j + \hat{\mathbf{s}} + \alpha \tilde{\mathbf{s}}) \geq 0$ e $\mathbf{b}^T (\hat{\mathbf{y}} + \alpha \tilde{\mathbf{y}}) = \gamma$. Portanto, $(\hat{\mathbf{y}} + \alpha \tilde{\mathbf{y}}, \alpha \mathbf{e}_j + \hat{\mathbf{s}} + \alpha \tilde{\mathbf{s}}) \in \mathbf{Y}^*$ e vale que $(\alpha \mathbf{e}_j + \hat{\mathbf{s}} + \alpha \tilde{\mathbf{s}})_j \geq \alpha > 0$;

(ii) Se $\tilde{\mathbf{y}}_{n+1} > 0$, então $\mathbf{A}^T \left(\frac{1}{\tilde{\mathbf{y}}_{n+1}} \tilde{\mathbf{y}} \right) + \frac{1}{\tilde{\mathbf{y}}_{n+1}} (\tilde{\mathbf{s}} + \mathbf{e}_j) = \mathbf{c}$, com $\frac{1}{\tilde{\mathbf{y}}_{n+1}} (\tilde{\mathbf{s}} + \mathbf{e}_j) \geq 0$ e $\mathbf{b}^T \left(\frac{1}{\tilde{\mathbf{y}}_{n+1}} \tilde{\mathbf{y}} \right) = \gamma$. Logo,

$$\left(\frac{1}{\tilde{\mathbf{y}}_{n+1}} \right) (\tilde{\mathbf{y}}, \tilde{\mathbf{s}} + \mathbf{e}_j) \in \mathbf{Y}^* \text{ e } \left(\frac{1}{\tilde{\mathbf{y}}_{n+1}} (\tilde{\mathbf{s}} + \mathbf{e}_j) \right)_j \geq \left(\frac{1}{\tilde{\mathbf{y}}_{n+1}} \right) > 0.$$

Nos dois casos acima, obtemos $(\lambda^{(j)}, \mathbf{s}^{(j)}) \in \mathbf{Y}^*$ tal que $\mathbf{s}_j^{(j)} > 0$, e segue o que queríamos. Sendo assim, para todo $j \in \{1, \dots, n\}$ obtemos $\mathbf{x}^{(j)} \in \mathbf{X}^*$ e $(\lambda^{(j)}, \mathbf{s}^{(j)}) \in \mathbf{Y}^*$ tais que $\mathbf{x}_j^{(j)} + \mathbf{s}_j^{(j)} > 0$. Então, definamos:

$$\mathbf{x}^* \equiv \frac{1}{n} \sum_{j=1}^n \mathbf{x}^{(j)},$$

$$(\lambda^*, \mathbf{s}^*) \equiv \left(\frac{1}{n} \sum_{j=1}^n \lambda^{(j)}, \frac{1}{n} \sum_{j=1}^n \mathbf{s}^{(j)} \right).$$

Pela convexidade destes conjuntos, temos que $\mathbf{x}^* \in \mathbf{X}^*$, $(\lambda^*, \mathbf{s}^*) \in \mathbf{Y}^*$ e ainda $\mathbf{x}^* + \mathbf{s}^* > 0$, provando assim o teorema. \square

Capítulo 2

Trajatórias Primal e Dual: Definições e Convergências

Neste capítulo, introduziremos as noções de trajetórias primal e dual associadas aos problemas de programação linear perturbados pela função de penalização *divergência de Kullback-Leibler*. *A priori* mostraremos a boa-definição e suavidade (C^∞) destas trajetórias. E por fim, mostraremos que estas convergem para o centro analítico de X^* e para o centróide s^c de S^* , respectivamente. Aqui, as principais referências usadas foram os trabalhos de Cominetti e San Martín em [4], Iusem em [7], Cruz Neto, Monteiro e Iusem em [8] e também de Iusem e Svaiter em [9] .

2.1 Trajetórias Primal e Dual

Seja o problema de programação linear

$$\min \{ \mathbf{c}^T \mathbf{x} : \mathbf{A} \mathbf{x} = \mathbf{b}, \mathbf{x} \geq 0 \}, \quad (2.1)$$

onde $\mathbf{c} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{A} \in \mathbb{R}^{m \times n}$ com $m \leq n$ e $\text{posto}(\mathbf{A}) = m$. Com a mesma notação da Secão 1.6 indicaremos por X^* o conjunto solução deste problema.

Para fazer sentido o estudo a ser feito, assumiremos que o problema (2.1) satisfaz as seguintes hipóteses não-degenerativas:

A1) $X^* \neq \emptyset$;

A2) $D^0 \equiv \{ \mathbf{x} \in \mathbb{R}_{++}^n : \mathbf{A} \mathbf{x} = \mathbf{b} \} \neq \emptyset$.

Associado ao problema (2.1), conforme a Seção 1.6, temos o seguinte problema dual:

$$\max \{ \mathbf{b}^\top \boldsymbol{\lambda} : \mathbf{A}^\top \boldsymbol{\lambda} + \mathbf{s} = \mathbf{c}, \mathbf{s} \geq 0 \}. \quad (2.2)$$

A hipótese **A1**) garante pelo Teorema 1.10 que o problema dual (2.2) possui solução. A saber, com a mesma notação da Seção 1.6 o conjunto solução deste problema será denotado por $Y^* = \{(\bar{\boldsymbol{\lambda}}, \bar{\mathbf{s}}) \in \Delta : \mathbf{b}^\top \bar{\boldsymbol{\lambda}} \geq \mathbf{b}^\top \boldsymbol{\lambda}, \forall (\boldsymbol{\lambda}, \mathbf{s}) \in \Delta\}$, onde $\Delta = \{(\boldsymbol{\lambda}, \mathbf{s}) \in \mathbb{R}^m \times \mathbb{R}^n : \mathbf{A}^\top \boldsymbol{\lambda} + \mathbf{s} = \mathbf{c}, \mathbf{s} \geq 0\}$.

As atenções serão voltadas para o conjunto S^* , das variáveis de folga dual de Y^* , abaixo definido:

$$S^* = \{ \mathbf{s} \in \mathbb{R}_+^n : \mathbf{s} \in \mathbf{c} + \text{Im } \mathbf{A}^\top, \bar{\mathbf{x}}^\top \mathbf{s} = 0 \},$$

onde $\bar{\mathbf{x}}$ é um elemento arbitrário de X^* .

Observação 2.1. *Da prova do Teorema 1.10, obtemos que $Y^* = M(\bar{\mathbf{x}})$ e $\bar{\mathbf{x}}^\top \mathbf{s} = 0$ para todos $\bar{\mathbf{x}} \in X^*$ e $(\boldsymbol{\lambda}, \mathbf{s}) \in M(\bar{\mathbf{x}})$. Logo a definição de S^* é consistente e independe do $\bar{\mathbf{x}} \in X^*$. A boa-definição de S^* também pode ser verificada pela Proposição 1.9.*

Por **A2**) a condição de Slater(**C3**) é satisfeita, logo pelo Teorema 1.8 temos que S^* é um conjunto limitado. E sendo S^* fechado segue que este é um conjunto compacto.

A importância de estudar S^* deve-se ao fato de que dado $\mathbf{s} \in S^*$ obtém-se uma única solução do problema dual, a saber $(\boldsymbol{\lambda}, \mathbf{s}) \in Y^*$. De fato, seja $\mathbf{s} = (\mathbf{c} + \mathbf{A}^\top \boldsymbol{\lambda}) \in \mathbf{c} + \text{Im } \mathbf{A}^\top$, para algum $\boldsymbol{\lambda} \in \mathbb{R}^m$. Pelo Lema 1.1 a matriz $\mathbf{A}\mathbf{A}^\top \in \mathbb{R}^{m \times m}$ é invertível, logo $\boldsymbol{\lambda}$ é unicamente determinado por:

$$\boldsymbol{\lambda} = (\mathbf{A}\mathbf{A}^\top)^{-1} \mathbf{A}(\mathbf{s} - \mathbf{c}).$$

Em todo o texto o estudo sobre problemas de programação linear penalizados é feito usando como penalidade a *Divergência de Kullback-Leibler* dada por:

$$D(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n \left(x_i \log \frac{x_i}{y_i} + y_i - x_i \right), \quad \text{onde } \mathbf{x} \in \mathbb{R}_+^n \text{ e } \mathbf{y} \in \mathbb{R}_{++}^n.$$

A divergência de Kullback-Leibler, que é muito útil em Estatística, será aqui vista como uma distância de Bregman induzida pela função entropia $\varphi : \mathbb{R}_+^n \rightarrow \mathbb{R}$, definida por $\varphi(\mathbf{x}) = \sum_{i=1}^n x_i \log x_i$. Desta maneira conforme o Exemplo 1.7, dados $\mathbf{x} \in \mathbb{R}_+^n$ e $\mathbf{y} \in \mathbb{R}_{++}^n$, temos:

$$D(\mathbf{x}, \mathbf{y}) = D_\varphi(\mathbf{x}, \mathbf{y}) \equiv \varphi(\mathbf{x}) - \varphi(\mathbf{y}) - \langle \nabla \varphi(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle.$$

Verifica-se que valem as condições **B1)** – **B7)** da definição de distâncias de Bregman. Analisar a penalidade sob este ponto de vista mostra-se muito eficiente já que é possível usar toda a maquinaria desenvolvida na Seção 1.4.

Agora consideremos a função barreira $D_\varphi^1 : \mathbb{R}_+^n \rightarrow \mathbb{R}$ definida por

$$D_\varphi^1(\mathbf{x}) \equiv D_\varphi(\mathbf{x}, \mathbf{x}^1) = \sum_{i=1}^n \left(x_i \log \frac{x_i}{x_i^1} + x_i^1 - x_i \right), \quad (2.3)$$

onde $\mathbf{x}^1 = (x_1^1, \dots, x_n^1)^\top \in D^0$.

A *Trajetória Primal* associada a barreira D_φ^1 é o conjunto $\{\mathbf{x}(\mu) : \mu > 0\}$, onde para cada $\mu > 0$ definimos $\mathbf{x}(\mu)$ como sendo a única solução do problema

$$\min \{ \mathbf{c}^\top \mathbf{x} + \mu D_\varphi^1(\mathbf{x}) : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq 0 \}. \quad (2.4)$$

Observe que até então não sabemos se $\mathbf{x}(\mu)$ está ao menos bem-definida. Para tanto, a proposição seguinte vem esclarecer esta situação mostrando que de fato a definição acima é legítima.

Proposição 2.1. *A trajetória primal está bem-definida, isto é, $\mathbf{x}(\mu)$ existe e é a única solução do problema (2.4), para todo $\mu > 0$. Mais ainda, vale que $\mathbf{x}(\mu) > 0$.*

Demonstração. Sejam $D = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq 0\}$ e $f_\mu : D \rightarrow \mathbb{R}$ definida por $f_\mu(\mathbf{x}) = \mathbf{c}^\top \mathbf{x} + \mu D_\varphi^1(\mathbf{x})$. Por **A1)** o problema primal (2.1) possui solução, cujo valor ótimo denotaremos por $\alpha = \min\{\mathbf{c}^\top \mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq 0\}$. Logo,

$$f_\mu(\mathbf{x}) = \mathbf{c}^\top \mathbf{x} + \mu D_\varphi^1(\mathbf{x}) \geq \alpha + \mu D_\varphi^1(\mathbf{x}), \quad \forall \mathbf{x} \in D.$$

Dado $\beta > \alpha$, usando a desigualdade anterior, temos $L_{f_\mu, D}(\beta) \subset L_{D_\varphi^1, D} \equiv \Gamma_1 \left(\mathbf{x}^1, \frac{\beta - \alpha}{\mu} \right)$. E pela condição **B3)** da definição de distância de Bregman, $L_{f_\mu, D}(\beta)$ é limitado. Como D é fechado e f_μ é contínua em D , obtemos que $L_{f_\mu, D}(\beta)$ é fechado. Então $L_{f_\mu, D}(\beta)$ é compacto e não-vazio, pois \mathbf{x}^1 pertence a este conjunto. Do Corolário 1.2 segue que o problema (2.1) possui solução e esta é única, já que f_μ é estritamente convexa (aqui estamos usando a condição **B2)** e o item ii) da Proposição 1.4). Concluimos assim, que $\mathbf{x}(\mu)$ está bem-definida para todo $\mu > 0$, como sendo:

$$\mathbf{x}(\mu) = \operatorname{argmin} \{ \mathbf{c}^\top \mathbf{x} + \mu D_\varphi^1(\mathbf{x}) : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq 0 \}.$$

Daí,

$$\mathbf{c}^\top \mathbf{x}(\mu) + \mu D_\varphi^1(\mathbf{x}(\mu)) \leq \mathbf{c}^\top \mathbf{x} + \mu D_\varphi^1(\mathbf{x}), \quad \forall \mathbf{x} \in D.$$

Pela definição de $D_\varphi^1(x)$, temos para todo $x \in D$ que:

$$c^T x(\mu) + \mu[\varphi(x(\mu)) - \varphi(x^1) - \langle \nabla \varphi(x^1), x(\mu) - x^1 \rangle] \leq c^T x + \mu[\varphi(x) - \varphi(x^1) - \langle \nabla \varphi(x^1), x - x^1 \rangle].$$

Seja $g_\mu : D \rightarrow \mathbb{R}$ dada por $g_\mu(x) = c^T x + \mu\varphi(x)$, segue que:

$$\langle u \nabla \varphi(x^1), x - x(\mu) \rangle \leq g_\mu(x) - g_\mu(x(\mu)), \quad \forall x \in D. \quad (2.5)$$

De (2.5), obtemos que:

$$\nabla \varphi(x^1) \in \partial g_\mu(x(\mu)). \quad (2.6)$$

Mostraremos que $\partial g_\mu(x) = \emptyset$, para todo $x \in D/D^0$. Este fato implicaria, em vista de (2.6), que $x(\mu) \in D^0$ e então $x(\mu) > 0$. Para tanto seja $x \in D/D^0$ e suponha que existe $\xi \in \partial g_\mu(x)$. Tomando $z \in D^0$, o que é possível por **A2**), definimos:

$$y^j = (1 - \varepsilon_j)x + \varepsilon_j z, \quad (2.7)$$

com $\lim_{j \rightarrow \infty} \varepsilon_j = 0$. Pela convexidade de D , obtemos que $y^j \in D^0$ e $\lim_{j \rightarrow \infty} y^j = x$. Daí,

$$\begin{aligned} \varepsilon_j \xi^T (z - x) &= \xi^T (y^j - x) \leq g_\mu(y^j) - g_\mu(x) \\ &= c^T y^j + \mu\varphi(y^j) - c^T x - \mu\varphi(x) \\ &\leq \varepsilon_j c^T (z - x) + \mu \nabla \varphi(y^j)^T (y^j - x) \\ &= \varepsilon_j c^T (z - x) + \mu \frac{\varepsilon_j}{1 - \varepsilon_j} \nabla \varphi(y^j)^T (z - y^j), \end{aligned}$$

onde a primeira desigualdade deve-se a definição de $\partial g_\mu(x)$ e a segunda tem-se graças a convexidade de φ em D . Logo,

$$\frac{1 - \varepsilon_j}{\mu} [(\xi - c)^T (z - x)] \leq \nabla \varphi(y^j)^T (z - y^j). \quad (2.8)$$

Desde que $\lim_{j \rightarrow \infty} y^j = x \in D/D^0 \subset \partial \mathbb{R}_+^n$ e φ é fronteira coerciva, o lado direito de (2.8) tende a $-\infty$ quando $j \rightarrow \infty$, enquanto o lado esquerdo tende a um limite finito. Desta contradição segue o resultado. \square

Observação 2.2. Na proposição acima a propriedade de φ ser fronteira coerciva (**B6**) foi fundamental na demonstração de que a trajetória primal $\{x(\mu) : \mu > 0\}$ está inteiramente contida em D^0 , isto exemplifica a praticidade em usarmos as distâncias de Bregman como penalizações em problema de minimização convexa, como comentado na Observação 1.2.

