UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE INFORMÁTICA

MARCOS VINÍCIUS LOPES

Trajetória Central Associada à Entropia e o Método do Ponto Proximal em Programação Linear

Goiânia
2007

Marcos Vinícius Lopes

Trajetória Central Associada à Entropia e o Método do Ponto Proximal em Programação Linear

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

Área de concentração: Otimização.

Orientador: Prof. Orizon Pereira Ferreira

Goiânia
2007

MARCOS VINÍCIUS LOPES

Trajetória Central Associada à Entropia e o Método do Ponto Proximal em Programação Linear

Dissertação defendida no Programa de Pós–Graduação do Instituto de Informática da Universidade Federal de Goiás como requisito parcial para obtenção do título de Mestre em Ciência da Computação, aprovada em 14 de Dezembro de 2007, pela Banca Examinadora constituída pelos professores:

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

Prof. Geci José Pereira da Silva
IME – UFG

Prof. Luis Mauricio Graña Drummond
FACC – UFRJ

Prof. Humberto José Longo
INF – UFG

Marcos Vinícius Lopes

Graduou–se em Matemática na UFG - Universidade Federal de Goiás. Durante sua graduação, foi monitor de Geometria Analítica e Álgebra Linear da UFG. Especializou-se em Educação Matemática pela mesma Universidade em 2001. Durante o Mestrado (2006-2007) no INF-UFG (Instituto de Informática - Universidade Federal de Goiás), esteve licenciado da cadeira de Matemática do Cepae-UFG do qual é professor desde 1992.

Dedico este trabalho principalmente a DEUS, aos meus pais, Benedita e Samuel, ao meu filho, Marco Túlio, bem como a todos que contribuiram direta ou indiretamente para a execução deste.

Agradecimentos

Agradeço acima de tudo a DEUS, sem o qual nada seria possível. Gostaria de agradecer aos meus pais, a todos professores, funcionários e alunos dos Institutos de Informática e de Matemática e Estatística. Agradeço aos professores Geci, Humberto, Hugo e Luís Roman pela amizade, pelos conselhos e ensinamentos. Em especial, gostaria de agradecer ao meu orientador, professor Orizon, pelo desvelo, empenho e paciência. Também não poderia deixar de citar e agradecer aos colegas do mestrado, pelos momentos de alegria e fraternidade. Agradeço a todos, em especial ao Luciano, que em momentos cruciais, ofertou apoio e perseverança.

A matemática, não é construída apenas por motivos práticos, mas também, e principalmente, pela liberdade de abstração científica...

Miguel Taube Netto,
Matemática para produtividade.

Resumo

Lopes, Marcos Vinícius. **Trajetória Central Associada à Entropia e o Método do Ponto Proximal em Programação Linear**. Goiânia, 2007. 91p. Dissertação de Mestrado. Instituto de Informática, Universidade Federal de Goiás.

Em nossa dissertação, estudamos a convergência das trajetórias central primal e dual associadas às funções entropia e exponencial, respectivamente, investigamos o comportamento assintótico das trajetórias central primal e dual associadas às funções penalidade entropia e exponencial, respectivamente, em programação linear. Estudamos também a convergência das trajetórias central primal e dual associadas à divergência de Kullback-Leibler e sua conjugada. Como aplicação do estudo das trajetórias central primal e dual, mostramos a convergência das seqüências proximal primal e dual ponderada associadas à distância Kullback-Leibler para os (PPL), no chamado Método de Ponto Proximal.

Palavras–chave

1. Problemas de Programação Linear. 2. Função Entropia. 3. Divergência de Kullback-Leibler. 4. Método de Ponto Proximal. 5. Trajetória Central

Abstract

Lopes, Marcos Vinícius. **Central trajectory Associated to Entropy and the Point Proximal Method in Linear Programming.**. Goiânia, 2007. 91p. MSc. Dissertation. Instituto de Informática, Universidade Federal de Goiás.

In ours dissertassion, we study the primal and dual central trajectory convergence associates to entropy and exponential functions , respectively, we investigate the assintótic behavior of the primal and dual central trajectory associates to entropy and exponential penalty functions, respectively, in linear programming. We also study the primal and dual central trajectory convergence associates to Kullback-Leibler divergence and your conjugated. As application of the studies of convergence of the primal and dual trajectories, we show to the convergency of the proximal primal and dual weighed associates to Kullback-Leibler distance for (PPL), in the call Method of Proximal Point.

Keywords

1. Linear Programming Problems. 2. Entropy Function. 3. Divergence of Kullback-Leibler. 4. Proximal Point's Method. 5. Central Trajetory

Sumário

Apresentação

O problema de otimizar uma função linear sujeita a restrições também lineares é chamado de Problema de Programação Linear (PPL). Neste trabalho estaremos apresentando o PPL no formato padrão, isto é, minimizar uma função linear sujeita a restrições lineares de igualdade bem como à restrições de não negatividade. Vários autores se concentraram no estudo e resolução destes problemas nas últimas décadas. Impulsionados, principalmente, pelo uso bélico e militar, uma gama de problemas de programação linear necessitavam ser resolvidos. Os esforços de guerra foram o principal combustível para o desenvolvimento desta área de pesquisa. Após o término da segunda grande guerra, as áreas econômica e industrial, predominantemente, passaram a demandar a utilização de métodos para resolução dos problemas de programação linear. Grande foi o número de métodos criados para a resolução de tais problemas. O método Simplex, que teve sua criação em meados da década de 40, portanto em pleno cenário de guerra, sem nenhuma dúvida, foi o mais famoso.

Mais tarde, provou-se que este método não possuía tempo algorítmico satisfatório. Ou seja, computacionalmente falando, na teoria o método demandava tempo exponencial, apesar de funcionar muito bem na prática. Preocupados com esse fato, vários grupos de pesquisa matemática, pesquisa operacional entre outros, envidaram esforços no sentido de encontrar algoritmos que pudessem resolver tais problemas em tempo polinomial. Neste contexto, advieram os chamados métodos de pontos interiores. Ao contrário do Simplex, que faz a busca da solução pela fronteira do conjunto viável, aqueles métodos prevêem uma trajetória pelo interior do conjunto viável. Muitos foram os métodos propostos. Dentre estes, podemos destacar os métodos de penalização. Nosso trabalho prevê uma breve apresentação histórica acerca da programação linear e a otimização.

O objetivo central do nosso trabalho é o estudo de propriedades da trajetória central primal-dual associada à entropia bem como a divergência de Kullback-Leibler. Uma outra parte importante é fazer uma aplicação desta última trajetória ao estudo do método de ponto proximal. Nosso estudo será restrito ao contexto da programação linear, por este motivo vamos aproveitar e estudar várias propriedades do problema de programação linear, bem como fazer um breve histórico. A dissetação é baseada nos artigos Cominetti e San Martín [3], Iusem, Svaiter e Cruz Neto [15], Iusem e Monteiro [14]. Nestes artigos, são adotados métodos de penalidades para a pertubação dos problemas primal e dual. No primeiro destes artigos, Cominetti e San Martín estudam o comportamento da trajetória central através da pertubação (penalização) dos problemas primal e dual com a utilização das funções entropia e sua conjugada, a função exponencial. Mais especificamente, o problema de programação linear (PPL) primal se escreve:

$$(P) \qquad \min\left\{c^T x : Ax = b, x \geq 0\right\},$$

onde $c \in I\!R^n$, $b \in I\!R^m$, a matriz A é de ordem $m \times n$ e a variável primal é $x \in I\!R^n$. Adicionando a função entropia à função objetivo de (P), obtemos sua versão penalizada:

$$(P_\mu) \qquad \min\left\{c^T x + \mu \sum_{i=1}^{n} x_i log x_i, Ax = b, x > 0\right\}, \ \mu > 0.$$

O problema dual associado a (P) é:

$$(D) \qquad \max\left\{b^T y : A^T y + s = c, s \geq 0\right\},$$

onde $A^T \in I\!R^{mxn}$ denota a matriz transposta de A e $(y,s) \in I\!R^m \times I\!R^n$ são as variáveis duais. Adicionando a função penalidade exponencial à função objetivo de (D), obtemos sua versão penalizada:

$$(D_\mu) \qquad \max\left\{b^T y - \mu \sum_{i=1}^{n} e^{-s_i/\mu - 1} : A^T y + s = c\right\}, \ \mu > 0.$$

Os conjuntos viáveis primal e dual são denotados por:

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

e

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

respectivamente. Os interiores relativos dos conjuntos viáveis primal e dual são denotados por $\mathcal{F}^0(P)$ e $\mathcal{F}^0(D)$, respectivamente. Denotamos também por $\mathcal{F}^*(P)$ e $\mathcal{F}^*(D)$, os conjuntos de soluções ótimas primal e dual, respectivamente. Supomos ainda que $\mathcal{F}^0(P) \neq \emptyset$ e $\mathcal{F}^0(D) \neq \emptyset$. Um dos nossos objetivos é provar que esta hipótese implica que os problemas (P_μ) e (D_μ) têm única solução $x(\mu)$ e $(s(\mu), y(\mu))$, respectivamente, ou, de forma equivalente, que as trajetórias central primal e dual, associadas às funções entropia e exponencial, denotadas por $\{x(\mu) : \mu > 0\}$ e $\{s(\mu) : \mu > 0\}$, respectivamente, estão bem definidas. Como consequência, $x(\mu)$ e $s(\mu)$ satisfazem a igualdade:

$$s(\mu) = -\mu \log(x(\mu)) - \mu e, \quad \mu > 0,$$

onde e é o vetor de números um. Provaremos que as trajetórias central primal e dual, associadas às funções entropia e exponencial, convergem para uma solução de (P) e (D), respectivamente, quando μ vai para 0. Desta forma, podemos pensar em (P_μ) e (D_μ) como métodos de penalidade para resolver problemas (PPL). Cominetti e San Martín em [3] investigaram a convergência das trajetórias central primal e dual associadas às funções penalidades entropia e exponencial, respectivamente, para problemas de programação linear. Uma de nossas propostas (Capítulo 4) será simplificar a prova de convergência, apresentando, acreditamos, num contexto mais didático, demosntrações e outros resultados de forma auto-contida.

No artigo de Iusem, Svaiter e Cruz Neto [15], os autores se preocupam em estudar a convergência primal do método do ponto proximal com a distância de Bregman. Iusem e Monteiro em [14], forneceram uma caracterização do ponto limite da trajetória central dual associada a uma grande classe de funções penalidades, incluindo a divergência de Kullback-Leibler, para problemas de programação convexa com restrições lineares.

Iremos utilizar os resultados obtidos, com respeito às trajetórias central primal e dual, através das pertubações dos problemas (P) e (D) com a funções entropia e exponencial, para demonstrar resultados de convergência para o método do ponto proximal generalizado aplicado à programação linear. Novamente aqui, nossa intenção é tentar simplificar as demonstrações, incentivando principalmente os leitores ainda não familiarizados com tais notações.

O método de ponto proximal gera uma seqüência $\{x^k\} \subset \mathbb{R}^n_{++}$ com ponto inicial $x_0 \in \mathcal{F}^0(P)$, de acordo com a iteração

$$x^{k+1} = \arg\min \left\{ c^T x + \lambda_k K_\varphi(x, x^k) : Ax = b, x > 0 \right\},$$

onde a seqüência $\{\lambda_k\} \subset \mathbb{R}_{++}$ satisfaz:

$$\sum_{k=0}^{\infty} {\lambda_k}^{-1} = +\infty$$

e $K : \mathbb{R}^n_{++} \times \mathbb{R}^n_{++} \to \mathbb{R}$ é a distância de Kullback-Leibler definida por:

$$K_\varphi(x, y) = \varphi(x) - \varphi(y) - \nabla\varphi(y)^T (x - y),$$

e $\varphi(x) = \sum_{i=1}^n x_i log x_i$ é a função entropia.

Na primeira fase de nosso estudo, fizemos um modesto levantamento das bibliografias sobre o assunto e os artigos estudados. Uma fundamentação teórica foi de suma importância para que conseguíssemos nos familiarizar com o assunto, suas notações, principais resultados e características. Nessecitávamos de um arcabouço teórico sobre o comportamento da trajetória central, associada à função entropia, e para tanto utilizamos [13], bem como [6], [14] e [15]. Nos importava ainda um embasamento acerca de métodos de penalidade, tais como as barreiras logarítmicas, exponenciais entre outras e nesse sentido usamos [18], [29], [33], além de outros anteriormente mencionados. Para o estudo das funções de Bregman, bem como do método do ponto proximal, adotamos [1], [2],[16], [11], além de outros acima já citados. Como não poderia deixar ser, uma série de outros "papers"e artigos foram ocasionalmente consultados via rede mundial de computadores sem que fizéssemos, aqui ou nas referências, uma menção explícita.

Nosso trabalho está estruturado da forma que se segue. No Capítulo 1 é apresentado um breve histórico acerca da programação linear e da otimização. O leitor já familiarizado com tal histórico, poderá, sem nenhum prejuízo, dispensar sua leitura e avançar ao Capítulo 2, onde mostramos uma série de resultados preliminares que embasarão uma sequência de demonstrações vindoras. Resultados clássicos com conjuntos e funções convexas, de álgebra linear, bem como exemplos de funções convexas, serão subsequentes. No Capítulo 3 os problemas de programação linear primal e dual, nos formatos padrão, bem como seus respectivos conjuntos viáveis e conjuntos solução, serão estudados. Nesse capítulo apresentamos ainda as condições de otimalidade dos problemas de programação linear. O Capítulo 4 foi reservado para a apresentação da função entropia e sua conjugada, a função exponencial. Ali, onde se encontra o cerne de nosso trabalho, fazemos a perturbação dos problemas primal e dual nos formatos padrão para obter as trajetórias central primal e dual, respectivamente, e daí demonstrar suas convergências. Como havíamos antecipado, o leitor atento, verificará uma série de similaridades entre os resultados do Capítulo 4 e os de seu subsequente. O fato é que os resultados obtidos para a função entropia em [3], podem ser obtidos para a divergência Kullback-Leibler, com isto, podemos fazer uma aplicação para o método do ponto proximal, exposta na última seção do capítulo derradeiro deste trabalho. O leitor poderá, portanto, fazer a leitura dos Capítulos 2 e 3, avançando ao Capítulo 4 ou optar por avançar, sem nenhum prejuízo, diretamente ao Capítulo 5. Ainda no Capítulo 5, na sua última seção, apresentaremos formalmente o método do ponto proximal generalizado com distância de Bregman associada à entropia, ou seja, a divergência de Kullback-Leibler. Na conclusão, apontamos as principais resultados que acreditamos ter alcançado, bem como as dificuldades encontradas, os futuros trabalhos e desafios que poderão suceder a este.

CAPÍTULO 1

Histórico de Programação Linear e Otimização

1.1 Introdução

Programação Linear (PL) é uma das linhas de pesquisa da área de programação matemática, termo este que representa a maximização ou minimização de funções objetivo lineares sujeitas a restrições lineares. Uma função objetivo é uma função de uma ou mais variáveis que desperte o interesse a alguém em descobrir o seu ponto de máximo ou de mínimo. A função pode representar muitas coisas, como por exemplo, o custo de algum processo de produção, o fluxo viário de alguma rodovia ou avenida, etc. A maximização ou minimização de uma função objetivo pode envolver limitações de igualdade e/ou de desigualdade nas variáveis de decisão, essas limitações são chamadas de restrições do problema de otimização.

Um modelo no qual a função objetivo e todas as restrições são funções lineares das variáveis de decisão é chamado Problema de Programação Linear(PPL). Se a função objetivo ou uma das restrições for uma função não-linear das variáveis de decisão, o modelo é chamado de Problema de Programação Não-Linear(PNL). Se o problema inclui restrições inteiras, ele é chamado de Problema de Programação Inteira Linear ou Não-Linear. Estes são alguns problemas de otimização existentes, os quais foram formulados e desenvolvidos dentro de uma área de pesquisa mais abrangente, chamada de Pesquisa Operacional(PO) de acordo com Murty em [27].

Otimização é um ramo da ciência de grande interesse desde os tempos pré-históricos. Planejar, programar e otimizar a obtenção e o consumo de alimentos afetou certamente a manutenção da espécime. Mais detidamente o termo Programação Linear, quando foi introduzido na década de 1940, era sinônimo de planejamento ou esquema e a disciplina de "planejamento e programação", onde esses métodos foram discutidos pela primeira vez, não tinha nenhuma relação com programação computacional. Com a realização de pesquisas desenvolvidas em várias áreas do conhecimento humano e com a magnitude de dados que deveriam ser trabalhados, verificou-se a importância computacional que esta linha de pesquisa viria a desempenhar.

1.2 Histórico

A pesquisa operacional e a programação matemática são áreas de pesquisa que tiveram um grande crescimento recentemente, impulsionadas principalmente, em meados e final do século passado por duas correntes:

- Esforços de guerra, principalmente pelos USA e Reino Unido;
- Modelagem matemática de sistemas econômicos (USA e URSS).

O termo Pesquisa Operacional surge em 1938, como resultado da aplicação da "pesquisa"em operações militares. O problema em questão era a coordenação das informações recebidas pelas estações de radar que faziam parte do sistema de defesa aéreo britânico (Royal Air Force Figher Command), instituído em 1936. Antevendo uma guerra eminente a RAF, força aérea britânica, começa a realizar testes a partir de 1937. Em 1938, um grande teste é realizado e verifica-se que as informações recebidas pelas diversas estações de radar, poderiam ser conflitantes. Para resolver este problema é formado um grande grupo de pesquisa com especialistas nas áreas de matemática, física e engenharia.

Já em 1939, às vésperas do início da 2ª Guerra Mundial, um último teste é realizado e constatou-se o grande êxito obtido pelo grupo de pesquisa operacional. Como resultado do grande sucesso obtido, o grupo é transferido para o quartel general da RAF. Em 1941, esse grupo recebe o nome de "Seção de Pesquisa Operacional". Merece destaque especial, entre várias pesquisas realizadas por esse grupo, o trabalho para aumentar o sucesso na localização e destruição de submarinos alemães através de ataques aéreos. Em meados de 1941 essa taxa de sucesso era de 2 a 3%, após a adoção das medidas sugeridas pelo grupo de Pesquisa Operacional, a probabilidade de ataque aéreo com sucesso subiu para 45%, resultado este que, devido às limitações tecnológicas da época, era fantástico.

Na área econômica as primeiras tentativas de modelagem matemática se remontam ao "Tableau Economique"de Quesnay segundo Dorfman em [10], que propunha um modelo linear para as relações entre o senhorio, camponeses e artesãos do século XVII. Em 1874, na obra [36], L. Walras propõe um modelo onde é introduzido o conceito de coeficiente tecnológico fixo. Entretanto, os modelos lineares não seriam explorados até meados dos anos 1930's, quando um grupo de economistas Austro-alemães tentou generalizar o modelo proposto por Walras. Segundo Dantzig em [7], este estudo teria influenciado o matemático americano Von Neumann, que em [35] de 1937, publicaria o trabalho "Modelo de Equilíbrio Econômico Geral". O modelo cosiderado por Von Neumann, que supunha taxa constante de expansão econômica e economia autosuficiente, não obteve grande repercussão. Uma das objeções levantadas pelos economistas de então, era o fato do modelo ser linear.

Posteriormente, outro modelo linear obteve maior sucesso. Esse novo modelo baseado em relações de entrada e saída desenvolvido por Leontif em [21], visava estabelecer um modelo para a economia americana, baseado em insumos e relações interempresariais. Um dos grandes méritos desse modelo foi o seu caráter quantitativo, propiciado em parte pelo grande desenvolvimento estatístico e de coleta de dados obtidos no início do século e com as novas pesquisas implantadas. Devido, principalmente, à quebra da bolsa de valores americana em 1929, o interesse na modelagem de problemas econômicos obtem um grande crescimento.

Em 1941, a Força Aérea Americana reúne um grupo de pesquisadores com o objetivo de aperfeiçoar e generalizar o método de Leontif. O objetivo era desenvolver um modelo automático para o planejamento de treinamento e suporte logístico. Este grupo tinha como consultor o matemático George Bernard Dantzig, que mais tarde seria o criador do método Simplex para Programação Linear. Este método, cujo desenvolvimento é iniciado em 1947 e finalizado em 1948, seria mantido em sigilo até a realização da 1ª Conferência de Programação Linear, patrocinada pela comissão Cowles. Os trabalhos apresentados nessa conferência, só viriam a público em 1951, em uma seleção editada por T. C. Koopmans, que juntamente com Dantzig, em 1948, cunharam o termo "Programação Linear".

O método Simplex de Dantzig da início a um grande esforço de pesquisa e aplicação. Nesse esforço há que se destacar duas correntes [Dorfman,1959]:

- Planejamento Gerencial ("Carnegie Institute of Technology")
- Implicações na teoria econômica (T.C.Koopmans)

Estas duas vertentes dão origem a uma série de resultados teóricos:

- Publicação do método da Projeção por C. B. Tompkins;
- Análise da questão da estabilidade numérica do Simplex por J. Hoffman;
- Publicação do método da relaxação para resolver Problemas de Programação Linear por Motzkin;
- Kuhn e Tucker publicam o trabalho "Programação Não-Linear" em [23], 1951.

Na área de aplicação econômica temos:

- Publicação do artigo "Mecanismos de mercado e maximização", por Samuelson em [31] de 1955, alterando a concepção de Leontif;
- Análise da Teoria Econômica em programação linear, por Dorfman em 1951, em [9].

A proposição da programação linear deu origem também a novos modelos de programação: a já citada programação não-linear (Kuhn-Tucker), a programação dinâmica (Bellman-1950), a programação inteira (Gomory-1958; Dantzig, Fulkerson e Johnson-1954). Este conjunto de novas técnicas deu origem à disciplina programação matemática, termo que tem sua origem em um congresso patrocinado pela Rand Composition em 1959. Culminando com a Teoria da Complexidade em 1974, conforme Minoux em [26].

1.3 Método Simplex

O método Simplex, proposto por Dantzig, foi influenciado por uma série de trabalhos anteriores. O próprio Dantzig reconheceu que os trabalhos matemáticos que mais o teriam influenciado seriam relativos à teoria dos jogos, desenvolvidos por Ville em 1938 e por Von Neumann em 1944. A grosso modo, o problema do Simplex resume-se a encontrar uma solução para um sistema de equações lineares com restrições de desigualdade.

Este problema foi inicialmente abordado por Fourier em [12] em 1826. Naquela oportunidade, Fourier investigava alguns aspectos da teoria de probabilidade. Esta abordagem o levou ao problema da solução de um sistema linear com restrições de desigualdade, que poderia ser reduzido à busca do ponto mais baixo de um conjunto poliedral. A solução proposta por Fourier, foi a busca seqüencial de vértice a vértice, até o encontro do mínimo do conjunto poliedral. Este princípio é a base do método Simplex. Este mesmo problema foi abordado por outro matemático francês, La Vallé de Poussin em [28], 1911, que propôs um método similar de resolução. Apesar disso este problema não despertou grande interesse nos demais matemáticos daquela época, sendo portanto, pouco explorado.

W. Karush, em sua tese de mestrado em [22], 1939, forneceu as condições de otimalidade para funções com restrições, contribuindo sobremaneira para a solução do problema de programação linear através do método Simplex.

Apenas para ilustrar a vasta capacidade matemática de Dantzig, relataremos um fato ocorrido em 1939. Dantzig acabara seu mestrado pela Universidade de Michigan em 1938 e fora aceito no programa de doutorado da Universidade de Berkeley na Califórnia, na área de estatística, sob orientação do professor Jerzy Neyman. Chegando atrasado para uma das aulas do professor Neyman, Dantzig observou a existência de dois problemas transcritos no quadro negro. Avaliou se tratar de uma "homework" e sem comentar com ninguém, tomou nota. Após uma semana, Dantzig compareceu ao gabinete do professor Neyman e lhe entregou a resolução dos dois problemas propostos. Somente naquele momento, diante da surpresa do mestre, ficou sabendo se tratar aqueles "exercícios" de dois problemas em aberto na área de estatística. Mais tarde, aqueles problemas se tornariam o foco central da Tese de doutorado de Dantzig. O leitor interessado, poderá verificar esta e outras histórias acerca de Dantzig em [5].

1.4 Método de Kantorovich

Um trabalho importante para a história da programação linear foi o trabalho desenvolvido pelo matemático soviético L.V. Kantorovich, que em 1939 demonstra o significado prático de uma restrita classe de modelos de programação linear para o planejamento da produção. Propõe um algoritmo rudimentar para a solução de tais problemas. Seu trabalho foi apresentado em uma monografia intitulada "Mathematical Methods in the Organization and Planning of Production".

Um grupo de pesquisadores soviéticos, entre eles Kantorovich, publicam uma série de trabalhos a este respeito. Porém estes trabalhos não tiveram grande repercussão na URSS e, devido à "guerra fria", permaneceram desconhecidos fora dela. O principal problema, segundo o próprio Kantorovich, era a desconfiança que os economistas tinham dos matemáticos. Desta forma, os trabalhos de Kantorovich só seriam divulgados no final da década de 1950. Como comenda e reconhecimento de suas contribuições à teoria da alocação de recursos, Kantorovich divide em 1975 o prêmio Nobel de Economia com T. C. Koopmans (um dos co-criadores do Simplex); este fato também pode ser verificado em [5].

