

UM ALGORITMO DE PLANOS DE CORTE BASEADO EM CENTROS PARA  
UM PROBLEMA DE ESTIMATIVA DE MÁXIMA VEROSSIMILHANÇA

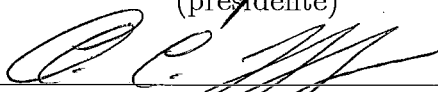
Fernanda Maria Pereira Raupp

TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS PRO-  
GRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE FE-  
DERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁ-  
RIOS PARA OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS EM ENGE-  
NHARIA DE SISTEMAS E COMPUTAÇÃO.

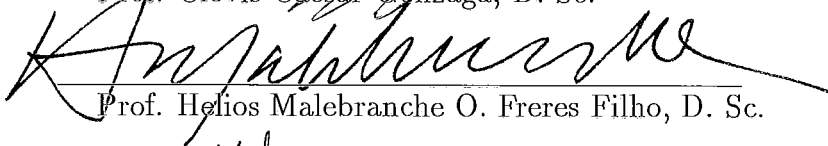
Aprovada por:



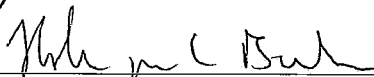
Prof. Nelson Maculan Filho, D. Sc.  
(presidente)



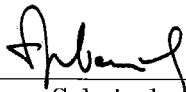
Prof. Clóvis Caesar Gonzaga, D. Sc.



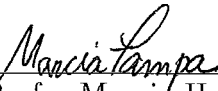
Prof. Helios Malebranche O. Freres Filho, D. Sc.



Prof. Helio Jose Correa Barbosa, D. Sc.



Profa. Susana Scheimberg de Makler, D. Sc.



Profa. Marcia Helena Costa Fampa, D. Sc.

RIO DE JANEIRO, RJ – BRASIL  
AGOSTO DE 1998

RAUPP, FERNANDA MARIA PEREIRA

Um Algoritmo de Planos de Corte Baseado  
em Centros para um Problema de Estimativa de  
Máxima Verossimilhança [Rio de Janeiro] 1998

VI, 104 p. 29.7 cm (COPPE/UFRJ, D. Sc.,  
Engenharia de Sistemas e Computação, 1998)

Tese – Universidade Federal do Rio de  
Janeiro, COPPE

- 1 – Algoritmo de Pontos Interiores
- 2 – Algoritmo de Planos de Corte
- 3 – Centro Analítico
- 4 – Verossimilhança

I. COPPE/UFRJ II. Título ( série ).

*Ao Augusto  
aos meus filhos  
aos meus pais  
à minha irmã  
e à memória de meu irmão*

# Agradecimentos

Ao Prof. Clóvis C. Gonzaga, pela orientação desta tese, pela contribuição prestada durante o seu desenvolvimento e por sua amizade. Agradeço-lhe sobretudo pela confiança em mim depositada e pela oportunidade de concretizar o projeto de doutorado sanduiche no exterior. Sem a sua indicação, o trabalho na Universidade de Iowa, que tanto enriqueceu esta tese, provavelmente não teria se concretizado.

Ao Prof. Nelson Maculan Filho, pela co-orientação desta tese e pela participação na banca examinadora desta tese e de qualificação. Agradeço-lhe também o interesse e a ajuda prestada durante todo o desenvolvimento deste trabalho.

Ao Prof. Kenneth Kortanek, por assentir na minha estadia na Universidade de Iowa através do programa de doutorado sanduiche no exterior.

Ao Prof Yinyu Ye, pela proposta do tema abordado nesta tese e pela orientação do trabalho desenvolvido durante a minha estadia na Universidade de Iowa, que resultou num dos exames para a minha qualificação. Agradeço-lhe também pelo incentivo e interesse demonstrado durante o desenvolvimento desta tese, sempre disposto a me oferecer material de pesquisa e a esclarecer as minhas dúvidas.

Ao Prof. Helios Malebranche, pela orientação na estatística e disposição para esclarecer as minhas dúvidas e por participar na banca examinadora desta tese.

Aos Profs. Susana Scheimberg de Makler, Helio Jose Correa Barbosa e Marcia Helena Costa Fampa, por assentirem em compor a banca examinadora.

Aos colegas de Otimização, em especial Xiong Zhang, pela ajuda e companheirismo durante a estadia na Universidade de Iowa e a Marcia Fampa que desde que comecei o curso tem me ajudado a esclarecer as minhas dúvidas, inclusive em Iowa.

Aos colegas do LNCC, em especial Ana Tereza que muito me ajudou sendo minha amiga-procuradora, Regina, Sandra e Helio, pelo incentivo. Aos demais colegas do LNCC, que no convívio diário contribuíram de diversas formas para esta tese.

Ao LNCC, por promover o meu crescimento enquanto profissional, dando-me a oportunidade de fazer o curso, pelas facilidades computacionais e pelo apoio administrativo.

Ao colega Ricardo Duarte, pelas informações e orientação no estudo da resolução de sistemas esparsos.

Aos professores na COPPE/UFRJ, pelos conhecimentos transmitidos durante a minha passagem pelo programa de pós-graduação. Aos funcionários técnicos e administrativos na COPPE/UFRJ, pela atenção e prestatividade.

À CAPES pelo suporte financeiro recebido antes de sair para o doutorado sanduiche. Ao CNPq, pelo suporte financeiro recebido durante o doutorado sanduiche.

Ao meu marido e aos meus filhos pelo apoio e solidariedade durante todo o desenvolvimento deste trabalho. Agradeço-lhes por me incentivarem a optar pelo doutorado sanduiche e por me acompanharem durante toda a minha estadia em Iowa City.

Aos meus pais e irmãos pelo incentivo, pelo exemplo de perseverança, de profissionalismo e pela ajuda prestada durante toda a minha pós-graduação, sem os quais a finalização deste trabalho se tornaria ao menos bem mais difícil. E à família do meu marido que desde o início me incentivou e me ajudou.

Resumo da Tese apresentada à COPPE como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D. Sc.)

UM ALGORITMO DE PLANOS DE CORTE BASEADO EM CENTROS PARA  
UM PROBLEMA DE ESTIMATIVA DE MÁXIMA VEROSSIMILHANÇA

Fernanda Maria Pereira Raupp

Agosto de 1998

Orientador: Clóvis Caesar Gonzaga

Programa: Engenharia de Sistemas e Computação

Neste trabalho desenvolvemos um algoritmo de planos de corte baseado em pontos centrais para resolver o problema de encontrar a estimativa de máxima verossimilhança de uma função distribuição com suporte finito. Começamos com o estudo do problema estatístico. Definimos um problema interno como sendo o problema de otimização resultante da modelagem do problema estatístico. O algoritmo de planos de corte resolve aproximadamente uma sequência de problemas internos utilizando um algoritmo de pontos interiores primais-duais inviáveis. Para compreensão, fazemos um paralelo do algoritmo que propomos com o algoritmo conceitual de planos de corte em pontos centrais para resolver um problema geral de maximização de uma função côncava com infinitas restrições lineares. Iniciamos a análise de complexidade do algoritmo conceitual e apresentamos alguns resultados computacionais.