A seguir, consideraremos a seguinte extensão da função barreira ao \mathbb{R}^n , que por abuso de notação ainda será denotada por D_φ^1 , como sendo $D_\varphi^1 : \mathbb{R}^n \rightarrow \mathbb{R}$ dada por:

$$D_\varphi^1(\mathbf{x}) = \begin{cases} D_\varphi(\mathbf{x}, \mathbf{x}^1), & \text{se } \mathbf{x} \in \mathbb{R}_+^n \\ +\infty, & \text{caso contrário} \end{cases}.$$

Desta maneira, para cada $\mu > 0$, o problema (2.4) equivale ao seguinte problema:

$$\min\{\mathbf{c}^\top \mathbf{x} + \mu D_\varphi^1(\mathbf{x}) : \mathbf{A}\mathbf{x} = \mathbf{b}\}. \quad (2.9)$$

Da Proposição 2.1 temos que a única solução do problema acima é $\mathbf{x}(\mu) > 0$. De sorte que em \mathbb{R}_{++}^n a função objetivo em questão é diferenciável, por **B1**) da definição de função de Bregman. Sendo assim, pelo Teorema 1.9 e pela Proposição 1.7 as condições de otimalidade de Karush-Kuhn-Tucker (KKT) para o problema (2.9) são necessárias e suficientes, ou seja, para cada $\mu > 0$ temos que $\mathbf{x}(\mu)$ é o único ponto em \mathbb{R}_{++}^n tal que $(\mathbf{x}(\mu), \boldsymbol{\lambda}(\mu))$, para algum $\boldsymbol{\lambda}(\mu) \in \mathbb{R}^m$, é solução do sistema:

$$\mathbf{c} + \mu(\nabla\varphi(\mathbf{x}) - \nabla\varphi(\mathbf{x}^1)) + \mathbf{A}^\top \boldsymbol{\lambda} = 0, \quad (2.10)$$

$$\mathbf{A}\mathbf{x} = \mathbf{b}. \quad (2.11)$$

Motivados por estas condições de otimalidades introduziremos a noção de *Trajetória Dual* associada a barreira D_φ^1 . Para cada $\mu > 0$, a *trajetória dual* denotada por $\mathbf{s}(\mu) \in \mathbb{R}^n$, é definida como sendo:

$$\mathbf{s}(\mu) \equiv -\mu(\nabla\varphi(\mathbf{x}(\mu)) - \nabla\varphi(\mathbf{x}^1)) = -\mu(\log \mathbf{x}(\mu) - \log \mathbf{x}^1). \quad (2.12)$$

De (2.10), segue que $\mathbf{s}(\mu) = \mathbf{c} + \mathbf{A}^\top \boldsymbol{\lambda}(\mu) \in (\mathbf{c} + \text{Im } \mathbf{A}^\top)$. E ainda, vale a relação:

$$\mathbf{x}(\mu) = \mathbf{x}^1 e^{-\mathbf{s}(\mu)/\mu}. \quad (2.13)$$

2.2 Suavidade das Trajetórias Primal e Dual

Nesta seção mostraremos que as trajetórias primal e dual definidas acima são curvas suaves, isto é, são de classe $C^\infty(\mathbb{R}_{++}, \mathbb{R}^n)$, como mostra a proposição abaixo.

Proposição 2.2. *A trajetória primal $\{\mathbf{x}(\mu) : \mu > 0\}$ é uma curva $C^\infty(\mathbb{R}_{++}, \mathbb{R}_+^n)$. Consequentemente, a trajetória dual $\{\mathbf{s}(\mu) : \mu > 0\}$ é também uma curva $C^\infty(\mathbb{R}_{++}, \mathbb{R}^n)$.*

Demonstração. Seja a função $F : \mathbb{R}_{++} \times \mathbb{R}^m \times \mathbb{R}_{++}^n \rightarrow \mathbb{R}^n \times \mathbb{R}^m$ dada por

$$F(s, \lambda, x) = (c + s(\nabla\varphi(x) - \nabla\varphi(x^1)) + A^T\lambda, Ax - b)^T.$$

Note que, F satisfaz as seguintes condições:

- i) $F \in C^\infty$, pois $\nabla\varphi(x) = (\log x + e) \in C^\infty$ em \mathbb{R}_{++}^n ;
- ii) Para todo $\mu > 0$, como $(x(\mu), \lambda(\mu))$ é solução do sistema (2.10)-(2.11), temos que $F(\mu, \lambda(\mu), x(\mu)) = (0, 0)$;
- iii) Pelo Lema 1.2, a matriz $J_{\lambda, x}F(\mu, \lambda(\mu), x(\mu)) = \begin{pmatrix} A^T & X^{-1}(\mu) \\ 0 & A \end{pmatrix} \in \mathbb{R}^{(m+n) \times (m+n)}$ é invertível para todo $\mu > 0$.

Sendo assim, para todo $\mu > 0$ o Teorema da Função Implícita (Teorema 1.3) garante que existem $Z = V \times W \subset \mathbb{R}_{++} \times (\mathbb{R}^m \times \mathbb{R}_{++}^n)$ aberto contendo $(\mu, \lambda(\mu), x(\mu))$, onde $V \subset \mathbb{R}_{++}$ é aberto contendo μ e uma única aplicação $\xi : V \rightarrow W \subset \mathbb{R}^m \times \mathbb{R}_{++}^n$ de classe C^∞ dada por $\xi(s) \equiv (\bar{\lambda}(s), \bar{x}(s))$, satisfazendo:

$$F(s, \xi(s)) = F(s, \bar{\lambda}(s), \bar{x}(s)) = (0, 0), \quad \forall s \in V.$$

Daí, dado $s \in V$ vale que:

$$\begin{cases} c + s(\nabla\varphi(\bar{x}(s)) - \nabla\varphi(x^1)) + A^T\bar{\lambda}(s) = 0 \\ A\bar{x}(s) = b \end{cases}.$$

Das condições de otimalidade para o problema (2.4), vimos que $x(s)$ é o único ponto tal que $(x(s), \lambda(s))$ é solução do sistema acima, para algum $\lambda(s) \in \mathbb{R}^m$. Desta maneira, $x(s) = \bar{x}(s)$ para todo $s \in V$. E portanto, $x : \mathbb{R}_{++} \rightarrow \mathbb{R}_{++}^n$ é de classe $C^\infty(\mathbb{R}_{++}, \mathbb{R}_{++}^n)$. Finalmente, como $\log : \mathbb{R}_{++} \rightarrow \mathbb{R}$ é C^∞ e $x(\mu) > 0$, concluímos que $s : \mathbb{R}_{++} \rightarrow \mathbb{R}$ dada por (2.12) é também $C^\infty(\mathbb{R}_{++}, \mathbb{R})$. □

2.3 Centróide de S^*

Nesta seção, definiremos um ponto especial do conjunto S^* , o *centróide*. Este ponto, bem como suas coordenadas positivas serão fundamentais no estudo que se segue, por exemplo, mostraremos que a trajetória dual converge para este ponto.

Dado $J \subset \{1, \dots, n\}$, definamos a função $\sigma_J : \mathbb{R}^n \rightarrow \mathbb{R}$ por $\sigma_J(s) = \min\{s_i : i \notin J\}$. Observamos que σ_J é contínua, pois o mínimo (finito) de funções contínuas ainda é uma função contínua.

Sendo X^* o conjunto solução do problema de programação linear (2.1), em todo o texto utilizaremos a seguinte partição do conjunto $\{1, \dots, n\}$:

$$B = \{i : \bar{x}_i > 0 \text{ para algum } \bar{x} \in X^*\} \text{ e } N = \{1, \dots, n\} \setminus B.$$

Desta partição e da definição de S^* temos que $s_B = 0$, para todo $s \in S^*$.

A seguir, definiremos o *centróide* s^c de S^* recursivamente. De início, consideremos o problema:

$$\max\{\sigma_B(s) : s \in S^*\}.$$

Como S^* é compacto e σ_B é contínua, segue do Teorema 1.4 que o problema acima possui solução. Sejam v_1^* e S_1^* respectivamente o valor ótimo e o conjunto solução deste problema, isto é,

$$v_1^* = \max\{\sigma_B(s) : s \in S^*\} \geq 0,$$

$$S_1^* = \{s \in S^* : \sigma_B(s) = v_1^*\} \neq \emptyset.$$

Valem as seguintes propriedades:

- (i) $v_1^* > 0$;
- (ii) S_1^* é compacto e convexo;
- (iii) $J_1 \equiv \{j \notin B : s_j = v_1^*, \forall s \in S_1^*\} \neq \emptyset$.

De fato, pelo Teorema 1.11 existem $x^* \in X^*$ e $(\lambda^*, s^*) \in Y^*$ tais que $x^* + s^* > 0$. Daí, como $s^* \in S^*$ e $x_N^* = 0$ temos que $s_N^* > 0$. Então, $v_1^* \geq \sigma_B(s^*) > 0$ e segue (i).

Para provar (ii), note que:

$$S_1^* = S^* \cap [v_1^*(0_B, e_N) + 0_B \times \mathbb{R}_+^{|N|}] \tag{2.14}$$

$$= S^* \cap \{s \in \mathbb{R}^n : \sigma_B(s) = v_1^*\}. \tag{2.15}$$

Por (2.14), S_1^* pode ser escrito como a interseção de dois conjuntos convexos, logo é convexo. E por (2.15), S_1^* é compacto, já que é a interseção de um fechado (aqui estamos usando que σ_B é contínua) com o compacto S^* .

E finalmente, suponha por absurdo que (iii) não vale. Daí, dado $i \in N$, temos que:

- a) Ou $s_i > v_1^*$, para todo $s \in S_1^*$. Denotaremos o conjunto destes indices por N^e ;

b) Ou existem $s^{(i)}, \bar{s}^{(i)}$ em S_1^* e $j \in N/\{i\}$ satisfazendo: $s_i^{(i)} = \sigma_B(s^{(i)}) = v_1^*, \bar{s}_j^{(i)} = \sigma_B(\bar{s}^{(i)}) = v_1^*$ e $\bar{s}_i^{(i)} > v_1^*$.

Sendo assim, dado $i \in N/N^e$ defina $\Sigma^{(i)} \equiv \frac{1}{2}(s^{(i)} + \bar{s}^{(i)})$, o qual pertence a S_1^* pela convexidade. E ainda, $\Sigma_i^{(i)} = \frac{1}{2}(s_i^{(i)} + \bar{s}_i^{(i)}) > v_1^*$. Novamente pela convexidade:

$$\Sigma \equiv \left(\frac{1}{|N/N^e|} \sum_{i \in N/N^e} \Sigma^{(i)} \right) \in S_1^*.$$

Portanto, $\Sigma_N > v_1^*e$ e então $\sigma_B(\Sigma) > v_1^*$, o que é um absurdo e segue a validade de (iii).

Se S_1^* for um conjunto com um único ponto, definiremos o centróide s^c como sendo este ponto. Caso contrário, $J_1 \neq N$ e seja $N_1 \equiv N/J_1 \neq \emptyset$. Dado $j \in N_1$, existe $s^{(j)} \in S_1^*$ tal que $s_j^{(j)} > v_1^*$. Como S_1^* é convexo vale que $\Lambda^1 \equiv \left(\frac{1}{|N_1|} \sum_{j \in N_1} s^{(j)} \right) \in S_1^*$ e $\Lambda_{N_1}^1 > v_1^*e$. Agora, considere o problema

$$\max\{\sigma_{B_1}(s) : s \in S_1^*\}, \text{ onde } B_1 = B \cup J_1.$$

Sejam v_2^* e S_2^* , respectivamente, o valor ótimo e o conjunto solução do problema acima. Observamos que $v_2^* \geq \sigma_{B_1}(\Lambda^1) > v_1^*$. E ainda, análogo à análise anterior concluímos que S_2^* é compacto e convexo e $J_2 \equiv \{j \notin B_1 : s_j = v_2^*, \forall s \in S_2^*\} \neq \emptyset$.

Se S_2^* for um conjunto com um único elemento, definimos s^c como sendo este elemento. Caso contrário, prosseguindo indutivamente obteremos a seguinte sequência de conjuntos $S^* \equiv S_0^* \supset S_1^* \supset S_2^* \cdots \supset S_r^* = \{s^c\}$, uma sequência de indices $B \equiv J_0, J_1, J_2, \dots, J_r$ e uma sequência de escalares

$$0 = v_0^* < v_1^* < v_2^* < \cdots < v_r^*, \quad (2.16)$$

tais que, para todo $\ell = 0, 1, \dots, r-1$:

- i) $v_{\ell+1}^*$ e $S_{\ell+1}^*$ são o valor ótimo e o conjunto solução, respectivamente, do problema $\max\{\sigma_{B_\ell}(s) : s \in S_\ell^*\}$, onde $B_\ell \equiv J_0 \cup J_1 \cup \cdots \cup J_\ell$;
- ii) $J_{\ell+1} \equiv \{j \notin B_\ell : s_j = v_{\ell+1}^*, \forall s \in S_{\ell+1}^*\}$.

Observação 2.3. Como cada $J_\ell \neq \emptyset$, após uma quantidade finita r de passos indutivos acima, teremos que $B_r = J_1 \cup \cdots \cup J_r = \{1, \dots, n\}$. E portanto, S_r terá realmente um único ponto, que será o s^c , satisfazendo:

$$s_{J_0}^c = 0e_{J_0}, s_{J_1}^c = v_1^*e_{J_1}, \dots, s_{J_r}^c = v_r^*e_{J_r}.$$

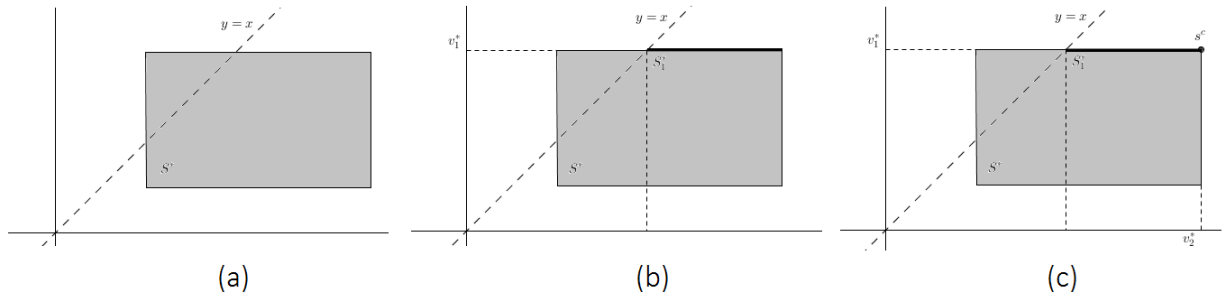


Figura 2.1: O centróide de $S_0 = S^*$ é obtido "empurrando as folgas" em (a) até colapsar em (b). Este processo continua até obtermos um único ponto s^c em (c).

A seguinte proposição caracteriza os conjuntos S_ℓ^* acima definidos.

Proposição 2.3. Para cada $\ell = 1, \dots, r$ tem-se:

$$S_\ell^* = \{s \in S^* : s_{J_k} = v_k^* e, \forall k \leq \ell, s_{J_k} \geq v_k^* e, \forall k > \ell\} \quad (2.17)$$

$$= \{s \in S^* : s_{J_k} \geq v_k^* e, \forall k \leq \ell, s_{J_k} \geq v_k^* e, \forall k > \ell\}. \quad (2.18)$$

Demonstração. Provaremos estas duas relações por indução em ℓ . A priori, assuma que (2.17) seja válido para $1, \dots, \ell - 1$. Dado $s \in S_\ell^* \subset S_{\ell-1}^*$ vale que $\sigma_{B_{\ell-1}}(s) = v_\ell^*$, pela definição de J_ℓ juntamente com a hipótese de indução obtemos $s_{J_k} = v_k^* e$ para todo $k \leq \ell$ e $\min\{s_i : i \notin B_{\ell-1}\} \geq v_\ell^*$. Portanto, s pertence ao conjunto do lado direito de (2.17). Reciprocamente, seja s no conjunto do lado direito de (2.17). Daí, pela hipótese de indução temos que $s \in S_{\ell-1}^*$ e ainda $s_{J_\ell} = v_\ell^* e$, o que nos dá $s \in S_\ell^*$. Provaremos a relação (2.18). Suponha que (2.18) vale para os índices $1, \dots, \ell - 1$. Seja $s \in S_\ell^* \subset S_{\ell-1}^*$, novamente pela hipótese de indução e pela definição de J_ℓ , análogo ao que fizemos acima, vemos que s pertence ao lado direito de (2.18). Reciprocamente, seja s no conjunto do lado direito de (2.18). Daí, pela hipótese de indução temos que $s \in S_{\ell-1}^*$. Como $s_{J_k} \geq v_k^* e$ para todo $k \geq \ell$, segue da definição de J_ℓ que $s \in S_\ell^*$. \square

2.4 Convergência das Trajetórias Primal e Dual

Agora, mostraremos a convergência das trajetórias primal e dual, a saber estas convergem para o centro analítico de X^* e para o centróide s^c de S^* , respectivamente. Os resultados a serem expostos são clássicos e foram extraídos de [4], [8] e [9].

Proposição 2.4. $\lim_{\mu \rightarrow 0} \mathbf{x}(\mu) = \mathbf{x}^*$, onde

$$\mathbf{x}^* \equiv \operatorname{argmin}_{\mathbf{x} \in X^*} \sum_{j \in B} \left(x_j \log \frac{x_j}{x_j^1} + x_j^1 - x_j \right).$$

Na literatura, \mathbf{x}^* é chamado de *Centro Analítico de X^* com relação a barreira D_φ* .