1.5 Aplicações Industriais da Programação Linear

A utilização de métodos de otimização no setor industrial é bastante vasta, e poderíamos citar como exemplos:

- A primeira, e talvez a mais importante aplicação industrial, foi nas refinarias de petróleo. Esta aplicação foi o resultado de um trabalho desenvolvido por Chanes, Cooper e Mellon em [4], 1952.
- Na indústia de alimentos, a primeira aplicação foi realizada em 1953. O problema consistia em abastecer 10 depósitos através de 6 fábricas de catchup.
- Outra grande aplicação industrial aparece, mais tarde, no problema conhecido como da "dieta". Este consiste no balanceamento da alimentação de um rebanho de vacas leiteras com intuito de aumentar a produção, minimizando os custos.

1.6 Aplicações Recentes

Após três décadas de absoluto domínio, a hegemonia do Simplex foi posta em questão por dois novos métodos. Em 1979, L. G. Kanchian, em [20], prova teoricamente que o método elipsóide, proposto por N. S. Shor, D. B. Youdin e S. Nemirovs, pode exibir melhor desempenho em sistemas de grande porte, por possuir convergência polinomial ao passo que o Simplex é de convergência exponencial. Apesar dessa superioridade teórica, as aplicações práticas apontam para um melhor desempenho do Simplex.

Já em 1984, Karmarkar em [19], propõe o método de pontos interiores que, por apresentar convergência polinomial, mostra-se mais eficiente que o Simplex para problemas de grande porte. Ao contrário do método elipsóide, esta superioridade teórica foi comprovada na prática. Este desempenho tornou o método Karmarkar o alvo de um crescente interesse por parte dos pesquisadores da atualidade, e tem sido aplicado com grande sucesso.

Atualmente, o planejamento através de técnicas de programação matemática é indispensável em qualquer setor produtivo. Em grandes organizações, as decisões são bastante complexas, dependentes usualmente de milhares de variáveis de decisão, tornando impossível que decisões baseadas somente na intuição humana sejam tomadas de forma a obter a eficiência máxima.

1.7 Otimização na Internet

Devido a seu caráter multidisciplinar e à utilização da computação como ferramenta, a programação matemática ou pesquisa operacional tem sido uma disciplina bastante divulgada através da Internet. Além das páginas dedicadas à divulgação de artigos e de associações internacionais, muitas têm se tornado fóruns de debate e troca de informações. Dentro desta perspectiva, duas páginas podem ser destacadas:

- RIOT(http://www.riot.ieor.berkeley.edu/c̃ander/riot/index.html)
- The Optimization Technology Center (http://www-c.mcs.anl.gov/home/otc)

A RIOT ("Remote Interactive Optimization Testbed") oferece ferramentas educacionais e de pesquisa para problemas de otimização, bem como apresenta uma série de novas aplicações de programação matemática.
O Centro Tecnológico de Otimização (OTC), mantido pelo governo norte-americano, através de um convênio entre o departamento de energia, o laboratório nacional de Argonne e a Universidade de Northwester, se propõe auxiliar os setores industrial, acadêmico e governamental na aplicação de ferramentas de Otimização. Este "site"está centrado no sistema NEOS ("Network-Enabled Optmization System"). Este sistema se propõe a

oferecer suporte teoórico e computacional na Internet. Através dele, problemas podem ser enviados ao OTC, sendo resolvidos pelos recursos computacionais do servidor NEOS. Além deste suporte, o NEOS oferece um guia da tecnologia de Otimização, com numerosos exemplos de aplicação, como, por exemplo, a aplicação do Simplex a problemas de dieta e carteira de ações. Uma biblioteca com programas avançados de otimização também está disponível.

Além dos "sites"descritos, existem muitos outros que oferecem rotinas computacionais, jornais eletrônicos, listas de discussão, etc. Despender algum tempo na consulta a estes "sites", antes de uma investigação científica, é um procedimento recomendável. Com isso, pode-se concluir que a utilização da Internet vem representando um novo marco na história da programação matemática.

CAPÍTULO 2

Resultados Preliminares

2.1 Resultados básicos

Neste capítulo, apresentaremos alguns resultados preliminares que apoiarão várias demonstrações no transcorrer de nosso trabalho. Operações básicas em álgebra linear, bem como desigualdades em conjuntos fechados e ou abertos, convexos ou não serão abordados. Apresentamos ainda, uma versão do Teorema da Função Implícita que dará suporte a demonstrações futuras. A bibliografia utilizada neste capítulo foi [3], [24], [25], [34] e [17].

2.1.1 Notações e terminologias

As seguintes notações e resultados serão usados ao longo desta dissertação. A partir de agora, $I\!R$ é o conjunto dos números reais, $I\!R^n$ o espaço euclidiano n-dimensional, ou seja, é o conjunto dos vetores reais da forma:

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

$I\!R^n_+ = \{(x_1,...,x_n)^T \in I\!R^n : \ x_i \geq 0 \ \forall \ i = 1,...,n\}$ denotará o ortante não-negativo e $I\!R^n_{++} = \{(x_1,...,x_n)^T \in I\!R^n : \ x_i > 0 \ \forall \ i = 1,...,n\}$ denotará o ortante positivo. $M_{m\times n}(I\!R)$ é o conjunto das matrizes reais de dimensão $m \times n$. O (i,j)-ésimo elemento de uma matriz $A \in I\!R^{n\times m}$ é denotado por A_{ij} e a j-ésima coluna é denotada por A_j. A transposta de $A \in I\!R^{n\times m}$ é denotado por A^T. Notamos A^{-1} à inversa da matriz $A \in M_{n\times n}(I\!R)$, quando existir. O vetor $x^T = (x_1, x_2, \ldots, x_n)$ é o transposto do vetor x. x_j é a j-ésima coordenada do vetor x e $||x||$ é a norma euclidiana de $x \in I\!R^n$. O produto interno euclidiano entre $x \in I\!R^n$ e $y \in I\!R^n$ é dado por $\langle x, y\rangle = \sum_{i=1}^n x_i y_i$. O vetor $e = (1, \ldots, 1)^T$, isto é, o vetor com todas as coordenadas iguais a um, com dimensão de acordo com o contexto. O vetor e^i

é um vetor de zeros, com o valor 1 na posição i. Notamos por $0_{m\times n}$ a matriz nula de dimensão $m\times n$ e I é a matriz identidade com ordem de acordo com o contexto. Teremos ainda que $X = diag(x)$ é a matriz diagonal formada pelas coordenadas do vetor x. Agora, dado $u \in I\!R^n_{++}$, definimos $logu = (logu_1,\dots,logu_n)^T \in I\!R^n$ e para $u \in I\!R^n$, definimos $e^u = (e^{u_1},\dots,e^{u_n})^T \in I\!R^n$.

2.1.2 Conjuntos Convexos e Cones

Apresentamos agora, algumas definições e proposições acerca de conjuntos convexos. Estes resultados se mostrarão cruciais para muitas demontrações no escopo do trabalho a ser apresentado.

Inicialmente, apresentamos um resultado que trata da decomposição de uma dada matriz em duas partes, uma básica (B) e outra não básica (N).

Lema 2.1.1 *Seja $A \in M_{m\times n}(I\!R)$ com posto (A) = m e $m < n$. Se o conjunto*

$$F = \{x \in I\!R^n : Ax = b,\ x \geq 0\},$$

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

$$\bar{x} = (A_B^{-1}b, 0) \in F.$$

Prova. Primeiro, note que sendo o posto (A) = $m > 0$, permutando eventualmente as colunas, existe uma submatriz inversível $A_B \in M_{m\times m}(I\!R)$ de A. Se b for o vetor nulo, então qualquer submatriz inversível $A_B \in M_{m\times m}(I\!R)$ de A satisfaz $A_B^{-1}b \geq 0$. Suponhamos que b não é o vetor nulo e seja $x = (x_1,...,x_n)^T \in F$. Então $Ax = b$, com x diferente do vetor nulo. Sem perda de generalidade, permutando eventualmente as coordenadas x pode ser escrito na forma $x = (x_{\tilde{B}}, 0)^T$, onde $x_{\tilde{B}} > 0$. Considere a submatriz $A_{\tilde{B}}$ de A, correspondente às coordenadas não nulas do vetor x. Deste modo,

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

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

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

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

Analisemos o caso i). Neste caso temos que $p \leq m$, pois posto(A) = m. Se $p = m$, a matriz $A_B = A_{\tilde{B}}$ é inversível. Se $p < m$, podemos adicionar $(m-p)$ colunas de A ao conjunto $\{a_1,...,a_p\}$, de modo que $\{a_1,...,a_p,a_{p+1},...,a_m\}$ seja um conjunto linearmente independente. Seja $A_B \in M_{m\times m}$ a submatriz de A, cuja as colunas sejam os vetores $\{a_1,...,a_p,a_{p+1},...,a_m\}$. Desde que $\{a_1,...,a_p,a_{p+1},...,a_m\}$ é um conjunto linearmente independente, A_B é inversível. Considere o vetor $\bar{x}_B = (x_1,...,x_p,0,...,0) = (x_{\tilde{B}},0) \in I\!R^m$ e note que (2-1) é equivalente a $A_B\bar{x}_B = b$, ou seja, $\bar{x}_B = A_B^{-1}b$. Desde que $x = (\bar{x}_B,0) \in F \subset I\!R^n$, a segunda condição que define F implica que $A_{(B)}^{-1}b = \bar{x}_B \geq 0$ e $x_i \geq 0$ para $i \notin B$. Analisemos o caso ii). Como o conjunto $\{a_1,...,a_p\}$ é linearmente dependente, então existe um vetor não nulo $y \in I\!R^p$, tal que $A_{\tilde{B}}y = 0$. Multiplicando esta última igualdade por ε e subtraindo em (2-1) temos que $A_{\tilde{B}}(x_{\tilde{B}} - \varepsilon y) = b$. Agora, tomando $\varepsilon = \min\{x_i/|y_i| : x_i > 0,\ y_i \neq 0\}$ teremos $x_{\tilde{B}} - \varepsilon y \geq 0$ e pelo menos uma de suas coordenadas igual a zero, pois $x_{\tilde{(B)}} > 0$. Seja $A_{\hat{B}}$ uma submatriz de $A_{\tilde{B}}$, que por sua vez também é submatriz de A, correspondente às coordenadas não nulas do vetor $x_{\tilde{B}} - \varepsilon y$. Assim

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

onde $x_{\hat{B}} > 0$ é o vetor formado pelas coordenadas não nulas do vetor $x_{\tilde{B}} - \varepsilon y$. Se as colunas de $A_{\hat{B}}$ são linearmente independente recaimos no caso i). Caso contrário, observando que b é não nulo e consequentemente $A_{\hat{B}}$ também é não nula, podemos repetir o processo não mais que $p-1$ vezes para então recair no caso i). Isto prova a primeira parte. A prova da segunda parte é imediata, pois $A\bar{x} = [A_B|A_N](A_B^{-1}b,0)^T = b$ e $\bar{x} = (A_B^{-1}b,0) \geq 0$. $\square$

Um conjunto C de um espaço vetorial euclidiano $I\!R^m$ é dito *convexo* se, dados dois pontos $x_1,x_2 \in C$, tivermos

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

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

Um subconjunto C de um espaço vetorial euclidiano $I\!R^m$ é chamado de cone quando satisfaz a seguinte condição:

$$\forall\, z \in C,\ \alpha \geq 0 \implies \alpha z \in C.$$

Proposição 2.1.2 *Seja $A \in M_{m\times n}(I\!R)$ com posto (A) = $m > 0$. Então o conjunto*

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

é um cone não vazio, fechado e convexo.

Prova. É imediato que C é não vazio, pois $0 \in C$. Para mostrarmos que C é cone, seja $z \in C$ e $\alpha > 0$. Agora, seja $x \geq 0$, tal que $Ax = z$. Por linearidade temos $\alpha z = \alpha Ax = A(\alpha x)$. Como $x \geq 0$ e $\alpha > 0$, segue que $\alpha x \geq 0$, assim $\alpha z \in C$ e portanto C é cone. Resta mostrar que C é fechado, ou seja, dada uma seqüência $\{z^k\} \subset C$ com $\lim z^k = \bar{z}$ então $\bar{z} \in C$. Como $\{z^k\} \subset C$, por definição existe uma sequência $\{x^k\} \subset \mathbb{R}^n$, tal que $Ax^k = z^k$ e $x^k \geq 0$, para todo k. Logo x^k pertence ao conjunto

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

Pelo Lema 2.1.1, existe uma partição da matriz A, na forma $A = [A_B : A_N]$, com $A_B \in M_{m\times m}(\mathbb{R})$ inversível, tal que $\bar{x}^k = (A_B^{-1}z^k, 0) \in F$, ou seja,

$$A\bar{x}^k = z^k, \qquad \bar{x}^k \geq 0.$$

Como $\lim z^k = \bar{z}$, temos que $\lim \bar{x}^k = (A_B^{-1}\bar{z}, 0) \geq 0$. Denote $(A_B^{-1}\bar{z}, 0) = \bar{x}$ e observe que

$$A\bar{x} = \bar{z}, \qquad \bar{x} \geq 0,$$

e isto implica que $\bar{z} \in C$. Logo C é fechado. Finalmente, mostremos que C é convexo. Tomemos x_1 , $x_2 \geq 0$, tais que $Ax_1 = z_1$ e $Ax_2 = z_2$, onde z_1 , $z_2 \in C$. Agora, tomando $\alpha \in (0,1)$, temos $\alpha z_1 + (1-\alpha)z_2 = A(\alpha x_1 + (1-\alpha)x_2)$, como $(\alpha x_1 + (1-\alpha)x_2) \geq 0$, segue que a combinação convexa $\alpha z_1 + (1-\alpha)z_2 \in C$, ou seja C é convexo. □

Dizemos que $d \in \mathbb{R}^n$ é uma *direção de recessão* do conjunto $D \subset \mathbb{R}^n$ quando

$$x + td \in D, \qquad \forall x \in D, \quad \forall t \in \mathbb{R}_+.$$

Proposição 2.1.3 *Seja $D \subset \mathbb{R}^n$ um conjunto convexo, fechado e não vazio. Então, se D é ilimitado existe pelo menos uma direção de recessão diferente de zero.*

Prova. Suponhamos D ilimitado e fixemos $x \in D$ arbitrário. Seja $\{x^k\} \subset D$ qualquer seqüência, com $x^k \neq x, \forall k$, tal que

$$||x^k - x|| \to \infty \ \ (k \to \infty).$$

Seja $d \in \mathbb{R}^n/\{0\}$, qualquer ponto de acumulação da sequência $\{d^k\}$, onde:

$$d^k = (x^k - x)/||x^k - x||.$$

Seja $t \in I\!R_+$ qualquer. Tem-se que:

$$x + td^k = (1 - t/||x^k - x||)x + (t/||x^k - x||)x^k \in D,$$

onde a pertinência vale para todo k, suficientemente grande (tal que $t/||x^k - x|| \leq 1$), pela convexidade de D. Como D é fechado, obtemos que

$$x + td = lim_{k_j \to \infty}(x + td^{k_j}) \in D, \qquad \forall\, t \in I\!R_+,$$

onde $lim_{k_j \to \infty} d^{k_j} = d$. Ou seja, existe pelo menos uma direção de recessão, d em D, que é diferente de zero, o que conclui nossa prova. □

2.1.3 Projeção em Conjuntos Convexos

Para esta subseção, reservamos a apresentação de resultados clássicos, tais como da projeção, o Teorema de Separação de Convexos, bem como o Lema de Farkas e o seu corolário.

Seja $C \neq \emptyset$ um conjunto convexo fechado em $I\!R^n$. Para $x \in I\!R^n$ fixado, consideremos o seguinte problema:

$$\inf\{\|y - x\| : y \in C\}. \tag{2-2}$$

Dado $c \in C$, considere um conjunto de sub-nível $S := \{y \in I\!R^n : \|y - x\| \leq \|c - x\|\}$. Então (2-2) é equivalente a

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

o qual possui uma solução, pois $y \mapsto \|y - x\|$ é contínua, S é compacto e $C \cap S$ é compacto. Portanto, deduzimos a *existência* de um ponto em C que minimiza a distância a x e desta forma o ínfimo em (2-2) é de fato um mínimo. Veremos que existe um único ponto que minimiza a distância de x ao convexo C.

Teorema 2.1.4 *Seja $x \in I\!R^n$. Se, para algum $y_x \in C$ tem-se $\|y_x - x\| \leq \|y - x\|$, para todo $y \in C$, então*

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

Prova. Tome y arbitrário em C, de modo que $y_x + \alpha(y - y_x) \in C$, para todo $\alpha \in]0,1[$. Então, podemos escrever

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

Esta desigualdade implica que

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

Dividindo por $\alpha > 0$ e fazendo $\alpha \to 0$ obtemos:

$$(x - y_x)^T(y - y_x) \leq 0.$$

$\square$

Teorema 2.1.5 *Seja $x \in I\!R^n$. Existe um único ponto $y_x \in C$, tal que $\|y_x - x\| \leq \|y - x\|$, para todo $y \in C$.*

Prova. Suponhamos que existam $y_x \in C$ e $y'_x \in C$ tais que:

$$\|y_x - x\| \leq \|y - x\| \quad \text{e} \quad \|y'_x - x\| \leq \|y - x\|,$$

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

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

e somando-as, obtemos:

$$0 \leq \|y'_x - y_x\|^2 = (y'_x - y_x)^T(y'_x - y_x) \leq 0.$$

Logo, $y'_x = y_x$. Portanto, tal ponto y_x é único. $\square$

Seja C um conjunto em $I\!R^n$, convexo e fechado, defina a função distância a C como sendo

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

A projeção ortogonal sobre C é a função $P_C : I\!R^n \to C$, definida para $x \in I\!R^n$ como sendo o único ponto $P_c(x) \in C$ tal que

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

Teorema 2.1.6 *Seja $C \subset I\!R^n$, um conjunto não vazio convexo e fechado. Então, para todo $b \notin C$ a projeção ortogonal $P_C(b)$ de b sobre C esta bem definida e satisfaz a seguinte* *desigualdade*

$$(x - P_C(b))^T (y - P_C(b)) \leq 0, \qquad \forall\, y \in C. \tag{2-3}$$

Prova. Segue imediatamente do Teorema 2.1.5 que a projeção esta bem definida. E o Teorema 2.1.4 implica a desigualdade (2-3). □

Vamos apresentar agora, o Teorema de Separação, outro resultado clássico e vital para futuras conclusões.

Teorema 2.1.7 *(Teorema de Separação) Seja C um conjunto convexo, fechado e não vazio em $\mathbb{R}^n$ e considere $b \notin C$. Então, existe um vetor $a \in \mathbb{R}^n$, $a \neq 0$ e um escalar $\delta \in \mathbb{R}$, tais que*

$$a^T x \leq \delta < a^T b, \qquad \forall\, x \in C.$$

Prova. Se C é um conjunto convexo fechado e não vazio e $b \notin C$. Pelo Teorema 2.1.6, a projeção $P_C(b) \in C$ está bem definida e satisfaz

$$(b - P_C(b))^T (x - P_C(b)) \leq 0, \tag{2-4}$$

para cada $x \in C$. Defina:

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

De (2-4) e definição de a temos $a^T (x - P_C(b)) \leq 0$ ou equivalentemente, $a^T x \leq a^T P_C(b)$. Esta equação, juntamente com a definição de δ, implica $a^T x \leq \delta$, que é a primeira desigualdade desejada.

Para demonstrar a segunda desigualdade note que pelas definições de a e δ temos:

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

onde a desigualdade estrita segue do fato que $b \notin C$. Assim, $a^T b > \delta$, o que conclui a demonstração. □

A seguir, apresentaremos dois resultados fundamentais para a seqüência de demonstrações que faremos nos próximos capítulos. Trata-se do lema de Farkas e de seu corolário. Devido a importância destes dois resultados, existe uma grande variedade de demonstrações na literatura. Escolhemos estas duas demonstrações por acreditarmos serem as mais claras e didáticas. Estes resultados podem ser encontrados em [34].

Lema 2.1.8 *(Lema de Farkas) Seja $A \in M_{m \times n}(\mathbb{R})$ com posto $(A) = m > 0$ e $b \in \mathbb{R}^m$. Então, exatamente um dos dois seguintes sistemas tem uma solução:*

(1) $Ax = b$ *e* $x \geq 0$*;*

(2) $A^T y \leq 0$ *e* $b^T y > 0$*.*

Prova. Suponha que (1) tenha uma solução $\hat{x}$, então $A\hat{x} = b$, e $\hat{x} \geq 0$. Assim, dado $y \in \mathbb{R}^m$, tal que $A^T y \leq 0$, temos:

$$b^T y = (A\hat{x})^T y = \hat{x}^T A^T y \leq 0,$$

isto implica que o sitema (2) não tem solução.

Agora, suponha que (1) não tenha solução. Pela Proposição 2.1.2, o seguinte conjunto

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

é um cone convexo e fechado. Pela nossa suposição $b \notin S$. Sabemos também que S é não vazio, pois $0 \in S$. Assim pelo Teorema de Separação 2.1.7, existem $y \in \mathbb{R}^m$ e um escalar $\delta \in \mathbb{R}$ tais que:

$$z^T y \leq \delta < b^T y, \qquad \forall\, z \in S. \tag{2-5}$$

Como $0 \in S$, segue da primeira desigualdade acima que $0 \leq \delta$, assim $b^T y > 0$. Agora, seja e^j é o j-ésimo vetor unitário em $\mathbb{R}^n$. Como S é um cone e $e^j \geq 0$ temos que:

$$A(\lambda e_j) \in S, \qquad \forall\, \lambda > 0.$$

Esta inclusão, juntamente com a primeira desigualdade em (2-5), implica que

$$\delta \geq \left(A(\lambda e^j)\right)^T y = \lambda (e^j)_j^T A^T y, \qquad \forall\, \lambda > 0.$$

Dividindo a última desigualdade por λ e fazendo $\lambda \to \infty$, temos

$$(e^j)_j^T A^T y \leq 0, \qquad j = 1, \ldots, n,$$

o que implica que $A^T y \leq 0$. Isto, juntamente com $b^T y > 0$, provado acima, implica que y é solução de (2). Portanto, o lema esta provado. $\square$

Corolário 2.1.9 *(Corolário de Farkas) Seja* $A \in M_{m\times n}(\mathbb{R})$ *e* $c \in \mathbb{R}^n$*. Então, exatamente um dos dois seguintes sistemas tem uma solução:*

(3) $A^T y \leq c$;

(4) $Ax = 0$, $x \geq 0$ *e* $c^T x < 0$.

Prova. Para demonstrar o Corolário de Farkas, reescresvemos o sistema do item (3) acima em um formato padrão, equivalente ao sistema do item (1) do Lema de Farkas e então utilizar a própria prova deste lema para concluir a demonstração. Primeiramente, vamos substituir no sistema do item (3):

$$y = y_1 - y_2, \qquad y_1 \geq 0 \ , \ y_2 \geq 0.$$

Então, o sistema do item (3) é equivalente ao seguinte sistema $A^T(y_1 - y_2) \leq c$, ou ainda,

$$(A^T \ -A^T) \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \leq c,$$

onde $(A^T \ -A^T) \in M_{n \times 2m}(\mathbb{R})$. Agora, fazendo $z = (y_1, y_2)^T \in \mathbb{R}^{2m}$ teremos:

$$(A^T \ -A^T)z \leq c, \qquad z \geq 0.$$

Acrescentando uma variável de folga $s \geq 0$, o sistema acima pode ser reescrito da seguinte forma:

$$(A^T \ -A^T)z + s = c, \qquad z \geq 0, \ s \geq 0.$$

Fazendo

$$\tilde{A} = (A^T \ -A^T \quad I) \in M_{n \times (2m+n)}(\mathbb{R}), \qquad \tilde{x} = \begin{pmatrix} z \\ s \end{pmatrix} \in \mathbb{R}^{2m+n}.$$

Segue-se que o sistema acima pode ser reescrito na seguinte forma standard:

$$\tilde{A}\tilde{x} = c, \qquad \tilde{x} \geq 0. \tag{2-6}$$

Portanto, o sistema do item (3) acima é equivalente ao último sistema. Por outro lado, consideremos o seguinte sistema:

$$\tilde{A}^T \tilde{y} \leq 0, \qquad c^T \tilde{y} > 0. \tag{2-7}$$

Agora, note que usando a definição de $\tilde{A}$, podemos reescrever o último sistema da seguinte forma:

$$A\tilde{y} \leq 0, \qquad A\tilde{y} \geq 0, \qquad \tilde{y} \leq 0, \qquad c^T \tilde{y} > 0.$$

Fazendo $-\tilde{y} = x$, temos das duas primeiras desigualdades que $Ax = 0$ e das duas últimas que $x \geq 0 \ \ e \ \ c^T x < 0$, ou seja, o sistema (2-7) é equivalente ao sistema do item (4). Veja que o sistema do item (3) é equivalente ao sistema (2-6) e o sistema do item (4) é equivalente ao sistema (2-7). Portanto, aplicando o Lema de Farkas aos sistemas (2-6) e (2-7), concluímos que exatamente um dos dois sistemas (item (3) ou item (4)) tem uma solução, o que conclui nossa prova. $\square$

2.1.4 Resultados importantes de Álgebra Linear

Na caracterização de nosso problema, necessitaremos de ferramentas de álgebra linear para a simplificação e suporte de provas de alguns resultados. Apresentamos então, as definições de espaço nulo e espaço linha de uma matriz, bem como uma proposição de ortogonalidade e um lema sobre a não singularidade de uma matriz específica.