Abstract of Thesis presented to COPPE as a partial fulfillment of the requirements for the degree of Doctor of Science (D. Sc. )

A CENTER CUTTING PLANE ALGORITHM FOR THE LIKELIHOOD  
ESTIMATE PROBLEM

Fernanda Maria Pereira Raupp

August 1998

Advisor: Clóvis Caesar Gonzaga

Department: Programa de Engenharia de Sistemas e Computação

In this work we develop a center cutting plane algorithm to solve the nonparametric maximum likelihood estimate problem of a mixing distribution. We start with the study of the statistical problem. We define an inner problem as the optimization problem derived by modelling the statistical problem. The cutting plane algorithm solves a sequence of inner problems approximately applying infeasible-interior-point primal-dual algorithm. In order to understand, we compare the algorithm we propose with a conceptual center cutting plane algorithm to maximize a general concave function subject to infinite linear constraints. We initiate the complex analysis of the conceptual algorithm and present some computational results.

# Índice

<b>I</b>	<b>Introdução</b>	<b>1</b>
<b>II</b>	<b>O Problema Estatístico</b>	<b>7</b>
<b>III</b>	<b>Modelagem do Problema</b>	<b>17</b>
III.1	O problema interno . . . . .	18
III.2	O Algoritmo de Otimização . . . . .	24
III.3	Introdução de um corte profundo: o Oráculo . . . . .	25
<b>IV</b>	<b>Otimização</b>	<b>28</b>
IV.1	Introdução a um Algoritmo de Pontos Interiores . . . . .	29
IV.2	Trajetória Central . . . . .	32
IV.3	O método preditor-corretor inviável . . . . .	38
IV.4	Um ponto central aproximado após a introdução de um ou mais cortes	45
IV.5	O ponto inicial após a introdução de um ou mais cortes . . . . .	48
IV.5.1	Atualização da Variável Dual . . . . .	50
IV.5.2	Atualização da Variável Primal . . . . .	52



<b>V O Algoritmo</b>	<b>61</b>
V.1 Geometria . . . . .	63
V.2 Algoritmo de Planos de Corte para Resolver o Problema de Maximização . . . . .	65
V.3 Introdução à Análise de Convergência . . . . .	69
<b>VI Experiência Computacional</b>	<b>83</b>
VI.1 Interpretação do problema estatístico . . . . .	83
VI.2 Implementação . . . . .	86
VI.3 Resultados Numéricos . . . . .	87
<b>VII Comentários Finais</b>	<b>97</b>

# Capítulo I

## Introdução

Nós propomos nesta tese um novo algoritmo de pontos interiores que utiliza o método de planos de corte para encontrar a estimativa de máxima verossimilhança de uma distribuição acumulada com suporte finito. Esta é uma das abordagens utilizadas pela estatística para estimar parâmetros populacionais desconhecidos a partir de um conjunto de observações.

Os algoritmos existentes podem ser classificados em três categorias. A primeira e mais antiga, introduzida na década de 60, fixa o número de pontos suportes como sendo o máximo possível, tornando a convergência do algoritmo muito lenta, Laird [29].

A segunda categoria de algoritmos está baseada na derivada direcional. O primeiro algoritmo deste tipo, a cada iteração, obtém um ponto que apresenta a máxima subida, porém não é um ponto na direção de máxima subida. Neste tipo de algoritmo somente um ponto por iteração é adicionado ao conjunto suporte, Böhning [3]. A mais recente variante, que apresenta vantagem em relação aos algoritmos anteriores, propõe que mais de um ponto seja adicionado ao conjunto suporte por iteração. Este último algoritmo considera todos os máximos locais positivos da derivada direcional e a atualização do ponto é feita na direção de máxima subida, Lesperance [30]. Após a interpretação geométrica dada ao problema estatístico feita por Lindsay [31], que inclusive permitiu explicar melhor os algoritmos

acima, o problema estatístico pode ser formulado como um problema de otimização. Como é um problema difícil (em geral as restrições não são facilmente estabelecidas), os pacotes de estatística falham em geral quando as amostras são grandes.

A terceira categoria resolve a formulação dual do problema de otimização originado pelo problema estatístico. Coope e Watson [5] interpretaram o problema de otimização como sendo um problema da programação semi-infinita e desenvolveram um algoritmo utilizando o método do Lagrangeano projetado. Para a sua interpretação não foi utilizada a idéia da derivada direcional e sua compreensão não ficou clara para os estatísticos.

Mais recentemente, outros algoritmos foram desenvolvidos por Terlaky e Vial [49] utilizando algoritmos de pontos interiores com abordagem primal e primal-dual. Porém dificuldades numéricas foram encontradas em amostras grandes.

Algoritmos de pontos interiores estão fundamentados em seguir aproximadamente uma trajetória no interior do conjunto viável do problema de otimização. Eles têm apresentado bastante sucesso na prática, em especial quando são utilizados para resolver alguns problemas da otimização convexa, como por exemplo: problemas de programação linear, problemas da programação quadrática, problemas de complementaridade linear, problemas da programação semi-infinita e problemas da programação semi-definida positiva. A pesquisa dos métodos de pontos interiores foi impulsionada a partir da publicação do algoritmo de complexidade polinomial para resolver um problema da programação linear de autoria de Karmarkar [22]. Verificou-se mais tarde que muitos de seus princípios já haviam sido introduzidos, como por exemplo a metodologia da barreira logarítmica [8]. À medida que a teoria avança, o limite de complexidade dos algoritmos de pontos interiores vem baixando. A evolução dessa área está descrita nos apanhados de Gonzaga [19], Todd [48] e den Hertog [9].

O problema de encontrar a estimativa de máxima verossimilhança pode ser formulado como um problema de maximização de uma função côncava

não-linear com restrições lineares que não podem ser explicitadas. Os algoritmos de planos de corte não precisam conhecer previamente as restrições do problema de otimização, eles adicionam uma ou mais restrições à medida que necessitam ao longo do processo. Nos algoritmos de planos de corte, uma sequência de problemas é examinada. Cada problema da sequência é originado a partir do anterior adicionando uma ou mais restrições.

O que determina o desenvolvimento dos algoritmos de planos de corte é a escolha do ponto a ser consultado pelo oráculo ou ponto teste. Este ponto define a nova aproximação poliédrica do conjunto solução. Dantzig e Wolfe [7] generalizaram o primeiro algoritmo de planos de corte para resolver um problema dual da programação linear utilizando o método simplex [6]. Neste caso o ponto teste é a solução de um problema de relaxação linear. O mesmo acontece com o método clássico desenvolvido por Kelley [24], que minimiza uma função linear num conjunto convexo. Os algoritmos de planos de corte tiveram bastante sucesso quando utilizaram como ponto teste o centro analítico sugerido por Sonnevend [47], isto é, utilizaram um algoritmo de pontos interiores para encontrar um ponto teste.

Vários autores têm desenvolvido algoritmos aplicando a metodologia de planos de corte em centro analítico para resolver diferentes tipos de problemas, para citar alguns: Goffin e Vial [13] resolvem um problema de programação linear usando uma abordagem primal, Ye [50] resolve um problema de viabilidade convexa com abordagem primal, Mitchell e Todd [38] resolvem um problema de programação combinatória com abordagem primal, Goffin et al. [14] resolvem um problema de multi-fluxo com abordagem primal e Bahn et al. [2] resolvem um problema de programação estocástica com abordagem primal. Recentemente Gondzio [16] resolve um problema da programação linear com abordagem primal-dual.

