Formação de professores: um exemplo de uso da otimização na preparação de novos professores

Marcos Vinícius Lopes*

Resumo

Com o intuito de reafirmar a importância do aprofundamento do rigor no estudo da matemática pura para a formação de novos professores de Matemática, propusemos o enfoque de uma área específica, a otimização, para assegurar a formação de novos professores de Matemática na UFG. Neste artigo estudamos a convergência das trajetórias central primal e dual associadas às funções entropia e exponencial, respectivamente, e investigamos o comportamento assintótico dessas trajetórias 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 e dual, mostramos a convergência das sequências proximal primal e dual ponderada, associadas à distância Kullback-Leibler para os Problemas de Programação Linear (PPL), no chamado Método de Ponto Proximal. Com uma linguagem mais acessível e palatável, focamos alunos de graduação no curso de Matemática, descrevendo uma das vertentes do estudo de otimização para a formação de novos professores.

Palavras-chave: problemas de programação linear, função entropia, divergência de Kullback-Leibler, método de ponto proximal, trajetória central.

The formation of teachers: an example of the use of otimization in the process

Abstract

In order to reaffirm the importance of an in-depth study of pure mathematics in the formation of new mathematics teachers, we have proposed to focus on a specific area, optimization, in order to provide training for new mathematics teachers at UFG. In this paper we study the convergence of primal and dual central paths associated to entropy and exponential functions, respectively, and we investigated the asymptotic behavior of these paths in linear programming. We also studied the convergence of primal and dual central paths associated to the Kullback-Leibler divergence and its conjugate. As an application of the study of central and dual paths, we have shown the convergence of the primal weighted dual proximal

* Professor Assistente de Matemática do Cepae/UFG. E-mail: marcoscoca@gmail.com

sequences to Kullback-Leibler distance for Problems of Linear Programming (PPL), in what is called the Proximal Point Method. Using more accessible and palatable language, we have undergraduate mathematics students in mind when describing one of the dimensions of the study of optimization for the formation of new mathematics teachers.
Keywords: linear programming problems, entropy function, Kullback-Leibler divergence, proximal point method, central path.

Introdução

Como professor do Centro de Ensino e Pesquisa Aplicada à Educação (Cepae) da Universidade Federal de Goiás há mais de quinze anos, venho acompanhando o processo de estágio e formação de novos professores, atuando como professor supervisor e, em algumas oportunidades, também como orientador no Instituto de Matemática e Estatística da UFG. O estágio em Matemática talvez seja hoje um dos mais elogiados e bem avaliados da UFG. Os alunos estagiários, futuros professores de Matemática, participam de todas as etapas do processo de ensino e aprendizagem do Cepae. Desde a elaboração do planejamento, antes mesmo do início das aulas, os estagiários têm contato com a nossa prática. É nesse ambiente de colaboração e coparticipação dos estagiários que ainda hoje verificamos um fato real na formação dos novos professores de Matemática.

Há que se destacar uma característica no curso de Matemática oferecido pela UFG e, por que não dizer, dos cursos de Matemática oferecidos pela maioria das universidades brasileiras: a questão da dicotomia entre as áreas de didática e metodologia de ensino (licenciatura), bem como da área de matemática pura (bacharelado). O fato é que muito se discute e uma pergunta ainda espera por resposta: "Qual a melhor maneira para se formar o professor de Matemática, o profissional de ensino em matemática?" O primeiro grupo citado não nega a importância do "saber matemática", da profunda preparação que é necessário oferecer aos novos professores, porém enfatiza a grande necessidade de assegurar uma vasta preparação na questão das didáticas e metodologias. O segundo grupo, em contrapartida, não nega a importância do "saber dar aula", porém acredita que um profissional conseguirá ser um bom professor de matemática se, acima de tudo, possuir um amplo, profundo e seguro conhecimento matemático dos conteúdos a serem ministrados. Considero que os dois grupos possuem suas razões e justificativas.