O espaço nulo de A é definido como $\text{Null}(A) = \{x \in \mathbb{R}^n; Ax = 0\}$ e o espaço linha de A é definido como $\text{Im}(A^T) = \{x \in \mathbb{R}^n; x = A^T z, z \in \mathbb{R}^m\}$.

Proposição 2.1.10 *Sejam $A \in M_{m\times n}(\mathbb{R})$, $\mathbb{N}ull(A)$, o espaço nulo de A e $\mathbb{I}m(A^T)$, o espaço linha de A. Então $\mathbb{N}ull(A)$ é ortogonal a $\mathbb{I}m(A^T)$*

Prova. Suponha que $x \in \mathbb{N}ull(A)$, então por definição temos que:

$$Ax = 0. \tag{2-8}$$

E suponha também que $y \in \mathbb{I}m(A^T)$, então por definição:

$$y = A^T z, \ \text{ para algum } z \in \mathbb{R}^m. \tag{2-9}$$

Agora, usando (2-8), (2-9) e realizando algumas manipulações matriciais, obtemos:

$$y^T x = (A^T z)x = z^T Ax = 0,$$

o que prova o resultado desejado. □

Lema 2.1.11 *Seja $M \in M_{m\times n}(\mathbb{R})$. Se o $Posto(M) = m$ e $D = diag(d)$, onde $d \in \mathbb{R}^n_{++}$, então a matriz*

$$\begin{pmatrix} M & 0 & 0 \\ 0 & M^T & I \\ D & 0 & I \end{pmatrix},$$

é não singular.

Prova. Para mostrar a não singularidade, basta mostrar que a seguinte equação matricial

$$\begin{pmatrix} M & 0 & 0 \\ 0 & M^T & I \\ D & 0 & I \end{pmatrix} \begin{pmatrix} u \\ v \\ w \end{pmatrix} = 0,$$

tem o vetor zero como solução única. Primeiro, note que esta equação é equivalente ao seguinte sistema:

$$Mu = 0, \qquad M^T v + w = 0, \qquad Du + w = 0. \tag{2-10}$$

Isolando o valor de w na última equação acima e substituindo na segunda, temos:

$$M^T v - Du = 0.$$

Como D é não singular, segue da última equação que $u = D^{-1}M^T v$. Então, usando a primeira equação de (2-10), teremos $MD^{-1}M^T v = 0$. Logo, pela definição de D, a equação acima pode ser escrita como:

$$(D^{-1/2}M^T)^T(D^{-1/2}M^T)v = 0,$$

o que implica que $D^{-1/2}M^T v = 0$, ou ainda, $M^T v = 0$. Como Posto$M = m$, segue-se que $v = 0$, que susbstituindo na segunda equação de (2-10), implica que $w = 0$. Usando esse fato na última equação de (2-10), concluímos que $u = 0$. □

2.1.5 Funções Convexas

Nesta subseção, apresentaremos alguns resultados, bem como alguns exemplos acerca de funções convexas, que serão utilizados no decorrer de nosso trabalho.

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

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

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

Proposição 2.1.12 *Seja $f : I \to \mathbb{R}$ uma função diferenciável no intervalo aberto I. A função f é convexa se, e somente se*

$$(y-x)f'(x) + f(x) \le f(y), \qquad \forall\, x, y \in I.$$

A demonstração da proposição acima pode ser encontrada no Corolário 2, página 108, em [24].

Proposição 2.1.13 *Seja $f : I \to \mathbb{R}$ uma função duas vezes diferenciável no intervalo aberto I. A função f é convexa se, e somente se*

$$f''(x) \geq 0, \qquad \forall\, x \in I. \tag{2-11}$$

Se a desigualdade em (2-11) é estrita para todo $x \in I$, então f é estritamente convexa.

A demonstração da proposição acima pode ser encontrada no Teorema 4, página 106, em [24].

Dada $f : \Omega \to \mathbb{R}$, onde Ω é um subconjunto de $\mathbb{R}^n$. O *epígrafo* de f é o conjunto não vazio

$$\mathrm{epi} f := \{(x,r) \in \Omega \times \mathbb{R} : r \geq f(x)\}.$$

Seja C um conjunto não vazio convexo em $\mathbb{R}^n$. Uma função $f : C \to \mathbb{R}$ é dita convexa em C quando, para todo par $x, y \in C$ e todo $\lambda \in [0,1]$

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

Note que esta definição estende à definição de funções convexas com domínio em um intervalo real.

Proposição 2.1.14 *Seja $f : C \to \mathbb{R}$ uma função, onde $C \subseteq \mathbb{R}^n$ é um conjunto convexo. As seguintes propriedades são equivalentes:*
i) f é convexa,
ii) o epígrafo de f é um conjunto convexo em $\mathbb{R}^n \times \mathbb{R}$.

Prova. $i) \Rightarrow ii)$. Sejam $(x, f(x))$ e $(y, f(y)) \in \mathrm{epi} f$ e $\lambda \in [0,1]$. Por hipótese, f é convexa. Logo, temos que para $\lambda \in (0,1)$,

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

Isto implica que $\lambda(x, f(x)) + (1-\lambda)(y, f(y)) \in \mathrm{epi} f$, para $\lambda \in [0,1]$. Portanto, $epi f$ é um subconjunto convexo de $\mathbb{R}^n \times \mathbb{R}$.
$ii) \Rightarrow i)$. Dados $x, y \in C$ e $\lambda \in [0,1]$. Note que $(x, f(x))$ e $(y, f(y))$ pertencem ao $\mathrm{epi} f$. Como o epígrafo de f é um subconjunto convexo de $\mathbb{R}^n \times \mathbb{R}$ e $\lambda \in [0,1]$, temos que:

$$(\lambda x + (1-\lambda)y, \lambda f(x) + (1-\lambda) f(y)) \in epi f.$$

Pela definição de $\mathrm{epi} f$, segue-se que $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$, mas como x, y são quaisquer pontos de C e λ é qualquer ponto de $[0,1]$, concluímos que f é convexa.
$\square$

Proposição 2.1.15 *Seja $f : C \to I\!R$ uma função, onde $C \subseteq I\!R^n$ é um conjunto convexo. A função f é convexa se, e somente se, a restrição de f a todo segmento de reta, que esta contido em C, é uma função convexa. Mais precisamente, a função f é convexa se, e somente se, para todo $x \in C$ e $v \in I\!R^n$, a função $h : I_{x,v} \to I\!R$ definida por*

$$h(t) = f(x+tv),$$

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

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

$$\begin{aligned} h\big(\lambda t_1 + (1-\lambda)t_2\big) &= f(x+[\lambda t_1 + (1-\lambda)t_2]v) \\ &= f\big(\lambda(x+t_1v) + (1-\lambda)(x+t_2v)\big) \\ &\leq \lambda f(x+t_1v) + (1-\lambda)f(x+t_2v) \\ &= \lambda h(t_1) + (1-\lambda)h(t_2). \end{aligned}$$

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

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

x e isto implica que f é convexa. □

A demonstração do resultado anterior pode ser encontrada no Teorema 8, na página 76, em [25]. Veja que deste resultado segue imediatamente que a restrição de uma função convexa a um conjunto convexo é convexa.

Proposição 2.1.16 *Seja $f : C \to I\!R$ uma função diferenciável, onde $C \subseteq I\!R^n$ é um conjunto aberto e convexo. A função f é convexa se, e somente se,*

$$f(y) \geq f(x) + \langle \nabla f(x), y-x \rangle,$$

para todo $x, y \in C$. A desigualdade acima é conhecida como "desigualdade do gradiente".

Prova. Dados $x, y \in C$. Tome $v = y - x$ e defina h como na Proposição 2.1.15. Note que I_v é aberto, pois C é aberto. Sabendo-se que f é convexa, pela Proposição 2.1.15, segue-se que h é convexa. Então, pela Proposição 2.1.13, temos que

$$h(1) \geq h(0) + h'(0)(1-0),$$

ou seja, que $f(y) \geq f(x) + \langle \bigtriangledown f(x), y-x \rangle$, e isto prova a primeira parte. Reciprocamente, dados $x \in C$ e $v \in \mathbb{R}^n$. Defina h, como na Proposição 2.1.15. Note que o intervalo $I_{x,v}$ é aberto e que $h'(t) = \langle \bigtriangledown f(x+tv), v \rangle$. Então, pela definição de h, temos que para todo $s,t \in I$:

$$\begin{aligned} h(s) &= f(x+sv) \\ &\geq f(x+tv) + \langle \bigtriangledown f(x+tv), (s-t)v \rangle \\ &= h(t) + h'(t)(s-t). \end{aligned}$$

Portanto, pela Proposição 2.1.13 segue-se que h é convexa e isto implica, pela Proposição 2.1.15, que f é convexa e a proposição está demonstrada. □

Veja no Teorema 3.4.7, na página 146, em [17] que, se uma função f é duas vezes diferenciável, ela pode ser caracterizada, da seguinte forma.

Proposição 2.1.17 *Seja $f : C \to \mathbb{R}$ uma função duas vezes diferenciável, onde $C \subseteq \mathbb{R}^n$ é um conjunto aberto e convexo. A função f é convexa se, e somente se, para todo $x \in C$, $v \in \mathbb{R}^n$,*

$$\langle \nabla^2 f(x)v, v \rangle \geq 0. \tag{2-12}$$

Se a desigualdade (2-12) é estrita , então f é estritamente convexa.

Prova. Dados $x \in C$ e $v \in \mathbb{R}^n$, defina h como na Proposição 2.1.15. Note que $I_{x,v}$ é aberto, pois C é aberto e que $0 \in I_{x,v}$. Sabendo-se que f é convexa, pela Proposição 2.1.15, segue-se que h é convexa. Então, pela Proposição 2.1.13, temos que

$$h''(0) \geq 0. \tag{2-13}$$

Pela definição de h e regra da cadeia, temos que $h''(t) = \langle \nabla^2 f(x+tv)v, v \rangle$, assim (2-13) é equivalente a

$$\langle \nabla^2 f(x)v, v \rangle \geq 0,$$

o que prova a primeira parte. Reciprocamente, dados $x \in C$ e $v \in \mathbb{R}^n$. Defina h, como na Proposição 2.1.15. Note que o intervalo $I_{x,v}$ é aberto e que

$$h''(t) = \langle \nabla^2 f(x+tv)v, v \rangle.$$

Então, pelas definições de h, de $I_{x,v}$ e pela equação (2-12), temos que para todo $t \in I_{x,v}$

$$h''(t) = \langle \nabla^2 f(x+tv)v, v\rangle \geq 0.$$

Portanto, pela Proposição 2.1.13, segue-se que h é convexa e isto implica, pela Proposição 2.1.15, que f é convexa e a proposição está demonstrada. □

Vamos exibir agora, alguns exemplos de funções convexas que brevemente figurarão no escopo dos problemas abordados.

Exemplo 2.1.18 *Seja a função $\varphi : \mathbb{R}^n_{++} \to \mathbb{R}$ definida por*

$$\varphi(x) = \sum_{i=1}^{n} x_i \log x_i, \tag{2-14}$$

Note que o gradiente de φ é dado por $\nabla\varphi(x) = (1+\log x_1, \ldots, 1+\log x_n)$ e sua hessiana por

$$\nabla^2\varphi(x) = \begin{pmatrix} \frac{1}{x_1} & 0 & .. & .. & 0 \\ 0 & \frac{1}{x_2} & .. & .. & 0 \\ . & . & .. & .. & . \\ 0 & 0 & .. & .. & \frac{1}{x_n} \end{pmatrix} = diag\, x^{-1}.$$

Vamos agora, denotar X^{-1} a matriz formada pelas entradas de x^{-1}. Logo

$$diag x^{-1} = X^{-1} > 0, \tag{2-15}$$

isto é, a hessiana de φ é positiva definida. A equação (2-15) e a Proposição 2.1.17, implicam que φ é estritamente convexa.

Segue outro importante exemplo de função convexa que será, posteriormente, usada em nosso trabalho.

Exemplo 2.1.19 *Seja a função $\tilde{K_\varphi} : \mathbb{R}^n_+ \times \mathbb{R}^n_{++} \to \mathbb{R}$, associada a função φ do exemplo acima, definida por*

$$\tilde{K_\varphi}(x,y) = \sum_{i=1}^{n} \left(x_i \log \frac{x_i}{y_i} + y_i - x_i \right). \tag{2-16}$$

Assim, como fizemos para a função φ, podemos mostrar que a hessiana da função K_φ é positiva definida, logo esta função também é estritamente convexa.

A seguinte proposição nos garante que os conjuntos dos sub-níveis de uma função convexa são convexos. Esse resultado é encontrado no Teorema 3.4.1, página 133, em [17].

Proposição 2.1.20 *Sejam $C \subset I\!R$ um conjunto convexo, i.e., um intervalo e $f : C \to I\!R$ uma função convexa. Então para todo $c \in I\!R$ o conjunto de sub-nível C*

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

é um conjunto convexo.

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

$$f(x) \leq c \quad \text{e} \quad f(y) \leq c. \tag{2-17}$$

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

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

Da desigualdade acima e de (2-17) obtemos que $f(x + (1-\alpha)y) \leq c$. Portanto $x + (1-\alpha)y \in L$, concluindo a demonstração. □

Proposição 2.1.21 *Sejam $C \subset I\!R$ um conjunto convexo e $f : C \to I\!R$ uma função convexa. Suponhamos que exista $c \in I\!R$ tal que o conjunto de sub-nível*

$$L_{f,I\!R^n}(c) = \{x \in C : f(x) \leq c\},$$

é não vazio e limitado. Então $L_{f,I\!R^n}(t)$ é limitado para todo $t \in I\!R$.

Prova. Seja $L_{f,I\!R^n}(c)$ um conjunto não vazio e limitado. Suponhamos, por absurdo, que exista $\bar{t}$, tal que $L_{f,I\!R^n}(\bar{t})$ seja ilimitado. Logo, pela Proposição 2.1.3, existe pelo menos uma direção de recessão d, diferente de zero, isto é, a semi-reta $\{x + td : t \in I\!R_+\}$ pertence a $L_{f,I\!R^n}(\bar{t})$.
Veja que $\bar{t} > c$, pois caso contrário, teríamos $L_{f,I\!R^n}(\bar{t}) \subset L_{f,I\!R^n}(c)$, o que seria absurdo pois $L_{f,I\!R^n}(c)$ é limitado e estamos supondo que $L_{f,I\!R^n}(\bar{t})$ é ilimitado. Conclusão $L_{f,I\!R^n}(c) \subset L_{f,I\!R^n}(\bar{t})$. Tome $\bar{x} \in L_{f,I\!R^n}(c)$ e $d \neq \{0\}$ uma direção de recessão para $L_{f,I\!R^n}(\bar{t})$. Fixe $t \in I\!R_+$ e tome $q > t$. Logo,

$$f(\bar{x} + td) = f((t/q)(\bar{x} + qd) + (1 - t/q)\bar{x}) \leq (t/q) f(\bar{x} + qd) + (1 - t/q) f(\bar{x}).$$

Donde utilizando as definições acima, chegamos a

$$f(\bar{x} + td) \leq (t/q)\bar{t} + (1 - t/q)c = c + (t/q)(\bar{t} - c).$$

Veja na última desigualdade que quando passarmos ao limite para $q \to \infty$, teremos $f(\bar{x}+td) \leq c$, para todo $t > 0$. Mas isso implicaria dizer que $\bar{x}+td \in L_{f,\mathbb{R}^n}(c)$, para todo $t > 0$. O que é absurdo, pois $d \neq \{0\}$ e $L_{f,\mathbb{R}^n}(c)$ é limitado. Logo $L_{f,\mathbb{R}^n}(t)$ é limitado para todo $t \in \mathbb{R}$, concluindo assim nossa prova. □

Esta proposição nos garante que havendo um conjunto de nível limitado, todos os demais também serão limitados. Este resultado pode ser encontrado no Teorema 3.4.4, página 140, em [17].

Corolário 2.1.22 *Seja $C \subset \mathbb{R}$ um conjunto convexo e fechado. Se $f : C \to \mathbb{R}$ é uma função estritamente convexa e possui um ponto de mínimo, então todo conjunto de sub-nível de f é compacto.*

Prova. Seja x^* um ponto de mínimo de f, que é único, pois f é estritamente convexa. Considere o conjunto de sub-nível

$$L_{f,\Omega}(f(x^*)) = \{x \in \Omega : f(x) \leq f(x^*)\} = \{x^*\}.$$

Este conjunto é claramente limitado e, pela Proposição 2.1.21 todo conjunto de sub-nível de f é limitado. Como Ω é fechado e f é contínua, segue-se que todo conjunto de sub-nível é fechado, logo todo conjunto de sub-nível de f é compacto, o que encerra nossa prova. □

2.1.6 Teorema da Função Implícita

Nesta seção, vamos enunciar uma versão do Teorema da Função Implícita que dará suporte a demonstrações futuras. Este resultado pode ser encontrado em [8].

Teorema 2.1.23 *(Teorema da Função Implícita) Seja $I \times \bar{I} \subset \mathbb{R} \times \mathbb{R}^p$ tal que $(w_0, z_0) \in I \times \bar{I}$, onde $I \subset \mathbb{R}$ e $\bar{I} \subset \mathbb{R}^p$ são conjuntos abertos. Suponhamos que $F : I \times \bar{I} \to \mathbb{R}^p$, dada por $F(w,z) = (f_1(w,z), \ldots, f_P(w,z))$, seja de classe C^1, que $F(w_0,z_0) = c \in \mathbb{R}^p$ e que a matriz*

$$\nabla_Z F(w_0,z_0) = \begin{bmatrix} \frac{\partial f_1}{\partial z_1}(w_0,z_0) & \ldots & \frac{\partial f_1}{\partial z_P}(w_0,z_0) \\ \vdots & & \vdots \\ \frac{\partial f_P}{\partial z_1}(w_0,z_0) & \ldots & \frac{\partial f_P}{\partial z_P}(w_0,z_0) \end{bmatrix} \in M_{P \times P}(\mathbb{R}),$$

seja inversível. Então existe um aberto $J \times U \subset I \times \bar{I}$ que contém (w_0,z_0), tal que $F(w,z) = c$ para $(w,z) \in J \times U$ se, e somente se, $z = \varphi(w)$, onde $\varphi : J \to U$ é a única função que satisfaz $F(w,\varphi(w)) = c, \forall w \in J$. Além disso, φ é de classe C^1.

CAPÍTULO 3

Condições de Otimalidade e o Centróide

3.1 Introdução

Neste capítulo, estabeleceremos as condições de otimalidade para o problema de programação linear (PPL). Com esta finalidade, demonstraremos alguns resultados importantes como os teoremas de Dualidade, Fraca e Forte. Para este último, utilizaremos alguns resultados apresentados no capítulo 2, tais como as definições de Projeção e Cone, o Teorema de Separação, bem como o Lema e o Corolário de Farkas. Finalizamos este capítulo com um resultado sob a hipótese de existência de ponto interior (não assumida até o momento) e a definição de centróide do conjunto solução dual.

Primeiro, vamos apresentar algumas notações importantes que utilizaremos no decorrer deste capítulo, tais como, as do (PPL) no formato padrão, tanto no formato Primal quanto Dual, bem como de seus conjuntos viáveis e de soluções. Para este capítulo utilizamos [3], [7], [17], [23], [34] e [6].

3.2 O Problema de Programação Linear

O *problema Primal* (PPL) no formato padrão é definido da seguinte forma:

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

onde os dados são a matriz A de ordem $m \times n$, os vetores c de ordem n e b de ordem m e a variável é o vetor x de ordem n.
Para facilitar os enunciados e demostrações de alguns resultados necessitaremos das seguintes notações:

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

o conjunto viável de (P), o seu interior relativo é dado por

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

e o conjunto de soluções ótimas de (P) por

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

Em todos os nossos resultados iremos utilizar, *às vezes sem menção explícita*, a seguinte hipótese:

$$\text{Posto}(A) = m < n.$$

O *problema Dual* associado ao problema Primal é o problema

$$\begin{array}{lrl} (D) & \textit{maximizar} & b^T y \\ & \textit{sujeito a} & A^T y + s = c \\ & & s \geq 0. \end{array}$$

Assim como no problema em formato Primal, no formato Dual utilizaremos as seguintes notações:

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

denotará o conjunto viável de (D), o seu interior relativo é dado por

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

e o conjunto de soluções ótimas de (D) por

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

3.3 Teoremas de Dualidade

O primeiro resultado que apresentaremos é o Teorema de Dualidade Fraca. Este teorema estabelece a relação entre as funções objetivo do (PPL) nos formatos Primal e Dual.

Teorema 3.3.1 *(Teorema de Dualidade Fraca) Suponha $x \in \mathcal{F}(P)$ e $(y,s) \in \mathcal{F}(D)$, então*

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

Em face disso, $c^T x \geq b^T y$.

Prova. Sejam $x \in \mathcal{F}(P)$ e $(y,s) \in \mathcal{F}(D)$. Das definições de $\mathcal{F}(P)$ e $\mathcal{F}(D)$, e após algumas manipulações algébricas simples, obtemos:

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

o que prova a primeira parte. Agora, como $x \geq 0$ e $s \geq 0$, então $x^T s \geq 0$. Concluímos, a partir da equação acima, que $c^T x \geq b^T y$, e isto prova o resultado desejado. $\square$

O Teorema de Dualidade Fraca, nos mostra que em todo PPL o valor da função objetivo primal é sempre maior ou igual que o valor da função objetivo dual. Futuramente, mostraremos que esses valores são de fato iguais, este resultado é conhecido como Teorema de Dualidade Forte. Para demonstrar este importante teorema, vamos utilizar alguns resultados preliminares, tais como o Lema e o Corolário de Farkas, apresentados no capítulo anterior. O Teorema de Dualidade Forte, devido a sua vasta utilização, possui vários formatos e várias formas de demonstração. Optamos pela abordagem exposta em [34], por apresentar uma demonstração com argumentos simples e de forma auto-contida. Esta demonstração consiste em aplicar várias vezes, de maneira conveniente, o Lema e o Corolário de Farkas.

Teorema 3.3.2 *(Teorema de Dualidade Forte) Considere os problemas (P) e (D). Apenas uma e somente uma das afirmações é correta:*

(a) *(P) e (D) são ambos inviáveis;*

(b) *(P) é inviável e (D) é viável e ilimitado;*

(c) *(D) é inviável e (P) é viável e ilimitado; ou*

(d) *(P) e (D) são ambos viáveis, $\mathcal{F}^*(P) \neq \emptyset$ e $\mathcal{F}^*(D) \neq \emptyset$. Além disso,*

$$\forall\, x^* \in \mathcal{F}^*(P), \quad \forall\, (y^*, s^*) \in \mathcal{F}^*(D) \quad \Rightarrow \quad c^T x^* = b^T y^*.$$

Prova. Primeiramente note que (P) e (D) são ambos viáveis ou pelo menos um dos problemas (P) ou (D) é inviável, note ainda que com a validade de uma das quatro possibilidades, as demais ficam excluídas.

Iniciamos considerando o caso em que (P) é inviável. Como a região viável de (P) é dada pelo sistema a seguir

$$Ax = b, \quad x \geq 0, \tag{3-1}$$

neste caso, nossa hipótese implica que o sistema anterior não tem solução. Então, pelo Lema de Farkas, existe $\tilde{y} \in R^m$ tal que:

$$A^T\tilde{y} \leq 0, \qquad b^T\tilde{y} > 0. \tag{3-2}$$

Todavia, nós necessitamos discutir a viabilidade de (D), temos duas possibilidades:

(i) (D) é inviável;

(ii) (D) é viável.

Ocorrendo **(i)**, teremos o caso **(a)**. Agora ocorrendo **(ii)**, existe $(\hat{y}, \hat{s}) \in \mathcal{F}(D)$. Note que a equação (3-2) pode ser reescrita na seguinte forma

$$A^T\tilde{y} + \tilde{s} = 0, \quad \tilde{s} \geq 0, \qquad b^T\tilde{y} > 0. \tag{3-3}$$

Portanto, para todo $\alpha > 0$, temos $(\hat{y} + \alpha\tilde{y}, \hat{s} + \alpha\tilde{s}) \in \mathcal{F}(D)$. De fato, usando a definição de $\mathcal{F}(D)$ e a igualdade em (3-3), temos

$$A^T(\hat{y} + \alpha\tilde{y}) + \hat{s} + \alpha\tilde{s} = A^T\hat{y} + \hat{s} + \alpha(A^T\tilde{y} + \tilde{s}) = c.$$

Logo, usando a última desigualdade em (3-3), concluímos que

$$\lim_{\alpha\to\infty} b^T(\hat{y} + \alpha\tilde{y}) = b^T\hat{y} + \lim_{\alpha\to\infty} \alpha b^T\tilde{y} = \infty.$$

Como $(\hat{y} + \alpha\tilde{y}, \hat{s} + \alpha\tilde{s}) \in \mathcal{F}(D)$ para todo $\alpha > 0$, a última igualdade implica que (D) é ilimitado e assim o caso **(b)** ocorre.

Agora, considere o caso que (D) é inviável. A região viável de (D) é dada pelo sistema de equações a seguir

$$A^Ty + s = c, \qquad s \geq 0.$$

Nossa hipótese implica então que este sistema não tem solução, ou equivalentemente, que o seguinte sistema

$$A^Ty \leq c,$$

não tem solução. Então pelo Corolário de Farkas, existe $\tilde{x} \in R^n$ tal que:

$$A\tilde{x} = 0, \;\; \tilde{x} \geq 0 \;\; e \;\; c^T\tilde{x} < 0. \tag{3-4}$$

Novamente necessitamos discutir a viabilidade de (P). Temos duas possibilidades:

(iii) (P) é inviável;

(iv) (P) é viável.