Demonstração. Iniciaremos mostrando que \mathbf{x}^* está bem-definido. Dado $\mathbf{b} > 0$, a função $h : \mathbb{R}_{++} \rightarrow \mathbb{R}$ contínua definida por $h(t) = (t \log \frac{t}{\mathbf{b}} + \mathbf{b} - t)$ é tal que $\lim_{t \rightarrow \infty} h(t) = \infty$. E sendo $H : X^* \rightarrow \mathbb{R}$ definida por $H(\mathbf{x}) = \sum_{j \in B} \left(x_j \log \frac{x_j}{x_j^1} + x_j^1 - x_j \right)$ contínua, para todo \mathbf{c} real temos que $L_{H, X^*}(\mathbf{c})$ é fechado e limitado, logo compacto. Então, segue do Corolário (1.2) que o problema de minimizar H em X^* possui solução global. Como $h''(t) = \frac{1}{t} > 0$, temos do Teorema 1.8 que h é estritamente convexa. Daí, H é estritamente convexa e \mathbf{x}^* é unicamente determinado. Pela definição de $\mathbf{x}(\mu)$, vale:

$$\mathbf{c}^T \mathbf{x}(\mu) + \mu D_\varphi(\mathbf{x}(\mu)) \leq \mathbf{c}^T \mathbf{x}^* + \mu D_\varphi(\mathbf{x}^*). \quad (2.19)$$

Desde que $\mathbf{x}^* \in X^*$, temos $\mu D_\varphi(\mathbf{x}(\mu)) \leq \mu D_\varphi(\mathbf{x}^*)$. Portanto,

$$\sum_{j=1}^n \left(x_j(\mu) \log \frac{x_j(\mu)}{x_j^1} + x_j^1 - x_j(\mu) \right) \leq \sum_{j=1}^n \left(x_j^* \log \frac{x_j^*}{x_j^1} + x_j^1 - x_j^* \right). \quad (2.20)$$

De (2.20) e do fato que $\lim_{t \rightarrow \infty} h(t) = \infty$, obtemos que $\{\mathbf{x}(\mu)\}$ é limitado quando $\mu \rightarrow 0$. Resta mostrarmos que \mathbf{x}^* é o único ponto de acumulação de $\{\mathbf{x}(\mu)\}$ quando $\mu \rightarrow 0$. Para tanto, seja $\mu_k \rightarrow 0$ com $\mathbf{x}(\mu_k) \rightarrow \tilde{\mathbf{x}}$. Afirmamos que $\tilde{\mathbf{x}} \in X^*$. De fato, como $A\mathbf{x}(\mu_k) = \mathbf{b}$ e $\mathbf{x}(\mu_k) \geq 0$, passando ao limite obtemos $A\tilde{\mathbf{x}} = \mathbf{b}$ e $\tilde{\mathbf{x}} \geq 0$. E de (2.19) com $\mu = \mu_k$, temos:

$$\mathbf{c}^T \mathbf{x}(\mu_k) + \mu_k \sum_{j=1}^n \left(x_j(\mu_k) \log \frac{x_j(\mu_k)}{x_j^1} + x_j^1 - x_j(\mu_k) \right) \leq \mathbf{c}^T \mathbf{x}^* + \mu_k \sum_{j=1}^n \left(x_j^* \log \frac{x_j^*}{x_j^1} + x_j^1 - x_j^* \right).$$

Fazendo $\mu_k \rightarrow 0$ na desigualdade acima, temos $\mathbf{c}^T \tilde{\mathbf{x}} \leq \mathbf{c}^T \mathbf{x}^*$ e segue que $\tilde{\mathbf{x}} \in X^*$.

Finalmente, fazendo $\mu_k \rightarrow 0$ em (2.20) e usando $\tilde{\mathbf{x}}_N = 0$, obtemos:

$$H(\tilde{\mathbf{x}}) = \sum_{j \in B} \left(\tilde{x}_j \log \frac{\tilde{x}_j}{x_j^1} + x_j^1 - \tilde{x}_j \right) \leq \sum_{j \in B} \left(x_j^* \log \frac{x_j^*}{x_j^1} + x_j^1 - x_j^* \right) = H(\mathbf{x}^*).$$

Concluimos, pela unicidade de \mathbf{x}^* que $\tilde{\mathbf{x}} = \mathbf{x}^*$. □

Em seguida, fornecemos propriedades e caracterizações para a trajetória dual, bem como o seu limite quando $\mu \rightarrow 0$.

Proposição 2.5. *Seja $\{s(\mu)\}$ a trajetória dual. Seja \tilde{x} um ponto qualquer do conjunto $\{x \in \mathbb{R}^n : Ax = b\}$. Então, para cada $\mu > 0$, tem-se que $s(\mu)$ é a única solução do problema*

$$\min \left\{ \tilde{x}^T s + \mu \sum_{i=1}^n x_i^1 e^{-s_i/\mu} : s \in c + \text{Im } A^T \right\}. \quad (2.21)$$

Demonstração. Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ definida por $f(s) = \tilde{x}^T s + \mu \sum_{i=1}^n x_i^1 e^{-s_i/\mu}$. Assim, o problema (2.21) equivale a

$$\min\{g(\lambda) : \lambda \in \mathbb{R}^m\}, \quad (2.22)$$

onde $g : \mathbb{R}^m \rightarrow \mathbb{R}$ é definida por $g(\lambda) = f(c + A^T \lambda)$. Na medida que, $\bar{\lambda}$ é solução de (2.22) se, e somente se, $\bar{x} = c + A^T \bar{\lambda}$ é solução de (2.21). Por (2.13), temos:

$$\nabla f(s(\mu)) = \tilde{x} + \mu \left(-\frac{1}{\mu} \right) x^1 e^{-s(\mu)/\mu} = \tilde{x} - x(\mu) \in \text{Null } A.$$

Sendo $\lambda(\mu) \in \mathbb{R}^m$ tal que $s(\mu) = c + A^T \lambda(\mu)$, obtemos:

$$\nabla g(\lambda(\mu)) = A \nabla f(c + A^T \lambda(\mu)) = A \nabla f(s(\mu)) = A(\tilde{x} - x(\mu)) = 0.$$

Pela Proposição 1.4, como f é estritamente convexa, temos que g também o é. Por fim, usando o Teorema 1.3 temos que $\lambda(\mu)$ é a única solução do problema (2.22). Portanto, $s(\mu) = c + A^T \lambda(\mu)$ é a única solução do problema (2.21). \square

Lema 2.1. *$\{s(\mu)\}$ permanece limitado quando $\mu \rightarrow 0$.*

Demonstração. Suponhamos que $\|s_k\| \rightarrow \infty$, com $s_k = s(\mu_k)$ e $\mu_k \rightarrow 0$. Como \mathbb{S}^{n-1} é compacto, pelo Teorema de Bolzano-Weierstrass (Teorema 1.2) a menos de subsequência, temos:

$$\frac{s_k}{\|s_k\|} \rightarrow d \in \mathbb{S}^{n-1}. \quad (2.23)$$

Afirmamos que d satisfaz: (i) $d \in \text{Im } A^T$, (ii) $d^T x^* = 0$ e (iii) $d \geq 0$. De fato, sendo $s_k = c + A^T \lambda_k$ com $\lambda_k \in \mathbb{R}^m$, temos:

$$\frac{s_k}{\|s_k\|} = \frac{c}{\|s_k\|} + \frac{A^T \lambda_k}{\|s_k\|}.$$

Tomando o limite na expressão acima, como $\frac{c}{\|s_k\|} \rightarrow 0$ e pela Proposição 1.1) $\text{Im } A^T$ é fechado, ganhamos que $d \in \text{Im } A^T$ e vale (i). Agora para mostrarmos (ii), usaremos que $\lim_{x \rightarrow 0} x \log x = 0$, a expressão (2.12) e a Proposição 2.4 para obter:

$$\begin{aligned} \lim_{k \rightarrow \infty} s(\mu_k)^T x(\mu_k) &= \lim_{k \rightarrow \infty} \sum_{j \in B} \mu_k (\log x_j^1 - \log x_j(\mu_k)) x_j(\mu_k) \\ &+ \lim_{k \rightarrow \infty} \sum_{j \in N} \mu_k (\log x_j^1 - \log x_j(\mu_k)) x_j(\mu_k) = 0. \end{aligned} \quad (2.24)$$

Daí, novamente pela Proposição 2.4:

$$0 = \lim_{k \rightarrow \infty} \frac{\mathbf{s}_k^T \mathbf{x}(\mu_k)}{\|\mathbf{s}_k\|} = \lim_{k \rightarrow \infty} \left(\frac{\mathbf{s}_k}{\|\mathbf{s}_k\|} \right)^T \mathbf{x}(\mu_k) = \mathbf{d}^T \mathbf{x}^*.$$

Finalmente, para verificarmos (iii), da definição de centróide é sabido que $\mathbf{s}^c \in \mathbf{c} + \text{Im } \mathbf{A}^T$ e $\mathbf{s}^c \geq 0$. Pela Proposição 2.5 com $\tilde{\mathbf{x}}$ não-nulo, temos:

$$\tilde{\mathbf{x}}^T \mathbf{s}_k + \mu_k \sum_{i=1}^n x_i^1 e^{-s_{k,i}/\mu_k} \leq \tilde{\mathbf{x}}^T \mathbf{s}^c + \mu_k \sum_{i=1}^n x_i^1 e^{-s_i^c/\mu_k} \leq \beta + nM\mu_k, \quad (2.25)$$

onde $\beta = \tilde{\mathbf{x}}^T \mathbf{s}$ e $M = \max_{1 \leq i \leq n} x_i^1$. Dado $i \in \{1, \dots, n\}$, de (2.25) obtemos:

$$e^{-s_{k,i}/\mu_k} \leq \frac{\beta + nM\mu_k - \tilde{\mathbf{x}}^T \mathbf{s}_k}{\mu_k x_i^1}.$$

Daí,

$$-\frac{s_{k,i}}{\mu_k} \leq \log \left(\frac{\beta + nM\mu_k - \tilde{\mathbf{x}}^T \mathbf{s}_k}{\mu_k x_i^1} \right).$$

Então,

$$\frac{s_{k,i}}{\|\mathbf{s}_k\|} \geq -\frac{\mu_k}{\|\mathbf{s}_k\|} \log \left(\frac{\beta + nM\mu_k - \tilde{\mathbf{x}}^T \mathbf{s}_k}{\mu_k x_i^1} \right). \quad (2.26)$$

Como $\mu_k \rightarrow 0$ e $\|\mathbf{s}_k\| \rightarrow \infty$, usando a desigualdade de Cauchy-Schwarz, para n suficientemente grande, temos que:

$$\frac{\beta + nM\mu_k - \tilde{\mathbf{x}}^T \mathbf{s}_k}{\mu_k x_i^1} \leq \frac{\|\tilde{\mathbf{x}}\| \|\mathbf{s}_k\|}{2\mu_k x_i^1}. \quad (2.27)$$

Agora, de (2.26) e (2.27) segue

$$\frac{s_{k,i}}{\|\mathbf{s}_k\|} \geq -\frac{\mu_k}{\|\mathbf{s}_k\|} \log \left(\frac{\|\tilde{\mathbf{x}}\| \|\mathbf{s}_k\|}{2\mu_k x_i^1} \right) = -\frac{\mu_k}{\|\mathbf{s}_k\|} \left[\log \left(\frac{\|\mathbf{s}_k\|}{\mu_k} \right) + \log \left(\frac{\|\tilde{\mathbf{x}}\|}{2x_i^1} \right) \right].$$

Tomando o limite com $k \rightarrow \infty$ nesta última expressão e sabendo que $\lim_{x \rightarrow 0} x \log \left(\frac{1}{x} \right) = 0$, obtemos de (2.23) que $\mathbf{d}_i \geq 0$ para todo i e portanto, $\mathbf{d} \geq 0$. Finalmente, tomando $\bar{\mathbf{s}} \in \mathbf{S}^* = \{\mathbf{s} \in \mathbb{R}_+^n : \mathbf{s} \in \mathbf{c} + \text{Im } \mathbf{A}^T, \mathbf{s}^T \mathbf{x}^* = 0\}$, usando (i), (ii) e (iii), concluímos que $\bar{\mathbf{s}} + t\mathbf{d} \in \mathbf{S}^*$, para todo $t > 0$. Desta maneira, como \mathbf{d} é não-nulo, temos um absurdo com o fato de \mathbf{S}^* ser limitado e segue o resultado. \square

Lema 2.2. *Se $\lim_{\mu_k \rightarrow 0} \mathbf{s}(\mu_k) = \tilde{\mathbf{s}}$, então $\tilde{\mathbf{s}} \in \mathbf{S}^* = \{\mathbf{s} \in \mathbb{R}_+^n : \mathbf{s} \in \mathbf{c} + \text{Im } \mathbf{A}^T, \mathbf{s}^T \mathbf{x}^* = 0\}$.*

Demonstração. Como $\mathbf{c} + \text{Im } \mathbf{A}^T$ é fechado e $\mathbf{s}(\mu_k) \rightarrow \tilde{\mathbf{s}}$ obtemos que $\tilde{\mathbf{s}} \in \mathbf{c} + \text{Im } \mathbf{A}^T$. De (2.24) e pela Proposição 2.4, segue que:

$$0 = \lim_{k \rightarrow \infty} \mathbf{x}(\mu_k)^T \mathbf{s}(\mu_k) = \tilde{\mathbf{s}}^T \mathbf{x}^*.$$

E por último afirmamos que $\tilde{s} \geq 0$. De fato, se para algum $i \in \{1, \dots, n\}$ tivéssemos $\lim_{k \rightarrow \infty} s_i(\mu_k) = \tilde{s}_i < 0$, chegaríamos a um absurdo por (2.13), pois

$$x_i^* = \lim_{k \rightarrow \infty} x_i(\mu_k) = \lim_{k \rightarrow \infty} x_i^1 e^{-s_i(\mu_k)/\mu_k} = \infty.$$

Portanto, $\tilde{s} \in S^*$. □

O teorema seguinte mostra, fazendo uso dos resultados técnicos acima expostos, a convergência de $\{s(\mu)\}$ quando $\mu \rightarrow 0$. A saber, a trajetória dual converge para o centróide do conjunto S^* . Além de ser esplêndido, este resultado será muito útil no restante deste trabalho. A seguinte demonstração foi dada por Cominetti e San Martín em [4].

Teorema 2.1. *A trajetória $\{s(\mu)\}$ converge para s^c , quando $\mu \rightarrow 0$.*

Demonstração. Pelo Lema 2.1, $s(\mu)$ permanece limitado quando $\mu \rightarrow 0$. Desta maneira, é suficiente mostrarmos que, se $\mu_k \rightarrow 0$ com $s_k \equiv s(\mu_k) \rightarrow \tilde{s}$, então $\tilde{s} = s^c$. Sendo assim, defina $s_k^* = s_k + (s^c - \tilde{s})$. Pelo Lema 2.2 temos $\tilde{s} \in S^*$, logo $(s^c - \tilde{s})_B = 0$, $(s^c - \tilde{s}) \in \text{Im } A^T$ e $s_k^* \in c + \text{Im } A^T$. Daí, pela Proposição 2.5 com $\tilde{x} = x^*$, temos:

$$(x^*)^T s_k + \mu_k \sum_{i=1}^n x_i^1 e^{-s_{k,i}/\mu_k} \leq (x^*)^T s_k^* + \mu_k \sum_{i=1}^n x_i^1 e^{-s_{k,i}^*/\mu_k}. \quad (2.28)$$

E ainda,

$$(x^*)^T (s_k - s_k^*) = (x^*)^T (\tilde{s} - s^c) = (x_B^*)^T (\tilde{s} - s^c)_B + (x_N^*)^T (\tilde{s} - s^c)_N = 0. \quad (2.29)$$

De (2.28) e (2.29) obtemos a desigualdade:

$$\sum_{i=1}^n x_i^1 e^{-s_{k,i}/\mu_k} \leq \sum_{i=1}^n x_i^1 e^{-s_{k,i}^*/\mu_k}. \quad (2.30)$$

De $(s^c - \tilde{s})_B = 0$ ganhamos $(s_k^*)_B = (s_k)_B$. Então, de (2.30) segue que:

$$\sum_{i \in N} x_i^1 e^{-s_{k,i}/\mu_k} \leq \sum_{i \in N} x_i^1 e^{-s_{k,i}^*/\mu_k}. \quad (2.31)$$

Sendo $\sigma_B(x) = \min\{x_j : j \notin B\}$, vale:

$$m e^{-\sigma_B(s_k)/\mu_k} \leq n M e^{-\sigma_B(s_k^*)/\mu_k},$$

onde $m = \min_{1 \leq i \leq n} x_i^1$ e $M = \max_{1 \leq i \leq n} x_i^1$. Daí,

$$\sigma_B(s_k) \geq -\mu_k \log \left(\frac{nM}{m} \right) + \sigma_B(s_k^*).$$

Como σ_B é contínua e $\mathbf{s}_k^* \rightarrow \mathbf{s}^c$, tomando o limite quando $k \rightarrow \infty$ na expressão anterior, obtemos:

$$\mathbf{v}_1^* = \sigma_B(\mathbf{s}^c) \leq \sigma_B(\tilde{\mathbf{s}}) \leq \mathbf{v}_1^*,$$

onde a última desigualdade deve-se da definição de \mathbf{v}_1^* . Logo, $\sigma_B(\tilde{\mathbf{s}}) = \mathbf{v}_1^*$ e então $\tilde{\mathbf{s}} \in \mathbf{S}_1^*$. Portanto, $\tilde{\mathbf{s}}_{J_1} = \mathbf{v}_1^* \mathbf{e} = \mathbf{s}_{J_1}^c$ e assim $(\mathbf{s}_k^*)_{J_1} = (\mathbf{s}_k)_{J_1}$. Agora, por (2.31), temos:

$$\sum_{i \notin B_1} x_i^1 e^{-s_{k,i}/\mu_k} \leq \sum_{i \notin B_1} x_i^1 e^{-s_{k,i}^*/\mu_k}. \quad (2.32)$$

Analogamente, concluímos que $\tilde{\mathbf{s}}_{J_2} = \mathbf{s}_{J_2}^c$. Indutivamente, obtemos $\tilde{\mathbf{s}} = \mathbf{s}^c$ e segue o resultado. \square

Capítulo 3

Comportamento limite das derivadas das trajetórias

Neste capítulo estudaremos o comportamento limite dos termos $\Delta s(\mu)$ e $\Delta x(\mu)$, definidos a seguir como funções das trajetórias primal e dual, respectivamente. Bem como analisaremos o limite de $\dot{s}(\mu)$, quando $\mu \rightarrow 0$. Este estudo será de suma importância para o capítulo seguinte, visto que será a base para a análise de convergência das sequências $\{x_k\}$ e $\{\bar{s}_k\}$, definidas a posteriori.