Quando optei pelo estudo na área de otimização, portanto, da matemática pura, para obtenção do título de mestre, não desmereci em nenhum momento as áreas de didática e metodologia no ensino de matemática nem delas desacreditei. Porém, compactuo com muitos de meus mestres a visão da importância do aprofundamento e do rigor matemático para a preparação do futuro professor de Matemática. Não obstante, o estudo de sistemas lineares, matrizes, determinantes requer um vasto e aprofundado rigor matemático na área de otimização. Pensando nesse enfoque, concluí minha dissertação, e utilizo a cada dia, na colaboração da formação de professores, as teorias da otimização.

O problema de otimizar uma função linear sujeita a restrições também lineares é chamado de Problema de Programação Linear (PPL). Neste trabalho apresentamos o PPL no formato padrão, isto é, minimizar uma função linear sujeita a restrições lineares de igualdade bem como a restrições de não negatividade. Vários autores se concentraram no estudo e resolução desses problemas nas últimas décadas, impulsionados principalmente pelo uso bélico e militar, pois uma gama de problemas de programação linear necessitava ser resolvida. Os esforços de guerra foram o principal combustível para o desenvolvimento dessa área de pesquisa. Após o término da Segunda Guerra Mundial, as áreas econômica e industrial, predominantemente, passaram a demandar a utilização de métodos para a 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 1940, portanto em pleno cenário de guerra, sem nenhuma dúvida foi o mais famoso.

Mais tarde provou-se que esse método não possuía tempo algorítmico satisfatório, ou seja, computacionalmente falando, na teoria o método demandava tempo exponencial (muito grande, inviável), apesar de funcionar muito bem na prática. Preocupados com esse fato, vários grupos de pesquisa matemática, pesquisa operacional, dentre outros, envidaram esforços no sentido de encontrar algoritmos que pudessem resolver tais problemas em tempo polinomial. Nesse 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 preveem uma trajetória pelo interior do conjunto viável. Muitos foram os métodos propostos e, dentre eles, podemos destacar os métodos de penalização. Nosso trabalho prevê uma breve apresentação histórica acerca da programação linear e da otimização.

O objetivo central do nosso trabalho foi o estudo de propriedades da trajetória central primal-dual associada à entropia, bem como à divergência de Kullback-Leibler. Uma outra parte importante foi fazer uma aplicação desta última trajetória ao estudo do método de ponto proximal. Este artigo é baseado em [1], nos artigos de Cominetti e San Martín [2], Iusem, Savaiter e Cruz Neto [3], Iusem e Monteiro [4]. Em minha dissertação e nesses artigos, são adotados métodos de penalidade para a perturbação dos problemas primal e dual. No artigo de Cominetti e San Martín, estes estudam o comportamento da trajetória central por meio da perturbação (penalização) dos problemas primal e dual, com a utilização das funções entropia e sua conjugada, a função exponencial.

Na primeira parte deste artigo, fazemos um modesto levantamento das bibliografias sobre o assunto e dos artigos estudados, bem como um relato histórico dos primeiros estudos acerca de PPLs. O leitor, já conhecedor dessa parte histórica, poderá avançar sem nenhum prejuízo diretamente ao item seguinte. Neste, passamos então ao delineamento de definições, preposições e teoremas, para chegarmos às demonstrações e aos resultados esperados. Dada a limitação de espaço e para não sobrecarregarmos a leitura, alertamos o leitor que as demonstrações das preposições e dos teoremas serão omitidas e poderão ser acompanhadas, de acordo com o seu interesse, em [1].

Histórico de programação linear e otimização

A 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 de alguém em descobrir o seu ponto de máximo ou de mínimo. A função pode representar muitas coisas, como 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, limitações que 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 de Problema de Programação Linear (PPL). Se a função objetivo, ou uma das restrições, for uma função não linear, esse problema é conhecido como Problema de Programação

Não Linear (PNL). Se o problema inclui restrições inteiras, ele é chamado de Problema de Programação Inteira Linear ou Não Linear. Esses são alguns dos 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 e [5].

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 do 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", que primeiramente discutiu esses métodos, não tinha nenhuma relação com a programação computacional. Com a realização de pesquisas em várias áreas do conhecimento humano e com a magnitude de dados que deveriam ser trabalhados, verificou-se a importância computacional que essa linha de pesquisa viria a desempenhar.