Ocorrendo **(iii)** obtemos, mais uma vez, o caso **(a)**. Ocorrendo **(iv)**, existe $\hat{x} \in \mathcal{F}(P)$. Isto posto, para todo $\alpha > 0$ temos que $\hat{x} + \alpha\tilde{x}$ também é viável para (P). De fato, usando a primeira igualdade em (3-4) e a definição de $\mathcal{F}(P)$, temos:

$$A(\hat{x} + \alpha\tilde{x}) = A\hat{x} + \alpha A\tilde{x} = b.$$

Logo, usando a última desigualdade em (3-4), concluímos que:

$$\lim_{\alpha\to\infty} c^T(\hat{x} + \alpha\tilde{x}) = c^T\hat{x} + \lim_{\alpha\to\infty} \alpha c^T\tilde{x} = -\infty.$$

Então (P) é ilimitado e temos o caso **(c)**.

Resta considerar o caso em que ambos, (P) e (D), são viáveis. Primeiramente, vamos construir um novo sistema com as restrições de (P) e (D), acrescidas de uma restrição:

$$Ax \leq b, \tag{3-5}$$
$$-Ax \leq -b, \tag{3-6}$$
$$-x \leq 0, \tag{3-7}$$
$$A^T y \leq c, \tag{3-8}$$
$$c^T x - b^T y \leq 0. \tag{3-9}$$

Note que se (x^*, y^*) é solução de (3-5)-(3-9), então x^* é solução de (P) e (y^*, s^*) é solução de (D), onde $s^* = c - A^T y^*$. De fato, (3-5)-(3-7) implica que x^* é viável para (P) e (3-8) implica que (y^*, s^*) é viável para (D). Agora, segue de (3-9) e do Teorema de Dualidade Fraca que, para todo $x \in \mathcal{F}(P)$ e $(y,s) \in \mathcal{F}(D)$ vale

$$b^T y \leq c^T x^* \leq b^T y^* \leq c^T x,$$

o que implica imediatamente que x^* é solução de (P) e (y^*, s^*) é solução de D, e mais ainda, que $c^T x^* = b^T y^*$. Portanto, basta mostrar que o sistema (3-5)-(3-9) tem solução viável. Observe que este sistema, pode ser reescrito na seguinte forma:

$$\bar{A}^T \bar{y} \leq \bar{c}, \tag{3-10}$$

onde a matriz $\bar{A}$, o vetor $\bar{c}$ e a variável $\bar{y}$ são definidas, respectivamente, por:

$$\bar{A} = \begin{pmatrix} A^T & -A^T & -I & 0 & c \\ 0 & 0 & 0 & A & -b \end{pmatrix}, \qquad \bar{y}^T = (x,y)^T, \quad \bar{c}^T = (b, -b, 0, c, 0)^T.$$

Se (3-10) é inviável, então pelo Corolário de Farkas, existe uma solução viável, $\bar{x} \in \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}$, para o sistema

$$\bar{A}\bar{x} = 0, \qquad \bar{x} \geq 0, \qquad \bar{c}^T\bar{x} < 0, \tag{3-11}$$

onde $\bar{x}^T = (u, v, w, z, \alpha)^T$. Este sistema pode ser reescrito no seguinte formato:

$$\begin{array}{cccccc} A^T u & -A^T v & -w & & +c\alpha & = 0 \\ & & & Az & -b\alpha & = 0 \\ b^T u & -b^T v & & +c^T z & & < 0 \\ u, & v, & w, & z, & \alpha & \geq 0. \end{array} \tag{3-12}$$

Desde que o sistema acima tenha solução viável, teremos uma das duas situações:

(v) $\alpha > 0$;

(vi) $\alpha = 0$.

Primeiramente, vamos considerar a possibilidade **(v)**. Neste caso, sem perda de generalidade, podemos assumir que $\alpha = 1$. Assim, a primeira igualdade e a última desigualdade em (3-12) implicam que

$$A^T(v-u) + w = c \quad e \quad w \geq 0. \tag{3-13}$$

Por outro lado, a segunda igualdade e a última desigualdade de (3-12), implicam que

$$Az = b \quad e \quad z \geq 0. \tag{3-14}$$

A equação (3-13) equivale a dizer que $(v-u, w)$ é viável para (D) e (3-14) equivale a dizer que z é viável para (P), assim pelo Teorema de Dualidade Fraca, temos $b^T(v-u) \leq c^T z$, o que contraria a estrita desigualdade em (3-12).
Agora, vamos considerar a possibilidade **(vi)**. Neste caso, o sistema (3-12) se reduz ao seguinte sistema:

$$\begin{array}{ccccc} A^T u & -A^T v & -w & & = 0 \\ & & & Az & = 0 \\ b^T u & -b^T v & & +c^T z & < 0 \\ u, & v, & w, & z & \geq 0. \end{array} \tag{3-15}$$

Finalmente, vamos considerar as seguintes possibilidades:

(vi-a) $c^T z < 0$;

(vi-b) $c^T z \geq 0$.

Vamos assumir que **(vi-a)** ocorre. Esta hipótese, juntamente com a segunda igualdade e a última desigualdade do sistema (3-15), implica que

$$Az = 0, \qquad z \geq 0, \qquad c^T z < 0.$$

Assim, pelos motivos acima expostos, o Corolário de Farkas garante que o sistema:

$$A^T(v-u) + w = c, \qquad w \geq 0,$$

não tem solução. Mas isto mostra que (D) é inviável, o que é absurdo, pois havíamos assumido que (D) era viável.
Agora, vamos assumir que **(vi-b)** ocorre, isto é, $c^T z \geq 0$. Esta desigualdade, juntamente com a desigualdade estrita, a primeira igualdade e a última desigualdade do sistema (3-15), implica que

$$A^T(v-u) + w = 0, \qquad w \geq 0, \qquad b^T(v-u) > 0.$$

Desde que $w \geq 0$, o sistema acima pode ser reescrito na seguinte forma equivalente:

$$A^T(v-u) \leq 0, \qquad b^T(v-u) > 0,$$

o que pelo Lema de Farkas, resulta em que (P) é inviável. Isto contraria a hipótese de viabilidade de (p) em (3-1).
Vemos assim que todas as duas situações conduzem a uma contradição, ou seja, o sistema (3-11) é inviável. Logo o sistema (3-10) é viável e portanto, (3-5)-(3-9) também, como queríamos demonstrar.
Note que se vale o teorema de dualidade fraco e vale a inequação (3-9), então

$$c^T x^* = b^T y^*,$$

para $\forall\, x^* \in \mathcal{F}^*(P), \quad \forall\, (y^*, s^*) \in \mathcal{F}^*(D)$. Isto conclui nossa prova. □

3.4 Condições de Otimalidade

Vimos que para cada Problema de Programação Linear Primal temos um outro problema associado, o chamado Problema Dual. Nesse par de problemas, existem várias propriedades e características envolvidas. Através dos teoremas de dualidade, comprovamos que uma vez encontrada a solução ótima de um deles, podemos, automaticamente, encontrar a solução ótima do outro e vice-versa. Computacionalmente falando, para problemas com grandes entradas, essa abordadogem pode se dar de forma mais rápida e objetiva. Através da análise das chamadas condições de otimalidade, que relacionam os conjuntos viáveis primal e dual, bem como o chamado *Gap* de dualidade, que seria a verificação de que $x^T s = 0$. Apresentamos agora, o teorema que estabelece tais condições.

Teorema 3.4.1 *(Condições de Otimalidade) Considere os problemas* (P) *e* (D)*. Um ponto* $x^* \in R^n$ *é uma solução ótima de* (P) *se e somente se, existe um par* $(y^*, s^*) \in R^m \times R^n$*, tal que* (x^*, y^*, s^*) *seja solução do sistema:*

$$\begin{cases} Ax = b \\ A^T y + s = c \\ x^T s = 0 \\ x, s \geq 0. \end{cases}$$

Prova. Seja $x^* \in \mathcal{F}^*(P)$. Nete caso, (P) é viável e limitado e assim pelo Teorema de Dualidade Forte (D) é viável e existe $(y^*, s^*) \in \mathcal{F}^*(D)$, tal que $c^T x^* = b^T y^*$ e pelo Teorema da Dualidade Fraca ${x^*}^T s^* = c^T x^* - b^T y^* = 0$. Logo (x^*, y^*, s^*) é solução do sistema.

A recíproca também é verdadeira, pois se $c^T x^* - b^T y^* = {x^*}^T s^* = 0$, então $c^T x^* = b^T y^*$ e, pelo Teorema de Dualidade Fraca, para qualquer $(y, s) \in \mathcal{F}(D)$, temos $c^T x^* \geq b^T y$, assim $(y^*, s^*) \in \mathcal{F}^*(D)$. Ainda temos que $c^T x^* = b^T y^*$ implica, também pelo Teorema de Dualidade Fraca, que $c^T x^* \leq c^T x$, para todo $x \in \mathcal{F}(P)$, logo $x^* \in \mathcal{F}^*(P)$. $\square$

3.5 Sob a Hipótese de Existência de Ponto Interior

Uma questão na qual devemos nos ater é a da hipótese da existência de um ponto interior no conjunto de pontos viáveis do problema dual, ou seja, na existência de $(y, s) \in \mathcal{F}^0(D)$. Com esta premissa, podemos expor uma proposição que servirá de base para futuras demonstrações.

Proposição 3.5.1 *Suponha* $\bar{x} \in \mathcal{F}(P)$*. Se existe* $(\bar{y}, \bar{s}) \in \mathcal{F}^0(D)$*, então o conjunto* $\mathfrak{L} = \{x \in \mathcal{F}(P);\ c^T x \leq c^T \bar{x}\}$ *é compacto.*

Prova. Para mostrar que o conjunto $\mathfrak{L}$ é compacto, devemos mostrar que ele é limitado e fechado. Para verificar que o conjunto $\mathfrak{L}$ é limitado, considere $x \in \mathfrak{L}$, pelo Teorema de Dualidade Fraca vale:

$$x^T \bar{s} = c^T x - b^T y$$

mas sabemos que $c^T x - b^T \bar{y} \leq c^T \bar{x} - b^T \bar{y} = \bar{x}^T \bar{s}$. Como $x^T \bar{s} \geq 0$, temos:

$$x_i \leq \frac{1}{\bar{s}_i} \bar{x}^T \bar{s} \Longrightarrow \|x\|_\infty \leq \left(max \left| \frac{1}{\bar{s}_i} \right| \right) \bar{x}^T \bar{s}.$$

Lembre-se que tomamos $x \in \mathfrak{L}$ arbitrário, logo $\mathfrak{L}$ é limitado.

Agora, verifiquemos que $\mathfrak{L}$ é fechado. Vamos imaginar uma seqüência $\{x^k\} \subset \mathfrak{L}$, tal que $lim_{k \longrightarrow \infty} x^k = \hat{x} \geq 0$. Logo, para qualquer que seja k, temos $c^T x^k \leq c^T \bar{x}$ e então

$$\lim_{k \longrightarrow \infty} c^T x^k = c^T \lim_{k \longrightarrow \infty} x^k = c^T \hat{x} \leq c^T \bar{x}.$$

O que implica dizer que $\hat{x} \in \mathfrak{L}$ ou ainda que $\mathfrak{L}$ é fechado. Como $\mathfrak{L}$ é fechado e limitado, $\mathfrak{L}$ é compacto. $\square$

Corolário 3.5.2 *Se $\mathcal{F}^0(D) \neq \emptyset \Rightarrow (P)$ possui ótimos.*

Prova. Da demonstração do Teorema de Dualidade Forte, o resultado segue. $\square$

3.6 Teorema da Complementaridade Estrita

Sejam $x^* \in \mathcal{F}^*(P)$ e $(y^*, s^*) \in \mathcal{F}^*(D)$, soluções ótimas para os problemas primal e dual respectivamente, dizemos que a igualdade $(x^*)^T s^* = 0$ é chamada condição de folga complementar. O Teorema das Folgas Complementares afirma que as soluções ótimas dos problemas primal e dual satisfazem a condição de folga complementar. A demosntração pode ser encontrada em [30].

Teorema 3.6.1 *Considere $x^* \in \mathcal{F}^*(P)$ e $(y^*, s^*) \in \mathcal{F}^*(D)$ soluções ótimas, respectivamente, para os problemas primal e dual. Então $x^* s^* = 0$.*

Prova. Considere as soluções ótimas $x^* \in \mathcal{F}^*(P)$ e $(y^*, s^*) \in \mathcal{F}^*(D)$, pelo Teorema de Dualidade Forte 3.3.2 temos que, $(x^*)^T s^* = 0$. Note que $x^* s^* = 0$ significa que $x_j^* s_j^* = 0$ para $j = 1, \cdots, n$. Que é equivalente à

$$(x^*)^T s^* = \sum_{j=1}^{n} x_j^* s_j^* = 0,$$

para $x \geq 0$ e $s \geq 0$. E o resultado se segue. $\square$

Considere novamente $x^* \in P^*$ e $(y^*, s^*) \in \mathcal{F}^*(D)$, soluções ótimas para os problemas primal e dual respectivamente, dizemos que a desigualdade $x^* + s^* > 0$ é chamada condição de folga complementar estrita. O Teorema de Complementaridade Estrita, afirma que qualquer PPL com solução ótima, possui uma solução ótima que satisfaz a condição de folga complementar estrita. Sua demosntração pode ser encontrada em [30].

Teorema 3.6.2 *Suponha que os problemas (P) e (D) admitam soluções ótimas. Então os problemas primal e dual têm um par de soluções que satisfaz $x^* + s^* > 0$.*

Prova. Queremos demonstrar que $x_j^* = 0$ para toda solução ótiuma primal se, e somente se, $s_j^* > 0$ para alguma solução ótima dual. Se $s_j^* > 0$ para alguma solução ótima dual, então pelo Teorema das Folgas Complementares 3.6.1, temos que $x_j^* = 0$ para toda solução ótima para o problema primal. Assim, suponha $x_j^* = 0$ para toda solução ótima primal, vamos demosntrar que $s_j^* > 0$ para alguma solução ótima para o dual. Para isso, considere v_P o valor ótimo primal, v_D o valor ótimo dual e $\kappa = v_P = v_D$. Seja o seguinte problema

$$\begin{aligned} (1)\ \ & minimizar - (u^j)^T x \\ & sujeito\ a : Ax = b \\ & \qquad -c^T x - t = -\kappa \\ & \qquad x \geq 0,\ \ t \geq 0, \end{aligned}$$

onde $u^j \in \mathbb{R}^n$ é um vetor de zeros com a j-ésima coordenada igual a 1. Note que $x_j^* = 0$ é equivalente ao problema acima admitir uma solução ótima, digamos $(\hat{x}, \hat{t})$, com valor ótimo $-(u^j)^T \hat{x} = 0$. O problema dual de (1) é o seguinte problema de otimização,

$$\begin{aligned} (2)\ \ & maximizar\ b^T y - \kappa\lambda \\ & sujeito\ a : A^T y - c\lambda + s = -u^j \\ & \qquad -t + s_{n+1} = 0 \\ & \qquad s \geq 0,\ \ s_{n+1} \geq 0. \end{aligned}$$

Pelo Teorema de Dualidade Forte 3.3.2 o problema (2) também admite solução ótima $(\hat{y}, \hat{\lambda}, \hat{s}, \hat{s}_{n+1})$ com valor ótimo

$$b^T \hat{y} - \kappa\hat{\lambda} = -(u^j)^T \hat{x} = 0. \tag{3-16}$$

Agora, defina $\hat{\lambda} = \hat{s}_{n+1} \geq 0$ e considere $(\bar{y}, \bar{s}) \in D^*$. Assim, segue-se que

$$(A^T \bar{y} + \bar{s}) + (A^T \hat{(y)} - c\hat{\lambda} + \hat{s}) = c - u^j$$

e com algumas manipulações algébricas na equação acima obtemos que,

$$\bar{s} + \hat{s} + u^j = (1+\hat{\lambda})c - A^T(\bar{y} - \hat{y}).$$

Dividindo ambos os lados da última expressão por $(1+\hat{\lambda}) > 0$ e tomando

$$y^* = \frac{\bar{y}+\hat{y}}{1+\hat{\lambda}} \qquad \text{e} \qquad s^* = \frac{\bar{s}+\hat{s}+u^j}{1+\hat{\lambda}}, \tag{3-17}$$

segue-se que

$$c - A^T y^* = s^* \qquad \text{e} \qquad s^* = \frac{\bar{s}+\hat{s}+u^j}{1+\hat{\lambda}} \geq \frac{1}{1+\hat{\lambda}} > 0. \tag{3-18}$$

Por outro lado, utilizando a definição de y^* em (3-17) e utilizando a igualdade (3-16) obtemos que

$$b^T y^* = b^T\left(\frac{\bar{y}+\hat{y}}{1+\hat{\lambda}}\right) = \left(b^T\bar{y} + \kappa\hat{\lambda}\right)\frac{1}{1+\hat{\lambda}}. \tag{3-19}$$

Como consideramos $(\bar{y},\bar{s}) \in \mathcal{F}^*(D)$ e κ sendo o valor ótimo dual, temos da última igualdade que

$$b^T y^* \geq (b^T y + b^T y\hat{\lambda})\frac{1}{1+\hat{\lambda}} = b^T y.$$

Da última desigualdade e de (3-18) obtemos que $(y^*,s^*) \in \mathcal{F}^*(D)$, logo concluímos que se $x_j^* = 0$ para toda solução ótima para o primal então, $s_j^* > 0$ para alguma solução ótima para o dual. Isto finaliza a demonstração. □

3.7 Centróide do Conjunto Solução Dual

Nesta seção vamos definir o centróide do conjunto solução dual o qual será de suma importância nos Capítulos 4 e 5. Começamos com um lema, que trata sobre a concavidade de um conjunto específico que aparecerá na dedução do Centróide do Conjunto Solução Dual.

Se $C \subset \mathbb{R}^n$ é um conjunto convexo, dizemos que $f : C \longrightarrow \mathbb{R}^n$ é uma função côncava em C, quando a função $(-f)$ é convexa em C.

Lema 3.7.1 *A função $\sigma_J : \mathbb{R}^n \longrightarrow \mathbb{R}$, definida por:*

$$\sigma_J(s) = min\{s_j : j \notin J\}$$

é côncava para todo $J \subset \{1,2,\ldots,n\}$.

Prova. Para provar a concavidade de σ_J, primeiro note que:

$$\sigma_J(s) := -max\{-s_j : j \notin J\}.$$

Assim, basta mostrar que a seguinte função $\bar{\sigma}_J(s) : I\!R^n \to I\!R$ definida por:

$$\bar{\sigma}_J(s) := max\{-s_j : j \notin J\}.$$

é convexa. Como para cada $j \in \bar{J} = \{1,2,...,n\}/J$ a função $I\!R^n \ni s \mapsto \varsigma_j(s) = s_j$ é linear e portanto convexa e, além disso,

$$\text{epi}\bar{\sigma}_J(s) = \cap_{j\in\bar{J}}\, \text{epi}\varsigma_j(s).$$

Temos pela Proposição 2.1.14 e o fato de que a intersecção de conjuntos convexos também é convexo, que $\bar{\sigma}_J$ é convexa. Portanto, como $\sigma_J = -\bar{\sigma}_J$, o resultado segue. □

O *centróide* s^c do conjunto $\mathcal{F}^*(D)$ é determinado recursivamente como segue. Primeiramente, considere o problema

$$\max\{\sigma_B(s) : \exists (y,s) \in \mathcal{F}^*(D), \;\; para\, algum\, y \in I\!R^m\}, \tag{3-20}$$

onde $B = \{j : \bar{x}_j > 0, \forall \quad \bar{x} \in \mathcal{F}^*(P)\}$ e $N = \{1,2,...,n\}/B$. Sejam v_1^* o valor ótimo e S_1^* o conjunto solução do problema (3-20). Definamos:

$$J_1 := \{j \notin B : s_j = v_1^*, \forall s \in S_1^*\}.$$

Como $s_j > 0$, para todo $j \in N$ temos $v_1^* > 0$. Desde que σ_B é uma função côncava, (veja o Lema 3.7.1) temos que S_1^* é um conjunto convexo, compacto, não vazio e $J_1 \neq \emptyset$. Se S_1^* possui um único ponto, então definimos s^c sendo $\{s^c\} = S_1^*$, caso contrário, consideramos o problema

$$\max\{\sigma_{B_1}(s) : s \in S_1^*\},$$

onde $B_1 \equiv B \cup J_1$. Seja v_2^* seu valor ótimo e S_2^* o conjunto solução ótimo. Definamos:

$$J_2 := \{j \notin B_1 : s_j = v_2^*, \forall s \in S_2^*\}.$$

Como S_1^* possui mais de um ponto e $J_1 \neq \emptyset$, segue-se que $v_2^* > v_1^*$. Também temos que S_2^* é um conjunto convexo, compacto, não vazio e $J_2 \neq \emptyset$. Se S_2^* possui um único ponto, então definimos s^c com $\{s^c\} = S_2^*$; caso contrário, continuando esse processo, recursivamente, obteremos uma seqüência de conjuntos

$$\mathcal{F}^*(D) \equiv S_0^* \supset S_1^* \supset S_2^* \supset \ldots \supset S_r^* = \{s^c\},$$

uma seqüência de conjuntos de índices disjuntos $B \equiv J_0, J_1, J_2, \ldots, J_r$ e uma sequência de escalares $0 \equiv v_0^* < v_1^* < v_2^* < \ldots < v_r^*$, tal que para cada $l = 0, 1, 2, \ldots, r-1$, temos:

i) v_{l+1}^* e S_{l+1}^* são o valor ótimo e conjunto solução do problema

$$\max\{\sigma_{B_l}(s) : s \in S_l^*\},$$

onde $B_l := J_0 \cup J_1 \cup J_2 \cup \ldots \cup J_l$;

ii) $J_{l+1} := \{j \notin B_l : s_j = v_{l+1}^*, \forall s \in s_{l+1}^*\}$.

CAPÍTULO 4

Trajetória Central Asssociada à Entropia - Convergências Primal e Dual

4.1 Introdução

Neste capítulo, estudamos as trajetórias central primal e dual associadas, respectivamente, às funções entropia e exponencial. Mais detidamente estudaremos a trajetória central primal no contexto da pertubação do problema primal através da função entropia, que é estritamente convexa e de classe C^{∞}. Provaremos a boa definição da trajetória central primal e sua convergência ao centro analítico da face ótima primal. Também provaremos a boa definição da trajetória central dual e sua convergência ao centróide da face ótima dual. Neste capítulo usamos como referências básicas Cominetti e San Martín em [3], bem como Iusem, Svaiter e Cruz Neto em [6].

4.2 A função entropia e a trajetória primal-dual

4.2.1 A função entropia

A noção de entropia, já foi objeto de muitas controvérsias e distintas formulações, teve sua origem nos estudos de termodinâmica (segunda lei da termodinâmica), onde foi introduzida para caracterizar o grau de desordem de um sistema. Assim, o valor da entropia é tanto mais baixo quanto menor for a agitação térmica da matéria e quanto mais ordenada e complexa for a configuração por ela assumida. E viceversa, a entropia é tanto maior quanto maior for a agitação térmica da matéria e quanto mais elementares sejam os modos pelos quais se estruturam as moléculas, os átomos e os núcleos atômicos.

Shannon, também conhecido como "pai" da teoria da informação, fora um famoso matemático americano que estabelecera a aplicação elétrica da álgebra booleana. O conceito proposto por Shannon, poderia ser chamado de entropia na teoria da informação e refere-se à incerteza de uma distribuição de probabilidade, o leitor interessado poderá obter outras informações em [32]. A entropia na teoria da informação corresponde à incerteza probabilística associada a uma distribuição de probabilidade. Cada distribuição reflete um certo grau de incerteza e diferentes graus de incerteza estão associados a diferentes distribuições (embora diferentes distribuições possam refletir o mesmo grau de incerteza). De um modo geral, quanto mais "espalhada" a distribuição de probabilidade, maior incerteza ela irá refletir.

Por exemplo, se alguém lança um dado de seis faces, sem saber se ele é viciado ou não, a probabilidade mais razoável a ser atribuída a cada resultado possível é 1/6, ou seja, representar a incerteza usando a distribuição uniforme. Esta atitude segue o conhecido princípio da razão insuficiente de Laplace, onde atribuir chances iguais aos eventos possíveis é a maneira mais razoável de alguém refletir sua ignorância (e sua incerteza) quanto às chances de ocorrência de cada evento. Por outro lado, prevendo-se a informação de que o dado é viciado e que ele dá números maiores (menores) que a média (=3,5, no caso uniforme) mais freqüentemente, então a pessoa naturalmente irá assumir uma distribuição alternativa à uniforme para expressar sua incerteza.

Shannon (1948) em [32], derivou uma medida para quantificar o grau de incerteza de uma distribuição de probabilidade. Denominando S à medida de entropia de Shannon, sua expressão formal para distribuições discretas de probabilidade é dada por:

$$S(p) = -\sum_{i=1}^{n} p_i log p_i. \tag{4-1}$$

Note que S é uma função estritamente côncava. Isto é importante, pois garante que S tenha um único máximo global, mesmo quando sujeita a restrições lineares.

Em nosso trabalho iremos utilizar a oposta da função entropia, ou seja, vamos utilizar a equação (4-1) sem o sinal de menos, portanto uma função estritamente convexa que possui um único mínimo global, o que será imprescindível para várias demonstrações em nosso trabalho.