3.1 Comportamento Limite de $\Delta s(\mu)$, $\Delta u(\mu)$ e $\dot{s}(\mu)$

Com o intuito de analisarmos o comportamento de $\dot{s}(\mu)$ quando μ positivo tende 0, estudaremos a priori o limite do quociente

$$\Delta s(\mu) \equiv \frac{s(\mu) - s^c}{\mu}, \quad (3.1)$$

quando $\mu \rightarrow 0+$.

Note que usando (2.12), a Proposição 2.4 e o fato que $s_B^c = 0$, temos:

$$\lim_{\mu \rightarrow 0} \Delta s_B(\mu) = \lim_{\mu \rightarrow 0} \frac{s_B(\mu)}{\mu} = \lim_{\mu \rightarrow 0} (\log x_B^1 - \log x_B(\mu)) = \log x_B^1 - \log x_B^*. \quad (3.2)$$

Sendo assim, resta provarmos a existência do limite $\lim_{\mu \rightarrow 0} \Delta s_N(\mu)$. Visando este objetivo, os próximos resultados fornecem um estudo sobre $\Delta s(\mu)$. O primeiro deles caracteriza $\Delta s_J(\mu)$ para qualquer $J \subset N$.

Lema 3.1. *Seja $J \subset \mathbb{N}$ dado e defina $\bar{J} = \{1, \dots, n\}/J$. Para $\mu > 0$, $\Delta s_J(\mu)$ é a única solução do problema*

$$\min_{\Delta s_J} \left\{ \sum_{j \in J} x_j^1 e^{-\Delta s_j - s_j^c / \mu} : \begin{pmatrix} \Delta s_{\bar{J}}(\mu) \\ \Delta s_J \end{pmatrix} \in \text{Im} \begin{pmatrix} A_{\bar{J}}^T \\ A_J^T \end{pmatrix} \right\}. \quad (3.3)$$

Demonstração. Seja a função $f : \mathbb{R}^{|\bar{J}|} \rightarrow \mathbb{R}$ definida por $f(x) = \sum_{j \in J} x_j^1 e^{-x_j - s_j^c / \mu}$. Como $s(\mu)$ e s^c pertencem a $\mathbf{c} + \text{Im } A^T$, existe $\lambda(\mu) \in \mathbb{R}^m$ tal que $\Delta s(\mu) = A^T(\lambda(\mu)) \in \text{Im } A^T$, e portanto $\Delta s_J(\mu) = A_J^T(\lambda(\mu))$ é viável para (3.3). Tomando $x^* \in X^*$, temos $x_j^* = 0$. E de (3.1), obtemos $-\frac{s(\mu)}{\mu} = -\Delta s(\mu) - \frac{s^c}{\mu}$. Então usando (2.13), segue que:

$$-x_j^1 e^{-\Delta s_j(\mu) - s_j^c / \mu} = -x_j^1 e^{-s_j(\mu) / \mu} = -x_j(\mu) = x_j^* - x_j(\mu), \quad \forall j \in J.$$

Daí, pela expressão anterior:

$$\nabla f(\Delta s_J(\mu)) = x_j^* - x_J(\mu). \quad (3.4)$$

Como $Ax^* = Ax(\mu) = \mathbf{b}$, segue que $x^* - x(\mu) \in \text{Null } A$, isto é:

$$\begin{aligned} A[x^* - x(\mu)] = 0 &\Leftrightarrow [A_J \quad A_{\bar{J}}] \begin{bmatrix} x_j^* - x_J(\mu) \\ x_{\bar{J}}^* - x_{\bar{J}}(\mu) \end{bmatrix} = 0 \\ &\Leftrightarrow A_J[x_j^* - x_J(\mu)] = A_{\bar{J}}[x_{\bar{J}}(\mu) - x_{\bar{J}}^*] \in \text{Im } A_{\bar{J}}. \end{aligned}$$

Portanto, por (3.4):

$$A_J[\nabla f(\Delta s_J(\mu))] \in \text{Im } A_{\bar{J}}.$$

Finalmente, pelo Corolário 1.5, concluímos que $\Delta s_J(\mu)$ é solução do problema (3.3) e esta é única, já que $\text{Hess}(f) > 0$. Então, do Teorema 1.8 segue que f é estritamente convexa. \square

Definamos agora para todo $\ell = 0, 1, \dots, r$ os seguintes conjuntos de índices:

$$N_\ell = J_\ell \cup \dots \cup J_r \quad \text{e} \quad B_{\ell-1} = \{1, \dots, n\}/N_\ell.$$

Lema 3.2. *O conjunto $\{\Delta s(\mu) : \mu \in (0, 1]\}$ é limitado.*

Demonstração. Por (3.2) temos que $\Gamma_0 = \{\Delta s_B(\mu) : \mu \in (0, 1]\}$ é limitado. Resta mostrarmos que $\Gamma_i = \{\Delta s_{J_i}(\mu) : \mu \in (0, 1]\}$ é limitado, para $i = 1, \dots, r$. Suponhamos o contrário

e seja $\ell \geq 1$ tal que $\Gamma_0, \dots, \Gamma_{\ell-1}$ são limitados, mas Γ_ℓ é ilimitado. Pelo Lema 3.1 com $J = N_\ell$, temos:

$$\Delta s_{N_\ell}(\mu) = \operatorname{argmin} \left\{ \sum_{j \in N_\ell} x_j^1 e^{-\Delta s_j - s_j^c / \mu} : \begin{pmatrix} \Delta s_{B_{\ell-1}}(\mu) \\ \Delta s_{N_\ell} \end{pmatrix} \in \operatorname{Im} \begin{pmatrix} A_{B_{\ell-1}}^\top \\ A_{N_\ell}^\top \end{pmatrix} \right\}. \quad (3.5)$$

Sendo $\Delta s(\mu) = (\Delta s_{B_{\ell-1}}(\mu), \Delta s_{N_\ell}(\mu))^\top \in \operatorname{Im} A^\top$, pelo Lema 1.3 existe $M = M(\operatorname{Im} A^\top)$ e um conjunto $\{\tilde{\Delta} s_{N_\ell}(\mu) : \mu \in (0, 1]\}$ tal que para todo $\mu \in (0, 1]$ vale

$$\begin{pmatrix} \Delta s_{B_{\ell-1}}(\mu) \\ \tilde{\Delta} s_{N_\ell}(\mu) \end{pmatrix} \in \operatorname{Im} A^\top = \operatorname{Im} \begin{pmatrix} A_{B_{\ell-1}}^\top \\ A_{N_\ell}^\top \end{pmatrix} \text{ e} \quad (3.6)$$

$$\|\tilde{\Delta} s_{N_\ell}(\mu)\| \leq M \|\Delta s_{B_{\ell-1}}(\mu)\|. \quad (3.7)$$

Como $\{\Delta s_{B_{\ell-1}}(\mu) : \mu \in (0, 1]\}$ é limitado, (3.7) implica que $\{\tilde{\Delta} s_{N_\ell}(\mu) : \mu \in (0, 1]\}$ é também limitada. Agora, segue por (3.5) e (3.6) que

$$\sum_{j \in J_\ell} x_j^1 e^{-\Delta s_j(\mu) - s_j^c / \mu} \leq \sum_{j \in N_\ell} x_j^1 e^{-\Delta s_j(\mu) - s_j^c / \mu} \leq \sum_{j \in N_\ell} x_j^1 e^{-\tilde{\Delta} s_j(\mu) - s_j^c / \mu}. \quad (3.8)$$

Multiplicando ambos os lados de (3.8) por $e^{v_\ell^* / \mu}$ e juntando com o fato de que $s_j^c = v_\ell^*$ para todo $j \in J_\ell$, obtemos:

$$\sum_{j \in J_\ell} x_j^1 e^{-\Delta s_j(\mu)} \leq \sum_{j \in J_\ell} x_j^1 e^{-\tilde{\Delta} s_j(\mu)} + \sum_{j \in N_\ell / J_\ell} x_j^1 e^{-\tilde{\Delta} s_j(\mu) - (s_j^c - v_\ell^*) / \mu}.$$

Da definição de centróide, segue que $s_j^c > v_\ell^*$ para todo $j \in N_\ell / J_\ell$, logo $(s_j^c - v_\ell^*) / \mu > 0$ para todo $j \in N_\ell / J_\ell$, daí:

$$\sum_{j \in J_\ell} x_j^1 e^{-\Delta s_j(\mu)} < \sum_{j \in J_\ell} x_j^1 e^{-\tilde{\Delta} s_j(\mu)} + \sum_{j \in N_\ell / J_\ell} x_j^1 e^{-\tilde{\Delta} s_j(\mu)} = \sum_{j \in N_\ell} x_j^1 e^{-\tilde{\Delta} s_j(\mu)}. \quad (3.9)$$

O fato de $\{\tilde{\Delta} s_{N_\ell}(\mu) : \mu \in (0, 1]\}$ ser limitado garante a existência de um $K > 0$ tal que $|\tilde{\Delta} s_j(\mu)| \leq K$, para todos $j \in N_\ell$ e $\mu \in (0, 1]$. Logo de (3.9), temos que:

$$\sum_{j \in J_\ell} x_j^1 e^{-\Delta s_j(\mu)} < \sum_{j \in N_\ell} x_j^1 e^{-K}.$$

Da última desigualdade, como $\lim_{x \rightarrow +\infty} e^x = +\infty$, temos que $\{\Delta s_j(\mu) : \mu \in (0, 1]\}$ é limitado inferiormente para todo $j \in J_\ell$. Resta mostrarmos que estes conjuntos são também limitados superiormente. De fato, suponhamos por contradição que exista $j \in J_\ell$, tal que $\{\Delta s_j(\mu) : \mu \in (0, 1]\}$ é ilimitado superiormente. Então existe uma sequência $\{\mu_k\}$ em $(0, 1]$

convergindo a 0, tal que $\lim_{k \rightarrow \infty} \|\Delta s_{J_\ell}(\mu_k)\| = \infty$. E $\frac{\Delta s_{J_\ell}(\mu_k)}{\|\Delta s_{J_\ell}(\mu_k)\|} \in \mathbb{S}^{|J_\ell|-1}$, como $\mathbb{S}^{|J_\ell|-1}$ é compacto pelo Teorema de Bolzano-Weierstrass, a menos de subsequência, temos:

$$\lim_{k \rightarrow \infty} \frac{\Delta s_{J_\ell}(\mu_k)}{\|\Delta s_{J_\ell}(\mu_k)\|} = \mathbf{d}_{J_\ell}^\infty \in \mathbb{S}^{|J_\ell|-1}. \quad (3.10)$$

Afirmamos que $\mathbf{d}_{J_\ell}^\infty$ é não-negativo. De fato, se existisse $j \in J_\ell$ tal que $\mathbf{d}_j^\infty < 0$, por (3.10) segue que, para n suficientemente grande, teríamos:

$$\frac{\Delta s_j(\mu_k)}{\|\Delta s_{J_\ell}(\mu_k)\|} < \frac{\mathbf{d}_j^\infty}{2} < 0 \text{ e}$$

$$\Delta s_j(\mu_k) < \frac{\mathbf{d}_j^\infty}{2} \|\Delta s_{J_\ell}(\mu_k)\|.$$

Como $\lim_{k \rightarrow \infty} \frac{\mathbf{d}_j^\infty}{2} \|\Delta s_{J_\ell}(\mu_k)\| = -\infty$, a última desigualdade geraria uma contradição com o fato já provado que o conjunto $\{\Delta s_{J_\ell}(\mu) : \mu \in (0, 1]\}$ é limitado inferiormente. Agora, usando que o conjunto $\{\Delta s_{B_{\ell-1}}(\mu) : \mu \in (0, 1]\}$ é limitado e que $\lim_{k \rightarrow \infty} \|\Delta s_{J_\ell}(\mu_k)\| = \infty$, obtemos que:

$$\lim_{k \rightarrow \infty} \frac{\Delta s_{B_{\ell-1}}(\mu_k)}{\|\Delta s_{J_\ell}(\mu_k)\|} = 0. \quad (3.11)$$

Novamente pelo Lema (1.3), como $\Delta s(\mu_k) = (\Delta s_{B_\ell}(\mu_k), \Delta s_{N_\ell/J_\ell}(\mu_k))^T \in \text{Im } \mathbf{A}^T$ existe $\tilde{\Delta} s_{N_\ell/J_\ell}(\mu_k)$ satisfazendo $\tilde{\Delta} s(\mu_k) \equiv (\Delta s_{B_\ell}(\mu_k), \tilde{\Delta} s_{N_\ell/J_\ell}(\mu_k))^T \in \text{Im } \mathbf{A}^T$ e

$$\|\tilde{\Delta} s_{N_\ell/J_\ell}(\mu_k)\| \leq M \|\Delta s_{B_\ell}(\mu_k)\| \leq M(\|\Delta s_{B_{\ell-1}}(\mu_k)\| + \|\Delta s_{J_\ell}(\mu_k)\|).$$

Logo,

$$\frac{\|\tilde{\Delta} s_{N_\ell/J_\ell}(\mu_k)\|}{\|\Delta s_{J_\ell}(\mu_k)\|} \leq M \left(\frac{\|\Delta s_{B_{\ell-1}}(\mu_k)\|}{\|\Delta s_{J_\ell}(\mu_k)\|} + 1 \right).$$

Dai, por (3.11) para n suficientemente grande, vale:

$$\frac{\|\tilde{\Delta} s_{N_\ell/J_\ell}(\mu_k)\|}{\|\Delta s_{J_\ell}(\mu_k)\|} \leq 2M.$$

Pelo Teorema de Bolzano-Weierstrass existe $\mathbf{d}_{N_\ell/J_\ell}^\infty \in \mathbb{R}^{|N_\ell/J_\ell|}$ tal que, a menos de subsequência, satisfaz:

$$\lim_{k \rightarrow \infty} \frac{\tilde{\Delta} s_{N_\ell/J_\ell}(\mu_k)}{\|\Delta s_{J_\ell}(\mu_k)\|} = \mathbf{d}_{N_\ell/J_\ell}^\infty. \quad (3.12)$$

Como $\tilde{\Delta} s(\mu_k) \in \text{Im } \mathbf{A}^T$, segue que:

$$\frac{\tilde{\Delta} s(\mu_k)}{\|\Delta s_{J_\ell}(\mu_k)\|} = \begin{pmatrix} \frac{\Delta s_{B_{\ell-1}}(\mu_k)}{\|\Delta s_{J_\ell}(\mu_k)\|} \\ \frac{\Delta s_{J_\ell} s(\mu_k)}{\|\Delta s_{J_\ell}(\mu_k)\|} \\ \frac{\Delta s_{N_\ell/J_\ell} s(\mu_k)}{\|\Delta s_{J_\ell}(\mu_k)\|} \end{pmatrix} \in \text{Im } \mathbf{A}^T. \quad (3.13)$$

Fazendo $k \rightarrow \infty$ em (3.13), e usando (3.10), (3.11) e (3.12) juntamente com a Proposição 1.1, concluímos que

$$\mathbf{d}^\infty = \lim_{k \rightarrow \infty} \frac{\tilde{\Delta} \mathbf{s}(\mu_k)}{\|\Delta \mathbf{s}_{J_\ell}(\mu_k)\|} \equiv \begin{pmatrix} 0 \\ \mathbf{d}_{J_\ell}^\infty \\ \mathbf{d}_{N_\ell/J_\ell}^\infty \end{pmatrix} \in \text{Im } \mathbf{A}^\top.$$

Considere o vetor $\mathbf{s}^\lambda \equiv \mathbf{s}^c + \lambda \mathbf{d}^\infty$, com $\lambda > 0$. Afirmamos que $\mathbf{s}^\lambda \in \mathbf{S}^*$, para λ suficientemente pequeno. De fato, vale que:

- i) Como $\mathbf{s}^c \in \mathbf{c} + \text{Im } \mathbf{A}^\top$ e $\mathbf{d}^\infty \in \text{Im } \mathbf{A}^\top$, segue que $\mathbf{s}^\lambda \in \mathbf{c} + \text{Im } \mathbf{A}^\top$;
- ii) Dado $\bar{\mathbf{x}} \in \mathbf{X}^*$ vale $\bar{\mathbf{x}}_N = 0$, $\mathbf{s}_B^c = 0$ e sendo $B \subset B_{\ell-1}$, temos $\mathbf{d}_B^\infty = 0$. Então, $\bar{\mathbf{x}}^\top \mathbf{s}^\lambda = \bar{\mathbf{x}}_B^\top \mathbf{s}_B^\lambda + \bar{\mathbf{x}}_N^\top \mathbf{s}_N^\lambda = \bar{\mathbf{x}}_B^\top \mathbf{s}_B^c = 0$;
- iii) Usando o fato de $\mathbf{d}_B^\infty = 0$, obtemos $\mathbf{s}_B^\lambda = \mathbf{s}_B^c = 0$. Pela Proposição 2.3, obtemos $\mathbf{s}_N^c \geq \mathbf{v}_1^* \mathbf{e} > 0$ e para $\lambda \in (0, \varepsilon)$ com ε suficientemente pequeno, temos $-\frac{\mathbf{v}_1^*}{2} \mathbf{e} < \lambda \mathbf{d}_N^\infty < \frac{\mathbf{v}_1^*}{2} \mathbf{e}$. Logo $\mathbf{s}_N^\lambda \geq \frac{\mathbf{v}_1^*}{2} \mathbf{e} > 0$ e daí, $\mathbf{s}^\lambda \geq 0$ para todo $\lambda \in (0, \varepsilon)$.
De i), ii) e iii) segue a afirmação.