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

- Esforços de guerra, principalmente pelos Estados Unidos da América (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 iminente, 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 esse 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 Segunda 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 as várias pesquisas realizadas por esse grupo, o trabalho para aumentar o sucesso na localização e destruição de submarinos alemães mediante ataques aéreos. Em meados de 1941, essa taxa de sucesso era de 2 a 3% e, após a adoção das medidas sugeridas pelo grupo de Pesquisa Operacional, a probabilidade de ataque aéreo com sucesso subiu para 45%, resultado este que, dadas as limitações tecnológicas da época, era fantástico.

Na área econômica, as primeiras tentativas de modelagem matemática remontam ao "Tableau Economique" de Quesnay, segundo Dorfman em [6], que propunha um modelo linear para as relações entre o senhorio, camponeses e artesãos do século XVIII. Em 1874, na obra [7], L. Walras propõe um modelo linear com coeficientes tecnológicos fixos. Entretanto, os modelos lineares não seriam explorados até meados dos anos 1930, quando um grupo de economistas austro-alemães tentou generalizar o modelo proposto por Walras. Segundo Dantzig em [8], esse estudo teria influenciado o matemático americano Von Neumann, que em [9] de 1937, publicaria o trabalho "Modelo de Equilíbrio Econômico Geral". O modelo considerado por Von Neumann, que supunha taxa constante de expansão econômica e economia autossuficiente, não obteve grande repercussão. Uma das objeções levantadas pelos economistas de então era o fato de o modelo ser linear.

Posteriormente, outro modelo linear obteve sucesso. Esse novo, modelo baseado em relações de entrada e saída e desenvolvido por Leontif em [10], visava a 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 obtém um grande crescimento.

Em 1941, a Força Aérea Americana reúne o 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. Esse grupo tinha como consultor o matemático George Bernard Dantzig, que mais tarde seria o criador do método Simplex para Programação Linear. Esse método, cujo desenvolvimento foi iniciado em 1947 e finalizado em 1948, seria mantido 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, cunhou em 1948 o termo "Programação Linear".

O método Simplex de Dantzig dá 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).

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

- Publicações do método da Projeção por C. B. Tompkins;
- Análise da questão da estabilidade numérica do Simplex, por J. Hoffman;
- Publicação do método de relaxação para resolver Problemas de Programação Linear, por Motzkin;
- Kuhn e Tucker publicam o trabalho "Programação Não-Linear" em [11], 1951.

Na área de aplicação econômica temos:

- Publicação de artigo "Mecanismos de mercado e maximização", por Samuelson em [12] de 1955, alterando a concepção de Leontif;
- Análise da Teoria Econômica em programação linear, por Dorfman em 1951, em [13].

A proposição da programação linear deu origem também a novos modelos de programação: a já citada programação não linear (KUHN-TUCKER, 1951), a programação dinâmica (BELLMAN, 1950) e a programação inteira (GOMORY, 1958; DANTZIG; FULKERSON; JOHNSON, 1954). Esse conjunto de novas técnicas deu origem à disciplina Programação Matemática, termo que tem sua origem em um congresso patrocinado pela Rand Comporation em 1959, culminando com a Teoria da Complexidade em 1974, conforme Minoux em [14].

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. Grosso modo, o problema do Simplex resume-se a encontrar uma solução para um sistema de equações lineares com restrições de desigualdade.

Esse problema foi inicialmente abordado por Fourier em [15] em 1826. Naquela oportunidade, Fourier investigava alguns aspectos da teoria da probabilidade. Essa 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 sequencial de vértice a vértice, até o encontro do mínimo do conjunto poliedral. Esse princípio é a base do método Simplex. Esse mesmo problema foi abordado por outro matemático francês, La Vallé de Poussion em [16], em 1891, que propôs um método similar de resolução. Apesar disso, esse problema não despertou grande interesse dos demais matemáticos daquela época, sendo, portanto, pouco explorado.

W. Karush, em sua tese de mestrado em [17], em 1939, forneceu as condições de otimalidade para funções com restrições, contribuindo sobremaneira para a solução do problema de programação linear com o 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 tratar-se de uma "homework", tarefa para casa, e sem comentar com ninguém, tomou nota. Após uma semana, Dantzig compareceu ao gabinete do professor Neyman e entregou-lhe 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 essa e outras histórias acerca de Dantzig em [18].