4.2.2 Os problemas (P) e (D) perturbados

Vamos utilizar a função entropia (segunda lei da termodinâmica), para perturbar o problema primal (P) e sua conjugada para perturbar o problema dual (D). Os problemas (P) e (D) perturbados com a função entropia e exponencial, respectivamente, são:

$$
\begin{aligned}
(P_\mu) \qquad & minimizar \;\; c^T x + \mu \sum_{i=1}^{n} x_i log x_i \\
& sujeito\ a \;\; Ax = b \\
& \qquad\qquad x \geq 0,
\end{aligned}
$$

$$
\begin{aligned}
(D_\mu) \qquad & maximizar \;\; b^T y - \mu \sum_{i=1}^{n} e^{-s_i/\mu - 1} \\
& sujeito\,a \;\; A^T y + s = c.
\end{aligned}
$$

Lembramos que por hipótese consideraremos que $\mathcal{F}^0(P) \neq \emptyset$ e também que $\mathcal{F}^0(D) \neq \emptyset$. Note que o problema D_μ, não possui a restrição de que $s \geq 0$. Ou seja, não existe a barreira para s e talvez resida aí a dificuldade em se resolver este problema.

4.2.3 Trajetória Central Primal-Dual Associada à Entropia

Nesta seção iremos provar alguns resultados, com a intenção de apresentar formalmente a Trajetória Central Primal-Dual. Trabalharemos com as versões perturbadas dos problemas primal e dual.

Seja φ a função definida na equação (2-14), apresentada no Capítulo 2. Vamos agora, entendê-la a $\mathbb{R}^n_+$ e por abuso de notação seguiremos chamando-a de φ. Seja $\varphi : \mathbb{R}^n_+ \to \mathbb{R}$ definida por:

$$
\varphi(x) = \sum_{i:x_i>0}^{n} x_i \log x_i, \text{ com a convenção de que } t \log t = 0, \; para\, t = 0, \; x > 0.
$$

Como já vimos anteriormente, φ é estritamente convexa em $\mathbb{R}^n_{++}$ e esse fato será importante no decorrer de algumas demonstrações. É fácil ver que φ é contínua.

Para simplificar as notações, definamos também a função $\varphi_\mu : \mathbb{R}^n_+ \to \mathbb{R}$, tal que

$$
\varphi_\mu(x) = c^T x + \mu\varphi(x). \tag{4-2}
$$

Note que φ_μ é estritamente convexa em $\mathbb{R}^n_{++}$ para $\mu > 0$, e possui um único ponto de mínimo em em $\mathbb{R}^n_+$.

Proposição 4.2.1 *Seja $\tilde{x} > 0$. Então o conjunto*

$$
\mathfrak{L} = \{x \in \mathbb{R}^n_+ : \;\; \varphi_\mu(x) \leq \varphi_\mu(\tilde{x})\},
$$

é não vazio e compacto.

Prova. A função φ é extritamente convexa e contínua. Note ainda que isto também vale para φ_μ. Temos que $\mathfrak{L}$ é claramente não vazio, pois $\tilde{x} \in \mathfrak{L}$. Como φ_μ possui um único ponto de mínimo em $\mathbb{R}^n_+$, segue do Corolário 2.1.22 o resultado desejado, ou seja, $\mathfrak{L}$ é compacto. $\square$

Teorema 4.2.2 *Para cada $\mu > 0$ o problema (P_μ) tem solução única, digamos, $x(\mu) > 0$.*

Prova. Dado $\tilde{x} \in \mathcal{F}^0(P)$. Pela Proposição 4.2.1, o conjunto

$$\mathfrak{L} = \{x \in \mathbb{R}^n_+ : \ \varphi_\mu(x) \leq \varphi_\mu(\tilde{x})\},$$

é não vazio e compacto. Logo, a convexidade estrita de φ_μ implica que ela tem um único minimizador no conjunto

$$\mathcal{F}(P) \cap \mathfrak{L} = \{x \in \mathbb{R}^n_+ : \ Ax = b,\ x \geq 0,\ \varphi_\mu(x) \leq \varphi_\mu(\tilde{x})\},$$

visto que este conjunto também é convexo e compacto. Denotamos $x(\mu)$ ao único minimizador de φ_μ em $\mathcal{F}(P) \cap \mathfrak{L}$. Vamos mostrar que $x(\mu) \in \mathcal{F}^0(P)$, isto é, $x(\mu) > 0$. Para isso, suponhamos por absurdo que

$$x(\mu) \in \partial\mathcal{F}(P) = \{x \in \mathbb{R}^n : Ax = b,\ x \geq 0,\ x_1 x_2 \ldots x_n = 0\}.$$

Agora, definimos o seguinte vetor:

$$z_\varepsilon = (1-\varepsilon)x(\mu) + \varepsilon\tilde{x},$$

onde $0 < \varepsilon < 1$. A igualdade acima é equivalente a

$$(z_\varepsilon - x(\mu)) = \frac{\varepsilon}{1-\varepsilon}(\tilde{x} - z_\varepsilon). \tag{4-3}$$

Por outro lado, como $z_\varepsilon \in \mathcal{F}(P) \cap \mathfrak{L}$, pela minimalidade de $x(\mu)$ temos que $\varphi_\mu(x(\mu)) \leq \varphi_\mu(z_\varepsilon)$, isto é,

$$0 \leq \varphi_\mu(z_\varepsilon) - \varphi_\mu(x(\mu)). \tag{4-4}$$

Pela desigualdade do gradiente, Proposição 2.1.16, obtemos da equação (4-4) que

$$0 \leq \varphi_\mu(z_\varepsilon) - \varphi_\mu(x(\mu)) \leq \langle \nabla\varphi_\mu(z_\varepsilon), z_\varepsilon - x(\mu)\rangle.$$

Donde, utilizando (4-3) e a equação acima, obteremos:

$$0 \leq \frac{\varepsilon}{1-\varepsilon}\langle \nabla\varphi_\mu(z_\varepsilon), \tilde{x} - z_\varepsilon\rangle \Rightarrow 0 \leq \langle \nabla\varphi_\mu(z_\varepsilon), \tilde{x} - z_\varepsilon\rangle.$$

Agora, utilizando a definição de φ_μ, a última inequação se reduz a

$$0 \leq \langle c + \mu\log(z_\varepsilon) + \mu e, \tilde{x} - z_\varepsilon\rangle.$$

Após algumas manipulações algébricas, chegamos a

$$0 \leq \langle \mu \log(z_\varepsilon), \tilde{x} \rangle - [\langle c, z_\varepsilon \rangle + \langle \mu \log(z_\varepsilon), z_\varepsilon \rangle] - \langle \mu e, z_\varepsilon \rangle + \langle c + \mu e, \tilde{x} \rangle.$$

Vemos que o termo que encontra-se entre colchetes, representa justamente o produto $\varphi_\mu(z_\varepsilon)$. Chegamos então a

$$0 \leq \langle \mu \log(z_\varepsilon), \tilde{x} \rangle - \varphi_\mu(z_\varepsilon) - \langle \mu e, z_\varepsilon \rangle + \langle c + \mu e, \tilde{x} \rangle.$$

Note que $\tilde{x} > 0$ e que $\lim_{\varepsilon \to 0} z_\varepsilon = x(\mu)$. Como estamos supondo que $x(\mu) \in \partial \mathcal{F}(P)$, existe pelo menos um $j \in \{1, 2, ..., n\}$, tal que $x_j(\mu) = 0$. Assim, usando o fato que φ_μ é contínua e que $\tilde{x} > 0$, $\mu > 0$, a equação acima nos leva a um absurdo, pois

$$\lim_{\varepsilon \to 0} \langle \log(z_\varepsilon), \tilde{x} \rangle = -\infty.$$

Logo $x(\mu) \in \mathcal{F}^0(P)$, ou seja, $x(\mu) > 0$. Agora, vamos mostrar que $x(\mu)$ minimiza φ_μ em $\mathcal{F}(P)$. Já que $x(\mu)$ minimiza φ_μ em $\mathcal{F}(P) \cap \mathfrak{L}$ resta mostrar que $x(\mu)$ minimiza φ_μ no seu complementar. Seja x pertencente ao complementar de $\mathcal{F}(P) \cap \mathfrak{L}$ em relação a $\mathcal{F}(P)$. Assim temos:

$$\varphi_\mu(x) > \varphi_\mu(\tilde{x}) \geq \varphi_\mu(x(\mu)),$$

logo $x(\mu)$ minimiza φ_μ em $\mathcal{F}(P)$. □

Também podemos mostrar que, para cada $\mu > 0$, o problema (D_μ) tem uma única solução, digamos, $(y(\mu), s(\mu))$. Vamos definir agora as Condições de Otimalidade dos Problemas Perturbados.

Proposição 4.2.3 *Para cada $\mu > 0$, o seguinte sistema:*

$$Ax = b, \qquad x > 0, \tag{4-5}$$

$$A^T y + s = c, \qquad s \in I\!R^n, \tag{4-6}$$

$$s + \mu \nabla \varphi(x) = 0, \qquad \mu > 0, \tag{4-7}$$

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

Prova. Pelo Teorema 4.2.2, a única solução de (P_μ) é $x(\mu) > 0$. Assim, a restrição $x \geq 0$ é inativa para o problema (P_μ), logo o Lagrangiano associado ao problema (P_μ) é dado por:

$$L_P(x, \lambda) := c^T x + \mu \varphi(x) - \langle \lambda, Ax - b \rangle, \qquad x > 0, \quad \mu > 0,$$

onde $\lambda \in I\!R^m$ é o vetor de multiplicadores de Lagrange. Por definição de $x(\mu)$, existe $\lambda(\mu)$, tal que o par $(x(\mu), \lambda(\mu))$ é o único mínimo de L_P, e ele anula seu gradiente, isto é

$$\nabla_x L_P(x(\mu), \lambda(\mu)) := c - A^T \lambda(\mu) + \mu \nabla \varphi(x(\mu)) = 0, \tag{4-8}$$

$$\nabla_\lambda L_P(x(\mu), \lambda(\mu)) := Ax(\mu) - b = 0. \tag{4-9}$$

Como o Posto$(A) = m$, existe um único $\lambda(\mu) \in I\!R^m$ satisfazendo a equação (4-8). Agora, defina o seguinte vetor $s(\mu) := -\mu \nabla \varphi(x(\mu))$ e note que esta igualdade é equivalente à

$$s(\mu) + \mu \nabla \varphi(x(\mu)) = 0. \tag{4-10}$$

Portanto, $s(\mu)$ e $x(\mu) > 0$ satisfazem a equação (4-7). Agora, fazendo $y(\mu) = \lambda(\mu)$ e como $s(\mu) = -\mu \nabla \varphi(x(\mu)))$, temos que a equação (4-8) pode ser reescrita da seguinte forma:

$$A^T y(\mu) + s(\mu) = c. \tag{4-11}$$

Segue das equações (4-9), (4-10) e (4-11) que a terna $(x(\mu), y(\mu), s(\mu))$ é a única solução do sistema (4-5)-(4-7). $\square$

As equações e inequações (4-5)-(4-7) são conhecidas como condições de Karush-Kuhn-Tucker (KKT) ou de otimalidade para o problema (P_μ). De modo análogo, o Lagrangiano para o problema (D_μ) é:

$$L_D(y, s, \lambda) := b^T y - \mu \sum_{i=1}^{n} e^{-s_i/\mu - 1} - \langle \lambda, A^T y + s - c \rangle, \quad \mu > 0,$$

onde λ é o multiplicador de Lagrange. Desta forma, podemos mostrar que as condições de otimalidade do problema (D_μ) são também dadas pelo sistema (4-5)-(4-7).

A trajetória central primal é o conjunto de pontos $\{x(\mu) : \mu > 0\}$, onde

$$x(\mu) = argmin \left\{ c^T x + \mu \sum_{i=1}^{n} x_i \log x_i : Ax = b, x > 0 \right\},$$

ou seja, para cada $\mu > 0$, o ponto $x(\mu)$ é a única solução do problema $P(\mu)$. De modo análogo, definimos a trajetória central dual como sendo o conjunto de pontos $\{(y(\mu), s(\mu)) : \mu > 0\}$, onde

$$s(\mu) = argmax \left\{ b^T y - \mu \sum_{i=1}^{n} e^{-s_i/\mu - 1} : A^T y + s = c \right\},$$

ou seja, para cada $\mu > 0$, o ponto $(y(\mu), s(\mu))$ é a única solução do problema (D_μ), para algum $y(\mu) \in I\!R^m$.

Portanto, definimos a *trajetória central primal-dual*, associada à entropia como sendo o conjunto de pontos

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

4.2.4 A Trajetória Central Associada à Entropia é uma curva diferenciável

Apresentamos agora, alguns resultados que mostram, de fato, que a Trajetória Central Primal-Dual associada à entropia é uma curva diferenciável.

Teorema 4.2.4 *As equações* (4-5)-(4-7) *definem uma curva diferenciável no conjunto* $I\!R^n_{++} \times I\!R^m \times I\!R^n_{++}$, *ou seja, a trajetória central é uma curva diferenciável.*

Prova. Primeiro, defina a seguinte função

$$\begin{aligned}\psi : I\!R^n_{++} \times I\!R^m \times I\!R^n \times I\!R &\rightarrow I\!R^m \times I\!R^n \times I\!R^n \\ (x,y,s,\mu) &\mapsto (Ax-b, A^T y + s - c, s + \mu\nabla\varphi(x)).\end{aligned}$$

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

$$\nabla_{(x,y,s)}\psi(x,y,s,\mu) = \begin{pmatrix} A & 0 & 0 \\ 0 & A^T & I \\ \mu\nabla^2\varphi(x) & 0 & I \end{pmatrix}.$$

Note que o sistema (4-5)-(4-7) é equivalente ao seguinte sistema

$$\psi(x,y,s,\mu) = (0_m,0_n,0_n), \qquad \mu > 0, \quad x > 0, \quad s \in I\!R^m. \tag{4-12}$$

Dado $\bar{\mu} > 0$, temos da Proposição 4.2.3 que o ponto $(x(\bar{\mu}),y(\bar{\mu}),s(\bar{\mu}),\bar{\mu})$ é solução do sistema acima. Então, como o Posto$(A) = m$ e $\mu\nabla^2\varphi(x)$ é não singular, temos pelo Lema 2.1.11 que a matriz jacobiana de ψ no ponto $(x(\bar{\mu}),y(\bar{\mu}),s(\bar{\mu}),\bar{\mu})$ é não-singular. Então, segue-se do Teorema da Função Implícita 2.1.23, que existe um aberto

$$U \times (\bar{\mu}-\varepsilon, \bar{\mu}+\varepsilon) \subset I\!R^n_{++} \times I\!R^m \times I\!R^n \times I\!R,$$

e uma única curva continuamente diferenciável $\eta : (\bar{\mu}-\varepsilon,\bar{\mu}+\varepsilon) \rightarrow U$, satisfazendo a seguinte igualdade $\eta(\bar{\mu}) = (x(\bar{\mu}),y(\bar{\mu}),s(\bar{\mu}))$ e além disso

$$\psi(\eta(\mu),\mu) = (0_m,0_n,0_n), \qquad \mu \in (\bar{\mu}-\varepsilon,\ \bar{\mu}+\varepsilon). \tag{4-13}$$

Pela Proposição 4.2.3, para cada $\mu > 0$, o sistema (4-5)-(4-7) tem uma única solução $(x(\mu),y(\mu),s(\mu)) \in I\!R^n_{++} \times I\!R^m \times I\!R^n$. Como $\eta(\mu) \in I\!R^n_{++} \times I\!R^m \times I\!R^n$ segue-se de (4-13), (4-12) que

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

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

Seja $\{(x(\mu), y(\mu), s(\mu)) : \mu > 0\}$ a trajetória associada à função entropia. A curva $\eta : (0, +\infty) \to I\!R$ definida por

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

tem como imagem a trajetória central primal-dual associada à entropiae como é usual também será chamada de trajetória central primal-dual associada à entropia.

4.3 Convergência da Trajetória Central Primal Associada à Entropia

Vamos demonstrar a convergência da Trajetória Central Primal associada à entropia. Como veremos, esta trajetória convergirá para um ponto denominado o centro analítico da face ótima primal.

Proposição 4.3.1 *Seja $\{x(\mu) : \mu > 0\}$ a trajetória central primal. As seguintes afirmações são verdadeiras:*

(i) *A função $0 < \mu \mapsto \varphi(x(\mu))$ é não-crescente;*

(ii) *O conjunto $\{x(\mu) : 0 < \mu < \bar{\mu}\}$ é limitado, para cada $\bar{\mu} > 0$;*

(iii) *Todos os pontos de acumulação da trajetória central primal são soluções do problema (P), quando $\mu \to 0$.*

Prova. Utilizando as equações definidas em (4-5)-(4-7), temos:

$$\mu \nabla \varphi(x(\mu)) = -c + A^T y(\mu), \qquad A(x(\mu)) = b, \qquad \mu > 0. \tag{4-14}$$

Sejam $0 < \mu_1 < \mu_2$. Como φ é convexa, pela desigualdade do gradiente, Proposição 2.1.16, temos:

$$\mu_1(\varphi(x(\mu_1)) - \varphi(x(\mu_2))) \le \langle \mu_1 \nabla \varphi(x(\mu_1)),\, x(\mu_1) - x(\mu_2) \rangle.$$

Usando a primeira igualdade de (4-14) e a equação acima, concluímos que

$$\mu_1(\varphi(x(\mu_1)) - \varphi(x(\mu_2))) \le \langle -c + A^T y(\mu_1),\, x(\mu_1) - x(\mu_2) \rangle.$$

Como a segunda igualdade de (4-14) implica que $(x(\mu_1) - x(\mu_2)) \in \mathbb{N}ullA$, então usando a Proposição 2.1.10, segue da última desigualdade que

$$\mu_1(\varphi(x(\mu_1)) - \varphi(x(\mu_2))) \leq \langle -c,\, x(\mu_1) - x(\mu_2)\rangle. \tag{4-15}$$

Podemos mostrar que a seguinte desigualdade também é válida

$$\mu_2(\varphi(x(\mu_2)) - \varphi(x(\mu_1))) \leq \langle -c\,,\;\; x(\mu_2) - x(\mu_1)\rangle. \tag{4-16}$$

Somando (4-15) e (4-16), obtemos

$$(\mu_1 - \mu_2)(\varphi(x(\mu_1)) - \varphi(x(\mu_2)) \leq 0.$$

E como $\mu_1 - \mu_2 < 0$, teremos que $\varphi(x(\mu_1)) \geq \varphi(x(\mu_2))$. Isto posto, concluímos que φ é não-crescente e a afirmação **(i)** está provada.
Agora, fixemos $\bar{\mu} > 0$. Usando o mesmo raciocíonio usado no item **(i)**, temos que

$$\mu(\varphi(x(\mu)) - \varphi(x(\bar{\mu})) \leq \langle -c\,,\, x(\mu) - x(\bar{\mu})\rangle, \tag{4-17}$$

para todo $0 < \mu < \bar{\mu}$. Do item **(i)**, sabemos que $0 \leq \varphi(x(\mu)) - \varphi(x(\bar{\mu}))$. Por outro lado, da equação (4-17), temos que $\langle c,\, x(\mu)\rangle \leq \langle c,\, x(\bar{\mu})\rangle$, para todo $0 < \mu < \bar{\mu}$. Com isso,

$$\{x(\mu) : 0 < \mu < \bar{\mu}\} \subset \{x \in \mathcal{F}(P) : c^T x \leq c^T x(\bar{\mu})\}.$$

Pela Proposicao 3.5.1, o conjunto de nível $\{x \in \mathcal{F}(P) : c^T x \leq c^T x(\bar{\mu})\}$ é compacto, logo a inclusão acima implica a afirmação do item **(ii)**.
Suponha agora que $\bar{x}$ seja um ponto de acumulação de $\{x(\mu) : \mu > 0\}$. Veja que $A\bar{x} = b \;\; e \;\; \bar{x} \geq 0$, isto é, $\bar{x} \in \mathcal{F}(P)$. Considere uma seqüência de números positivos $\{\mu_k\}$, tal que $\lim_{k\to\infty} \mu_k = 0$ e que $\lim_{k\to\infty} x(\mu_k) = \bar{x}$. Seja x^* uma solução de (P) e $x \in \mathcal{F}^0(P)$. Como $x^* \in \partial\mathcal{F}^0(P)$, $x \in \mathcal{F}^0(P)$ e $\mathcal{F}^0(P)$ é convexo, temos que

$$Z(\varepsilon) = (1-\varepsilon)x^* + \varepsilon x \in \mathcal{F}^0(P), \qquad \forall\, \varepsilon \in (0,1).$$

Da minimalidade de $x(\mu)$, segue-se que $c^T x(\mu_k) + \mu_k \varphi(x(\mu_k)) \leq c^T Z(\varepsilon) + \mu_k \varphi(Z(\varepsilon))$, ou ainda que

$$\mu_k(\varphi(x(\mu_k)) - \varphi(Z(\varepsilon))) \leq \langle c,\, Z(\varepsilon) - x(\mu_k)\rangle.$$

Como φ é contínua na fronteira $\partial\mathcal{F}^0(P)$, $\lim_{k\to\infty} x(\mu_k) = \bar{x}$ e $Z(\varepsilon) \to x^*$ quando $\varepsilon \to 0$, concluímos da última equação que

$$0 \leq \langle c,\, x^* - \bar{x}\rangle,$$

pois $\mu_k \to 0$. A última desigualdade equivale a $c^T\bar{x} \le c^T x^*$. Como x^* é uma solução do problema (P) e $\bar{x} \in \mathcal{F}(P)$, segue-se que $\bar{x}$ também é solução de (P), o que prova a afirmação **(iii)**. $\square$

Teorema 4.3.2 *Seja $x^c \in I\!R^n_{++}$, o centro analítico de $\mathcal{F}^*(P)$, isto é, o único ponto que satisfaz*

$$x^c = arg\ min\left\{\sum_{i=1}^{n} x_i \log x_i : x \in \mathcal{F}^*(P)\right\}. \tag{4-18}$$

Então

$$\lim_{\mu\to 0} x(\mu) = x^c.$$

Prova. Sejam $\bar{x}$ um ponto de acumulação da trajetória central primal, que existe pela Proposição 4.3.1, e uma seqüência de números positivos $\{\mu_k\}$, tal que $\lim_{k\to\infty}\mu_k = 0$ e $\lim_{k\to\infty} x(\mu_k) = \bar{x}$. De maneira análoga ao feito na prova da Proposição 4.3.1, podemos mostrar que

$$\mu_k(\varphi(x(\mu_k)) - \varphi(x)) \le \langle c,\, x - x(\mu_k)\rangle,$$

para todo $x \in \mathcal{F}(P)$. Tomando $x \in \mathcal{F}^*(P)$, temos $c^T x \le c^T x(\mu_k)$, para todo k. Assim, teremos da equação acima que, $\varphi(x(\mu_k)) \le \varphi(x)$, para todo k. Agora, como φ é contínua, podemos tomar o limite quando $k \to \infty$, e concluir que $\varphi(\bar{x}) \le \varphi(x)$, isto é, que

$$\sum_{i=1}^{n} \bar{x}_i \log \bar{x}_i \le \sum_{i=1}^{n} x_i \log x_i,$$

para todo $x \in \mathcal{F}^*(P)$. Desse modo, qualquer ponto de acumulação da trajetória central (4-18), ou seja, é solução do problema. Por outro lado, x^c é o único ponto satisfazendo (4-18), uma vez que a função objetivo deste problema é estritamente convexa e $\mathcal{F}^*(P)$ é um conjunto convexo e fechado. Portanto, a trajetória central converge a x^c. $\square$

4.4 Convergência da Trajetória Central Dual Associada à Entropia

Mostraremos agora, a convergência da Trajetória Central Dual. Como veremos, esta trajetória convergirá ao centróide do conjunto solução dual. Começamos com o seguinte lema.