Usando que $\mathbf{d}_{B_{\ell-1}}^\infty = 0$, $\mathbf{d}_{J_\ell}^\infty \geq 0$ e $\mathbf{v}_k^* > \mathbf{v}_\ell^*$ para todo $k > \ell$, concluímos, pela Proposição 2.3, que:

$$\begin{aligned} \mathbf{s}_{J_k}^\lambda &= \mathbf{s}_{J_k}^c \geq \mathbf{v}_k^* \mathbf{e}, \quad \forall k < \ell; \\ \mathbf{s}_{J_\ell}^\lambda &= \mathbf{s}_{J_\ell}^c + \lambda \mathbf{d}_{J_\ell}^\infty \geq \mathbf{s}_{J_\ell}^c \geq \mathbf{v}_\ell^* \mathbf{e}; \\ \mathbf{s}_{J_k}^\lambda &= \mathbf{s}_{J_k}^c + \lambda \mathbf{d}_{J_k}^\infty \geq \mathbf{s}_{J_k}^c \geq \mathbf{v}_\ell^* \mathbf{e}, \quad \forall k > \ell. \end{aligned}$$

Então, $\mathbf{s}^\lambda \in \mathbf{S}_\ell^*$ para todo $\lambda \in (0, \varepsilon)$. Novamente pela Proposição 2.3 segue, para todo $\lambda \in (0, \varepsilon)$, que: $\mathbf{v}_\ell^* = \mathbf{s}_{J_\ell}^\lambda = \mathbf{s}_{J_\ell}^c + \lambda \mathbf{d}_{J_\ell}^\infty = \mathbf{v}_\ell^* + \lambda \mathbf{d}_{J_\ell}^\infty$. Portanto, devemos ter obrigatoriamente $\mathbf{d}_{J_\ell}^\infty = 0$, o que gera um absurdo com (3.10), e segue o lema. \square

O próximo teorema, finalmente, mostra a convergência de $\Delta \mathbf{s}(\mu)$ quando $\mu \rightarrow 0+$, e ainda caracteriza o limite.

Teorema 3.1. $\lim_{\mu \rightarrow 0} \Delta \mathbf{s}(\mu) = \Delta \mathbf{s}^\infty$, onde:

- i) $\Delta \mathbf{s}_B^\infty \equiv \log \mathbf{x}_B^1 - \log \mathbf{x}_B^*$;

ii) para todo $\ell = 1, \dots, r$, $\Delta s_{J_\ell}^\infty$ é indutivamente a única solução do problema

$$\min_{\Delta s_{J_\ell}} \left\{ \sum_{j \in J_\ell} x_j^1 e^{-\Delta s_j} : \begin{pmatrix} \Delta s_{B_{\ell-1}}^\infty \\ \Delta s_{J_\ell} \end{pmatrix} \in \text{Im} \begin{pmatrix} A_{B_{\ell-1}}^\top \\ A_{J_\ell}^\top \end{pmatrix} \right\}. \quad (3.14)$$

Demonstração. A assertiva i) segue imediatamente por (3.2). Resta provarmos ii), o que será feito por indução em ℓ . De fato, assumamos que ii) vale para os índices $1, \dots, \ell - 1$. Portanto,

$$\lim_{\mu \rightarrow 0} \Delta s_{B_{\ell-1}}(\mu) = \Delta s_{B_{\ell-1}}^\infty. \quad (3.15)$$

Pelo Lema 3.2 o conjunto $\{\Delta_{J_\ell} s(\mu) : \mu \in (0, 1]\}$ é limitado. Seja $d_{J_\ell}^\infty$ tal que, existe $\{\mu_k\}$ com $\lim_{k \rightarrow \infty} \mu_k = 0$ e $\lim_{k \rightarrow \infty} \Delta s_{J_\ell}(\mu_k) = d_{J_\ell}^\infty$ (observe que a existência de $d_{J_\ell}^\infty$ é garantida pelo Teorema 1.2). Como $\Delta s(\mu_k) \in \text{Im } A^\top$ temos $\begin{pmatrix} \Delta s_{B_{\ell-1}}(\mu_k) \\ \Delta s_{J_\ell}(\mu_k) \end{pmatrix} \in \text{Im} \begin{pmatrix} A_{B_{\ell-1}}^\top \\ A_{J_\ell}^\top \end{pmatrix}$, fazendo $k \rightarrow \infty$ e usando (3.15) obtemos, pela Proposição 1.1, que $\begin{pmatrix} \Delta s_{B_{\ell-1}}^\infty \\ d_{J_\ell}^\infty \end{pmatrix} \in \text{Im} \begin{pmatrix} A_{B_{\ell-1}}^\top \\ A_{J_\ell}^\top \end{pmatrix}$. Portanto, $d_{J_\ell}^\infty$ é viável ao problema (3.14). Pelo Lema 3.1, $\Delta s_{N_\ell}(\mu)$ é a solução ótima do problema (3.3) com $J = N_\ell$. Então, das condições de otimalidade deste problema, dadas pelo Corolário 1.5, temos:

$$A_{N_\ell} \left(x_{N_\ell}^1 e^{-\Delta s_{N_\ell}(\mu) - s_{N_\ell}^c / \mu} \right) \in \text{Im } A_{B_{\ell-1}}. \quad (3.16)$$

Multiplicando (3.16) por $e^{v_i^* / \mu}$, como $s_{J_i}^c = v_i^* e$ para todo $i = 1, \dots, r$, ganhamos que:

$$A_{J_\ell} \left(x_{J_\ell}^1 e^{-\Delta s_{J_\ell}(\mu)} \right) + \sum_{i=\ell+1}^r A_{J_i} \left(x_{J_i}^1 e^{-\Delta s_{J_i}(\mu) - (v_i^* - v_\ell^*) / \mu} \right) \in \text{Im } A_{B_{\ell-1}}.$$

Fazendo $\mu = \mu_k$ na expressão acima e usando o Lema 3.2 mais o fato que $(v_i^* - v_\ell^*) > 0$ para todo $i > \ell$, obtemos que $\lim_{k \rightarrow \infty} \left(x_{J_i}^1 e^{-\Delta s_{J_i}(\mu_k) - (v_i^* - v_\ell^*) e / \mu_k} \right) = 0$ quando $i > \ell$. Agora, fazendo $k \rightarrow \infty$ pela continuidade das A_{J_i} e pela Proposição 1.1, obtemos:

$$A_{J_\ell} \left(x_{J_\ell}^1 e^{-d_{J_\ell}^\infty} \right) \in \text{Im } A_{B_{\ell-1}}. \quad (3.17)$$

Sendo $f_\ell : \mathbb{R}^{|J_\ell|} \rightarrow \mathbb{R}$ definida por $f_\ell(x) = \sum_{j \in J_\ell} x_j^1 e^{-x_j}$. Daí, $A_{J_\ell}(\nabla f_\ell(d_{J_\ell}^\infty)) \in \text{Im } A_{B_{\ell-1}}$ por (3.17). Então, pelo Corolário 1.5, segue que $d_{J_\ell}^\infty$ é solução do problema (3.14). Como $\text{Hess}(f_\ell) > 0$, o Teorema 1.8 implica que f_ℓ é estritamente convexa e obtemos assim a unicidade de $d_{J_\ell}^\infty$. Portanto, $\lim_{\mu \rightarrow 0} \Delta s_{J_\ell}(\mu) = \Delta s_{J_\ell}^\infty \equiv d_{J_\ell}^\infty$. \square

O objetivo agora é mostrar que o limite da derivada $\dot{s}(\mu)$ quando $\mu \rightarrow 0$ existe e é igual ao limite de $\Delta s(\mu)$ quando $\mu \rightarrow 0$. Com este intuito, os próximos dois lemas tratam

sobre a diferença

$$\Delta \mathbf{u}(\mu) = \dot{\mathbf{s}}(\mu) - \Delta \mathbf{s}(\mu), \quad (3.18)$$

e sobre seu limite quando $\mu \rightarrow 0$.

Lema 3.3. Para cada $\mu > 0$ e $\ell = 0, \dots, r$, $\Delta \mathbf{u}_{N_\ell}(\mu)$ é a única solução do problema

$$\min_{\mathbf{p}_{N_\ell}} \left\{ \frac{1}{2} \left\| \mathbf{X}_{N_\ell}^{1/2}(\mu) (\mathbf{p}_{N_\ell} - (\mathbf{s}_{N_\ell}^c - \mathbf{v}_\ell^* \mathbf{e}) / \mu) \right\|^2 : \begin{pmatrix} \Delta \mathbf{u}_{B_{\ell-1}}(\mu) \\ \mathbf{p}_{N_\ell} \end{pmatrix} \in \text{Im } \mathbf{A}^\top \right\}, \quad (3.19)$$

onde $B_{-1} = \emptyset$.

Demonstração. Como $\mathbf{s}(\mu) \in \text{Im } \mathbf{A}^\top$, para todo $\mu > 0$, segue que $\dot{\mathbf{s}}(\mu) \in \text{Im } \mathbf{A}^\top$, pelo Corolário 1.1. Daí, de (3.18) obtemos que $\Delta \mathbf{u}(\mu) = (\Delta \mathbf{u}_{B_{\ell-1}}(\mu), \Delta \mathbf{u}_{N_\ell}(\mu))^\top \in \text{Im } \mathbf{A}^\top$, pois $\Delta \mathbf{s}(\mu) \in \text{Im } \mathbf{A}^\top$. Deste modo, $\Delta \mathbf{u}_{N_\ell}(\mu)$ é viável para o problema (3.19), para todo $\ell = 0, \dots, r$. Agora, derivando (2.12), ganhamos que:

$$\dot{\mathbf{s}}(\mu) = \frac{\mathbf{s}(\mu)}{\mu} - \mu \mathbf{X}^{-1}(\mu) \dot{\mathbf{x}}(\mu),$$

e então

$$\dot{\mathbf{s}}(\mu) - \left(\frac{\mathbf{s}(\mu)}{\mu} - \frac{\mathbf{s}^c}{\mu} \right) - \frac{\mathbf{s}^c}{\mu} = -\mu \mathbf{X}^{-1}(\mu) \dot{\mathbf{x}}(\mu).$$

De (3.1) e (3.18) obtemos:

$$\Delta \mathbf{u}(\mu) - \frac{\mathbf{s}^c}{\mu} = -\mu \mathbf{X}^{-1}(\mu) \dot{\mathbf{x}}(\mu).$$

Daí,

$$\Delta \mathbf{u}(\mu) - \frac{(\mathbf{s}^c - \mathbf{v}_\ell^* \mathbf{e})}{\mu} = -\mu \mathbf{X}^{-1}(\mu) \dot{\mathbf{x}}(\mu) + \frac{\mathbf{v}_\ell^* \mathbf{e}}{\mu}.$$

Sendo assim, dado $\mathbf{x}^* \in \mathbf{X}^*$ segue que:

$$\begin{aligned} \mathbf{X}(\mu) \left(\Delta \mathbf{u}(\mu) - \frac{(\mathbf{s}^c - \mathbf{v}_\ell^* \mathbf{e})}{\mu} \right) &= -\mu \dot{\mathbf{x}}(\mu) + \frac{\mathbf{v}_\ell^*}{\mu} \mathbf{x}(\mu) \\ &= -\mu \dot{\mathbf{x}}(\mu) + \frac{\mathbf{v}_\ell^*}{\mu} (\mathbf{x}(\mu) - \mathbf{x}^*) + \frac{\mathbf{v}_\ell^*}{\mu} \mathbf{x}^*. \end{aligned}$$

Como $\mathbf{A} \mathbf{x}^* = \mathbf{A} \mathbf{x}(\mu) = \mathbf{b}$, temos $(\mathbf{x}(\mu) - \mathbf{x}^*)$ pertencente a $\text{Null}(\mathbf{A})$. Novamente pelo Corolário 1.1, vale que $\dot{\mathbf{x}}(\mu) \in \text{Null}(\mathbf{A})$. Estes fatos implicam em:

$$\mathbf{X}(\mu) \left(\Delta \mathbf{u}(\mu) - \frac{(\mathbf{s}^c - \mathbf{v}_\ell^* \mathbf{e})}{\mu} \right) \in \text{Null}(\mathbf{A}) + \frac{\mathbf{v}_\ell^*}{\mu} \mathbf{x}^*.$$

Visto que $\mathbf{x}_N^* = 0$ e $\mathbf{v}_0^* = 0$ temos $\mathbf{v}_\ell^* \mathbf{x}^* = 0$ para cada $\ell = 0, 1, \dots, r$. Portanto,

$$\Gamma(\mu) \equiv \mathbf{X}(\mu) \left(\Delta \mathbf{u}(\mu) - \frac{(\mathbf{s}^c - \mathbf{v}_\ell^* \mathbf{e})}{\mu} \right) \in \text{Null}(\mathbf{A}).$$

Daí,

$$A\Gamma(\mu) = [A_{B_{\ell-1}} \quad A_{N_\ell}] \begin{bmatrix} \Gamma_{B_{\ell-1}}(\mu) \\ \Gamma_{N_\ell}(\mu) \end{bmatrix} = 0 \text{ e}$$

$$A_{N_\ell} \left[X_{N_\ell}(\mu) \left(\Delta u_{N_\ell}(\mu) - \frac{(s_{N_\ell}^c - v_\ell^* e)}{\mu} \right) \right] = A_{N_\ell}[\Gamma_{N_\ell}(\mu)] = A_{B_{\ell-1}}[-\Gamma_{B_{\ell-1}}] \in \text{Im } A_{B_{\ell-1}}.$$

Sendo $f_\ell : \mathbb{R}^{N_\ell} \rightarrow \mathbb{R}$ definida por $f_\ell(x) = \frac{1}{2} \left\| X_{N_\ell}^{1/2}(\mu) (x - (s_{N_\ell}^c - v_\ell^* e) / \mu) \right\|^2$. Pela expressão acima, segue que $A_{N_\ell}(\nabla f_\ell(\Delta u_{N_\ell}(\mu))) \in \text{Im } A_{B_{\ell-1}}$ (no caso $\ell = 0$, temos $A(\nabla f_0(\Delta u(\mu))) = 0$) e então, pelo Corolário 1.5 segue que $\Delta u_{N_\ell}(\mu)$ é solução do problema (3.19). Como $\text{Hess}(f_\ell) > 0$, do Teorema 1.8 obtemos que f_ℓ é estritamente convexa, garantindo assim a unicidade de $\Delta u_{N_\ell}(\mu)$ como solução do problema. \square

O próximo resultado mostra que $\dot{s}(\mu)$ e $\Delta s(\mu)$ possuem o mesmo limite quando $\mu \rightarrow 0$.

Lema 3.4. $\lim_{\mu \rightarrow 0} \Delta u(\mu) = 0$.

Demonstração. Para cada $\ell \in \{0, \dots, r\}$, supondo que

$$\lim_{\mu \rightarrow 0} \Delta u_{B_{\ell-1}}(\mu) = 0, \quad (3.20)$$

concluiremos que $\lim_{\mu \rightarrow 0} \Delta u_{J_\ell}(\mu) = 0$ e daí, segue o lema. (Para $\ell = 0$, (3.20) vale por vacuidade, muito embora para mostrar que $\lim_{\mu \rightarrow 0} \Delta u_B(\mu) = 0$ não usaremos (3.20)).