Após três décadas de absoluto domínio, a hegemonia do Simplex foi posta em questão por dois novos métodos. Em 1979, L. G. Kanchian, em [19], 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 [20] 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, essa superioridade teórica foi comprovada na prática. Esse 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 mediante 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 eficiência máxima.

Cominetti e San Martín em [2] investigaram a convergência das trajetórias central primal e dual associadas às funções penalidades entropia e exponencial, respectivamente, para problemas de programação linear.

No artigo de Iusem, Svaiter e Cruz Neto [3], os autores se preocupam em estudar a convergência primal do método do ponto proximal com a distância de Bregman. Iusem e Monteiro em [4] 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.

Utilizamos os resultados obtidos, com respeito às trajetórias central primal e dual, através das perturbações 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. Nossa intenção foi simplificar as demonstrações, incentivando principalmente o leitor ainda não familiarizado com tais notações. Os resultados apresentados no próximo item são bastante similares aos apresentados no seguinte. O leitor já familiarizado com os resultados com a perturbação feita pela função entropia poderá dispensar, sem prejuízos, o próximo item, indo diretamente para o seguinte.

Trajetória central associada à entropia – convergências primal e dual

Nesta seção apresentamos definições, proposições, corolários e teoremas que estabelecem suporte a demonstrações e provas dos resultados por nós apresentados. Enfatizamos que as demonstrações são aqui omitidas, mas que poderão ser facilmente comprovadas pelo leitor em [1].

O problema de programação linear

O problema de programação linear (PPL) primal se escreve:

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

Onde c $\in IR^n$ e b $\in IR^m$, a matriz A é de ordem m x n e a variável primal é $x \in IR^n$.

O problema dual associado a (P) é:

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

onde $A^T \in IR^{mxn}$ denota a matriz transposta de A e $(y,s) \in IR^m \times IR^n$ são as variáveis duais.

Função entropia

Chamamos de S a medida de entropia de Shannon, "pai" da teoria da informação, sua expressão formal para distribuições discretas de probabilidade. Sua expressão é dada por:

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

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.

Utilizamos a oposta da função entropia, ou seja, utilizamos a equação acima sem o sinal de menos, portanto, uma função estritamente convexa que possui um único mínimo global, o que foi imprescindível para várias demonstrações em nosso trabalho.

Os problemas (P) e (D) perturbados

Utilizamos 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:

$$(P_\mu) \qquad minimizar\ c^T x + \mu \sum_{i=l}^{n} x_i \log x_i$$

$$sujeito\ a\ Ax = b$$

$$x \geq 0$$

$$(D_\mu) \qquad maximar\ b^T y - \mu \sum_{i=l}^{n} e^{-\frac{s_i}{\mu}-1}$$

$$sujeito\ a\ A^T y + s = c$$

Lembramos que por hipótese consideraremos que $F^0(P) \neq \emptyset$ e também que $F^0(D) \neq \emptyset$. 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 o problema.

Trajetória central primal-dual associada à entropia

Seja $\varphi : IR^n_+ \to IR$ definida por:

$$\varphi(x) = \sum_{i=l}^{n} x_i \log x_i$$

com a convenção de que $t \log t = 0$ para $t = 0$.

A função φ é estritamente convexa em IR^n_{++} e esse fato será importante no decorrer de algumas demonstrações. É fácil a verificação de que φ é contínua.

Para simplificar as notações, definamos também a função φμ : IRn+ → IR, tal que

$$\varphi_\mu(x) = c^T x + \mu\varphi(x)$$

Note que φμ é estritamente convexa em IRn++ para μ > 0 e possui um único ponto de mínimo em IRn+.

Preposição

Para cada μ > 0, o seguinte sistema:

$$Ax = b, \qquad x > 0$$

$$A^T y + s = c, \qquad s \in IR^n$$

$$s + \mu\nabla\varphi(x) = 0, \qquad \mu > 0$$

tem solução única, digamos (x(μ), y(μ), s(μ)). Verifique prova em [1], p. 56.