O algoritmo que propomos utiliza o método de planos de corte em centro analítico para resolver o dual do problema de otimização originado pelo problema estatístico. Cada problema não-linear definido com a adição de um ou mais cortes é resolvido por um algoritmo de pontos interiores com abordagem primal-

dual. Escolhemos a abordagem primal-dual, porque ela tem apresentado a maior eficiência quando comparada com as demais desde que foi introduzida por Kojima, Mizuno e Yoshida [26].

Ye [51] mostrou ser polinomial a complexidade do algoritmo de múltiplos planos de corte em centro analítico que resolve um problema de viabilidade convexa, supondo que o oráculo gera um número finito de cortes por iteração, que o algoritmo não faz nenhum descarte das restrições desnecessárias e que os cortes passam pelo centro analítico. Recentemente, Goffin e Vial [15] obtiveram o mesmo resultado quando relaxaram a exigência dos cortes passarem pelo centro analítico, podendo ser pouco profundos ou profundos. Como o problema de viabilidade convexa é equivalente ao problema de maximizar uma função côncava em um conjunto convexo, o mesmo limite de complexidade pode ser considerado para o algoritmo de planos de corte em centro analítico que resolve o problema de encontrar a estimativa de máxima verossimilhança de uma distribuição com suporte finito.

A seguir passamos à descrição dos capítulos. O capítulo II apresenta a descrição do problema de encontrar a estimativa de máxima verossimilhança de uma distribuição com suporte finito. No capítulo III fazemos a modelagem do problema estatístico que resulta na definição de um problema interno. A formulação dual do problema interno é apresentada e estudam-se as condições de otimalidade deste problema. Apresentamos brevemente o nosso algoritmo de planos de corte em centro analítico para resolver a formulação dual do problema estatístico e definimos o oráculo e o tipo de corte que ele fornece. No capítulo IV apresentamos a metodologia da otimização para resolver o problema estatístico. Apresentamos os algoritmos de pontos interiores com abordagem primal-dual para resolver cada problema interno e encontrar um novo ponto teste quando um ou mais cortes são adicionados a um problema interno. Por fim fazemos um estudo do ponto inicial para o algoritmo de centralização após a adição de um ou mais cortes. O capítulo V apresenta o algoritmo de múltiplos planos de corte baseado em centros para resolver o problema estatístico com mais detalhes. Como parte de nossa contribuição, apre-

sentamos também o algoritmo conceitual de planos de corte em centro analítico para resolver um problema geral de maximização de uma função côncava com infinitas restrições lineares. Fazemos um paralelo entre os dois algoritmos com o objetivo de entender o primeiro. Em seguida apresentamos a complexidade dos algoritmos de planos de corte em centro analítico existente. Damos início à análise de convergência deste algoritmo quando resolve um problema de maximizar uma função côncava com infinitas restrições, sem recorrer ao caso da equivalência deste algoritmo com o algoritmo que resolve o problema de viabilidade convexa e obtemos o mesmo resultado para a variação do potencial. Nosso algoritmo conceitual assim como a introdução à análise é mais simples do que aqueles que encontramos em Mokhtarian e Goffin [42]. O capítulo VI apresenta a interpretação do problema estatístico para iniciar o algoritmo de otimização e depois para entender os resultados do algoritmo. Descrevemos a implementação do algoritmo e apresentamos alguns resultados computacionais. Finalmente, fazemos alguns comentários a respeito do nosso trabalho.

## Notação

Dentro do contexto da estatística, adotamos a seguinte notação. Usaremos letras maiúsculas para variáveis aleatórias enquanto que minúsculas representam possíveis valores realizados por elas.

A seguir apresentamos a notação que é encontrada com maior frequência na literatura a respeito de algoritmos de pontos interiores. Letras minúsculas como  $x$ ,  $s$ ,  $y$ , ... são vetores, enquanto que letras gregas como  $\alpha$ ,  $\beta$ ,  $\gamma$ , ... são escalares.

A representação de conjuntos se faz com letras maiúsculas, podendo ser gregas ou não. O interior de um conjunto  $F$  denotamos por  $\overset{\circ}{F}$ . Também utilizaremos letras maiúsculas, como por exemplo,  $X$  e  $S$ , para denotar matrizes diagonais, cujos elementos da diagonal são as componentes dos vetores  $x$  e  $s$ , respectivamente.

Conseqüentemente, o vetor formado pelo produto entre os vetores  $x$  e  $s$  componente a componente é  $Xs$ . Tal produto é conhecido como produto de Hadamard.

O símbolo  $e$  representa o vetor com componentes todas iguais a 1 e dimensão de acordo com o contexto. A transposição da matriz  $X$  denota-se por  $X^T$ . A matriz identidade denotamos por  $I$ . A inversa da matriz  $M$  representamos por  $M^{-1}$ .

Operações componente a componente entre vetores são muito comuns na teoria de algoritmos de pontos interiores e sua notação simplifica as expressões matemáticas. A operação exponenciação componente a componente é denotada por um vetor com expoente, como exemplo, temos  $x^2 = Xx$ , assim como  $x^{-1} = X^{-1}e$ .

O produto escalar entre os vetores  $x$  e  $s$  é representado pela notação usual  $x^T s$ . E a norma euclidiana do vetor  $x$  denota-se por  $\|x\|$ . A desigualdade vetorial  $x \geq y$  significa  $x_j \geq y_j \quad \forall j = 1, \dots, n$ .

Sendo  $\mathbb{R}$  o conjunto dos números reais, denota-se por  $\mathbb{R}_+$  o conjunto  $\{x \in \mathbb{R} : x \geq 0\}$  e por  $\mathbb{R}_{++}$  o conjunto  $\{x \in \mathbb{R} : x > 0\}$ .

As funções exponencial e logaritmo natural são denotadas, respectivamente, por  $\exp(\cdot)$  e  $\log(\cdot)$ .

## Capítulo II

# O Problema Estatístico

Dentre as técnicas estatísticas existentes, passamos a destacar aquela que a partir de um conjunto de observações realizadas infere sobre alguns parâmetros da população desconhecidos. Uma das abordagens utilizadas na estimação de parâmetros é maximizar a verossimilhança. O princípio da estimação pelo método da máxima verossimilhança é geralmente superior ao princípio da estimação pelo método dos momentos. Além de satisfazer as propriedades básicas de um bom estimador, o estimador da máxima verossimilhança apresenta como propriedade o fato de ser normalmente distribuído, isto é, de amostra para amostra, o valor estimado varia segundo uma distribuição normal.

Encontrar a estimativa de máxima verossimilhança na Teoria de Modelos Mistos é um problema clássico na estatística. A seguir passamos a descrever o problema estatístico tendo como referência alguns livros introdutórios tais como [54, 44] e os artigos [30, 31].

Suponha que  $V$  é um vetor  $n$ -dimensional de variáveis aleatórias contínuas e independentes. Suponha também que cada componente é uma variável aleatória  $V_i$  possuindo densidade de probabilidade condicionada a um parâmetro populacional desconhecido. Suponha que este parâmetro seja dado por uma variável aleatória  $\Theta$ , contínua no intervalo  $\Gamma \subset \mathbb{R}$ , com distribuição acumulada  $F_\Theta$  e densidade de probabilidade  $f_\Theta$ . Por hipótese, todos os demais parâmetros populacionais



são considerados fixos e conhecidos.

Observa-se na prática que pontos relevantes de um dado evento estão sempre condicionados a alguma restrição característica da população, daí o fato de considerarmos probabilidades condicionadas.