Pelo Lema 1.3, existe $M = M(\text{Im } A^T) > 0$ e $\{\tilde{\Delta} u_{N_\ell}(\mu) : \mu > 0\}$, tais que:

$$\begin{pmatrix} \Delta u_{B_{\ell-1}}(\mu) \\ \tilde{\Delta} u_{N_\ell}(\mu) \end{pmatrix} \in \text{Im } A^T \text{ e } \|\tilde{\Delta} u_{N_\ell}(\mu)\| \leq M \|\Delta u_{B_{\ell-1}}(\mu)\|, \quad \forall \mu > 0. \quad (3.21)$$

Por (3.20) e (3.21), obtemos:

$$\lim_{\mu \rightarrow 0} \tilde{\Delta} u_{N_\ell}(\mu) = 0. \quad (3.22)$$

Para $\ell = 0$, sem usar o Lema 1.3, podemos tomar simplesmente $\tilde{\Delta} u_{N_0}(\mu) = \tilde{\Delta} u(\mu) = 0$ para todo $\mu > 0$. De (3.21) temos que $\tilde{\Delta} u_{N_\ell}(\mu)$ é viável ao problema (3.19), então pelo Lema 3.3 vale:

$$\left\| X_{N_\ell}^{1/2}(\mu) (\Delta u_{N_\ell}(\mu) - (s_{N_\ell}^c - v_\ell^* e) / \mu) \right\|^2 \leq \left\| X_{N_\ell}^{1/2}(\mu) (\tilde{\Delta} u_{N_\ell}(\mu) - (s_{N_\ell}^c - v_\ell^* e) / \mu) \right\|^2.$$

Desde que $s_{J_\ell}^c = v_\ell^* e$ e $x(\mu) \in \mathbb{R}_{++}^n$, como consequência da última equação, para $k \in J_\ell$, ganhamos:

$$\begin{aligned} x_k(\mu) (\Delta u_k(\mu))^2 &\leq \sum_{i \in J_\ell} x_i(\mu) (\Delta u_i(\mu))^2 + \sum_{i \in N_{\ell+1}} x_i(\mu) \left(\Delta u_i(\mu) - \frac{s_i^c - v_\ell^* e}{\mu} \right)^2 \\ &\leq \sum_{i \in J_\ell} x_i(\mu) (\tilde{\Delta} u_i(\mu))^2 + \sum_{i \in N_{\ell+1}} x_i(\mu) \left(\tilde{\Delta} u_i(\mu) - \frac{s_i^c - v_\ell^* e}{\mu} \right)^2. \end{aligned}$$

De (2.13) e (3.1) temos $x(\mu) = x^1 e^{-\Delta s(\mu) - s^c/\mu}$. Multiplicando a expressão acima por $e^{v_\ell^*/\mu}$ obtemos:

$$\begin{aligned} x_k^1 e^{-\Delta s_k(\mu)} (\Delta u_k(\mu))^2 &\leq \sum_{i \in J_\ell} x_i^1 e^{-\Delta s_i(\mu)} (\tilde{\Delta} u_i(\mu))^2 \\ &+ \sum_{i \in N_{\ell+1}} x_i^1 e^{-\Delta s_i(\mu)} e^{-(s_i^c - v_\ell^*)/\mu} \left(\tilde{\Delta} u_i(\mu) - \frac{s_i^c - v_k^*}{\mu} \right)^2. \end{aligned}$$

Daí,

$$\begin{aligned} 0 \leq (\Delta u_k(\mu))^2 &\leq \frac{1}{x_k^1 e^{-\Delta s_k(\mu)}} \left[\sum_{i \in J_\ell} x_i^1 e^{-\Delta s_i(\mu)} (\tilde{\Delta} u_i(\mu))^2 \right. \\ &\left. + \sum_{i \in N_{\ell+1}} x_i^1 e^{-\Delta s_i(\mu)} e^{-(s_i^c - v_\ell^*)/\mu} \left(\tilde{\Delta} u_i(\mu) - \frac{s_i^c - v_k^*}{\mu} \right)^2 \right]. \end{aligned}$$

Para todo $i \in N_{\ell+1}$, sendo $v_\ell^* < s_i^c$, vale que $(s_i^c - v_\ell^*) > 0$. Fazendo $\mu \rightarrow 0$ na última desigualdade usando o Teorema do Confronto, o Teorema 3.1, a relação (3.22) e o fato que para todo k natural vale $\lim_{x \rightarrow 0^+} \left(\frac{1}{x} \right)^k e^{-(\frac{1}{x})} = \lim_{y \rightarrow \infty} y^k e^{-y} = 0$, concluímos que $\lim_{\mu \rightarrow 0} (\Delta u_k(\mu))^2 = 0$. Portanto, $\lim_{\mu \rightarrow 0} \Delta u_k(\mu) = 0$ para cada $k \in J_\ell$ e daí segue que $\lim_{\mu \rightarrow 0} \Delta u_{J_\ell}(\mu) = 0$. \square

Um dos principais resultados deste capítulo segue diretamente do lema anterior, e é o conteúdo do próximo teorema.

Teorema 3.2. $\lim_{\mu \rightarrow 0} \dot{s}(\mu) = \Delta s^\infty$.

Demonstração. Este resultado segue da relação (3.18), do Teorema 3.1 e do Lema 3.4. \square

3.2 Comportamento Limite de $\Delta x(\mu)$

Agora, o objetivo é estudar o comportamento limite do seguinte quociente

$$\Delta x(\mu) = \frac{x(\mu) - x^*}{e^{-v_1^*/\mu}}, \quad (3.23)$$

onde v_1^* é dado da definição de centróide. O estudo feito aqui terá grande relevância na análise de convergência da sequência proximal primal com distância de Bregman que será definida no próximo capítulo.

Com o intuito de obter resultados sobre a expressão definida acima, estudaremos primeiramente o comportamento limite do seguinte caminho $\mathbf{q} : (0, 1) \rightarrow \mathbb{R}^n$ definido por

$$\mathbf{q}(\mu) = \mu^2 e^{v_1^*/\mu} \dot{\mathbf{x}}(\mu).$$

O resultado seguinte relaciona os limites, quando $\mu \rightarrow 0$, dos dois caminhos definidos acima.

Lema 3.5. *Se $\lim_{\mu \rightarrow 0} \mathbf{q}(\mu) = \mathbf{q}^\infty$, então $\lim_{\mu \rightarrow 0} \Delta \mathbf{x}(\mu) = \mathbf{q}^\infty / v_1^*$.*

Demonstração. Dado $i \in \{1, 2, \dots, n\}$, sejam $f_i, g_i : [0, \mu] \rightarrow \mathbb{R}$ definidas por

$$f_i(r) = \begin{cases} x_i(r) - x_i^*, & \text{se } r \in (0, \mu] \\ 0, & \text{se } r = 0 \end{cases} \quad \text{e} \quad g_i(r) = \begin{cases} e^{-v_1^*/r}, & \text{se } r \in (0, \mu] \\ 0, & \text{se } r = 0 \end{cases}.$$

Fazendo uso das Proposições 2.2 e Proposição 2.4, temos que f_i e g_i são contínuas em $[0, \mu]$ e diferenciáveis em $(0, \mu)$. Pelo Teorema 1.1, existe $\xi_i = \xi_i(\mu) \in (0, \mu)$ tal que:

$$f_i'(\xi_i)[g_i(\mu) - g_i(0)] = g_i'(\xi_i)[f_i(\mu) - f_i(0)],$$

ou seja,

$$\dot{x}_i(\xi_i) e^{-v_1^*/\mu} = \frac{v_1^*}{\xi_i^2} (x_i(\mu) - x_i^*) e^{-v_1^*/\xi_i}.$$

Dai,

$$\frac{(x_i(\mu) - x_i^*)}{e^{-v_1^*/\mu}} = \frac{1}{v_1^*} \xi_i^2 e^{v_1^*/\xi_i} \dot{x}_i(\xi_i).$$

Portanto, $\Delta x_i(\mu) = \frac{q_i(\xi_i)}{v_1^*}$. Sendo $\mu \rightarrow 0$ temos $\xi_i \rightarrow 0$, então se $\lim_{\mu \rightarrow 0} q_i(\mu) = q_i^\infty$, concluímos que $\lim_{\mu \rightarrow 0} \Delta x_i(\mu) = \frac{q_i^\infty}{v_1^*}$ e segue o resultado. \square

Vamos mostrar que $\mathbf{q}(\mu)$ possui limite não-nulo quando $\mu \rightarrow 0$. Para tanto, primeiro enunciaremos uma proposição que caracteriza $\mathbf{q}(\mu)$.

Proposição 3.1. *As seguinte afirmações são verdadeiras:*

i) $\lim_{\mu \rightarrow 0} \mathbf{q}_N(\mu) = \mathbf{q}_N^\infty$, onde $\mathbf{q}_{N/J_1}^\infty = 0$ e $\mathbf{q}_{J_1}^\infty = v_1^* \mathbf{x}_{J_1}^1 e^{-\Delta s_{J_1}^\infty}$;

ii) para cada $\mu > 0$ temos que $\mathbf{q}_B(\mu)$ é a única solução do problema

$$\min \left\{ (1/2) \|X_B^{-1/2}(\mu) \mathbf{u}_B\|^2 : \begin{pmatrix} \mathbf{u}_B \\ \mathbf{q}_N(\mu) \end{pmatrix} \in \text{Null } A \right\}; \quad (3.24)$$

iii) o conjunto $\{\mathbf{q}(\mu) : \mu \in (0, 1]\}$ permanece limitado quando $\mu \rightarrow 0$.

Demonstração. Para provar o item i), derivando (2.13) obtemos:

$$\mu^2 \dot{x}(\mu) = x^1 e^{-\frac{s(\mu)}{\mu}} (s(\mu) - \mu \dot{s}(\mu)).$$

Multiplicando esta expressão por $e^{\frac{v_1^*}{\mu}}$, temos:

$$q(\mu) = x^1 e^{-\frac{s(\mu)}{\mu}} e^{\frac{s^c}{\mu}} e^{-\frac{(s^c - v_1^*)}{\mu}} (s(\mu) - \mu \dot{s}(\mu)).$$

Daí,

$$q(\mu) = x^1 e^{-\Delta s(\mu)} e^{-(s^c - v_1^*)/\mu} (s(\mu) - \mu \dot{s}(\mu)). \quad (3.25)$$

Como $0 = v_0^* < v_1^* < v_2^* \cdots < v_r^*$, segue que:

$$\lim_{\mu \rightarrow 0} e^{-\frac{(s_i^c - v_1^*)}{\mu}} = \begin{cases} 1, & \text{se } i \in J_1 \\ 0, & \text{se } i \in N/J_1 \end{cases}.$$

Assim, da expressão (3.25) usando o Teorema 3.1, o Teorema 2.1 e o Teorema 3.2, obtemos:

$$\lim_{\mu \rightarrow 0} q_i(\mu) = \begin{cases} v_1^* x_i^1 e^{-\Delta s_i^\infty}, & \text{se } i \in J_1 \\ 0, & \text{se } i \in N/J_1 \end{cases}.$$

Com o intuito de provar ii) note que dado $s \in S^*$ temos $s \in c + \text{Im } A^T$ e $s_B = 0$, logo $c_B \in \text{Im } A_B^T$. Como $s(\mu) \in c + \text{Im } A^T$, segue que $s_B(\mu) \in c_B + \text{Im } A_B^T \subset \text{Im } A_B^T$, logo:

$$\Gamma(\mu) \equiv (\log x_B(\mu) - \log x_B^1) = -\frac{1}{\mu} s_B(\mu) \in \text{Im } A_B^T.$$

Pelo Corolário 1.1 obtemos que $\dot{\Gamma}(\mu) = X_B^{-1}(\mu) \dot{x}_B(\mu) \in \text{Im } A_B^T$. E então,

$$X_B^{-1}(\mu) q_B(\mu) = (\mu^2 e^{\frac{v_1^*}{\mu}}) X_B^{-1}(\mu) \dot{x}_B(\mu) \in \text{Im } A_B^T. \quad (3.26)$$

Observe que $q_B(\mu)$ é viável ao problema (3.24), já que $\dot{x}(\mu) \in \text{Null } A$ e assim temos $q(\mu) \in \text{Null } A$. Seja $f : \mathbb{R}^{|B|} \rightarrow \mathbb{R}$ dada por $f(x) = (1/2) \|X_B^{-1/2}(\mu)x\|^2$. Da expressão (3.26), segue que $\nabla f(q_B(\mu)) \in \text{Im } A_B^T$, então o Corolário 1.4 garante que $q_B(\mu)$ é solução do problema. Como $\text{Hess}(f) = X_B^{-1} > 0$, pelo Teorema 1.8 temos que f é estritamente convexa, garantindo assim a unicidade de $q_B(\mu)$ como solução do problema.

Finalmente, para provar iii), resta mostrarmos que $\{q_B(\mu) : \mu \in (0, 1]\}$ permanece limitado quando $\mu \rightarrow 0$. Sabemos que $(q_B, q_N)^T \in \text{Null } A$ e usando o Lema 1.3 no subespaço $\text{Null } A$, existem $M > 0$ e $\tilde{q}_B(\mu)$ tal que $(\tilde{q}_B, q_N)^T \in \text{Null } A$, satisfazendo a desigualdade

$$\|\tilde{q}_B(\mu)\| \leq M \|q_N(\mu)\|, \text{ para todo } \mu \in (0, 1]. \quad (3.27)$$

Agora, por (3.24) e (3.27), temos:

$$\begin{aligned} \|\mathbf{q}_B(\mu)\| &= \|\mathbf{X}_B^{1/2}(\mu)\mathbf{X}_B^{-1/2}(\mu)\mathbf{q}_B(\mu)\| \leq \|\mathbf{X}_B^{1/2}(\mu)\| \|\mathbf{X}_B^{-1/2}(\mu)\mathbf{q}_B(\mu)\| \\ &\leq \|\mathbf{X}_B^{1/2}(\mu)\| \|\mathbf{X}_B^{-1/2}(\mu)\tilde{\mathbf{q}}_B(\mu)\| \leq M \|\mathbf{X}_B^{1/2}(\mu)\| \|\mathbf{X}_B^{-1/2}(\mu)\| \|\mathbf{q}_N(\mu)\|, \end{aligned}$$

para todo $\mu \in (0, 1]$. Pelo Teorema 2.4 vale que $\lim_{\mu \rightarrow 0} \mathbf{x}_B(\mu) = \mathbf{x}_B^* \neq 0$ e deste fato juntamente com i), segue que o lado direito da expressão acima permanece limitado quando $\mu \rightarrow 0$, e segue o resultado. \square

Observação 3.1. Na demonstração do item iii) da proposição acima usamos, sem fazer menção explícita, que no conjunto das matrizes $\mathbb{R}^{m \times n}$ está bem-definida a seguinte norma:

$$\|\mathbf{C}\| = \sup_{\mathbf{x} \in \mathbb{R}^n / \{0\}} \frac{\|\mathbf{C}\mathbf{x}\|}{\|\mathbf{x}\|}, \text{ para toda } \mathbf{C} \in \mathbb{R}^{m \times n}.$$

E ainda, que dadas $\mathbf{C} \in \mathbb{R}^{m \times n}$ e $\mathbf{D} \in \mathbb{R}^{m \times k}$ valem as seguintes propriedades:

a) $\|\mathbf{C}\mathbf{D}\| \leq \|\mathbf{C}\| \|\mathbf{D}\|;$

b) $\|\mathbf{C}\mathbf{x}\| \leq \|\mathbf{C}\| \|\mathbf{x}\|$, para todo $\mathbf{x} \in \mathbb{R}^n$.

Teorema 3.3. $\lim_{\mu \rightarrow 0} \mathbf{q}(\mu) = \mathbf{q}^\infty$, onde $\mathbf{q}_N^\infty \neq 0$ e \mathbf{q}_B^∞ é a única solução do problema

$$\min \left\{ (1/2) \|(\mathbf{X}_B^*)^{-1/2} \mathbf{u}_B\|^2 : \begin{pmatrix} \mathbf{u}_B \\ \mathbf{q}_N^\infty \end{pmatrix} \in \text{Null } \mathbf{A} \right\}. \quad (3.28)$$

Consequentemente, $\lim_{\mu \rightarrow 0} \Delta \mathbf{x}(\mu) = \mathbf{q}^\infty / \mathbf{v}_1^*$.

Demonstração. Pelo item i) da Proposição 3.1, temos $\lim_{\mu \rightarrow 0} \mathbf{q}_N(\mu) = \mathbf{q}_N^\infty \neq 0$. Seja \mathbf{d}_B^∞ um ponto de acumulação do conjunto $\{\mathbf{q}_B(\mu) : \mu \in (0, 1]\}$, tal que existe uma sequência $\{\mu_k\} \subset \mathbb{R}_{++}^n$, com $\lim_{k \rightarrow \infty} \mu_k = 0$ e $\lim_{k \rightarrow \infty} \mathbf{q}_B(\mu_k) = \mathbf{d}_B^\infty$ (A existência de um tal ponto é garantida pelo item iii) da mesma Proposição 3.1 juntamente com o Teorema 1.2). Desde que $\mathbf{q}(\mu_k) = (\mathbf{q}_B(\mu_k), \mathbf{q}_N(\mu_k))^T \in \text{Null } \mathbf{A}$, fazendo $k \rightarrow \infty$ obtemos que $(\mathbf{d}_B^\infty, \mathbf{q}_N^\infty) \in \text{Null } \mathbf{A}$ e portanto, \mathbf{d}_B^∞ é viável ao problema (3.28). Passando o limite em (3.26) pelo Teorema 2.4 e pelo Corolário 1.1, temos:

$$(\mathbf{X}_B^*)^{-1} \mathbf{d}_B^\infty \in \text{Im } \mathbf{A}_B^T. \quad (3.29)$$

Sendo $f : \mathbb{R}^{|\mathbf{B}|} \rightarrow \mathbb{R}$ definida por $f(\mathbf{x}) = (1/2) \|(\mathbf{X}_B^*)^{-1/2}(\mathbf{x})\|^2$. De (3.29), vale que $\nabla f(\mathbf{d}_B^\infty) \in \text{Im } \mathbf{A}_B^T$, então pelo Corolário 1.4 temos que \mathbf{d}_B^∞ é solução do problema (3.28). Sendo $\text{Hess}(f) = (\mathbf{X}_B^*)^{-1} > 0$, segue mais uma vez do Teorema 1.8 que f é estritamente convexa e portanto, \mathbf{d}_B^∞ é a única solução. Assim, concluímos que $\lim_{\mu \rightarrow 0} \mathbf{q}_B(\mu) = \mathbf{q}_B^\infty \equiv \mathbf{d}_B^\infty$. A parte final do teorema segue do Lema 3.5. \square

Capítulo 4

Método do ponto proximal com distância de Bregman

Finalmente, neste capítulo definiremos e estudaremos o método do ponto proximal associado à barreira de Kullback-Leibler. Usando as ferramentas desenvolvidas nos capítulos anteriores para as trajetórias primal e dual mostraremos a convergência da sequência proximal primal (definida a posteriori) para uma solução do problema de programação linear e que as sequências medial dual e proximal dual (definidas a seguir) convergem ambas para o centróide de S^* . Por fim, encerraremos fazendo um estudo sobre as velocidades destas convergências.