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

$$x(\mu) = argmin\left\{c^T + \mu\sum_{i=l}^{n} e^{-\frac{s_i}{\mu}-1} \colon Ax = b, x > 0\right\}$$

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

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

ou seja, para cada μ > 0, o ponto (y(μ), s(μ)) é a única solução do problema (Dμ) para algum y(μ) ∈ IR^m.

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

{(x(μ), y(μ), s(μ)): μ > 0}.

Convergência da trajetória central primal associada à entropia

Teorema

Seja $x^c \in IR^n_{++}$ o centro analítico de $F^*(P)$, isto é, o único ponto que satisfaz

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

Então

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

Ver prova em [1], p. 61.

Preposição

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

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

Ver prova em [1], p. 65.

Apresentaremos agora um teorema central para a demonstração dos resultados esperados. A convergência da trajetória para o centróide quando μ tende a zero ratifica o bom funcionamento desse método de pontos interiores.

Teorema

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

Ver prova em [1], p. 65.

Trajetória central associada à divergência de Kullback-Leibler e o método do ponto proximal

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 por:

$$D_{KL}(p//q) = (p//q) = \sum_{x} p(x) log\left(\frac{p(x)}{q(x)}\right)$$

Dessa medida, é importante destacar duas propriedades:
(i) ela é sempre positiva ou nula, isto é, $K(p\backslash\backslash q) \geq 0$;
(ii) ela é nula, $K(p\backslash\backslash q) = 0$, se e só se p(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 a 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 de fonte é dada por p(A), e o código é estabelecido assumindo a distribuição de probabilidade q(A). Nessa situação em que o comprimento médio do código obtido é penalizado em função da divergência (afastamento) entre essas duas distribuições de probabilidade iremos obter um valor para

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

A entropia relativa $K(p\backslash\backslash q)$ mede a ineficiência de se assumir que a distribuição de probabilidade é q, quando na verdade é p.

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 $\varphi : IR^n_{++} \rightarrow IR$ definida por $\varphi(x) = \sum_{i=l}^{n} x_i log x_i$, com a convenção de que t log t=0, para t=0, x > 0.

A distância de Bregman $K_\varphi : IR^n_+ \times IR^n_{++} \rightarrow IR$, associada à função φ, é definida por

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

Assim, da igualdade acima podemos dizer que a função φ nos oferece a distância de Bregman ou divergência de Kullback-Leibler,

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

Por hipótese, iremos considerar que

$$F^0(P) \neq \emptyset$$

Fixe x ϵ $F^0(P)$. seja $K_\varphi: IR^n_+ \rightarrow IR^n$ é definida como:

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

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

$$K_{\mu(x)} := c^T x + \mu K_\varphi(x)$$

Note que Kμ é também estritamente convexa para μ > 0, e possui um único ponto de mínimo.

O problema primal (P) perturbado com a divergência de Kullback--Leibler é:

$$(D_\varphi) \qquad maximar\ b^T y - \mu \sum_{i=l}^{n} x_i e^{-s_i/\mu}$$

$$sujeito\ a\ Ax = b$$

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

$$(D_\mu) \qquad maximizar\ b^T y - \mu \sum_{i=l}^{n} x_i e^{-s_i/\mu}$$

$$sujeito\ a\ A^T y + s = c$$

Por hipótese, também iremos considerar

$$F^0(D) \neq \emptyset$$

Denote por F (P_φ) o conjunto viável primal de (P_φ) e por $F(D_\varphi)$ o conjunto viável dual de (D_φ).

Preposição

Para cada $\mu > 0$ seguinte sistema

$$\begin{array}{ll} Ax=b, & x > 0, \\ A^T y + s = c & s \in IR^m, \\ s + \varphi \quad K_\varphi(x) = 0, & \mu > 0 \end{array}$$

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

Ver prova em [1], p. 73.

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

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

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

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

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

Portanto, definimos a trajetória central primal – dual associada a divergência de Kullback -Leibler como sendo o conjunto de pontos.

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

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

As equações acima definem uma curva diferenciável no conjunto IR^n_{++} x IR^m x IR^n_{++}, ou seja, a Trajetória Central KL é uma curva diferenciável.