As funções densidades de probabilidade conjuntas de cada  $V_i$  e  $\Theta$ ,  $i = 1, \dots, n$ , são dadas por

$$f_{V_i, \Theta}(v_i, \theta) = f_{V_i | \Theta}(v_i | \theta) f_{\Theta}(\theta), \quad (\text{II.1})$$

onde  $f_{V_i | \Theta}$  é a função de densidade de probabilidade de  $V_i$  condicionada a  $\Theta$ . A função densidade de probabilidade (marginal) de cada variável aleatória  $V_i$  é dada por

$$f_{V_i}(v_i) = \int_{\Gamma} f_{V_i, \Theta}(v_i, \theta) d\theta,$$

e a função densidade de probabilidade conjunta das variáveis aleatórias  $V_i$ ,  $i = 1, \dots, n$ , é

$$f_V(v) = \prod_{i=1}^n \int_{\Gamma} f_{V_i, \Theta}(v_i, \theta) d\theta.$$

Observa-se que esta função pode ser reescrita como

$$f_V(v) = \prod_{i=1}^n \int_{\Gamma} f_{V_i | \Theta}(v_i | \theta) dF_{\Theta}(\theta), \quad (\text{II.2})$$

pois, por definição, temos, fazendo  $f_{\Theta}(\lambda) = 0$  para  $\lambda \in \mathbb{R} \setminus \Gamma$ ,

$$Pr(\Theta \leq \theta) = F_{\Theta}(\theta) = \int_{-\infty}^{\theta} f_{\Theta}(\lambda) d\lambda,$$

onde  $Pr(\Theta \leq \theta)$  denota a probabilidade da variável aleatória  $\Theta$  assumir valores menores ou iguais a  $\theta$ .

## Verossimilhança

Sejam  $\mathcal{F}_{\Theta}$  a família de todas as distribuições acumuladas definidas em  $\Gamma$  e  $F_{\Theta} \in \mathcal{F}_{\Theta}$ . Se fixarmos o vetor  $v$ , o valor de  $f_V(v)$  acima é a verossimilhança da amostra

associada à função distribuição acumulada  $F_{\Theta}(\cdot)$ . Definimos por *estimador da verossimilhança de  $F_{\Theta}(\cdot)$*  a função

$$\mathcal{L}(F_{\Theta}) = f_V(\cdot).$$

## O Problema

Dada uma observação  $v \in \mathbb{R}^m$  e dadas as funções  $f_{V_i|\Theta}(\cdot)$  para  $i = 1, \dots, m$ , onde cada  $v_i$  apresenta multiplicidade  $t_i$ , como veremos abaixo,  $i = 1, \dots, m$ , queremos encontrar uma função distribuição acumulada  $F_{\Theta}(\cdot) \in \mathcal{F}_{\Theta}$  que maximiza  $f_V(v)$ .

Vamos desenvolver um pouco mais a expressão para  $f_V(\cdot)$ . Seja  $V$  um vetor  $n$ -dimensional de variáveis aleatórias. Por definição, para cada  $i = 1, \dots, n$ ,  $f_{V_i|\Theta}(\cdot)$  é uma função não-negativa.

Agora, suponha que o vetor das observações  $(v_1, v_2, \dots, v_n)$  apresente  $m \leq n$  dados distintos. Sem perda de generalidade, dividimos as variáveis aleatórias em  $m$  grupos, com multiplicidades  $t_1, t_2, \dots, t_m$  tais que  $t_1 + t_2 + \dots + t_m = n$ , de modo que as variáveis de cada grupo  $r$ ,  $r = 1, \dots, m$ , apresentem a mesma função densidade de probabilidade condicionada. Estas funções são chamadas de núcleos distintos do estimador da verossimilhança e denotadas por  $f_{V_r|\Theta}$ ,  $r = 1, \dots, m$ . Com esta notação, a expressão (II.2) fica

$$\mathcal{L}(F_{\Theta}) = \prod_{r=1}^m (f_{V_r}(v_r))^{t_r}.$$

Tomando o logaritmo, o problema é escrito como

$$(EMV) \quad \underset{F_{\Theta} \in \mathcal{F}_{\Theta}}{\text{maximizar}} \quad \mathcal{L}(F_{\Theta}) = \sum_{r=1}^m t_r \log \int_{\Gamma} f_{V_r|\Theta}(v_r|\theta) dF_{\Theta}(\theta).$$

## Notação

Dado  $\theta \in \Gamma$ , denotamos para  $i = 1, \dots, m$

$$\begin{aligned} l_i(\theta) &= f_{V_i|\Theta}(v_i|\theta) \text{ e} \\ l(\theta) &= (l_1(\theta), \dots, l_m(\theta)). \end{aligned} \tag{II.3}$$

Seja  $l(\theta) = (l_1(\theta), \dots, l_m(\theta))$  o vetor que representa os  $m$  núcleos distintos do estimador da verossimilhança que têm origem nos dados também distintos  $v_1, v_2, \dots, v_m$ . Sendo  $t_i$  a multiplicidade de  $l_i(\theta)$  para  $i = 1, \dots, m$  e dada  $F_\Theta \in \mathcal{F}_\Theta$ , podemos escrever

$$\mathcal{L}(F_\Theta) = \sum_{i=1}^m t_i \log \int_{\Gamma} l_i(\theta) dF_\Theta(\theta)$$

e fazendo para  $i = 1, \dots, m$

$$L_i(F_\Theta) = \int_{\Gamma} l_i(\theta) dF_\Theta(\theta), \quad (\text{II.4})$$

temos agora

$$\mathcal{L}(F_\Theta) = \sum_{i=1}^m t_i \log L_i(F_\Theta).$$

De acordo com a notação acima, o problema *EMV* pode ser substituído por encontrar  $\hat{F}_\Theta \in \mathcal{F}_\Theta$  tal que

$$\mathcal{L}(\hat{F}_\Theta) = \max_{F_\Theta \in \mathcal{F}_\Theta} \sum_{i=1}^m t_i \log L_i(F_\Theta).$$

## O Espaço de Soluções

Seja  $\theta \in \Gamma$  dado. A função degrau  $U(\theta_j) \in \mathcal{F}_\Theta$  é definida por: dado  $\theta_j \in \Gamma$

$$U(\theta_j)(\theta) = \begin{cases} 0, & \text{se } \theta < \theta_j \\ 1, & \text{se } \theta \geq \theta_j \end{cases},$$

cujos gráficos podemos ver na figura (II.1). Observe que  $U(\theta_j)$  é a distribuição acu-

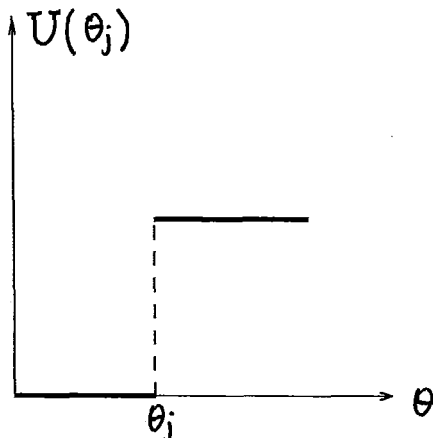


Figura II.1: Função  $U(\theta_j)$

mulada trivial definida em  $\Gamma$ ,  $Pr(\Theta \in \Gamma) = 1$ .