Lema 4.4.1 *Sejam $(x,\ (y,s)) \in \mathcal{F}(P)\times\mathcal{F}(D)$ e $(x^0,\ (y^0,s^0)) \in \mathcal{F}(P)\times\mathcal{F}(D)$. Então vale a relação de ortogonalidade:*

$$\langle x - x^0,\ s - s^0\rangle = 0.$$

Prova. Pela viabilidade primal de x e x^0 temos que $Ax = b$ e $Ax^0 = b$. Logo, $A(x - x^0) = Ax - Ax^0 = 0$. De onde deduzimos que

$$x - x^0 \in \mathbb{N}ullA. \tag{4-19}$$

Do mesmo modo, como (y,s) e (y^0,s^0) são viáveis, temos que $A^T y + s = c$, $A^T y^0 + s^0 = c$ e $A^T(y - y^0) + (s - s^0) = c - c = 0$, o que implica que

$$s - s^0 \in \mathbb{I}mA^T. \tag{4-20}$$

Portanto, usando (4-19), (4-20) e a Proposição 2.1.10, obtemos o resultado desejado. □

Proposição 4.4.2 *As afirmações abaixo são verdadeiras:*

(i) *O conjunto $\{s(\mu) : 0 < \mu < \bar{\mu}\}$ é limitado, para cada $\bar{\mu} > 0$;*

(ii) *Todos os pontos de acumulação da trajetória central dual são soluções do problema (D), quando $\mu \to 0$.*

Prova. Para provar **(i)**, consideramos $x^0 \in \mathcal{F}^0(P)$ e $(y^0,s^0) \in \mathcal{F}^0(D)$. Então, de acordo com o lema anterior, temos que:

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

Como $x(\mu) > 0$ e $s^0 > 0$, temos que $\langle x(\mu),\ s^0 \rangle > 0$, logo

$$\langle x^0,\ s(\mu) \rangle \leq \langle x(\mu),\ s(\mu) \rangle + \langle x^0,\ s^0 \rangle,$$

substituindo o valor de $s(\mu) = -\mu \nabla \varphi(x(\mu)) = -\mu \log x(\mu) - \mu e$, nessa desigualdade, obtemos:

$$\langle x^0,\ s(\mu) \rangle \leq -\mu \langle x(\mu),\ \log x(\mu) + e \rangle + \langle x^0,\ s^0 \rangle.$$

Mas $-\mu \langle x(\mu),\ e \rangle < 0$, logo temos da equação acima que:

$$\langle x^0,\ s(\mu) \rangle \leq -\mu \langle x(\mu),\ \log x(\mu) \rangle + \langle x^0,\ s^0 \rangle.$$

Como $\varphi(x(.))$ é não-crescente no intervalo $(0,\bar{\mu})$, usando sua definição, podemos escrever a partir da última desigualdade que:

$$\langle x^0,\ s(\mu) \rangle \leq -\mu \varphi(x(\bar{\mu})) + \langle x^0,\ s^0 \rangle. \tag{4-21}$$

Considere x^c, o centro analítico da face ótima primal, e note que pelo Teorema 4.3.2, $\lim_{\mu\to 0} x(\mu) = x^c$. Considere também os seguintes conjuntos:

$$B := \{j : x_j^c > 0\}, \qquad N := \{j : x_j^c = 0\}.$$

Note que $s(\mu)$ pode ser dividido em duas partes:

$$s_B(\mu) = -\mu(\log x_B(\mu) + e_B), \qquad s_N(\mu) = -\mu(\log x_N(\mu) + e_N). \tag{4-22}$$

Da definição do conjunto B e da primeira igualdade em (4-22) , temos em vista do Teorema 4.3.2:

$$\lim_{\mu\to 0} s_B(\mu) = 0. \tag{4-23}$$

Veja ainda que, pela definição de N e a segunda equação em (4-22), existe $\tilde{\mu} < \bar{\mu}$ tal que:

$$s_N(\mu) = -\mu(\log x_N(\mu) + e_N) > 0, \qquad 0 < \mu < \tilde{\mu}.$$

Logo, concluímos que:

$$\min\left\{x_i^0 : \ i \in N\right\} ||s_N(\mu)|| \leq \langle x_N^0, \, s_N(\mu)\rangle,$$

pois, nesse caso, $||s_N(\mu)|| \leq \sum_{i\in B} s_i(\mu)$. Combinando (4-21), a definição dos conjuntos B e N e a desigualdade acima temos

$$||s_N(\mu)|| \leq [\langle x^0 \ , \ s^0\rangle - \mu\varphi(x(\bar{\mu})) - \langle x_B^0, \, s_B(\mu)\rangle] / \min\left\{x_i^0 : \ i \in N\right\},$$

para $0 < \mu < \tilde{\mu}$, pois $x^0 > 0$ e $s(\mu) < 0$. Obtemos de (4-23) e da última inequação que o conjunto $\{s(\mu) : 0 < \mu < \bar{\mu}\}$ é limitado e o item **(i)** está provado.

Agora, vamos demonstrar o item **(ii)**. Pelo Teorema 3.4.1 (vide demonstração), basta mostrarmos que todo ponto de acumulação da trajetória dual satisfaz às seguinte equações:

$$A^T y + s = c, \tag{4-24}$$
$$(x^*)^T s = 0, \tag{4-25}$$
$$s \geq 0, \tag{4-26}$$

onde x^* é qualquer solução do problema (P), isto é $x^* \in \mathcal{F}^*(P)$. Sejam $(\bar{y}, \bar{s})$ um ponto de acumulação da trajetória dual e $\{\mu_k\}$ uma sequência de números positivos tais que

$$\bar{s} = \lim_{k\to\infty} s(\mu_k), \qquad \lim_{k\to\infty} \mu_k = 0.$$

Note que $A^T y(\mu_k) + s(\mu_k) = c$ e passando o limite, temos que $A^T\bar{y} + \bar{s} = c$ e a equação (4-24) se verifica.

Veja que $\lim_{\mu\to 0} x(\mu) = x^c \in \mathcal{F}^*(P)$, de onde podemos escrever que:

$$(x^c)^T\bar{s} = \lim_{k\to\infty}(x(\mu_k))^T \lim_{k\to\infty} s(\mu_k) = \lim_{k\to\infty}(^T(\mu_k)s(\mu_k).$$

Observe que de (4-7) temos $s(\mu_k) = -\mu_k\nabla\varphi(x(\mu_k))$. Assim, substituindo esta igualdade na equação acima e usando a definição de φ obtemos

$$(x^c)^T\bar{s} = -\lim_{k\to\infty}\mu_k[\varphi(x(\mu_k)) + x(\mu_k)^T e].$$

Como $\lim_{k\to\infty}\mu_k = 0$, $\lim_{\mu\to 0} x(\mu) = x^c$ e φ é contínua, a última igualdade equivale

$$(x^c)^T\bar{s} = 0,$$

assim (4-25) está provada, para $x^* = x^c$.

Resta mostrar que (4-26) vale ou, mais especificamente, que $\bar{s} \geq 0$. Primeiro considere os seguintes conjuntos

$$N = \{i : x_i^c = 0\}, \qquad B = \{1,\ldots,n\}/N.$$

Desde que $s(\mu_k) = -\mu_k\nabla\varphi(x(\mu_k))$ e $\bar{s} = \lim_{k\to\infty} s(\mu_k)$, pela definição de φ, podemos escrever

$$\bar{s}_B = -\lim_{k\to\infty}\mu_k[(\log x_B(\mu_k) + e], \qquad \bar{s}_N = \lim_{k\to\infty}\mu_k[(-\log x_N(\mu_k) - e]. \tag{4-27}$$

Agora, como $\lim_{\mu\to 0} x_N(\mu_k) = x_N^c = 0$ e $\mu_k > 0$ existe k_0 tal que

$$\mu_k(-\log x_N(\mu_k) - e) > 0, \quad k > k_0.$$

Logo, da inequação anterior e da segunda igualdade em (4-27), temos que $\bar{s}_N \geq 0$. Por outro lado, $\lim_{\mu\to 0} x_B(\mu_k) = x_B^c > 0$, logo concluimos da primeira igualdade em (4-27) que $\bar{s}_B = 0$. Portanto (4-26) vale e o item **(ii)** está provado. $\square$

Proposição 4.4.3 *Seja $\{x(\mu) : \mu > 0\}$ a trajetória central primal. Se $\mu > 0$ e $\bar{x} \in \mathcal{F}(P)$, então $s(\mu) = -\mu\log x(\mu) - \mu e$ é a única solução de*

$$min\left\{\bar{x}^T s + \mu\sum_{i=1}^{n} exp^{-s_i/\mu - 1} : s \in c + I\!\!ImA^T\right\}. \tag{4-28}$$

Prova. Pelas condições de otimalidade de (4-28), devemos mostrar que:

$$\nabla_s\left(\bar{x}^T s+\mu\sum_{i=1}^{n} exp^{-s_i/\mu-1}\right)=\bar{x}-exp^{-s/\mu-e},$$

avaliado em $s(\mu)$, pertence ao complemento ortogonal de $\mathbb{I}mA^T$, isto é, ao conjunto $\mathbb{N}ullA$. Como $s(\mu)=-\mu\log x(\mu)-\mu e$, da equação acima temos:

$$\nabla_s\left(\bar{x}^T s+\mu\sum_{i=1}^{n} exp^{-s_i/\mu-1}\right)\Big|_{s=s(\mu)}=\bar{x}-x(\mu).$$

Portanto o resultado segue-se, pois $\bar{x}-x(\mu)\in\mathbb{N}ullA$. □

Seja o s^c centróide do conjunto $\mathcal{F}^*(D)$, definido na Seção 3.7 do Capítulo 3.

Teorema 4.4.4 *A trajetória dual $\{s(\mu):\mu>0\}$ converge ao centróide s^c do conjunto solução dual $\mathcal{F}^*(D)$, quando μ tende a zero.*

Prova. Pela Proposição 4.4.2, temos que para cada $\bar{\mu}>0$, o conjunto $\{s(\mu):\bar{\mu}>\mu>0\}$, é limitado. Portanto, é suficiente provar que s^c é o único ponto de acumulação do conjunto $\{s(\mu):0<\mu<\bar{\mu}\}$ quando μ tende a zero, para concluir que

$$\lim_{\mu\to 0} s(\mu)=s^c.$$

Seja $\bar{s}$ um ponto de acumulação do conjunto $\{s(\mu):0<\mu<\bar{\mu}\}$ quando μ vai para zero e $\{\mu_k\}$ uma seqüência tal que:

$$\lim_{k\to\infty}\mu_k=0,\qquad \lim_{\mu_k\to 0}s(\mu_k)=\bar{s},\qquad \lim_{\mu_k\to 0}y(\mu_k)=\bar{y}, \tag{4-29}$$

onde $A^Ty(\mu_k)+s(\mu_k)=c$ e $-A^T\bar{y}-\bar{s}=-c$. Para simplificar a notação, definamos

$$s^k=s(\mu_k),\qquad y^k=y(\mu_k),\qquad k=0,1,\ldots.$$

Definamos ainda

$$\tilde{s}^k=s^k+(s^c-\bar{s}),\qquad k=0,1,\ldots. \tag{4-30}$$

Como $s^c\in\mathcal{F}^*(D)$, existe y^* tal que $A^Ty^*+s^c=c$. Note que da definição acima, temos que

$$\lim_{k\to\infty}\tilde{s}^k=s^c. \tag{4-31}$$

Efetuando uma soma termo a termo, é imediato concluir, das definições acima que

$$A^T(-\bar{y}+y^*+y^k)+s^k+(s^c-\bar{s})=c.$$

Donde podemos deduzir que $\bar{s}^k \in c + I\!Im A^T$.
Pela minimalidade de s^k, para cada μ_k, podemos usar a Proposição 4.4.3, com $\bar{x}$ sendo uma solução do problema primal (P), para obtermos:

$$\bar{x}^T s^k + \mu_k \sum_{j=1}^{n} e^{-s_j^k/\mu_k - 1} \le \bar{x}^T \bar{s}^k + \mu_k \sum_{j=1}^{n} e^{-\bar{s}_j^k/\mu_k - 1}.$$

Note que pela Proposição 4.4.2 (ii) e pelo Lema 4.4.1, $\bar{x}^T\bar{s}^k = \bar{x}^T s^k + \bar{x}^T s^c - \bar{x}^T \bar{s} = \bar{x}^T s^k$, isto é,

$$\bar{x}^T\bar{s}^k = \bar{x}^T s^k.$$

Logo, usando a igualdade acima, a última desigualdade pode ser reescrita como

$$\sum_{j\in B} e^{-s_j^k/\mu_k-1} + \sum_{j\notin B} e^{-s_j^k/\mu_k-1} \le \sum_{j\in B} e^{-\bar{s}_j^k/\mu_k-1} + \sum_{j\notin B} e^{-\bar{s}_j^k/\mu_k-1}.$$

Para $j \in B$, da igualdade (4-30) concluímos que $\bar{s}_j^k = s_j^k$. Com efeito, pelo que vimos no fim da demonstração da parte (i) da Proposição 4.4.2, $\bar{s_B} = 0$. Por outro lado, como $x^c \ge 0$, $s^c \ge 0$ e $x_B^c = 0$, do Lema 4.4.1 segue que $s_B^c = 0$. Então podemos simplificar a última desigualdade para estabelecermos que:

$$\sum_{j\notin B} e^{-s_j^k/\mu_k} \le \sum_{j\notin B} e^{-\bar{s}_j^k/\mu_k}. \tag{4-32}$$

Agora, utilizando a definição da função σ_B (vide seção 3.7), o lado esquerdo de (4-32), pode ser limitado inferiormente como

$$e^{-\sigma_B(s^k)/\mu_k} \le \sum_{j\notin B} e^{-s_j^k/\mu_k},$$

já que todos os termos da soma são positivos e $e^{-\sigma_B(s^k)/\mu_k}$ é um deles. Por outro lado, utilizando novamente a definição de σ_B, de (4-32), obtemos

$$\sum_{j\notin B} e^{\bar{s_j}^k/\mu_k} \le (n-|B|)e^{-\sigma_B(\bar{s}^k)/\mu_k},$$

onde $|B|$ denota a cardinalidade do conjunto B. Combinando as duas últimas desigualdades com a desigualdade (4-32), segue que

$$e^{-\sigma_B(s^k)/\mu_k} \leq (n-|B|)e^{-\sigma_B(\bar{s}^k)/\mu_k}.$$

Aplicando o logaritmo à última desigualdade, e após algumas manipulações algébricas, chegamos a

$$\sigma_B(s^k) \geq \sigma_B(\bar{s}^k) - \mu_k \log(n-|B|).$$

Tomando o limite quando k tende para ∞ na última desigualdade, utilizando a continuidade de σ_B e as equações (4-29) e (4-31), obtemos

$$\sigma_B(\bar{s}) \geq \sigma_B(s^c) = v_1^*,$$

o que implica dizer que $\bar{s} \in S_1^*$ (vide seção 3.7). Portanto $\bar{s}_j = s_j^c$, para todo $j \in J_1$. Desta última igualdade e de (4-30), concluímos que:

$$s_j^k = \bar{s}_j^k, \qquad \forall\, j \in J_1.$$

Utilizando esta última igualdade, podemos eliminar os termos correspondentes em (4-32), para obter:

$$\sum_{j \notin B_1} e^{-s_j^k/\mu_k} \leq \sum_{j \notin B_1} e^{-\bar{s}_j^k/\mu_k},$$

onde $B_1 = B \cup J_1$. Usando argumento similar, iremos obter:

$$\sigma_{B_1}(s^k) \geq \sigma_{B_1}(\bar{s}^k) - \mu_k \log(n-|B_1|),$$

onde $|B_1|$ denota a cardinalidade do conjunto B_1. Portanto,

$$\sigma_{B_1}(\bar{s}) \geq \sigma_{B_1}(s^c) = v_2^*,$$

o que implica dizer que $\bar{s} \in S_2^*$. Continuando esse processo, concluímos que

$$\bar{s} \in S_r^* = \{s^c\},$$

para algum $0 < r < n$, terminando assim a nossa prova. □

Trajetória Central Associada à Divergência de Kullback-Leibler e o Método do Ponto Proximal

5.1 Introdução

Neste capítulo, estudamos as trajetórias central primal e dual associadas, respectivamente, à divergência de Kullback-Leibler (que é também uma distância de Bregman) e sua conjugada. Mais detidamente estudaremos a trajetória central primal no contexto da perturbação do problema primal através da divergência de Kullback-Leibler, que é estritamente convexa e de classe C^{∞}. Provaremos a boa definição da trajetória central primal e sua convergência ao centro analítico da face ótima primal associado a divergência de Kullback-Leibler. Também provaremos a boa definição da trajetória central dual e sua convergência ao centróide da face ótima dual. Verifica-se dessa forma uma grande similaridade ao exposto no capítulo anterior. Neste momento é de bom alvitre enfatizar que todos resultados, exceto os da última seção, têm provas bastante similares às do capítulo anterior, entretanto, vamos repeti-las aqui, para tornar os capítulos independentes. Finalizamos este capítulo apresentando uma aplicação dos resultados obtidos no estudo do método de ponto proximal. Aqui, novamente, usamos como referências básicas Cominetti e San Martín em [3], Iusem, Svaiter e Cruz Neto em [6], além de Iusem e Monteiro em [14].

5.2 Divergência de Kullback-Leibler e a trajetória primal-dual

5.2.1 A divergência de Kullback-Leibler

A divergência de Kullback-Leibler, ou entropia relativa, mede a diferença entre duas distribuições de probabilidade. Considerando $p(x)$ e $q(x)$ duas distribuições de probabilidade discretas e distintas, a entropia relativa entre ambas é dada por:

$$D_{KL}(p||q) = K(p||q) = \sum_x p(x)\log(\frac{p(x)}{q(x)}).$$

Desta medida, é importante destacar duas propriedades:

(i) ela é sempre positiva ou nula, i.e., $K(p||q) \geq 0$;

(ii) ela é nula, $K(p||q) = 0$, se e só se $p(x) = q(x)$, para todo x.

O teorema da codificação de fonte, também conhecido por Primeiro Teorema de Shannon, indica que o comprimento médio do código (L) utilizado para codificar, os símbolos produzidos pela fonte de entropia $H(A)$ é maior ou igual à entropia da fonte:

$$L \geq H(A).$$

Nos casos em que o comprimento médio coincide com a entropia da fonte, o código é chamado de ideal. A distribuição de probabilidades da fonte é dada por $p(A)$, e o código é estabelecido assumindo a distribuição de probabilidade $q(A)$. Nesta situação, em que o comprimento médio do código obtido é penalizado em função da divergência (afastamento) entre estas duas distribuições de probabilidade iremos obter um valor para

$$L = H(A) + K(p||q).$$

A entropia relativa $K(p||q)$ mede a ineficiência de se assumir que a distribuição de probabilidades é q quando na verdade é p.

5.2.2 Os problemas (P) e (D) perturbados com a divergência de Kullback-Leibler

Vamos utilizar a divergência de KullBack-Leibler para perturbar os problemas primal (P) e dual (D). Primeiro vamos introduzir algumas notações: Seja $\varphi : \mathbb{R}^n_{++} \to \mathbb{R}$ definida por

$$\varphi(x) = \sum_{i=1}^{n} x_i \log x_i, \text{ com a convenção de que } t\log t = 0, \; para\, t = 0, \; x > 0.$$

A *distância de Bregman* $\tilde{K_\varphi}$: $\mathbb{R}^n_+ \times \mathbb{R}^n_{++} \to \mathbb{R}$, associada a função φ, é definida por

$$\tilde{K_\varphi}(x,y) = \varphi(x) - \varphi(y) - \nabla\varphi(y)^T(x-y). \tag{5-1}$$

Assim, da igualdade acima podemos dizer que a função φ nos fornece a *distância de Bregman* ou divergência de KullBack-Leibler,

$$\tilde{K_\varphi}(x,y) = \sum_{i=1}^{n}\left(x_i \log \frac{x_i}{y_i} + y_i - x_i\right),$$

já citada anteriormente em (2-16) como exemplo. Por hipótese, iremos considerar que

$$\mathcal{F}^0(P) \neq \emptyset.$$

Fixe $\tilde{x} \in \mathcal{F}^0(P)$. Seja $\tilde{K_\varphi} : \mathbb{R}^n_+ \to \mathbb{R}^n$ é definida como:

$$\tilde{K_\varphi}(0) = 0, \qquad \tilde{K_\varphi}(x) := \tilde{K_\varphi}(x,\tilde{x}).$$

Note que a função $\tilde{K_\varphi}$ é estritamente convexa e esse fato será importante no decorrer de algumas demonstrações. É fácil ver que $\tilde{K_\varphi}$ é contínua com a convenção de que $t\log t = 0$, para $t = 0$. Para simplificar as notações, definamos também a função $K_\mu : \mathbb{R}^n_{++} \to \mathbb{R}$, tal que

$$K_\mu(x) := c^T x + \mu\tilde{K_\varphi}(x). \tag{5-2}$$

Note que K_μ também é estritamente convexa para $\mu > 0$, e possui um único ponto de mínimo.
O problema primal (P) perturbado com a divergência de Kullback-Leibler é:

$$\begin{aligned} (\tilde{P}_\mu) \qquad & minimizar \;\; c^T x + \mu\tilde{K_\varphi}(x), \\ & sujeito\ a \;\; Ax = b, \\ & \qquad\qquad x \geq 0. \end{aligned}$$

O problema dual (D) perturbado com a conjugada da divergência de Kullback-Leibler é:

$$\begin{aligned} (\tilde{D}_\mu) \qquad & maximizar \;\; b^T y - \mu\sum_{i=1}^{n}\tilde{x}_i e^{-s_i/\mu} \\ & sujeito\ a \;\; A^T y + s = c. \end{aligned}$$

Por hipótese, também iremos considerar

$$\mathcal{F}^0(D) \neq \emptyset.$$

Denote por $\mathcal{F}(\tilde{P}_\mu)$ o *conjunto viável primal* de $(\tilde{P}_\mu)$ e por $\mathcal{F}(\tilde{D}_\mu)$ o *conjunto viável dual* de $(\tilde{D}_\mu)$.

5.2.3 Trajetória Central Primal-Dual

Nesta seção, iremos provar alguns resultados com a intenção de apresentar formalmente a trajetória central primal-dual. Trabalharemos com as versões perturbadas dos problemas primal e dual pela divergência de Kullback-Leibler e sua conjugada, respectivamente. Começamos com a seguinte proposição.

Proposição 5.2.1 *Seja $\tilde{x} \in \mathcal{F}^0(P)$. Então o conjunto*

$$\mathfrak{L} = \{x \in I\!R^n_+ : \ K_\mu(x) \leq K_\mu(\tilde{x})\},$$

é não vazio e compacto.

Prova. A função $\tilde{K_\varphi}$ é extritamente convexa e contínua. Note ainda que estes resultados também valem para K_μ. Temos que $\mathfrak{L}$ é claramente não vazio, pois $\tilde{x} \in \mathfrak{L}$. Como K_μ possui um único ponto de mínimo em $I\!R^n_+$, segue do Corolário 2.1.22 o resultado desejado, ou seja, $\mathfrak{L}$ é compacto. □

Teorema 5.2.2 *Para cada $\mu > 0$ o problema $(\tilde{P}_\mu)$ tem solução única, $x(\mu) > 0$.*

Prova. Primeiro note que $\mathcal{F}(\tilde{P}_\mu) = \mathcal{F}(P)$. Dado $\tilde{x} \in \mathcal{F}^0(P)$. Pela Proposição 5.2.1, o conjunto

$$\mathfrak{L} = \{x \in I\!R^n_+ : \ K_\mu(x) \leq K_\mu(\tilde{x})\},$$

é não vazio e compacto. Logo a convexidade estrita de K_μ, implica que ela tem um único minimizador no conjunto

$$\mathcal{F}(P) \cap \mathfrak{L} = \{x \in I\!R^n_+ : \ Ax = b, \ K_\mu(x) \leq K_\mu(\tilde{x})\},$$

visto que este conjunto também é compacto. Denotamos $x(\mu)$ o único minimizador de K_μ em $\mathcal{F}(\tilde{P}_\mu) \cap \mathfrak{L}$. Vamos mostrar que $x(\mu) \in \mathcal{F}^0(P)$. Para isso, suponhamos por absurdo que

$$x(\mu) \in \partial\mathcal{F}(P) = \{x \in I\!R^n_+ : Ax = b, \ x_1.x_2...x_n = 0\}.$$

Agora, definimos o seguinte vetor:

$$z_\varepsilon = (1-\varepsilon)x(\mu) + \varepsilon\tilde{x},$$

onde $0 < \varepsilon < 1$. A igualdade acima pode ser escrita de forma equivalente

$$(z_\varepsilon - x(\mu)) = \frac{\varepsilon}{1-\varepsilon}(\tilde{x} - z_\varepsilon). \tag{5-3}$$

Por outro lado, pela minimalidade de $x(\mu)$ temos que $K_\mu(x(\mu)) \leq K_\mu(z_\varepsilon)$, isto é

$$0 \leq K_\mu(z_\varepsilon) - K_\mu(x(\mu)). \tag{5-4}$$

Pela desigualdade do gradiente, Proposição 2.1.16, temos da equação acima que

$$0 \leq K_\mu(z_\varepsilon) - K_\mu(x(\mu)) \leq \langle \nabla K_\mu(z_\varepsilon), z_\varepsilon - x(\mu)\rangle,$$

donde, utilizando (5-3) e a equação acima, obteremos

$$0 \leq \frac{\varepsilon}{1-\varepsilon}\langle \nabla K_\mu(z_\varepsilon), \tilde{x} - z_\varepsilon\rangle \Rightarrow 0 \leq \langle \nabla K_\mu(z_\varepsilon), \tilde{x} - z_\varepsilon\rangle.$$

Agora, utilizando a definição de K_μ, a última inequação se reduz a

$$0 \leq \langle c + \mu[\log(z_\varepsilon) - \log \tilde{x}], \tilde{x} - z_\varepsilon\rangle.$$

Após algumas manipulações algébricas chegamos a

$$0 \leq \langle \mu \log(z_\varepsilon), \tilde{x}\rangle - \langle \mu \log(z_\varepsilon), z_\varepsilon\rangle + \langle c - \mu \log \tilde{x}, \tilde{x} - z_\varepsilon\rangle.$$

Vemos que o segundo termo é justamente $-\mu\varphi(z_\varepsilon)$, logo chegamos a

$$0 \leq \langle \mu \log(z_\varepsilon), \tilde{x}\rangle - \mu\varphi(z_\varepsilon) + \langle c - \mu \log \tilde{x}, \tilde{x} - z_\varepsilon\rangle.$$

Note que $\tilde{x} > 0$ e que $\lim_{\varepsilon\to 0} z_\varepsilon = x(\mu)$. Como supomos que $x(\mu) \in \partial\mathcal{F}(P)$, temos que existe pelo menos um $j \in \{1,2,...,n\}$, tal que $x_j(\mu) = 0$. Assim, usando o fato de que φ_μ é contínua e que $\tilde{x} > 0$, $\mu > 0$ a inequação acima nos leva a um absurdo, pois

$$\lim_{\varepsilon\to 0}\langle \log(z_\varepsilon), \tilde{x}\rangle = -\infty.$$

Logo $x(\mu) \in \mathcal{F}^0(P)$, ou seja, $x(\mu) > 0$. Agora, vamos mostrar que $x(\mu)$ minimiza K_μ em $\mathcal{F}(P)$. Dado x pertencente ao complementar de $\mathcal{F}(P) \cap \mathfrak{L}$ em relação a $\mathcal{F}(P)$, temos que

$$K_\mu(x) > K_\mu(\tilde{x}) \geq K_\mu(x(\mu)),$$

pois $x(\mu)$ minimiza $\mathcal{F}(P) \cap \mathfrak{L}$. Logo as desigualdades acima implicam que $x(\mu)$ minimiza K_μ em $\mathcal{F}(P)$. $\square$

Também podemos mostrar que, para cada $\mu > 0$, o problema $(\tilde{D}_\mu)$ tem uma única solução, digamos, $(y(\mu), s(\mu))$. Vamos definir agora as Condições de Otimalidade dos problemas perturbados.