Seja

$$\Psi : IR^n_{++}\ x\ IR^m\ x\ IR^n\ x\ IR \rightarrow IR^m\ x\ IR^n\ x\ IR^n$$

$$(x, u, s, \mu) \rightarrow (Ax - b, A^T y + s - c, s + \mu \quad K_\mu(x)).$$

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

$$\nabla_{(x,y,s)}\Psi(x, y, s, \mu) = \begin{pmatrix} A & 0 & 0 \\ 0 & A^T & 1 \\ \mu\nabla^2 K_\mu(x) & 0 & 1 \end{pmatrix} = \begin{pmatrix} A & 0 & 0 \\ 0 & A^T & 1 \\ \mu\nabla^2 \varphi(x) & 0 & 1 \end{pmatrix}$$

Note que o sistema acima é equivalente ao seguinte sistema

$\Psi(x,y,s,\mu) = (0_m, 0_n, 0_n)$, $\mu > 0$, $x > 0$, $s \in IR^m$.

Seja $\{(x(\mu), y(\mu), s(\mu)) : \mu > 0\}$ a trajetória primal-dual. A curva diferenciável definida por

$$n : (0, +\infty) \to IR$$

$$\mathbf{n}\ (\mu) = (x(\mu), y(\mu), s(\mu)),$$

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

Teorema

Seja $x^c \in IR^n{}_{++}$, o centro analítico de F^* (P), isto é, o único ponto satisfazendo

$$x^c = \arg\min \left\{ \sum_{i=l}^{n} \left(x_i \, log \frac{x_i}{x_i} + x_i - x_i \right) : x \; F^*(P) \right\}$$

Ver prova em [1], p. 77.

Preposição

Seja $\{x(\mu) : \mu > 0\}$ a trajetória central KL primal. Se $\mu > 0$ e x E {x E IR^n : $Ax = b$, $x \geq 0\}$, então $s(\mu) = - \mu \quad K_\varphi(x(\mu)) = - \mu \log x\ (\mu) + \mu \log x$ é a única solução de

$$\min \left\{ x^T s + \mu \sum_{i=1}^{n} x_i e^{-\frac{s_i}{\mu}} : s \in c + II\ mA^T \right\}$$

Ver prova em [1], p. 81.

Teorema

Sejam $\{x(\mu) : \mu > 0\}$ e $\{ s(\mu)) : \mu > 0\}$ as trajetórias central primal e dual associadas a K_φ, respectivamente. Suponhamos dada uma sequência $\{Y_k\}$ C IR_{++}, satisfazendo (5-36) e a sequência $\{\mu_k\}$ definida como:

$$u_k = \left(\sum_{j=0}^{k} Y_j^{-1} \right)^{-1}, para\ k = 0,1,2 \ldots$$

Então $x^{k+1} + x\ (\mu_k)$ e $s^{-k} = s\ (\mu_k)$ para k = 0,1,2..., onde $\{x^k\}$ e $\{s^{-k}\}$ são as sequências de ponto proximal primal e dual ponderada associadas a $\{\mu_k\}$, respectivamente. Consequentemente,

$$\lim_{k \to +\infty} (x^k, x^{-k}) = (x^c, s^c)$$

onde x^c é o centro analítico da face ótima e s^c é o centróide do conjunto ótimo dual.

Ver prova em [1], p. 85.

Conclusão

Estudamos a convergência das trajetórias central primal e dual associadas às funções entropia exponencial, respectivamente, para os problemas de programação linear. Os autores Cominetti e San Martin em [2] obtiveram uma caracterização dos seus pontos limites. Partimos das hipóteses de que $F^0(P) \neq \emptyset$ e $F^0(D) \neq \emptyset$ e isso 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 essa função é contínua na fronteira, pudemos 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, dentre 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, pudemos provar a convergência da trajetória central primal-dual.