4.1 Sequências Proximais Primal e Dual

O método do ponto proximal com distância de Bregman D_φ para resolver o problema (2.1) gera uma sequência $\{x^k\}$, definida como sendo:

$$x^1 \in D^0, \quad (4.1)$$

$$x^{k+1} = \operatorname{argmin} \{c^T x + \lambda_k D_\varphi(x, x^k) : Ax = b, x \geq 0\}, \quad (4.2)$$

onde neste trabalho φ é dada particularmente no Exemplo 1.7, ou seja, a função barreira D_φ será a divergência de Kullback-Leibler e a sequência $\{\lambda_k\}$ satisfaz

$$\lambda_k > 0 \quad \text{e} \quad \sum_{k=0}^{\infty} \lambda_k^{-1} = \infty. \quad (4.3)$$

Neste caso, para todo $k \geq 1$, temos

$$\mathbf{x}^{k+1} = \operatorname{argmin} \left\{ \mathbf{c}^\top \mathbf{x} + \lambda_k \sum_{j=1}^n \left(x_j \log \frac{x_j}{x_j^k} + x_j^k - x_j \right) : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq 0 \right\}.$$

A saber, a sequência $\{\mathbf{x}^k\}$ é chamada de *Sequência Proximal Primal (SPP)* associada a D_φ . Usando o raciocínio da Proposição 2.1 indutivamente, concluímos que $\{\mathbf{x}^k\}$ está bem-definida e inteiramente contida em D^0 .

Assim, como no Capítulo 3, consideraremos a seguinte extensão da função barreira $D_\varphi(\cdot, \mathbf{x}^k)$ ao \mathbb{R}^n que ainda será representada pela mesma notação:

$$D_\varphi(\mathbf{x}, \mathbf{x}^k) = \begin{cases} \varphi(\mathbf{x}) - \varphi(\mathbf{x}^k) - \nabla \varphi(\mathbf{x}^k)^\top (\mathbf{x} - \mathbf{x}^k), & \text{se } \mathbf{x} \in \mathbb{R}_+^n \\ +\infty, & \text{caso contrário} \end{cases}.$$

Daí, o problema (4.2) equivale ao problema abaixo:

$$\mathbf{x}^{k+1} = \operatorname{argmin} \{ \mathbf{c}^\top \mathbf{x} + \lambda_k D_\varphi(\mathbf{x}, \mathbf{x}^k) : \mathbf{A}\mathbf{x} = \mathbf{b} \}. \quad (4.4)$$

Como $\mathbf{x}^{k+1} > 0$ é a única solução do problema (4.4), e de sorte que em \mathbb{R}_{++}^n a função objetivo em questão é diferenciável, por **B1**) da definição de função de Bregman. Sendo assim, pelo Teorema 1.9 e pela Proposição 1.7 as condições de otimalidade de Karush-Kuhn-Tucker (KKT) para o problema (4.4) são necessárias e suficientes, ou seja, \mathbf{x}^{k+1} é o único ponto em \mathbb{R}_{++}^n que para algum $\mathbf{w}^{k+1} \in \mathbb{R}^m$, satisfaz:

$$\mathbf{c} + \lambda_k (\nabla \varphi(\mathbf{x}^{k+1}) - \nabla \varphi(\mathbf{x}^k)) + \mathbf{A}^\top \mathbf{w}^{k+1} = \mathbf{0} \quad \text{e} \quad (4.5)$$

$$\mathbf{A}\mathbf{x}^{k+1} = \mathbf{b}. \quad (4.6)$$

Motivados por estas condições de otimalidade, definimos a sequência $\{\mathbf{s}_k\}$, chamada de *Sequência Proximal Dual (SPD)* associada a D_φ , como sendo

$$\mathbf{s}^k \equiv \lambda_k [\nabla \varphi(\mathbf{x}^k) - \nabla \varphi(\mathbf{x}^{k+1})] = \lambda_k [\log \mathbf{x}^k - \log \mathbf{x}^{k+1}].$$

Por (4.5), temos que $\mathbf{s}^k \in \mathbf{c} + \operatorname{Im} \mathbf{A}^\top$, mas não necessariamente $\mathbf{s}^k \geq 0$.

Neste trabalho, estamos interessados nas propriedades e na convergência da sequência proximal dual, mas ao invés de analisar estes fatos diretamente, analisaremos primeiro o comportamento limite da *Sequência Média Dual (SMD)* denotada por $\{\bar{\mathbf{s}}^k\}$ e definida como sendo a média ponderada abaixo:

$$\bar{\mathbf{s}}^k = \sum_{i=1}^k \mu_k \lambda_i^{-1} \mathbf{s}^i, \quad \text{onde } \mu_k = \left(\sum_{i=1}^k \lambda_i^{-1} \right)^{-1}. \quad (4.7)$$

A próxima proposição faz uma importante conexão entre as trajetórias $x(\mu)$ e $s(\mu)$ com as sequências $\{x^k\}$ e $\{\bar{s}^k\}$. A saber, como mostrado pela primeira vez por Cruz Neto, Iusem e Svaiter em [9], temos que a sequência proximal primal está contida na trajetória primal, assim como a sequência média dual está contida na trajetória dual. Sendo assim, podemos utilizar a maquinaria desenvolvida nos Capítulos 2 e 3, para analisar o comportamento destas sequências acima definidas.

Proposição 4.1. *As sequências $\{x^n\}$ e $\{\bar{s}^n\}$ satisfazem*

$$x^{n+1} = x(\mu_n) \quad e \quad \bar{s}^n = s(\mu_n), \quad \text{para todo } n = 1, 2, \dots \quad (4.8)$$

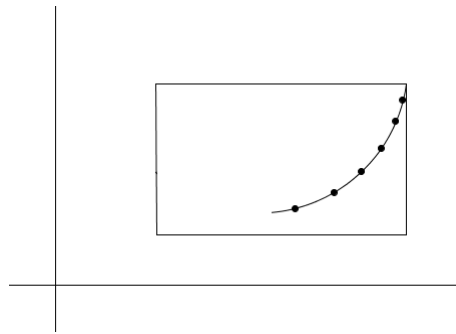


Figura 4.1: A sequência proximal primal e a média dual estão contidas, reespectivamente, na trajetória primal e dual.

Demonstração. De (4.5), segue que:

$$\frac{1}{\lambda_k} (c + A^T w^{k+1}) + [\nabla \varphi(x^{k+1}) - \nabla \varphi(x^k)] = 0.$$

Somando a expressão acima com k variando de 1 a n , obtemos:

$$\left(\sum_{k=1}^n \frac{1}{\lambda_k} \right) c + [\nabla \varphi(x^{n+1}) - \nabla \varphi(x^1)] + A^T \left(\sum_{k=1}^n \frac{1}{\lambda_k} w^{k+1} \right) = 0.$$

Daí, multiplicando por $\mu_n = \left(\sum_{i=1}^n \lambda_i^{-1} \right)^{-1}$, temos:

$$c + \mu_n [\nabla \varphi(x^{n+1}) - \nabla \varphi(x^1)] + A^T \bar{w}^{n+1} = 0, \quad (4.9)$$

onde $\bar{w}^{n+1} = \mu_n \left(\sum_{k=1}^n \frac{1}{\lambda_k} w^{k+1} \right)$. Como $Ax^{n+1} = b$, por (4.9) segue que (x^{n+1}, \bar{w}^{n+1}) é solução do sistema (2.10)-(2.11) com $\mu = \mu_n$, daí $x^{n+1} = x(\mu_n)$. E ainda,

$$\begin{aligned} \bar{s}^n &= \sum_{k=1}^n \mu_n \frac{1}{\lambda_k} s^k = \mu_n \sum_{k=1}^n \frac{1}{\lambda_k} \lambda_k (\log(x^k) - \log(x^{k+1})) \\ &= \mu_n (\log x^1 - \log x^{n+1}) = -\mu_n (\log x(\mu_n) - \log x^1) = s(\mu_n). \end{aligned}$$

□

4.2 Resultados de Convergência

Nesta seção, mostraremos a convergência das sequências proximal primal, proximal dual e média dual. Mostraremos inicialmente a convergência de $\{\mathbf{x}^k\}$. Este clássico resultado que pode ser encontrado em Iusem [7], sobre a hipótese restritiva de que $0 < \lambda_k \leq \tilde{\lambda} < \infty$. Muito embora, a demonstração aqui presente deva-se a Chen e Teboulle em [3], que mostraram o resultado sobre a hipótese mais fraca (4.3). Para auxiliar na demonstração deste fato, faremos primeiro um lema técnico.

Lema 4.1. *Seja $D = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq 0\}$ e $\{\mathbf{x}^k\}$ a sequência proximal primal definida acima, então:*

$$(i) \ D_\varphi(\mathbf{u}, \mathbf{x}^k) - D_\varphi(\mathbf{u}, \mathbf{x}^{k+1}) - D_\varphi(\mathbf{x}^{k+1}, \mathbf{x}^k) = \frac{1}{\lambda_k} \langle \mathbf{c}, \mathbf{x}^{k+1} - \mathbf{u} \rangle, \text{ para todo } \mathbf{u} \in D;$$

(ii) $\{\mathbf{c}^\top \mathbf{x}^k\}$ é não-crescente;

(iii) $\{D_\varphi(\bar{\mathbf{x}}, \mathbf{x}^k)\}$ é não-crescente, sempre que $\bar{\mathbf{x}} \in X^*$;

(iv) $\sigma_{n+1} \mathbf{c}^\top (\mathbf{x}^{n+1} - \mathbf{u}) \leq D_\varphi(\mathbf{u}, \mathbf{x}^1) - D_\varphi(\mathbf{u}, \mathbf{x}^{n+1}) - \sum_{k=1}^n \lambda_k \sigma_{k+1} D_\varphi(\mathbf{x}^{k+1}, \mathbf{x}^k)$, onde

$$\sigma_1 = 0, \sigma_n = \sum_{k=1}^n \lambda_k^{-1} \text{ para } n > 1 \text{ e } \mathbf{u} \in D.$$

Demonstração. (i) Pela Proposição 1.6(i) com $\mathbf{x} = \mathbf{u}$, $\mathbf{y} = \mathbf{x}^k$ e $\mathbf{z} = \mathbf{x}^{k+1}$, temos:

$$D_\varphi(\mathbf{u}, \mathbf{x}^k) - D_\varphi(\mathbf{u}, \mathbf{x}^{k+1}) - D_\varphi(\mathbf{x}^{k+1}, \mathbf{x}^k) = \langle \nabla \varphi(\mathbf{x}^k) - \nabla \varphi(\mathbf{x}^{k+1}), \mathbf{x}^{k+1} - \mathbf{u} \rangle. \quad (4.10)$$

De (4.5) em (4.10), como $A\mathbf{x}^{k+1} = A\mathbf{u} = \mathbf{b}$ e $(\text{Null } A)^\perp = \text{Im } A^\top$, temos:

$$\begin{aligned} D_\varphi(\mathbf{u}, \mathbf{x}^k) - D_\varphi(\mathbf{u}, \mathbf{x}^{k+1}) - D_\varphi(\mathbf{x}^{k+1}, \mathbf{x}^k) &= \frac{1}{\lambda_k} \langle \mathbf{c} + A^\top \mathbf{w}^{k+1}, \mathbf{x}^{k+1} - \mathbf{u} \rangle \\ &= \frac{1}{\lambda_k} \langle \mathbf{c}, \mathbf{x}^{k+1} - \mathbf{u} \rangle. \end{aligned}$$

(ii) De (4.4), obtemos que:

$$\mathbf{c}^\top \mathbf{x}^{k+1} + \lambda_k D_\varphi(\mathbf{x}^{k+1}, \mathbf{x}^k) \leq \mathbf{c}^\top \mathbf{x}^k + \lambda_k D_\varphi(\mathbf{x}^k, \mathbf{x}^k).$$

Portanto,

$$\frac{1}{\lambda_k} (\mathbf{c}^\top \mathbf{x}^k - \mathbf{c}^\top \mathbf{x}^{k+1}) \geq D_\varphi(\mathbf{x}^{k+1}, \mathbf{x}^k) \geq 0. \quad (4.11)$$

Como $\lambda_k > 0$, segue o resultado.

(iii) Dado $\bar{x} \in X^*$, vale que $c^T(x^{k+1} - \bar{x}) \geq 0$. Então, pelo item (i) temos:

$$D_\varphi(\bar{x}, x^k) - D_\varphi(\bar{x}, x^{k+1}) - D_\varphi(x^{k+1}, x^k) = \frac{1}{\lambda_k} \langle c, x^{k+1} - \bar{x} \rangle \geq 0.$$

Logo,

$$D_\varphi(\bar{x}, x^{k+1}) \leq D_\varphi(\bar{x}, x^k) - D_\varphi(x^{k+1}, x^k) \leq D_\varphi(\bar{x}, x^k).$$

(iv) Multiplicando (4.11) por σ_k e usando que $\sigma_{k+1} = \sigma_k + \lambda_k^{-1}$ obtemos:

$$\sigma_k c^T x^k - \sigma_{k+1} c^T x^{k+1} + \lambda_k^{-1} c^T x^{k+1} \geq \sigma_k \lambda_k D_\varphi(x^{k+1}, x^k).$$

Somando estas desigualdades com $k = 1, \dots, n$, temos:

$$-\sigma_{n+1} c^T x^{n+1} + \sum_{k=1}^n \lambda_k^{-1} c^T x^{k+1} \geq \sum_{k=1}^n \sigma_k \lambda_k D_\varphi(x^{k+1}, x^k). \quad (4.12)$$

Somando a identidade do item (i) com k variando de 1 a n :

$$-\sigma_{n+1} c^T u + \sum_{k=1}^n \lambda_k^{-1} c^T x^{k+1} = D_\varphi(u, x^1) - D_\varphi(u, x^{n+1}) - \sum_{k=1}^n D_\varphi(x^{k+1}, x^k) \quad (4.13)$$

Finalmente, subtraindo (4.12) de (4.13) e usando que $1 + \sigma_k \lambda_k = \lambda_k \sigma_{k+1}$, concluímos que:

$$\sigma_{n+1} c^T (x^{n+1} - u) \leq D_\varphi(u, x^1) - D_\varphi(u, x^{n+1}) - \sum_{k=1}^n \lambda_k \sigma_{k+1} D_\varphi(x^{k+1}, x^k). \quad \square$$

Sendo $X^* \neq \emptyset$ por **A1**), denotaremos $\alpha = \min\{c^T x : Ax = b, x \geq 0\}$.

Proposição 4.2. *A sequência $\{x^k\}$ gerada por (4.4) converge para uma solução do problema (2.1) e satisfaz:*

$$c^T x^k - \alpha \leq \sigma_k^{-1} D_\varphi(\bar{x}, x^1), \quad (4.14)$$

para todos $\bar{x} \in X^*$, $k > 1$ e onde σ_k é definida no Lema 4.1.

Demonstração. Pelo Lema 4.1 (ii), $\{c^T x^k\}$ é uma sequência não-crescente e sendo $\alpha \leq c^T x^k$ para todo k , concluímos que esta é uma sequência convergente e vale $\alpha \leq \lim_{k \rightarrow \infty} c^T x^k$. Dado $\varepsilon > 0$, pela definição de α , existe $v \in D$ tal que $c^T v < \alpha + \varepsilon$. E pelo Lema 4.1 (iv) com $u = v$, temos:

$$c^T x^k - c^T v \leq \sigma_k^{-1} D_\varphi(v, x^1). \quad (4.15)$$

Como $\sigma_k \rightarrow \infty$ obtemos $\lim_{k \rightarrow \infty} c^T x^k \leq c^T v < \alpha + \varepsilon$. Desde que o $\varepsilon > 0$ é arbitrário segue que $\lim_{k \rightarrow \infty} c^T x^k = \alpha$. E fazendo $v = \bar{x} \in X^*$ em (4.15) obtemos (4.14). Agora

pelo Lema 4.1(iii), $\{D_\varphi(\bar{x}, x^k)\}$ é não-crescente, logo $D_\varphi(\bar{x}, x^k) \leq D_\varphi(\bar{x}, x^1)$. Por **B3** da definição de função de Bregman, temos que $\{x^k\}$ é limitada. Seja z um ponto de acumulação de $\{x^k\}$, então existe uma subsequência $x^{k_j} \rightarrow z \in D$, pois D é fechado. E por continuidade $c^\top z = \alpha$, já que $c^\top x^{k_j} \rightarrow \alpha$. Portanto, $z \in X^*$. Como $\{D_\varphi(z, x^k)\}$ é não-crescente e possui, por **B4**) da definição de função de Bregman, uma subsequência $D_\varphi(z, x^{k_j}) \rightarrow 0$, obtemos que $D_\varphi(z, x^k) \rightarrow 0$. Para finalizar mostraremos que $\{x^k\}$ tem apenas um ponto de acumulação. Suponha que seja z^* outro ponto de acumulação da sequência, tal que $x^{k_i} \rightarrow z^*$. Como $D_\varphi(z, x^{k_i}) \rightarrow 0$, concluímos por **B5**) que $z = z^*$. Portanto, $x^k \rightarrow z \in X^*$. \square

Proposição 4.3. *A sequência $\{\bar{s}^k\}$ definida por (4.7) converge ao centróide s^c de S^* , isto é, $\lim_{k \rightarrow \infty} \bar{s}^k = s^c \in S^*$.*

Demonstração. Como $\mu_k \rightarrow 0$ o resultado segue aplicando o Teorema 2.1 em (4.8). \square

Observação 4.1. *O resultado da proposição anterior pode ser obtido diretamente, sem recorrer a Proposição 4.1, como é mostrado por Montero e Iusem em [8].*

O próximo teorema é o principal resultado deste trabalho, visto que estabelece a convergência da sequência proximal dual, utilizando a sequência média dual e outros resultados.