Proposição 5.2.3 *Para cada $\mu > 0$ o seguinte sistema*

$$\begin{aligned} Ax = b, &\qquad x > 0, &\quad (5\text{-}5)\\ A^T y + s = c, &\qquad s \in \mathbb{R}^m, &\quad (5\text{-}6)\\ s + \mu \nabla \tilde{K}_\varphi(x) = 0, &\qquad \mu > 0, &\quad (5\text{-}7) \end{aligned}$$

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

Prova. Pelo Teorema 5.2.2 , a solução de $(\tilde{P}_\mu)$ é $x(\mu) > 0$. Assim a restrição $x > 0$ é inativa, logo a função de Lagrange associada ao problema $(\tilde{P}_\mu)$ é dada por:

$$L_P(x,\lambda) := c^T x + \mu \tilde{K}_\varphi(x) - \langle \lambda, Ax - b \rangle, \qquad x > 0, \quad \mu > 0,$$

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

$$\nabla_x L_P(x(\mu), \lambda(\mu)) := c - A^T \lambda(\mu) + \mu \nabla \tilde{K}_\varphi(x(\mu)) = 0, \qquad (5\text{-}8)$$

$$\nabla_\lambda L_P(x(\mu), \lambda(\mu)) := Ax(\mu) - b = 0. \qquad (5\text{-}9)$$

Como Posto$(A) = m$, existe um único $\lambda(\mu) \in \mathbb{R}^m$ satisfazendo a equação (5-8). Agora defina o seguinte vetor: $s(\mu) := -\mu \nabla \tilde{K}_\varphi(x(\mu))$ e note que esta igualdade é equivalente a

$$s(\mu) + \mu \nabla \tilde{K}_\varphi(x(\mu)) = 0. \qquad (5\text{-}10)$$

Portanto $s(\mu)$ e $x(\mu) > 0$ satisfazem a equação (5-7). Agora, fazendo $y(\mu) = \lambda(\mu)$ e como $s(\mu) = -\mu \nabla \tilde{K}_\varphi(x(\mu))$, temos que a equação (5-8) pode ser reescrita da seguinte forma

$$A^T y(\mu) + s(\mu) = c. \qquad (5\text{-}11)$$

Logo, segue das equações (5-9), (5-10) e (5-11) que a terna $(x(\mu), y(\mu), s(\mu))$ é a única solução do sistema (5-5)-(5-7). $\square$

O sistema (5-10) e (5-11) são as condições de Karush-Kuhn-Tucker (KKT) ou de otimalidade para o problema $(\tilde{P}_\mu)$.

De modo análogo, o Lagrangiano para o problema $(\tilde{D}_\mu)$ é:

$$L_D(y,s,\lambda) := b^T y - \mu \sum_{i=1}^{n} x_i^1 e^{s_i/\mu} - \langle \lambda, A^T y + s - c \rangle, \quad \mu > 0,$$

onde λ é o multiplicador de Lagrange. Desta forma, podemos mostrar que as condições de otimalidade do problema $(\tilde{D}_\mu)$ são também dadas pelo sistema (5-5)-(5-7). A *trajetória central KL primal* é o conjunto de pontos $\{x(\mu) : \mu > 0\}$, onde

$$x(\mu) = argmin\{c^T x + \mu \tilde{K_\varphi}(x) : Ax = b, x > 0\},$$

ou seja, para cada $\mu > 0$, o ponto $x(\mu)$ é a única solução do problema $(\tilde{P}_\mu)$. De modo análogo, definimos a *trajetória central KL dual* como sendo o conjunto de pontos $\{(y(\mu), s(\mu)) : \mu > 0\}$, onde

$$s(\mu) = argmax\left\{b^T y - \mu \sum_{i=1}^{n} x_i^1 e^{s_i/\mu} : A^T y + s = c\right\},$$

ou seja, para cada $\mu > 0$, o ponto $(y(\mu), s(\mu))$ é a única solução do problema $(\tilde{D}_\mu)$, para algum $y(\mu) \in I\!R^m$.
Portanto, definimos a *trajetória central primal-dual* associada a divergência de KullBack-Leibler como sendo o conjunto de pontos

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

5.2.4 A Trajetória Central é uma curva diferenciável

Vamos agora, apresentar o resultado que mostra que a Trajetória Central KL Primal-Dual é uma curva diferenciável.

Teorema 5.2.4 *As equações* (5-5)-(5-7) *definem uma curva diferenciável no conjunto* $I\!R^n_{++} \times I\!R^m \times I\!R^n_{++}$*, ou seja, a trajetória central KL é uma curva diferenciável.*

Prova. Primeiro, defina a seguinte função

$$\begin{aligned} \psi : I\!R^n_{++} \times I\!R^m \times I\!R^n \times I\!R &\rightarrow I\!R^m \times I\!R^n \times I\!R^n \\ (x, y, s, \mu) &\mapsto (Ax - b, A^T y + s - c, s + \mu \nabla K_\mu(x)). \end{aligned}$$

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

$$\nabla_{(x,y,s)} \psi(x, y, s, \mu) = \begin{pmatrix} A & 0 & 0 \\ 0 & A^T & I \\ \mu \nabla^2 K_\mu(x) & 0 & I \end{pmatrix} = \begin{pmatrix} A & 0 & 0 \\ 0 & A^T & I \\ \mu \nabla^2 \varphi(x) & 0 & I \end{pmatrix}.$$

Note que o sistema (5-5)-(5-7) é equivalente ao seguinte sistema

$$\psi(x, y, s, \mu) = (0_m, 0_n, 0_n), \qquad \mu > 0, \quad x > 0, \quad s \in I\!R^m. \tag{5-12}$$

Dado $\bar{\mu} > 0$, temos de (5-5)-(5-7) que o ponto $(x(\bar{\mu}), y(\bar{\mu}), s(\bar{\mu}), \bar{\mu})$ é solução do sistema acima. Então, como o Posto$(A) = m$ e $\mu\nabla^2\varphi(x) = \mu\text{diag}(1/x_1, \ldots, 1/x_n)$ é não singular, temos pelo Lema 2.1.11 que a matriz jacobiana de ψ no ponto $(x(\bar{\mu}), y(\bar{\mu}),\ s(\bar{\mu}), \bar{\mu})$ é não-singular. Então segue-se do Teorema da Função Implícita (Teorema 2.1.23) que existe um aberto

$$U \times (\bar{\mu} - \varepsilon, \bar{\mu} + \varepsilon) \subset I\!R^n_{++} \times I\!R^m \times I\!R^n \times I\!R,$$

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

$$\psi(\eta(\mu), \mu) = (0_m, 0_n, 0_n), \qquad \mu \in (\bar{\mu} - \varepsilon,\ \bar{\mu} + \varepsilon). \tag{5-13}$$

Pela Proposição 5.2.3, para cada $\mu > 0$, o sistema (5-5)-(5-7) tem uma única solução $(x(\mu), y(\mu), s(\mu)) \in I\!R^n_{++} \times I\!R^m \times I\!R^n$. Assim, como $\eta(\mu) \in I\!R^n_{++} \times I\!R^m \times I\!R^n$ segue-se de (5-13), (5-12) e do sistema (5-5)-(5-7) que

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

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

Seja $\{(x(\mu), y(\mu), s(\mu)) : \mu > 0\}$ a trajetória primal-dual. A curva diferenciável $\eta : (0, +\infty) \to I\!R$ definida por

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

tem como imagem a trajetória central primal-dual e como é usual também será chamada de trajetória central primal-dual.

5.3 Convergência da Trajetória Central KL Primal

Vamos demonstrar a convergência da trajetória central KL primal. Como veremos, esta trajetória convergirá para um ponto denominado o centro analítico da face ótima primal.

Proposição 5.3.1 *Seja $\{x(\mu) : \mu > 0\}$ a trajetória central KL primal. As seguintes afirmações abaixo são verdadeiras:*

(i) *A função $0 < \mu \mapsto \tilde{K}_\varphi(x)$ é não crescente;*

(ii) *O conjunto $\{x(\mu) : 0 < \mu < \bar{\mu}\}$ é limitado, para cada $\bar{\mu} > 0$;*

(iii) *Todos os pontos de acumulação da trajetória central primal são soluções do problema $(\tilde{P}_\mu)$, quando $\mu \to 0$.*

Prova. Utilizando as equações definidas em (5-5)-(5-7), teremos:

$$\mu \nabla \tilde{K}_\varphi(x(\mu)) = -c + A^T y(\mu), \qquad A(x(\mu)) = b, \qquad \mu > 0. \tag{5-14}$$

Sejam $0 < \mu_1 < \mu_2$. Como $\tilde{K}_\varphi$ é convexa, segue da desigualdade do gradiente (Proposição 2.1.16) que

$$\mu_1(\tilde{K}_\varphi(x(\mu_1)) - \tilde{K}_\varphi(x(\mu_2)) \leq \langle \mu_1 \nabla \tilde{K}_\varphi(x(\mu_1)),\, x(\mu_1) - x(\mu_2) \rangle.$$

Usando a primeira igualdade de (5-14), com $\mu = \mu_1$ e a equação acima, concluímos que

$$\mu_1(\tilde{K}_\varphi(x(\mu_1)) - \tilde{K}_\varphi(x(\mu_2)) \leq \langle -c + A^T y(\mu_1),\, x(\mu_1) - x(\mu_2) \rangle.$$

Como a segunda igualdade de (5-14) implica que $(x(\mu_1) - x(\mu_2)) \in \mathbb{N}ullA$, então segue da última desigualdade que

$$\mu_1(\tilde{K}_\varphi(x(\mu_1)) - \tilde{K}_\varphi(x(\mu_2)) \leq \langle -c,\, x(\mu_1) - x(\mu_2) \rangle. \tag{5-15}$$

Da mesma forma, podemos mostrar que a seguinte desigualdade também é válida:

$$\mu_2(\tilde{K}_\varphi(x(\mu_2)) - \tilde{K}_\varphi(x(\mu_1)) \leq \langle -c,\, x(\mu_2) - x(\mu_1) \rangle. \tag{5-16}$$

Somando (5-15) e (5-16), obtemos

$$(\mu_1 - \mu_2)(\tilde{K}_\varphi(x(\mu_1)) - \tilde{K}_\varphi(x(\mu_2)) \leq 0.$$

E como $\mu_1 - \mu_2 < 0$, teremos que $\tilde{K}_\varphi(x(\mu_1)) \geq \tilde{K}_\varphi(x(\mu_2)$. Isto posto, concluímos que $\tilde{K}_\varphi$ é não-crescente e a afirmação **(i)** está provada.
Agora, fixemos $\bar{\mu} > 0$. Usando o mesmo raciocínio adotado para o item **(i)**, temos que

$$\mu(\tilde{K}_\varphi(x(\mu)) - \tilde{K}_\varphi(x(\bar{\mu})) \leq \langle -c,\, x(\mu) - x(\bar{\mu}) \rangle, \tag{5-17}$$

para todo $0 < \mu < \bar{\mu}$. Do item **(i)**, sabemos que $0 \leq \tilde{K}_\varphi(x(\mu)) - \tilde{K}_\varphi(x(\bar{\mu})$. Logo, da equação (5-17), temos que $\langle c,\, x(\mu) \rangle \leq \langle c,\, x(\bar{\mu}) \rangle$, para todo $0 < \mu < \bar{\mu}$. Com isso

$$\{x(\mu) : 0 < \mu < \bar{\mu}\} \subset \{x \in \mathcal{F}(\tilde{P}_\mu) : c^T x \leq c^T x(\bar{\mu})\}.$$

Pela Proposicao 3.5.1, o conjunto de nível $\{x \in \mathcal{F}(\tilde{P}_\mu) : c^T x \leq c^T x(\bar{\mu})\}$ é compacto, logo, a inclusão acima implica a afirmação do item **(ii)**.

Suponha agora que $\bar{x}$ seja um ponto de acumulação de $\{x(\mu) : \mu > 0\}$. Veja que $A\bar{x} = b$ e $\bar{x} \geq 0$, isto é, $\bar{x} \in \mathcal{F}(\tilde{P}_\mu)$. Considere uma sequência de números positivos $\{\mu_k\}$, tal que $\lim_{k\to\infty} \mu_k = 0$ e que $\lim_{k\to\infty} x(\mu_k) = \bar{x}$. Sejam x^* uma solução de $(\tilde{P}_\mu)$ e $x \in \mathcal{F}^0(\tilde{P}_\mu)$. Como $x^* \in \partial\mathcal{F}^0(\tilde{P}_\mu)$, $x \in \mathcal{F}^0(\tilde{P}_\mu)$ e $\mathcal{F}^0(\tilde{P}_\mu)$ é convexo, temos que:

$$Z(\varepsilon) = (1-\varepsilon)x^* + \varepsilon x \in \mathcal{F}^0(\tilde{P}_\mu), \qquad \forall\, \varepsilon > 0.$$

Da minimalidade de $x(\mu)$, segue-se que $c^T x(\mu_k) + \mu_k \tilde{K}_\varphi(x(\mu_k)) \leq c^T Z(\varepsilon) + \mu_k \tilde{K}_\varphi(Z(\varepsilon))$, ou ainda que

$$\mu_k(\tilde{K}_\varphi(x(\mu_k)) - \tilde{K}_\varphi(Z(\varepsilon)) \leq \langle c,\, Z(\varepsilon) - x(\mu_k)\rangle.$$

Como K_μ é contínua na fronteira $\partial\mathcal{F}^0(\tilde{P}_\mu)$ e $\lim_{k\to\infty} x(\mu_k) = \bar{x}$ e $Z(\varepsilon) \to x^*$ quando $\varepsilon \to 0$, concluímos da última equação que

$$0 \leq \langle c,\, x^* - \bar{x}\rangle.$$

A última desigualdade pode ser escrita como $c^T\bar{x} \leq c^T x^*$. Como x^* é uma solução do problema $(\tilde{P}_\mu)$ e $\bar{x} \in \mathcal{F}(\tilde{P}_\mu)$, segue-se que $\bar{x}$ também é solução de $(\tilde{P}_\mu)$, o que prova a afirmação **(iii)**. □

Teorema 5.3.2 *Seja $x^c \in I\!R^n_{++}$, o centro analítico de $\mathcal{F}^*(P)$, isto é, o único ponto satisfazendo*

$$x^c = arg\ min \left\{ \sum_{i=1}^{n} \left(x_i \log\frac{x_i}{\tilde{x}_i} + \tilde{x}_i - x_i \right) : x \in \mathcal{F}^*(P) \right\}. \tag{5-18}$$

Então

$$\lim_{\mu\to 0} x(\mu) = x^c.$$

Prova. Sejam $\bar{x}$ um ponto de acumulação da trajetória central KL primal e uma seqüência de números positivos $\{\mu_k\}$, tal que $\lim_{k\to+\infty} \mu_k = 0$ e $\lim_{k\to+\infty} x(\mu_k) = \bar{x}$. De maneira análoga ao realizado na prova da Proposição 5.3.1, podemos mostrar que

$$\mu_k(\tilde{K}_\varphi(x(\mu_k)) - \tilde{K}_\varphi(x)) \leq \langle c,\, x - \, x(\mu_k)\rangle,$$

para todo $x \in \mathcal{F}(\tilde{P}_\mu)$. Note que $\mathcal{F}^*(P) \subseteq \mathcal{F}(P) = \mathcal{F}(\tilde{P}_\mu)$. Tomando $x \in \mathcal{F}^*(P)$, temos $c^T x \leq c^T x(\mu_k)$, para todo k. Assim, teremos da inequação acima que $\tilde{K}_\varphi(x(\mu_k)) \leq \tilde{K}_\varphi(x)$, para todo k. Agora, como $\tilde{K}_\varphi$ é contínua, podemos tomar o limite quando $k \to \infty$, para concluir que $\tilde{K_\varphi}(\bar{x}) \leq \tilde{K_\varphi}(x)$, isto é, que

$$\sum_{i=1}^{n} \bar{x}_i \log \frac{\bar{x}_i}{\tilde{x}_i} + \tilde{x}_i - \bar{x}_i \leq \sum_{i=1}^{n} x_i \log \frac{x_i}{\tilde{x}_i} + \tilde{x}_i - x_i,$$

para todo $x \in \mathcal{F}^*(P)$. Desse modo, qualquer ponto de acumulação da trajetória central satisfaz (5-18). Por outro lado, x^c é o único ponto satisfazendo (5-18), uma vez que a função objetivo deste problema é estritamente convexa e $\mathcal{F}^*(P)$ é um conjunto convexo e fechado. Portanto a trajetória central converge a x^c. □

5.4 Convergência da Trajetória Central KL Dual

Vamos agora, mostrar a convergência da Trajetória Central KL Dual. Como veremos, esta trajetória converge para um ponto denominado centróide.

Lema 5.4.1 *Sejam* $(x,\ (y,s)) \in \mathcal{F}(\tilde{P}_\mu) \times \mathcal{F}(\tilde{D}_\mu)$ *e* $(x^0,\ (y^0,s^0)) \in \mathcal{F}(\tilde{P}_\mu) \times \mathcal{F}(\tilde{D}_\mu)$. *Então vale a relação de ortogonalidade:*

$$\langle x - x^0,\ s - s^0 \rangle = 0.$$

Prova. Pela viabilidade primal de x e x^0 temos que $Ax = b$ e $Ax^0 = b$. Logo, $A(x - x^0) = Ax - Ax^0 = 0$. De onde deduzimos que

$$x - x^0 \in \mathbb{N}ull A. \tag{5-19}$$

Do mesmo modo, como (y,s) e (y^0,s^0) são viáveis, temos que $A^T y + s = c$, $A^T y^0 + s^0 = c$ e que $A^T(y - y^0) + (s - s^0) = c - c = 0$, o que implica que

$$s - s^0 \in \mathbb{I}m A^T. \tag{5-20}$$

Portanto, usando (5-19), (5-20) e a Proposição 2.1.10, obtemos o resultado desejado. □

Proposição 5.4.2 *As afirmações abaixo são verdadeiras:*

(i) *O conjunto* $\{s(\mu) : 0 < \mu < \bar{\mu}\}$ *é limitado, para cada* $\bar{\mu} > 0$*;*

(ii) *Todos os pontos de acumulação da trajetória central dual são soluções do problema* (D)*, quando* $\mu \to 0$.

Prova. Para provar **(i)**, iniciamos com $x^0 \in \mathcal{F}^0(P) = \mathcal{F}(\tilde{P}_\mu)$ e $(y^0, s^0) \in \mathcal{F}^0(D) \subseteq \mathcal{F}(\tilde{D}_\mu)$. Então, de acordo com o lema anterior, temos que:

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

Como $x(\mu) > 0$ e $s^0 > 0$, temos que $\langle x(\mu)\ ,\ s^0 \rangle > 0$, logo

$$\langle x^0\ ,\ s(\mu) \rangle \leq \langle x(\mu)\ ,\ s(\mu) \rangle + \langle x^0\ ,\ s^0 \rangle,$$

substituindo agora o valor de $s(\mu) = -\mu \nabla K_\mu(x(\mu))$, no lado direito dessa desigualdade, obteremos

$$\langle x^0,\ s(\mu) \rangle \leq -\mu \langle x(\mu),\ \log x(\mu) - \log \tilde{x} \rangle + \langle x^0,\ s^0 \rangle.$$

Usando a definição da função φ (2-14), obtemos:

$$\langle x^0,\ s(\mu) \rangle \leq -\mu \varphi(x(\mu)) + \mu \langle x(\mu),\ \log \tilde{x} \rangle + \langle x^0,\ s^0 \rangle.$$

Como a $\varphi(x(.))$ é não-crescente no intervalo $(0, \bar{\mu})$, podemos escrever a partir da última desigualdade que

$$\langle x^0,\ s(\mu) \rangle \leq -\mu \varphi(x(\bar{\mu})) + \langle x(\mu),\ \log \tilde{x} \rangle + \langle x^0,\ s^0 \rangle. \tag{5-21}$$

Considere $x^c > 0$, o centro analítico da face ótima primal e note que pelo Teorema 5.3.2 que $\lim_{\mu \to 0} x(\mu) = x^c$. Definimos também, os seguintes conjuntos:

$$B := \{j : x_j^c > 0\}, \qquad N := \{j : x_j^c = 0\}. \tag{5-22}$$

Note que $s(\mu)$ pode ser dividida em duas partes:

$$s_B(\mu) = -\mu(\log x_B(\mu) + \log \tilde{x}_B) \quad e \quad s_N(\mu) = -\mu(\log x_N(\mu) + \log \tilde{x}_N). \tag{5-23}$$

Da definição do conjunto B e da primeira igualdade na equação acima, temos

$$\lim_{\mu \to 0} s_B(\mu) = 0. \tag{5-24}$$

Veja ainda que, pela definição de N e equação (5-23), existe $\tilde{\mu} < \bar{\mu}$ tal que:

$$s_N(\mu) = -\mu(\log x_N(\mu) + \log \tilde{x}_N) > 0, \qquad 0 < \mu < \tilde{\mu}. \tag{5-25}$$

Como $s_N(\mu) > 0$, para $0 < \mu < \tilde{\mu}$, concluímos que

$$\min\left\{x_i^0 : i \in N\right\} ||s_N(\mu)|| \leq \langle x_N^0,\, s_N(\mu)\rangle.$$

Agora, note que $\langle x^0,\, s(\mu)\rangle = \langle x_N^0,\, s_N(\mu)\rangle + \langle x_B^0,\, s_B(\mu)\rangle$, assim combinando as equações (5-21), (5-25) e a última desigualdade, resulta

$$||s_N(\mu)|| \leq [-\mu\varphi(x(\bar{\mu})) + \mu\langle x(\mu),\, \log\tilde{x}\rangle + \langle x^0\,,\, s^0\rangle - \langle x_B^0,\, s_B(\mu)\rangle] / \min\left\{x_i^0 : i \in N\right\},$$

para $0 < \mu < \tilde{\mu}$. Como $\lim_{\mu\to 0}\mu\varphi(x(\bar{\mu})) = 0$, $\lim_{\mu\to 0}\langle x(\mu),\, \log\tilde{x}\rangle = 0$, e por (5-24)$\lim_{\mu\to 0}\langle x_B^0,\, s_B(\mu)\rangle = 0$, obtemos da última inequação que o conjunto $\{s(\mu) : 0 < \mu < \bar{\mu}\}$ é limitado e o item **(i)** está provado.