Como aplicação do estudo de trajetórias central prima e dual, mostramos a convergência das sequências proximal primal e dual ponderada associadas à distância Kullback-Leibler para os (PPL). Isto é exatamente o obtido por Iusem et al. [3] e Iusem e Monteiro [4], respectivamente, em programação linear. Embora tenhamos provado que a sequência proximal dual é um problema em aberto, esse pode ser um desafio a ser enfrentado. Além disso, apresentamos a convergência generalizada. Para futuras pesquisas, apontamos então a prova da convergência e caracterização do ponto limite de trajetória central para uma classe de funções mais gerais, na difícil tarefa de mostrar a convergência da sequência proximal dual plena, e não apenas ponderada, bem como estudar os (PPL) para funções objetivas convexas e quaisquer. Reafirmamos a importância de uma segura e vasta formação dos novos profissionais do ensino em matemática. Sem desmerecer as áreas da didática e da metodologia de ensino, acreditamos que o novo professor de Matemática só terá segurança e desenvoltura em sua profissão se possuir uma sólida preparação, de forma aprofundada, nos mais variados conteúdos matemáticos. Esperamos estar, assim, colaborando para a formação e o rigor matemático dos futuros colegas professores.

Referências

[1] LOPES, M. V. Trajetória central associada à entropia e o método do ponto proximal em programação linear. 2007. Dissertação (Mestrado em Ciência da Computação) - Instituto de Informática da Universidade Federal de Goiás, Goiânia, GO, 2007.

[2] COMINETTI, R.; SAN MARTIN, J. Asymptotic Analysis of the exponential penalty trajectory in linear programming. *Mathematical Programming*. v. 67, n. 2, p. 169-187, 1994.

[3] IUSEM, A. N.; SVAITER, B. F.; CRUZ NETE, J.X. da. Central paths, generalized proximal point method and cauchy trajetories in riemannian manifolds. SIAM - *Journal on Control and Optimization*, v. 37, n. 2, p. 566-588, 1999.

[4] IUSEM, A. N.; MONTEIRO, R. D. C. On dual convergence of the generalized proximal point method with Bregman distances. *Mathematcs of Operations Research.* v. 25, n. 4, p. 606-624, 2000.

[5] MURTY, K. G. *Operacions research,* Princeton Hall, Englewood Cliffs, New Jersey, 1995.

[6] DORFMAN, R. The conception of the Tableau Economique. *The Economist Journal*, v. 107, p. 491-495, 1959.

[7] WALRAS, L. *Principe d'une théorie mathematique de l'échange,* Journal des Economistes, Elements of Pure Economics, or the theory of social wealth, 1874.

[8] DANTZIG, G. B. *Linear programming and extensions*. Princeton University Press, Princeton, New Jersey, 1963.

[9] NEUMANN, John von. A model of general economic equilibrium. In: MENGER, K. (editor). *Ergebnisse eines mathematischen kolloquiums*, 1937.

[10] LEONTIF, W. Input an output economics. *Scientific American Journal*, v. 35, 1936.

[11] KUHN, H. W.; TUCKER, A. W. *Nonlinear programming,* Proceedings of 2nd Berkeley Symposium: p. 481-492, Berkeley: University of California Press, 1951.

[12] LEONTIF, W. Input an output economics. *Scientific American Journal*, v. 35, 1936.

[13] DORFMAN, R. and others. *Linear programming and economic analysis.* Dover Books on Mathematics an production to linear algebra, 1951.

[14] MINOUX, M. *Mathematical programming* – Theory and algorithms, Wiley Press, Chichester, 1986.

[15] FOURIER, J. B. Solution d'une question particuliere du calcul des inegalites. *Nouveau Bulletin des Sciences par la Societe philomathique*, 1826.

[16] POUSSIN, L. V. de. *Sur la method de l'approximation minimum,* Ann. Soc. Sci. Bruxelles, p. 1-16. English translation by H. E. Salger. National Bureau of Standards, USA, 1911.

[17] KARUSH, W. *Minima of functions of several variables with inequalities as side constraints*, M. Sc. Dissertation. Dept. of Mathematics, Univ. of Chicago, Chicago, lllinois, 1939.

[18] COTTLE, R.; JOHNSON, E.; WETS, R. George B. Dantzig (1914 – 2005). *Notices of the American Mathematical Society*, v. 54, n. 3, p. 334-362, march 2007.

[19] KHACHIYAN, L. G. *A polynomial algorithm in linear programming.* Soviet Mathematics Doklady 20, p. 191-194, 1979.

[20] KARMARKAR, N. *A new polynominal time algorithm for linear programming.* Combinatorica 4, p. 373-395, 1984.

Recebido em: 25 out. 2010
Aceito em: 23 nov. 2010