Teorema 4.1. $\lim_{k \rightarrow \infty} s^k = s^c$.

Demonstração. Por (4.7) e (4.8) obtemos que:

$$\begin{aligned} s(\mu_{k+1}) + \frac{\lambda_{k+1}}{\mu_k} (s(\mu_{k+1}) - s(\mu_k)) &= \bar{s}^{k+1} + \frac{\lambda_{k+1}}{\mu_k} (\bar{s}^{k+1} - \bar{s}^k) \\ &= \mu_{k+1} \left(\sum_{i=1}^k \frac{1}{\lambda_i} s^i + \frac{s^{k+1}}{\lambda_{k+1}} \right) + \frac{\lambda_{k+1}}{\mu_k} \left[(\mu_{k+1} - \mu_k) \left(\sum_{i=1}^k \frac{1}{\lambda_i} s^i \right) + \frac{\mu_{k+1}}{\lambda_{k+1}} s^{k+1} \right] \\ &= \left[\frac{\lambda_{k+1}}{\mu_k} (\mu_{k+1} - \mu_k) + \mu_{k+1} \right] \left(\sum_{i=1}^k \frac{1}{\lambda_i} s^i \right) + \left(\frac{\mu_{k+1}}{\mu_k} + \frac{\mu_{k+1}}{\lambda_{k+1}} \right) s^{k+1} \\ &= 0 \left(\sum_{i=1}^k \frac{1}{\lambda_i} s^i \right) + s^{k+1} = s^{k+1}. \end{aligned}$$

Usando o Teorema do Valor Médio de Cauchy (Teorema 1.1) na última expressão, para cada k existe $\xi^k \in \mathbb{R}^n$, com $\xi_i^k \in (\mu_{k+1}, \mu_k)$ tal que:

$$s_i^{k+1} = s_i(\mu_{k+1}) + \frac{\lambda_{k+1}}{\mu_k} (\mu_{k+1} - \mu_k) \dot{s}_i(\xi_i^k), \quad i = 1, \dots, n. \quad (4.16)$$

Como $\left(\frac{\lambda_{k+1}}{\mu_k}\right) (\mu_{k+1} - \mu_k) = -\mu_{k+1}$, temos de (4.16) que:

$$s_i^{k+1} = s_i(\mu_{k+1}) - \mu_{k+1} \dot{s}_i(\xi_i^k), \quad i = 1, \dots, n. \quad (4.17)$$

Desde que $\lim_{k \rightarrow \infty} \mu_k = 0$, pelo Teorema 3.2 obtemos $\lim_{k \rightarrow \infty} \mu_{k+1} \dot{s}_i(\xi_i^k) = 0$. Portanto, tomando o limite em (4.17) e usando o Teorema 2.1 concluímos o teorema. \square

4.3 Análise de Convergência

Agora, o objetivo é estudar as velocidades de convergência das sequências proximal primal e média dual. Para tanto, introduziremos a seguinte notação:

$$\bar{\lambda} = \limsup_{k \rightarrow \infty} \lambda_k, \quad (4.18)$$

onde definimos $e^{-v_1^*/\bar{\lambda}} = 0$, quando $\bar{\lambda} = 0$ e $e^{-v_1^*/\bar{\lambda}} = 1$, quando $\bar{\lambda} = \infty$.

E antes de prosseguirmos com o estudo, revisaremos as definições de algumas formas de convergência:

a) A sequência $(x^k) \subset \mathbb{R}^n$ converge linearmente para $\bar{x} \in \mathbb{R}^n$, com razão de convergência $r \in (0, 1)$, quando

$$\limsup_{k \rightarrow \infty} \frac{\|x^{k+1} - \bar{x}\|}{\|x^k - \bar{x}\|} = r;$$

b) A sequência $(x^k) \subset \mathbb{R}^n$ converge sublinearmente para $\bar{x} \in \mathbb{R}^n$, quando $r = 1$;

c) A sequência $(x^k) \subset \mathbb{R}^n$ converge superlinearmente para $\bar{x} \in \mathbb{R}^n$, quando $r = 0$.

Teorema 4.2. *Seja $\{x^k\}$ a sequência definida em (4.4) e seja v_1^* da definição de centróide em (2.16). Então,*

$$\limsup_{k \rightarrow \infty} \frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|} = e^{-v_1^*/\bar{\lambda}}.$$

Demonstração. Pela Proposição 4.1 e equação (3.23) temos, após algumas manipulações algébricas, que:

$$\begin{aligned} \frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|} &= \left\| \frac{x^{k+1} - x^*}{e^{-v_1^*/\mu_k}} \right\| \left\| \frac{x^k - x^*}{e^{-v_1^*/\mu_{k-1}}} \right\|^{-1} e^{-v_1^*/\lambda_k} \\ &= \left\| \frac{x(\mu_k) - x^*}{e^{-v_1^*/\mu_k}} \right\| \left\| \frac{x(\mu_{k-1}) - x^*}{e^{-v_1^*/\mu_{k-1}}} \right\|^{-1} e^{-v_1^*/\lambda_k} \\ &= \frac{\|\Delta x(\mu_k)\|}{\|\Delta x(\mu_{k-1})\|} e^{-v_1^*/\lambda_k}. \end{aligned}$$

Pelo Teorema 3.3, temos:

$$\lim_{k \rightarrow \infty} \frac{\|\Delta x(\mu_k)\|}{\|\Delta x(\mu_{k-1})\|} = 1.$$

Portanto, tomando o \limsup , quando k tendendo ao ∞ , na equação inicial e usando as propriedades de \limsup , segue que:

$$\limsup_{k \rightarrow \infty} \frac{\|x^{k+1} - x^*\|}{\|x^k - x^*\|} = \lim_{k \rightarrow \infty} \frac{\|\Delta x(\mu_k)\|}{\|\Delta x(\mu_{k-1})\|} \limsup_{k \rightarrow \infty} e^{-v_1^*/\lambda_k} = e^{-v_1^*/\bar{\lambda}}.$$

□

Observação 4.2. *No teorema acima, bem como no seguinte, usamos duas propriedades operacionais do \limsup :*

a) *Sejam duas seqüências (a_n) e (b_n) tais que $\lim_{n \rightarrow \infty} a_n = a > 0$ e $b_n \geq 0$, então*

$$\limsup_{n \rightarrow \infty} (a_n b_n) = a \cdot \limsup_{n \rightarrow \infty} b_n.$$

b) *Seja $f: \mathbb{R} \rightarrow \mathbb{R}$ crescente e contínua e (a_n) uma seqüência de números reais, então*

$$\limsup_{n \rightarrow \infty} f(a_n) = f(\limsup_{n \rightarrow \infty} a_n)$$

As demonstrações destes fatos podem ser encontradas respectivamente em [11] e [12].

Finalmente, analisaremos o tipo de convergência da seqüência média dual $\{\bar{s}^k\}$. A análise será feita sobre a dependência de $\Delta s^\infty \neq 0$, onde Δs^∞ é dado no Teorema 3.1. Uma condição suficiente para isto é que $x_B^1 \neq x_B^*$.

Teorema 4.3. *Seja $\{\bar{s}^k\}$ a seqüência definida em (4.7). Assumindo que $\Delta s^\infty \neq 0$. Então,*

$$\limsup_{k \rightarrow \infty} \frac{\|\bar{s}^{k+1} - s^c\|}{\|\bar{s}^k - s^c\|} = \limsup_{k \rightarrow \infty} \frac{\mu_{k+1}}{\mu_k}. \quad (4.19)$$

Como consequência, se $\bar{\lambda} > 0$ então a seqüência $\{\bar{s}^k\}$ converge sublinearmente.

Demonstração. Pela Proposição 4.1 e pela relação (3.1) obtemos, após algumas manipulações algébricas, que:

$$\begin{aligned} \frac{\|\bar{s}^{k+1} - s^c\|}{\|\bar{s}^k - s^c\|} &= \frac{\|s(\mu_{k+1}) - s^c\|}{\mu_{k+1}} \left(\frac{\|s(\mu_k) - s^c\|}{\mu_k} \right)^{-1} \frac{\mu_{k+1}}{\mu_k} \\ &= \frac{\|\Delta s(\mu_{k+1})\|}{\|\Delta s(\mu_k)\|} \frac{\mu_{k+1}}{\mu_k}. \end{aligned}$$

Pelo Teorema 3.1, temos:

$$\lim_{k \rightarrow \infty} \frac{\|\Delta s(\mu_{k+1})\|}{\|\Delta s(\mu_k)\|} = 1.$$

Portanto, tomando o \limsup , quando k tende a ∞ , na equação inicial segue que:

$$\limsup_{k \rightarrow \infty} \frac{\|\bar{s}^{k+1} - s^c\|}{\|\bar{s}^k - s^c\|} = \lim_{k \rightarrow \infty} \frac{\|\Delta s(\mu_{k+1})\|}{\|\Delta s(\mu_k)\|} \limsup_{k \rightarrow \infty} \frac{\mu_{k+1}}{\mu_k} = \limsup_{k \rightarrow \infty} \frac{\mu_{k+1}}{\mu_k}.$$

Pela definição de $\{\mu_k\}$, obtemos:

$$\limsup_{k \rightarrow \infty} \frac{\mu_{k+1}}{\mu_k} = \limsup_{k \rightarrow \infty} \frac{1}{1 + \frac{\mu_k}{\lambda_{k+1}}}.$$

Se $\bar{\lambda} > 0$, segue da última linha que $\limsup_{k \rightarrow \infty} \frac{\mu_{k+1}}{\mu_k} = 1$, já que $\lim_{k \rightarrow \infty} \mu_k = 0$. E portanto, neste caso em especial a convergência é sublinear, pois

$$\limsup_{k \rightarrow \infty} \frac{\|\bar{s}^{k+1} - s^c\|}{\|\bar{s}^k - s^c\|} = 1.$$

□

Teorema 4.4. *Seja $\{s^k\}$ a sequência proximal dual. Então,*

$$\lim_{k \rightarrow \infty} \frac{s^k - s^c}{\mu_k} = 0.$$

Demonstração. Pela equação (4.17) temos, para todo $i \in \{1, \dots, n\}$, que

$$\frac{s_i^{k+1} - s_i^c}{\mu_{k+1}} = \frac{s_i(\mu_{k+1}) - s_i^c}{\mu_{k+1}} - \dot{s}_i(\xi_i^k) = \Delta s(\mu_{k+1}) - \dot{s}_i(\xi_i^k).$$

Como $\lim_{k \rightarrow \infty} \xi_i^k = 0$, concluímos o resultado usando o Teorema 3.2 e o Teorema 3.1. □

SPP \ SMD	Superlinear	Linear	Sublinear
Superlinear	$\lambda_k = 1/k!$	$\lambda_k = a^k, a \in (0, 1)$	$\lambda_k = 1/k$
Linear	Impossível	Impossível	$\lambda_k = \lambda$
Sublinear	Impossível	Impossível	$\lambda_k = k$

De posse destes resultados é possível fazermos a seguinte análise sobre os tipos de convergências das sequências proximal primal e média dual. Segue do Teorema 4.2 e da expressão (4.18), que $\{x^k\}$ converge:

- superlinearmente se, e só se, $\limsup_{k \rightarrow \infty} \lambda_k = 0$;
- linearmente se, e só se, $0 < \limsup_{k \rightarrow \infty} \lambda_k < \infty$;
- sublinearmente se, e só se, $\limsup_{k \rightarrow \infty} \lambda_k = \infty$.

Do Teorema 4.3, obtemos que $\{\bar{s}^k\}$ converge sublinearmente sempre que $\limsup_{k \rightarrow \infty} \lambda_k > 0$. Para outras possibilidades da sequência $\{\lambda_k\}$, a sequência $\{\bar{s}^k\}$ pode convergir superlinearmente, linearmente ou sublinearmente, como mostram os exemplos da tabela abaixo.

Na tabela acima, as situações impossíveis de acontecer também são justificadas pelo Teorema 4.2 e pelo Teorema 4.3. De fato, se $\{x^k\}$ convergir linearmente, então temos $0 < \limsup_{k \rightarrow \infty} \lambda_k < \infty$ e daí, $\{\bar{s}^k\}$ convergirá sublinearmente. E se $\{x^k\}$ convergir sublinearmente, então $\limsup_{k \rightarrow \infty} \lambda_k = \infty$ e portanto, $\{\bar{s}^k\}$ convergirá sublinearmente.

Referências Bibliográficas

- [1] BARTLE, R. G. The elements of real analysis. 1. ed. New York: Wiley, 1967.
- [2] BREGMAN, L. M. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *Zh. Vychisl. Mat. Mat. Fiz.*, v. 7, n. 3, p. 620-631, 1967.
- [3] CHEN, G.; M. TEBoulLE. Convergence analysis of a proximal-like minimization algorithm using Bregman functions, *SIAM- Journal on Control and Optimization*, v.3, n. 3, p. 538-543, 1993.
- [4] COMINETTI, R.; SAN MARTÍN, J. (1994), Asymptotic Analysis of the exponential penalty trajectory in linear programming, *Mathematical Programming*, v. 67, n. 2, p. 169-187, 1994.
- [5] ECKSTEIN, J. Nonlinear proximal point algorithms using Bregman Functions, with applications to convex programming, *Mathematical Sciences Research*, p. 202-226, 1993.
- [6] GONZAGA, C. C. Algoritmos de pontos interiores para programação linear. 1. ed. Rio de Janeiro: Impa, 1989. 116 p.
- [7] IUSEM, A. N. Métodos de ponto proximal em otimização. 1.ed. Rio de Janeiro: Impa, 1995. 56 p.
- [8] IUSEM, A. N.; MONTEIRO, R. D. C. On dual convergence of the generalized proximal point method with Bregman distance. *Mathematics of Operations Research*, v. 25, n. 4, p. 606-624, 2000. no. 4, pp. 606–624.

- [9] IUSEM, A. N.; SVAITER, B. F.; CRUZ NETO, J.X. Central paths, generalized proximal point methods and Cauchy trajectories in Riemannian manifolds. *SIAM-Journal on Control and Optimization*, v. 37, n. 2, p. 566-588, 1999.
- [10] IZMAILOV, A.; SOLODOV, M. Otimização: Condições de otimalidade, elementos de análise convexa e de dualidade. 2. ed. Rio de Janeiro: IMPA, 2009. 270 p. vol. 1.
- [11] KACZOR, J. W.; NOWAK, M. T. Problems in Mathematical Analysis I: real numbers, sequences and series. 1. ed. Lublin: AMS, 2000. 380 p.
- [12] KACZOR, J. W.; NOWAK, M. T. Problems in Mathematical Analysis II: continuity and differentiation. 1. ed. Lublin: AMS, 2000. 380 p.
- [13] LEMARÉCHAL, C.; HIRIANT-URRUTY, J.-P. Convex analysis and minimization algorithms I. 2. ed. [s.n]: Springer-Verlag, 1996. 418 p. vol. 1
- [14] LIMA, E. L. Análise real: Funções de n Variáveis. 5.ed. Rio de Janeiro: IMPA, 2010. 202 p. v.2.
- [15] LIMA, E. L. Análise real: Funções de Uma Variável. 5.ed. Rio de Janeiro: IMPA, 2010. 202 p. v.1.
- [16] LIMA, E. L. Álgebra Linear.8. ed. Rio de Janeiro: IMPA, 2011. 357p.
- [17] LOPES, M. V. Trajetória central associada à entropia e o método do ponto proximal em programação linear. 2007. 91 f.. Dissertação (Mestrado em Ciência da Computação) - Instituto de Informática, Universidade Federal de Goiás, Goiânia, 2007.
- [18] RIBEIRO, A. A.; KARAS, E. W. Um curso de otimização. Curitiba: [s.n], 2011.
- [19] SAIGAL, R. Linear programming: a modern integrated analysis. 1. ed. New York: Springer Science and Business Media New York , 1995. 342p.