Vamos demonstrar o item **(ii)**. Pelo Teorema 3.4.1, basta mostrarmos que todo ponto de acumulação da trajetória KL dual satisfaz às seguintes equações:

$$A^T y + s = c \tag{5-26}$$

$$(x^*)^T s = 0 \tag{5-27}$$

$$s \geq 0, \tag{5-28}$$

onde x^* é uma solução do problema (P), isto é $x^* \in \mathcal{F}^*(P)$. Seja $(\bar{y},\bar{s})$ um ponto de acumulação da trajetória dual e $\{\mu_k\}$ uma seqüência de números positivos tais que:

$$(\bar{y},\bar{s}) = \lim_{k\to\infty}(y(\mu_k, s(\mu_k)), \qquad \lim_{k\to\infty}\mu_k = 0.$$

Observe que $A^T y(\mu_k) + s(\mu_k) = c$ e que todos os termos dessa equação são lineares. Logo, passando ao limite, teremos que $A^T\bar{y} + \bar{s} = c$ e a equação (5-26) está provada para $(\bar{y},\bar{s})$. Note que $\lim_{\mu\to 0} x(\mu) = x^c \in \mathcal{F}^*(P)$, de onde podemos escrever que

$$(x^c)^T\bar{s} = \lim_{k\to\infty}(x(\mu_k))^T \lim_{k\to\infty} s(\mu_k) = \lim_{k\to\infty} x^T(\mu_k)s(\mu_k).$$

Veja que de (5-7), temos $s = -\mu\nabla K_\varphi(x)$. Assim, substituindo esta igualdade na equação acima e usando a definição de φ obtemos

$$(x^c)^T\bar{s} = -\lim_{k\to\infty}\mu_k[\varphi(x(\mu_k)) - x(\mu_k)^T\log\tilde{x}].$$

Como $\lim_{k\to\infty}\mu_k = 0$, $\lim_{\mu\to 0} x(\mu) = x^c$ e φ é contínua, a última igualdade equivale

$$(x^c)^T\bar{s} = 0,$$

assim, (5-27) está provada, fazendo $x^* = x^c$.
Resta mostrar que (5-28), i.e., que $\bar{s} \geq 0$. Primeiro, considere os conjuntos B e N, definidos em (5-22). Desde que $s(\mu_k) = -\mu_k \nabla K_\varphi(x(\mu_k))$ e $\bar{s} = \lim_{k\to\infty} s(\mu_k)$, pela definição de K_φ e dos conjuntos B e N, podemos escrever

$$\bar{s}_B = -\lim_{k\to\infty} \mu_k[\log x_B(\mu_k) - \log \tilde{x}_B], \qquad \bar{s}_N = \lim_{k\to\infty} \mu_k[-\log x_N(\mu_k) + \log \tilde{x}_N]. \tag{5-29}$$

Agora, como $\lim_{\mu\to 0} x_N(\mu_k) = 0$ e $\mu_k > 0$, exite k_0, tal que:

$$\mu_k(-\log x_N(\mu_k) + \log \tilde{x}_N) > 0, \quad k > k_0.$$

Logo, da equação anterior e da segunda igualdade em (5-29), temos que $\bar{s}_N \geq 0$. Por outro lado, $\lim_{\mu\to 0} x_B(\mu_k) = x_B^c > 0$, então concluímos da primeira igualdade em (5-29) que $\bar{s}_B = 0$. Portanto, o item **(ii)** está provado. $\square$

Proposição 5.4.3 *Seja $\{x(\mu) : \mu > 0\}$ a trajetória central KL primal. Se $\mu > 0$ e $\bar{x} \in \{x \in \mathbb{R}^n : Ax = b, x \geq 0\}$, então $s(\mu) = -\mu\nabla K_\varphi(x(\mu)) = -\mu \log x(\mu) + \mu \log \tilde{x}$ é a única solução de*

$$min \left\{ \bar{x}^T s + \mu \sum_{i=1}^{n} \tilde{x}_i e^{-s_i/\mu} : s \in c + \mathbb{I}mA^T \right\}. \tag{5-30}$$

Prova. Pelas condições de otimalidade de (5-30), devemos mostrar que

$$\nabla_s \left(\bar{x}^T s + \mu \sum_{i=1}^{n} \tilde{x}_i e^{-s_i/\mu} \right) = \bar{x} - \tilde{x}e^{-s/\mu},$$

avaliado em $s(\mu)$, pertence ao complemento ortogonal de $\mathbb{I}mA^T$, isto é, ao conjunto $\mathbb{N}ullA$. Como $s(\mu) = -\mu \log x(\mu) + \mu \log \tilde{x}$, da equação acima temos

$$\nabla_s \left(\bar{x}^T s + \mu \sum_{i=1}^{n} \tilde{x}_i e^{-s_i/\mu} \right)_{|_{s=s(\mu)}} = \bar{x} - x(\mu).$$

Portanto o resultado segue, pois $\bar{x} - x(\mu) \in \mathbb{N}ullA$. $\square$

Seja s^c, o centróide do conjunto $\mathcal{F}^*(D)$, como definido na Seção 3.7 do Capítulo 3.

Proposição 5.4.4 *A trajetória dual $\{s(\mu) : \mu > 0\}$ converge para o centróide s^c do conjunto solução dual $\mathcal{F}^*(D)$, quando μ tende a zero.*

Prova. Pela Proposição 5.4.2, temos que para cada $\bar{\mu} > 0$, o conjunto $\{s(\mu) : 0 < \mu < \bar{\mu}\}$, é limitado. Portanto, é suficiente provar que s^c é o único ponto de acumulação do conjunto $\{s(\mu) : 0 < \mu < \bar{\mu}\}$, quando μ tende a zero, para concluir que

$$\lim_{\mu \to 0} s(\mu) = s^c.$$

Seja $\bar{s}$ um ponto de acumulação do conjunto $\{s(\mu) : \bar{\mu} > \mu > 0\}$, quando μ vai para zero e $\{\mu_k\}$ uma seqüência tal que

$$\lim_{k \to \infty} \mu_k = 0, \qquad \lim_{\mu_k \to 0} s(\mu_k) = \bar{s}, \qquad \lim_{\mu_k \to 0} y(\mu_k) = \bar{y}. \tag{5-31}$$

Note $A^T y(\mu_k) + s(\mu_k) = c$ e $A^T \bar{y} + \bar{s} = c$. Para simplificar a notação definamos:

$$s^k = s(\mu_k), \qquad y^k = y(\mu_k), \qquad k = 0, 1, \ldots.$$

Definamos ainda o seguinte vetor

$$\tilde{s}^k = s^k + (s^c - \bar{s}), \qquad k = 0, 1, \ldots. \tag{5-32}$$

Como $s^c \in \mathcal{F}^*(\tilde{D}_\mu)$, tome y^* tal que $A^T y^* + s^c = c$. Note que da definição acima temos

$$\lim_{k \to \infty} \tilde{s}^k = s^c.$$

É imediato concluir, efetuando uma soma termo a termo das definições acima, que

$$A^T(-\bar{y} + y^* + y^k) + s^k + (s^c - \bar{s}) = c.$$

Donde podemos deduzir que $\tilde{s}^k \in c + \mathbb{I}m A^T$.
Pela minimalidade de s^k, para cada μ_k, podemos usar a Proposição 5.4.3, com $\bar{x}$ sendo uma solução do problema primal (P), para obtermos:

$$\bar{x}^T s^k + \mu_k \sum_{j=1}^{n} \tilde{x}_j e^{-s_j^k/\mu_k} \leq \bar{x}^T \tilde{s}^k + \mu_k \sum_{j=1}^{n} \tilde{x}_j e^{-\tilde{s}_j^k/\mu_k}.$$

Note que, $\bar{x}^T \tilde{s}^k = \bar{x}^T s^k + \bar{x}^T s^c - \bar{x}^T \bar{s} = \bar{x}^T s^k$, isto é,

$$\bar{x}^T \tilde{s}^k = \bar{x}^T s^k.$$

Logo, a última desigualdade poderia ser reescrita como

$$\sum_{j\in B}\tilde{x}_j e^{-s_j^k/\mu_k}+\sum_{j\notin B}\tilde{x}_j e^{-s_j^k/\mu_k}\leq\sum_{j\in B}\tilde{x}_j e^{-\bar{s}_j^k/\mu_k}+\sum_{j\notin B}\tilde{x}_j e^{-\bar{s}_j^k/\mu_k}.$$

Para $j\in B$, da igualdade (5-32), concluímos que $\bar{s}_j^k = s_j^k$. Então, a desigualdade acima se reduz a

$$\sum_{j\notin B}\tilde{x}_j e^{-s_j^k/\mu_k}\leq\sum_{j\notin B}\tilde{x}_j e^{-\bar{s}_j^k/\mu_k}. \tag{5-33}$$

Segue imediatamente da última desigualdade que

$$\min\{\tilde{x}_j : j\notin B\}\sum_{j\notin B} e^{-s_j^k/\mu_k}\leq\max\{\tilde{x}_j : j\notin B\}\sum_{j\notin B} e^{-\bar{s}_j^k/\mu_k}. \tag{5-34}$$

Agora, utilizando a definição da função σ_B, o lado esquerdo de (5-34), pode ser limitado inferiormente como

$$\min\{\tilde{x}_j : j\notin B\}e^{-\sigma_B(s^k)/\mu_k}\leq\max\{\tilde{x}_j : j\notin B\}\sum_{j\notin B} e^{-\bar{s}_j^k/\mu_k}.$$

Por outro lado, utilizando novamente a definição de σ_B, de (5-34), obtemos

$$\min\{\tilde{x}_j : j\notin B\}e^{-\sigma_B(s^k)/\mu_k}\leq\max\{\tilde{x}_j : j\notin B\}(n-|B|)e^{-\sigma_B(\bar{s}^k)/\mu_k},$$

onde $|B|$ denota a cardinalidade do conjunto B. Aplicando o logaritmo à última desigualdade e após algumas manipulações algébricas chegamos a

$$\log\min\{\tilde{x}_j : j\notin B\}-\frac{1}{\mu_k}\sigma_B(s^k)\leq\log\left(\max\{\tilde{x}_j : j\notin B\}(n-|B|)\right)-\frac{1}{\mu_k}\sigma_B(\bar{s}^k).$$

Agora, multiplicando ambos os lados da última desigualdade por μ_k, obteremos

$$\mu_k\log\min\{\tilde{x}_j : j\notin B\}-\sigma_B(s^k)\leq\mu_k\log\left(\max\{\tilde{x}_j : j\notin B\}(n-|B|)\right)-\sigma_B(\bar{s}^k).$$

Tomando o limite quando k tende a ∞ na última desigualdade, utilizando a continuidade de σ_B e as equações (5-31) e (5-32), temos

$$\sigma_B(\bar{s})\geq\sigma_B(s^c)=v_1^*,$$

o que implica dizer que $\bar{s}\in S_1^*$. Portanto $\bar{s}_j = s_j^c$, para todo $j\in J_1$. Desta última igualdade e de (5-32), concluímos que

$$s_j^k=\bar{s}_j^k,\qquad\forall\ j\in J_1.$$

Utilizando esta última igualdade, podemos eliminar os termos correspondentes em (5-33), para obter

$$\sum_{j \notin B_1} e^{-s_j^k/\mu_k} \leq \sum_{j \notin B_1} e^{-\bar{s}_j^k/\mu_k},$$

onde $B_1 = B \cup J_1$. Usando argumento similar, iremos obter

$$\mu_k \log \min\{\tilde{x}_j : j \notin B_1\} - \sigma_{B_1}(s^k) \leq \mu_k \log\left(\max\{\tilde{x}_j : j \notin B_1\}(n - |B_1|)\right) - \sigma_{B_1}(\bar{s}^k),$$

onde $|B_1|$ denota a cardinalidade do conjunto $|B_1|$. Portanto,

$$\sigma_{B_1}(\bar{s}) \geq \sigma_{B_1}(s^c) = v_2^*,$$

o que implica dizer que $\bar{s} \in S_2^*$. Continuando esse processo, concluiremos que

$$\bar{s} \in S_r^* = \{s^c\},$$

para algum $0 < r < n$, isto é, $\bar{s} = s^c$, terminando assim a nossa prova. □

5.5 Método do Ponto Proximal Generalizado

Nesta seção, vamos apresentar uma aplicação dos resultados obtidos neste capítulo ao estudo do método de ponto proximal. As referências que utilizamos foram Iusem, Svaiter e Cruz Neto [6] e Iusem e Monteiro [14].

Seja $x_0 \in \mathcal{F}^0(P)$. O *método do ponto proximal generalizado* com a distância de Bregman ou divergência de Kullback-Leibler $\tilde{K}_\varphi$, gera uma seqüência $\{x_k\} \subset \mathcal{F}^0(P)$ com ponto inicial $x_0 \in \mathcal{F}^0(P)$ da seguinte forma:

$$x^{k+1} = \arg\min\left\{c^T x + \lambda_k \tilde{K}_\varphi(x, x^k) : Ax = b, x > 0\right\}, \tag{5-35}$$

onde a seqüência $\{\lambda_k\} \subset \mathbb{R}_{++}$, é tomada satisfazendo à seguinte condição

$$\sum_{k=0}^{\infty} {\lambda_k}^{-1} = +\infty. \tag{5-36}$$

Proposição 5.5.1 *A seqüência $\{x_k\}$ está bem definida.*

Prova. Primeiro note que $x_0 \in \mathcal{F}^0(P)$. Para cada k, suponha que $x^k \in \mathcal{F}^0(P)$. Agora tomando $\tilde{x} = x^k$ e $\mu = \mu_k$, segue imediatamente do Teorema 5.2.2, do problema (PP_μ) e da definição de $\tilde{K}_\varphi$, que existe um único ponto no conjunto $\mathcal{F}^0(P)$, digamos x^{k+1}, satisfazendo (5-35). Isto prova a boa definição da seqüência $\{x^k\}$. $\square$

De agora em diante referiremos à seqüência acima $\{x^k\}$, como a *seqüência de ponto proximal primal* com respeito a $\tilde{K}_\varphi$, associada a $\{\mu_k\}$ e ponto inicial x^0. De (5-35), segue-se que $\{x^k\}$ satisfaz

$$c + \lambda_k \left(\log(x^{k+1}) - \log(x^k) \right) = A^T z^k, \tag{5-37}$$

para alguma seqüência $\{z^k\}$ em $I\!R^m$ e $k = 0, 1, 2, \cdots$.

As condições de otimalidade para (5-35) determinam também $\{s^k\}$ a seqüência de ponto proximal dual com respeito a $\tilde{K}_\varphi$, como

$$s^k := -\lambda_k \nabla K_\varphi(x^{k+1}) = \mu_k(\log(x^k) - \log(x^{k+1})), \qquad k = 0, 1, 2, \cdots. \tag{5-38}$$

Da seqüência dual $\{s^k\}$ definimos a *seqüência proximal dual ponderada* $\{\bar{s}^k\}$ como

$$\bar{s}^k = \sum_{j=0}^{k} \lambda_j{}^{-1} \mu_k s_j, \qquad \mu_k = \left(\sum_{j=0}^{k} \lambda_j{}^{-1} \right)^{-1}, \qquad k = 0, 1, 2, \cdots. \tag{5-39}$$

Teorema 5.5.2 *Sejam $\{x(\mu) : \mu > 0\}$ e $\{s(\mu)) : \mu > 0\}$ as trajetórias central primal e dual associadas a $\tilde{K}_\varphi$, respectivamente. Suponhamos dada uma seqüência $\{\lambda_k\} \subset I\!R_{++}$, satisfazendo (5-36), e a seqüência $\{\mu_k\}$ definida como*

$$\mu_k = \left(\sum_{j=0}^{k} \lambda_j{}^{-1} \right)^{-1}, \quad para \quad k = 0, 1, 2 \cdots. \tag{5-40}$$

Então $x^{k+1} = x(\mu_k)$ e $\bar{s}^k = s(\mu_k)$ para $k = 0, 1, 2 \cdots$, onde $\{x^k\}$ e $\{\bar{s}^k\}$ são as seqüências de ponto proximal primal e dual ponderada associadas a $\{\mu_k\}$, respectivamente. Conseqüentemente,

$$\lim_{k \to +\infty} (x^k, \bar{s}^k) = (x^c, s^c),$$

onde x^c é o centro analítico da face ótima primal e s^c é o centróide do conjunto ótimo dual.

Prova. Sejam $\{x^k\}$ e $\{s^k\}$ as seqüências de ponto proximal primal e dual, respectivamente. De (5-35), (5-37) e (5-38), temos x^k e s^k satisfazendo

$$\begin{aligned}
&Ax^{k+1} = b, \quad x^{k+1} > 0, \\
&A^T z^k + s^k = c, \\
&s^k = \lambda_k(\log(x^k) - \log(x^{k+1})), \quad \lambda_k > 0
\end{aligned}$$

para alguma seqüência $\{z^k\}$ em $I\!R^m$ e $k = 0, 1, 2, \cdots$. Da última equação do sistema anterior, segue que $\sum_{j=0}^{k}(1/\lambda_j)s_j = \log(x^0) - \log(x^{k+1})$. Deduz-se da última expressão, juntamente com (5-39) e (5-40), que

$$\bar{s}^k = -\mu_k(\log(x^{k+1}) - \log(x^0)).$$

Assim, é fácil concluir que $\{x^k\}$ e $\{\bar{s}^k\}$ satisfazem:

$$\begin{aligned} &Ax^{k+1} = b,\, x^{k+1} > 0,\\ &A^T\bar{y}^k + \bar{s}^k = c,\\ &\bar{s}^k + \mu_k(\log(x^{k+1}) - \log(x^0)) = 0, \quad \mu_k > 0. \end{aligned} \tag{5-41}$$

para $\bar{y}^k = \mu_k \sum_{j=0}^{k}(1/\lambda_j)z_j$, $k = 0, 1, 2, \cdots$. Dessa forma, dos sistemas (5-5)-(5-7) e (5-41) obtemos:

$$x^{k+1} = x(\mu_k), \qquad \bar{y}^k = y(\mu_k), \qquad \bar{s}^k = s(\mu_k).$$

Como $\{\lambda_k\}$ satisfaz (5-36), é claro que $\lim_{k\to+\infty}\mu_k = 0$. Agora, use o fato que $\lim_{\mu\to 0}(x(\mu), s(\mu)) = (x^c, s^c)$ para concluir que $\lim_{k\to+\infty}(x^k, \bar{s}^k) = (x^c, s^c)$, e a prova está completa. $\square$

Conclusão

Em nossa dissertação, estudamos a convergência das trajetórias central primal e dual associadas às funções entropia e exponencial, respectivamente, para os problemas de programação linear. Os autores, Cominetti e San Martín em [3], investigaram o comportamento assintótico das trajetórias central primal e dual associadas às funções penalidade entropia e exponencial, respectivamente, em programação linear. Em particular, eles obtiveram uma caracterização dos seus pontos limites. Partimos da hipótese de que $\mathcal{F}^0(P) \neq \emptyset$ e $\mathcal{F}^0(D) \neq \emptyset$ e isto foi essencial para o desenvolvimento das provas. Estabelecemos as propriedades e a convergência da trajetória central primal associada a uma função satisfazendo algumas hipóteses específicas. Uma vez que esta função é contínua na fronteira, podemos provar que a trajetória central primal converge ao centro analítico do conjunto solução do problema (P). Em um contexto mais amplo, Iusem, Monteiro, et al. em [14], caracterizam o ponto limite da trajetória central dual associada a uma grande classe de funções de penalidade, entre elas a função penalidade exponencial, para problemas de programação convexa com restrições lineares, ou seja, para os (PPL). Vale ressaltar que se a trajetória central dual associada ao problema (P) é limitada e o sistema que determina a trajetória central primal-dual é de classe C^1, e podemos provar a convergência da trajetória central primal-dual.

Como aplicação do estudo das trajetórias central primal e dual, mostramos a convergência das seqüências proximal primal e dual ponderada associadas à distância Kullback-Leibler para os (PPL). Isto é exatamente o obtido por Iusem et al. [15] e Iusem e Monteiro [14], respectivamente, em programação linear. Embora tenhamos provado que a seqüência dual ponderada converge, vale a pena mencionar que a convergência completa da seqüência proximal dual é um problema em aberto. Esse pode ser um desafio a ser enfrentado. Além disso, apresentamos a convergência do método do ponto proximal generalizado associado a uma distância generalizada. O trabalho de pesquisa bibliográfica, bem como a familiarização com novas notações, talvez tenham sido as principais dificuldades enfrentadas para o desenvolvimento desta dissertação. O interesse por apresentar um texto auto-contido, de fácil leitura, até mesmo para alunos de graduação, fora a mola propulsora para a execução deste trabalho. Destacamos entre as principais contribuições, que acreditamos ter apresentado, as demontrações do Lema e do Corolário de Farkas, bem como dos Teoremas de Dualidade Fraca e Forte. Para futuras pesquisas, apontamos então a prova da convergência e caracterização do ponto limite da trajetória central para uma classe de funções mais gerais, a difícil tarefa de mostrar a convergência da seqüência proximal dual plena, e não apenas ponderada, bem como estudar os (PPL) para funções objetivo convexas e quaisquer.

Referências Bibliográficas

[1] Censor, Y. and Zenios, S., *The Proximal Minimization Algorithm with D-Functions*, Journal of Optimization Theory and Applications, Vol. 73, No. 3, pp. 451-464, 1992.

[2] Chen, G. and Teboulle, M., *Convergence Analysis of a Proximal-Like Minimization Algorithm using Bregman Functions*, SIAM Journal on Optimization Vol. 3, No. 3, pp. 538-543, 1993.

[3] Cominetti, R. and San Martín, J., *Asymptotic Analysis of the Exponential Penalty Trajectory in Linear Programming*, Mathematical Programming Vol. 67, No. 2, pp. 169-187, 1994.

[4] Cooper, W., Charnes, A. e Mellon W. B. *Blending Aviation Gasolines–A Study in Programming Interdependent Activities in an Integrated Oil Company Econometrica* , Vol. 20, pp. 135-159, No. 2, 1952.

[5] Cottle, R., Johnson, E. and Wets, R., *George B. Dantzig (1914-2005)*, Notices of the American Mathematical Society, Vol. 54, No. 3, pp. 344-362, march 2007.

[6] da Cruz Neto, J. X., Ferreira, O. P., Iusem, A. N., and Monteiro, R. D. C., Dual convergence of the proximal point method with Bregman distances for linear programming . Optimization Methods and Software, England, v. Online, pp. 1-23, 2006.

[7] Dantzig, G. B., *Linear Programming and Extensions*, Princeton University Press, Princeton, New Jersey, 1963.

[8] J. Dieudonné. *Foundations of Modern Analisys*. Academic Press, New York and London, 1960.

[9] Dorfman, R. and others, *Linear Programming and Economic Analysis* , Dover Books on mathematics an prodution to linear algebra, 1951.

[10] Dorfman, R. *The Conception of th Tableau Economique*, De Economist Journal, vol. 107, pp. 491-495, 1959.

[11] Eckstein, J., *Nonlinear Proximal Point Algorithm using Bregman Functions, with Applications to Convex Programming*, Mathematics of Operations Research, Vol. 18, No. 1, pp. 202-226, 1993.

[12] Fourier, J.B., *Solution d'une question particuliere du calcul des inegalites*, Nouveau Bulletin des Sciences par la Societe philomathique, 1826.

[13] Güler, O., *Limiting Behavior of Weighted Central Paths in Linear Programming*, Mathematical Programming, Vol. 65, pp. 347-363, 1994.

[14] Iusem, A. N., and Monteiro, R. D. C., *On Dual Convergence of the Generalized Proximal Point Method with Bregman Distances*, Mathematics of Operations Research Vol.25, No. 4, pp. 606-624, 2000.

[15] Iusem, A. N., Svaiter, B. F. and da Cruz Neto, J.X., *Central Paths, Generalized Proximal Point Methods and Cauchy Trajectories in Riemannian Manifolds*, SIAM Journal on Control and Optimization, Vol. 37, No. 2, pp. 566-588, 1999.

[16] Iusem, A. N., *On Some Properties of Generalized Proximal Point Methods for Variational Inequalities*, Journal of Optimization Theory and Applications, Vol. 96, No. 2, pp. 337-362, 1998.

[17] Izmailov, A. and Solodov, M., *Otimização - Volume 1: Condições de Otimalidade, Elementos de Análise Convexa e de Dualidade*. IMPA, 2005.

[18] Jensen, D. L. and Polyak, R. A., *The Convergence of a Modified Barrier Method for Convex Programming*, IBM Journal of Research and Development, Vol. 38, No.3, pp. 307-321, 1994.

[19] Karmarkar, N. *A new polynomial time algorithm for linear programming*. Combinatorica 4, pp. 373-395, 1984.

[20] Khachiyan, L.G., *A polynomial algorithm in linear programming*. Soviet Mathematics Doklady 20, pp. 191-194, 1979.

[21] Leontif, W., *Input an Output Economics*, Scientific American Journal, vol.35, 1936.

[22] Karush, W., *Minima of Functions of Several Variables with Inequalities as Side Constraints*, M.Sc. Dissertation. Dept. of Mathematics, Univ. of Chicago, Chicago, Illinois, 1939.

[23] Kuhn, H. W.; Tucker, A. W., *Nonlinear programming*,Proceedings of 2nd Berkeley Symposium: p. 481-492, Berkeley: University of California Press, 1951.

[24] Lima, E. L., *Análise Real: Funções de Uma Variável - Volume 1*. IMPA, 8ª edição, 2006.

[25] Lima, E. L., *Análise Real, Vol. 2*, Coleção Matemática Universitária, IMPA, 2004.

[26] Minoux, M. *Mathematical Programming- Theory and Algorithms*, Wiley Press, Chichester, 1986.

[27] Murty, K. G., *Operacions Research*, Princeton Hall, Englewood Cliffs, New Jersey, 1995.

[28] de Poussin, L. V.,*Sur la méthode de l'approximation minimum*, Ann. Soc. Sci. Bruxelles, pp. 1–16 (English translation by H. E. Salger, National Bureau of Standards, USA, 1911.

[29] Powell, M. J. D., *Some Convergence Properties of the Modified Log Barrier Method for Linear Programming*, SIAM Journal on Optimization, Vol.5, No. 4, pp. 695-739, 1995.

[30] R. Saigal. *Linear Programming: A Modern Integrated Analysis*. Kluwer Academic Publishers, 2ª edição, 1997.

[31] Samuelson, P.A., *Foundations of Economic Analysis* , Review of Economics and Statistics, Cambridge, Massachussets: pp. 350-356, 1955.

[32] Shannon, C. E., *A Mathematical Theory of Communication*, Bell System Technical Journal, vol. 27, pp. 379-42 and 623-656, 1948.

[33] Tseng, P. and Bertsekas, D. P., *On the Convergence of the Exponential Multiplier Method for Convex Programming*, Mathematical Programming, Vol. 60, No. 1, pp. 1-19, 1993.

[34] Todd, M. J.,*Mathematical Programming*, OR 630, Lecture 1 - 9, 2005.

[35] von Neuman, Jonh, *A Model of General Economic Equilibrium*, in K. Menger, editor, Ergebnisse eines mathematischen Kolloquiums, 1937.

[36] Walras, L., *Principe d'une Théorie Mathématique de l'Échange*, Journal des Economistes, Elements of Pure Economics, or the theory of social wealth, 1874.