A função degrau  $U(\theta_j)$  apresenta como “derivada” a “função” delta de Dirac,  $\delta : \mathbb{R} \rightarrow \mathbb{R} \cup \{+\infty\}$ , que tem a seguinte propriedade: dado  $\bar{t} \in \mathbb{R}$  e uma função  $f : \mathbb{R} \rightarrow \mathbb{R}$ , então

$$\int_{-\infty}^{+\infty} f(t)\delta(\bar{t})dt = f(\bar{t}).$$

Seja  $F_\Theta \in \mathcal{F}_\Theta$  uma distribuição acumulada definida em  $\Gamma$ ,  $Pr(\Theta \in \Gamma) = 1$ .  $F_\Theta$  diz-se finitamente discretizada se é expressa como uma combinação convexa finita do tipo

$$F_\Theta = \sum_{j=1}^p x_j U(\theta_j), \quad (\text{II.5})$$

onde, para  $j = 1, \dots, p$ ,  $\theta_j \in \Gamma$ ,  $U(\theta_j) \in \mathcal{F}_\Theta$ ,  $\sum_{j=1}^p x_j = 1$ , e  $x_j \geq 0$ , com  $p > 0$  inteiro.

Denotamos cada valor discretizado  $\theta_j$  de  $\Gamma$ ,  $j = 1, \dots, p$ , por ponto suporte de  $F_\Theta$ . E o conjunto dos pontos suportes chamamos de conjunto suporte ou simplesmente suporte.

Um exemplo de  $F_\Theta$  finitamente discretizada ou com suporte finito apresentamos em (II.2)

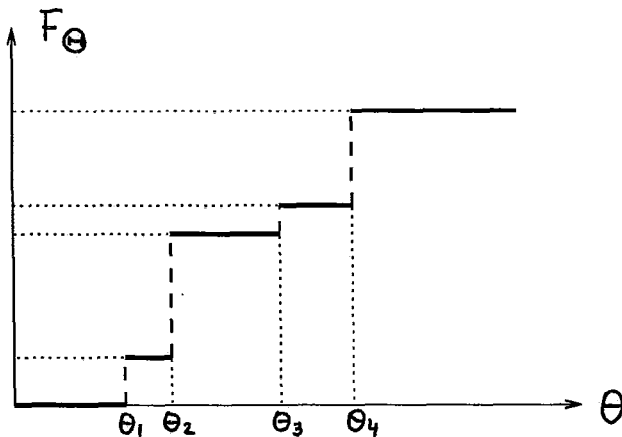


Figura II.2: Exemplo de  $F_\Theta$  com suporte finito

Devido à integração em (II.4), o problema EMV pode ser escrito

como: encontrar  $\hat{F}_\Theta \in \mathcal{F}_\Theta$  tal que

$$\mathcal{L}(\hat{F}_\Theta) = \sup \left\{ \sum_{i=1}^m t_i \log L_i(F_\Theta) : F_\Theta \in \mathcal{F}_\Theta, F_\Theta \text{ com suporte finito} \right\}.$$

Vamos construir um conjunto no espaço  $\mathbb{R}^m$  de elementos  $l(\theta)$  (como definido em (II.3)) cujas componentes têm origem em todas as possíveis distribuições acumuladas triviais em  $\Gamma$ :

$$C = \{l(\theta) : \theta \in \Gamma\}. \quad (\text{II.6})$$

No lema a seguir veremos como deve ser um elemento do conjunto  $C$ .

**Lema II.1** *Dado  $\bar{\theta} \in \Gamma$ , se  $F_\Theta = U(\bar{\theta}) \in \mathcal{F}_\Theta$ , então  $l(\bar{\theta}) = L(U(\bar{\theta})) \in C$ .*

*Prova.* De acordo com (II.3),

$$\begin{aligned} l(\bar{\theta}) &= (l_1(\bar{\theta}), \dots, l_m(\bar{\theta})) \\ &= \left( \int_{\Gamma} l_1(\theta) \delta(\bar{\theta}) d\theta, \dots, \int_{\Gamma} l_m(\theta) \delta(\bar{\theta}) d\theta \right) \\ &= \left( \int_{\Gamma} l_1(\theta) dU(\bar{\theta})(\theta), \dots, \int_{\Gamma} l_m(\theta) dU(\bar{\theta})(\theta) \right) \\ &= (L_1(U(\bar{\theta})), \dots, L_m(U(\bar{\theta}))) \\ &= L(U(\bar{\theta})). \end{aligned}$$

■

Geometricamente o conjunto  $C$  representa uma curva limitada sobre a qual estão todos os valores possíveis de  $l(\theta)$  para  $\theta \in \Gamma$  [32].

Uma combinação convexa finita de elementos de  $C$  é expressa por  $\sum_{j=1}^p x_j l(\theta_j)$ , onde  $\theta_j \in \Gamma$ ,  $\sum_{j=1}^p x_j = 1$ , e  $x_j \geq 0$ ,  $j = 1, \dots, p$ , com  $p > 0$  inteiro. A envoltória convexa de  $C$  é o conjunto de todas as combinações convexas finitas de elementos de  $C$ .

**Lema II.2** *Considere o conjunto  $C$  definido como em (II.6). A envoltória convexa de  $C$  é dada por*

$$\text{conv}(C) = \{L(F_\Theta) : F_\Theta \in \mathcal{F}_\Theta \text{ com suporte finito}\},$$

onde  $L(F_\Theta) = (L_1(F_\Theta), \dots, L_m(F_\Theta))$ , definido em (II.4).

*Prova.* Dado  $\theta \in \Gamma$ , se  $F_\Theta(\theta) = \sum_{j=1}^p x_j U(\theta_j)(\theta)$ , então

$$\frac{dF_\Theta}{d\theta}(\theta) = \sum_{j=1}^p x_j \frac{dU(\theta_j)}{d\theta}(\theta) = \sum_{j=1}^p x_j \delta(\theta_j).$$

Seja  $\sum_{j=1}^p x_j l(\theta_j)$  uma combinação convexa finita de elementos de  $C$ . Temos

$$\begin{aligned} \sum_{j=1}^p x_j l(\theta_j) &= \sum_{j=1}^p x_j (l_1(\theta_j), \dots, l_m(\theta_j)) \\ &= \left( \sum_{j=1}^p x_j l_1(\theta_j), \dots, \sum_{j=1}^p x_j l_m(\theta_j) \right) \\ &= \left( \sum_{j=1}^p x_j \int_{\Gamma} l_1(\theta) \delta(\theta_j) d\theta, \dots, \sum_{j=1}^p x_j \int_{\Gamma} l_m(\theta) \delta(\theta_j) d\theta \right) \\ &= \left( \int_{\Gamma} l_1(\theta) \left[ \sum_{j=1}^p x_j \delta(\theta_j) d\theta \right], \dots, \int_{\Gamma} l_m(\theta) \left[ \sum_{j=1}^p x_j \delta(\theta_j) d\theta \right] \right) \\ &= \left( \int_{\Gamma} l_1(\theta) \left[ \sum_{j=1}^p x_j dU(\theta_j)(\theta) \right], \dots, \int_{\Gamma} l_m(\theta) \left[ \sum_{j=1}^p x_j dU(\theta_j)(\theta) \right] \right) \\ &= \left( \int_{\Gamma} l_1(\theta) dF_\Theta(\theta), \dots, \int_{\Gamma} l_m(\theta) dF_\Theta(\theta) \right) \\ &= (L_1(F_\Theta), \dots, L_m(F_\Theta)) = L(F_\Theta). \end{aligned}$$

■

Logo, o problema de maximização pode ser escrito como encontrar  $\hat{L}$  tal que

$$\hat{L} = \sup \left\{ \sum_{i=1}^m t_i \log L_i : L \in \text{conv}(C) \right\} \text{ e}$$

$\hat{F}_\Theta$  tal que  $\hat{L} = L(\hat{F}_\Theta)$ , onde  $\hat{F}_\Theta \in \mathcal{F}_\Theta$  possui suporte finito e  $L = (L_1, \dots, L_m)$ .

## Existência e Unicidade

Antes de prosseguirmos, vamos considerar que  $C$  seja um subconjunto fechado do  $\mathbb{R}^m$ . Com isto,  $\text{conv}(C)$  é um conjunto compacto, já que  $C$  é limitado.

Vamos denotar a função objetivo do problema acima por  $f : \text{conv}(C) \subset \mathbb{R}^m \rightarrow \mathbb{R}$  tal que para  $L \in \text{conv}(C)$

$$f(L) = \sum_{i=1}^m t_i \log L_i.$$

De acordo com as propriedades da função  $\log$ , a função  $f(\cdot)$  é contínua, estritamente côncava e monotonamente crescente. Devido a estas propriedades e mais ainda considerando  $\text{conv}(C) \neq \emptyset$ , podemos afirmar que existe um único  $\hat{L}$  na fronteira de  $\text{conv}(C)$  tal que

$$\hat{L} := \underset{L \in \text{conv}(C)}{\text{argmax}} \sum_{i=1}^m t_i \log L_i \quad (\text{II.7})$$

(Ver Teorema 18 em [32].)

### Existência e Discretização

Sendo  $C \subset \mathbb{R}^m$ , pelo Teorema de Caratheodory, qualquer  $L \in \partial(\text{conv}(C))$  pode ser representado como uma combinação convexa de no máximo  $m$  elementos de  $C$  (ver [46]).

Seja  $\hat{L}$  a solução ótima de (II.7). Portanto,  $\hat{L}$  pode ser expresso como uma combinação convexa de no máximo  $m$  elementos de  $C$  que, de acordo com o lema II.2, é igual a  $L(F_\Theta)$  onde  $F_\Theta \in \mathcal{F}_\Theta$  possui no máximo  $m$  pontos suportes. (Ver Teorema 3.1 em [31].)

Note que  $\hat{L}$  é único, porém podem existir mais de uma função  $\hat{F}_\Theta \in \mathcal{F}_\Theta$  tal que  $L(\hat{F}_\Theta) = \hat{L}$ . Os métodos clássicos da estatística encontram  $\hat{L}$  aproximadamente a partir de  $\hat{F}_\Theta$ . O primeiro método [29] fixa o suporte de tamanho  $m$  e  $\hat{F}_\Theta$  é expresso como uma combinação de  $m$  funções degraus. O segundo método, mais eficiente e que possui mais de uma variante, utiliza um algoritmo que a cada iteração adiciona uma ou mais funções degraus à aproximação de  $\hat{F}_\Theta$  e atualiza os coeficientes  $x_j$ ,  $j = 1, \dots, p$ ,  $0 < p \leq m$ , respectivamente em [3] e [30]. Observe que neste método, o tamanho do suporte é no máximo  $m$ .

O algoritmo que propomos no próximo capítulo segue a lógica do se-

gundo método estatístico, porém adotamos uma abordagem diferente para atualizar os coeficientes da combinação convexa (II.5), que na estatística são denotados por massas. Esta etapa envolve a resolução de um problema de maximização restrita. Nossa abordagem encontra uma solução aproximada através do dual do problema de maximização.

### Derivada Direcional

Dadas  $F_{\Theta}^0, F_{\Theta}^1 \in \mathcal{F}_{\Theta}$  com suporte finito, a derivada direcional de  $f$  do ponto  $L^0 = L(F_{\Theta}^0)$  para o ponto  $L^1 = L(F_{\Theta}^1)$  com  $L^0, L^1 \in \text{conv}(C)$  (ou seja, a derivada direcional de  $f(\cdot)$  na direção  $L^0 - L^1$ ) é definida por

$$\begin{aligned}
D(L(F_{\Theta}^0), L(F_{\Theta}^1)) &= \lim_{\varepsilon \rightarrow 0^+} \frac{f[(1-\varepsilon)L^0 + \varepsilon L^1] - f(L^0)}{\varepsilon} \\
&= \lim_{\varepsilon \rightarrow 0^+} \frac{\sum_{i=1}^m t_i \log [(1-\varepsilon)L_i^0 + \varepsilon L_i^1] - \sum_{i=1}^m t_i \log L_i^0}{\varepsilon} \\
&= \lim_{\varepsilon \rightarrow 0^+} \frac{1}{\varepsilon} \left\{ \sum_{i=1}^m t_i \log [(1-\varepsilon)L_i^0 + \varepsilon L_i^1] - \sum_{i=1}^m t_i \log L_i^0 \right\} \\
&= \lim_{\varepsilon \rightarrow 0^+} \frac{1}{\varepsilon} \left\{ \sum_{i=1}^m t_i \log \left[ \frac{(1-\varepsilon)L_i^0 + \varepsilon L_i^1}{L_i^0} \right] \right\} \\
&= \lim_{\varepsilon \rightarrow 0^+} \sum_{i=1}^m \frac{1}{\varepsilon} t_i \log \left[ (1-\varepsilon) + \varepsilon \frac{L_i^1}{L_i^0} \right] \\
&= \lim_{\varepsilon \rightarrow 0^+} \sum_{i=1}^m t_i \log \left[ 1 + \varepsilon \left( \frac{L_i^1}{L_i^0} - 1 \right) \right]^{1/\varepsilon} \\
&= \sum_{i=1}^m t_i \log \left\{ \lim_{\varepsilon \rightarrow 0^+} \left[ 1 + \varepsilon \left( \frac{L_i^1}{L_i^0} - 1 \right) \right]^{1/\varepsilon} \right\} \\
&= \sum_{i=1}^m t_i \log \left[ \exp \left( \frac{L_i^1}{L_i^0} - 1 \right) \right] \\
&= \sum_{i=1}^m t_i \left( \frac{L_i^1}{L_i^0} - 1 \right).
\end{aligned}$$

Dada uma distribuição  $F_{\Theta}$  com suporte finito com  $L(F_{\Theta}) \in \text{conv}(C)$ , sabemos calcular a derivada direcional de  $f$  ao longo de qualquer direção a partir de  $L(F_{\Theta})$ . Dado  $\bar{\theta}$ , uma direção de subida para  $f(\cdot)$  a partir de  $L(F_{\Theta})$  para um ponto  $L(U(\bar{\theta}))$  pode ser obtida quando  $D(L(F_{\Theta}), L(U(\bar{\theta}))) > 0$ . Escolhendo em particular



$l(\bar{\theta}) \in C$ , o que podemos fazer devido ao lema II.1, teremos uma direção de subida quando  $D(L(F_\Theta), l(\bar{\theta})) > 0$ , onde

$$D(L(F_\Theta), l(\bar{\theta})) = \sum_{i=1}^m t_i \left( \frac{l_i(\bar{\theta})}{L_i(F_\Theta)} - 1 \right).$$

Denotamos essas derivadas direcionais por

$$g(\bar{\theta}, F_\Theta) = D(L(F_\Theta), l(\bar{\theta})). \quad (\text{II.8})$$

Observe que se  $F_\Theta$  é a solução ótima então  $g(\theta, F_\Theta) = 0$  em qualquer direção para todo  $\theta \in \Gamma$ .

A seguir anunciamos o Teorema que caracteriza a solução ótima  $\hat{L} \in \text{conv}(C)$  e sua correspondente  $\hat{F}_\Theta \in \mathcal{F}_\Theta$  com suporte finito tal que  $\hat{L} = L(\hat{F}_\Theta)$ .

**Teorema II.1** *A. As três afirmativas abaixo são equivalentes:*

1.  $\hat{F}_\Theta$  maximiza  $\mathcal{L}(F_\Theta)$ ,
2.  $\hat{F}_\Theta$  minimiza o  $\sup_{\theta \in \Gamma} g(\theta, F_\Theta)$ ,
3.  $\sup_{\theta \in \Gamma} g(\theta, \hat{F}_\Theta) = 0$ .

*B. Seja  $\hat{F}_\Theta$  a solução ótima de (II.7). O suporte de  $\hat{F}_\Theta$  está contido no conjunto  $\{\theta : g(\theta, \hat{F}_\Theta) = 0\}$ .*

A demonstração deste teorema é feita por analogia ao Teorema 1 de Whittle [53] e se encontra em [32]. Outro resultado importante do Teorema acima, cuja prova encontra-se em [32], diz respeito ao intervalo onde estão os pontos suportes de  $\hat{F}_\Theta$ .

**Corolário II.2** *Sejam  $\theta \in \Gamma \subset \mathbb{R}$  e cada núcleo do estimador da verossimilhança  $l_i(\theta)$  unimodal em  $\theta$ . Então se  $\tilde{\theta}_i$  é a única moda de  $l_i(\theta)$  temos que todos os pontos suportes de  $\hat{F}_\Theta$  estão no intervalo*

$$I = [\min_i \tilde{\theta}_i, \max_i \tilde{\theta}_i].$$

# Capítulo III

## Modelagem do Problema

Conforme capítulo anterior, o problema EMV pode ser escrito como

$$(EMV) \quad \begin{array}{l} \text{maximizar} \quad \sum_{i=1}^m t_i \log L_i \\ L \in \text{conv}(C) \end{array}$$

### Notação

Dada uma discretização  $\theta_1, \theta_2, \dots, \theta_p$  de  $\Gamma$ , definimos para  $i = 1, \dots, m$  e  $j = 1, \dots, p$

$$\begin{aligned} l_{ij} &= l_i(\theta_j) = f_{V_i|\Theta}(v_i|\theta_j), \\ z_i &= L_i(F_\Theta) = \int_{\Gamma} l_i(\theta) dF_\Theta(\theta) = \int_{\Gamma} f_{V_i|\Theta}(v_i|\theta) dF_\Theta(\theta). \end{aligned}$$

Por definição de  $\text{conv}(C)$  (II.2), o problema acima fica

$$(EMV) \quad \begin{array}{l} \text{maximizar} \quad \sum_{i=1}^m t_i \log z_i \\ \text{sujeito a} \quad z_i = \sum_{j=1}^p l_{ij} x_j \\ \quad \quad \quad \sum_{j=1}^p x_j = 1 \\ \quad \quad \quad x_j \geq 0, \quad j = 1, \dots, p \\ \quad \quad \quad z_i > 0, \quad i = 1, \dots, m \\ \quad \quad \quad l_i \in C, \quad i = 1, \dots, m \\ \quad \quad \quad p \in \mathbb{N}. \end{array} \tag{III.1}$$

Se fosse possível conhecer a discretização ótima de  $\Gamma$ , encontraríamos a solução do problema EMV resolvendo o problema acima. Sabemos pelo capítulo anterior que existe uma solução ótima com  $p = m$ , mas não vamos procurá-la diretamente. Nosso algoritmo gera discretizações de tamanho  $p \geq m$  crescente, adicionando um ou mais novos pontos por iteração. Contudo, quando o algoritmo converge a cardinalidade

do conjunto suporte ótimo é no máximo  $m$ . Veremos adiante que a adição de um ponto à discretização corresponde à adição de um corte em um polítopo.

A idéia geral do algoritmo que resolve o problema EMV, que veremos mais adiante neste capítulo, é resolver uma sucessão de problemas definidos a partir de discretizações de tamanhos crescentes até encontrar uma solução do problema original.

Fixando uma discretização  $\theta_1, \theta_2, \dots, \theta_p$  de  $\Gamma$ , define-se o seguinte problema interno de EMV

$$\begin{array}{ll} \text{maximizar} & \sum_{i=1}^m t_i \log z_i \\ \text{sujeito a} & Ax = z \\ & e^T x = 1 \\ & x \geq 0 \\ & z > 0, \end{array}$$

onde  $A \in \mathbb{R}^{m \times p}$  é a matriz cujos elementos são  $l_{ij} = f_{v_i|\Theta}(\theta_j)$ , para  $i = 1, \dots, m$  e  $j = 1, \dots, p$ ,  $t_i$  é a multiplicidade de cada observação distinta da amostra, para  $i = 1, \dots, m$ ,  $x \in \mathbb{R}^p$  e  $z \in \mathbb{R}^m$ . De acordo com a definição de probabilidade, temos que todos os elementos da matriz  $A$  são não negativos e menores ou iguais a 1.

### III.1 O problema interno

Nesta seção vamos estudar o problema interno. O algoritmo que veremos na próxima seção, obtém a cada iteração uma solução aproximada de um problema interno.

Deste modo, a cada iteração do algoritmo, o problema de otimização é resolver um problema interno de EMV apresentado no formato primal por

$$\begin{array}{ll} (P^k) \text{ maximizar} & \sum_{i=1}^m t_i \log z_i \\ \text{sujeito a} & A^k x = z \\ & e^T x = 1 \\ & x \geq 0 \\ & z > 0, \end{array}$$

onde  $A^k \in \mathbb{R}^{m \times p}$  é conhecida na iteração  $k$  assim como o vetor  $t \in \mathbb{R}^m$ , cujas componentes são as multiplicidades das observações distintas da amostra, tal que

$\sum_{i=1}^m t_i = n$ , conforme apresentado no capítulo anterior. As variáveis primais são  $x \in \mathbb{R}^p$  e  $z \in \mathbb{R}^m$ . E  $p$  é a cardinalidade do conjunto suporte de  $F_{\Theta}^k$  na iteração  $k$ . Vamos denotar o conjunto dos pontos viáveis do problema  $P^k$  por

$$\mathcal{F}_p^k = \{(x, z) \in \mathbb{R}^{p+m} : A^k x = z, e^T x = 1, x \geq 0, z > 0\}$$

e o conjunto viável interior de  $P^k$  por

$$\overset{\circ}{\mathcal{F}}_p^k = \{(x, z) \in \mathcal{F}_p^k : x > 0\}.$$

Note que em algoritmos de pontos interiores a definição de conjunto viável interior refere-se ao conjunto de pontos interiores do ortante positivo do conjunto viável e não a definição topológica de conjunto interior.

Vamos destacar as primeiras hipóteses para o problema de otimização:  $A^k$  possui rank máximo e  $\overset{\circ}{\mathcal{F}}_p^k$  é não vazio.

## O Formato Dual

No contexto da Teoria da Dualidade Convexa, é possível definir um problema equivalente ao formato primal do problema interno de EMV através de sua formulação dual, que anunciamos a seguir.

Antes de prosseguirmos vamos considerar mais uma hipótese: *o interior do conjunto viável dual é não-vazio.*

**Lema III.1** *Seja  $P^k$  o problema interno de EMV na iteração  $k$  com conjunto viável interior não-vazio. Considere o interior do conjunto viável dual não-vazio. O formato dual de  $P^k$  é*

$$(D^k) \quad \begin{array}{l} \text{maximizar} \quad \sum_{i=1}^m t_i \log y_i \\ \text{sujeito a} \quad (A^k)^T y + s = c^k \\ \quad \quad \quad s \geq 0 \\ \quad \quad \quad y > 0, \end{array}$$

onde  $c^k = (\sum_{i=1}^m t_i) e$ , e as variáveis duais são  $y \in \mathbb{R}^m$  e  $s \in \mathbb{R}^p$ . Na iteração  $k$ , são conhecidos  $t \in \mathbb{R}^m$ ,  $c^k \in \mathbb{R}^p$  e  $A^k \in \mathbb{R}^{m \times p}$ .

*Prova.* Vamos construir  $D^k$  segundo a formulação dual de Wolfe [33].

Considere o seguinte problema de programação convexa com restrições lineares

$$(P) \quad \begin{array}{ll} \text{minimizar} & f(x) \\ \text{sujeito a} & Ax = b \\ & x \geq 0, \end{array}$$

onde  $f : \mathbb{R}_+^n \rightarrow \mathbb{R}$  é uma função convexa diferenciável e  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ ,  $x \in \mathbb{R}^n$ . Então a formulação dual do problema acima, também chamada de dual de Wolfe, se apresenta como

$$(D) \quad \begin{array}{ll} \text{maximizar} & b^T y - (\nabla f(x))^T x + f(x) \\ \text{sujeito a} & A^T y - \nabla f(x) + s = 0 \\ & x > 0 \\ & s \geq 0, \end{array}$$

onde  $y \in \mathbb{R}^m$  é livre e  $s \in \mathbb{R}^n$  é o vetor das folgas duais ou vetor dos multiplicadores de Lagrange.

Vamos colocar o problema  $P^k$  num formato apresentado como em  $P$ , fazendo

$$(\bar{P}^k) \quad \begin{array}{ll} \text{minimizar} & \bar{f}(\bar{x}) \\ \text{sujeito a} & \bar{A}\bar{x} = \bar{b} \\ & \bar{x} \geq 0, \end{array}$$

onde  $\bar{x}$  é o vetor das variáveis primais dado por

$$\bar{x} = \begin{bmatrix} x \\ z \end{bmatrix} \in \mathbb{R}^{(p+m)},$$

a matriz  $\bar{A}$  dos coeficientes das variáveis primais das restrições de igualdade é representada por

$$\bar{A} = \begin{bmatrix} A^k & -I \\ e^T & 0^T \end{bmatrix} \in \mathbb{R}^{(m+1) \times (p+m)},$$

o vetor  $\bar{b}$  é o vetor dos termos independentes das restrições de igualdade

$$\bar{b} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \in \mathbb{R}^{(m+1)}$$

e

$$\bar{f}(\bar{x}) = -f(x, z) = -\sum_{i=1}^m t_i \log z_i, \quad \text{com} \quad \nabla \bar{f}(\bar{x}) = \begin{bmatrix} 0 \\ -Z^{-1}t \end{bmatrix} \in \mathbb{R}^{(p+m)}.$$

Assim posto, o formato dual de  $\bar{P}^k$  é

$$\begin{aligned}
 (\bar{D}^k) \quad & \text{maximizar} \quad \bar{b}^T \bar{y} - (\nabla \bar{f}(\bar{x}))^T \bar{x} + \bar{f}(\bar{x}) \\
 & \text{sujeito a} \quad \bar{A}^T \bar{y} - \nabla \bar{f}(\bar{x}) + \bar{s} = 0 \\
 & \quad \bar{x} > 0 \\
 & \quad \bar{s} \geq 0,
 \end{aligned}$$

onde as variáveis duais  $\bar{y}$  e  $\bar{s}$  são, conforme a estrutura e dimensão do problema  $\bar{P}$ , respectivamente,

$$\bar{y} = \begin{bmatrix} y \\ \tilde{y} \end{bmatrix} \in \mathbb{R}^{(m+1)} \quad \text{e} \quad \bar{s} = \begin{bmatrix} s \\ \tilde{s} \end{bmatrix} \in \mathbb{R}^{(p+m)}.$$

Porém, vamos reescrever  $\bar{s}$  como

$$\bar{s} = \begin{bmatrix} s \\ 0 \end{bmatrix} \in \mathbb{R}^{(p+m)},$$

pois sabemos que os multiplicadores de Lagrange associados à restrição  $z > 0$  em  $P^k$  são nulos. Então, substituindo as constantes e variáveis e efetuando os produtos internos, temos para o novo formato dual de  $P^k$ :

$$(D^k) \quad \text{maximizar} \quad \tilde{y} + \sum_{i=1}^m t_i - \sum_{i=1}^m t_i \log z_i \tag{III.2}$$

$$\text{sujeito a} \quad (A^k)^T y + \tilde{y} e + s = 0$$

$$-y + Z^{-1} t = 0 \tag{III.3}$$

$$s \geq 0$$

$$z > 0,$$

onde  $y$  e  $\tilde{y}$  são livres. Por (III.3) sabemos que  $z_i = t_i/y_i$ , onde  $y_i > 0$ ,  $\forall i = 1, \dots, m$ .

Assim, eliminando a variável  $z$  no formato acima, ficamos com

$$\begin{aligned}
 (D^k) \quad & \text{maximizar} \quad \tilde{y} + \sum_{i=1}^m t_i - \sum_{i=1}^m t_i \log t_i + \sum_{i=1}^m t_i \log y_i \\
 & \text{sujeito a} \quad (A^k)^T y + \tilde{y} e + s = 0 \\
 & \quad s \geq 0 \\
 & \quad y > 0.
 \end{aligned}$$

Seja o conjunto viável dual interior denotado por  $\overset{o}{\mathcal{F}}_d^k$ . Por hipótese,  $\overset{o}{\mathcal{F}}_d^k$  é não vazio. Logo, pelo Teorema da Dualidade Forte [33], temos que os valores das funções objetivo nas soluções ótimas dos problemas primal e dual são iguais.

Observando a função objetivo em (III.2) e comparando-a com a formulação primal  $P^k$ , podemos fazer

$$\tilde{y} = - \sum_{i=1}^m t_i.$$

Com isto, chegamos à formulação como no enunciado deste lema. ■

Explicitamos o conjunto viável dual por

$$\mathcal{F}_d^k = \{(y, s) \in \mathbb{R}^{m+p} : s = c^k - (A^k)^T y \geq 0, \ y > 0\}$$

e o conjunto viável interior de  $D^k$  por

$$\overset{\circ}{\mathcal{F}}_d^k = \{(y, s) \in \mathcal{F}_d^k : s > 0\}.$$

Para relacionar os problemas de otimização  $P^k$  e  $D^k$ , sejam  $(x, z)$  uma solução primal viável e  $(y, s)$  uma solução dual viável. Define-se por gap de complementaridade o escalar  $x^T s$ . Segue-se o fato que

$$x^T s = c^T x - y^T z \geq 0. \tag{III.4}$$

Isto sai direto por substituição:  $x^T s = x^T (c - A^T y) = c^T x - (Ax)^T y = c^T x - y^T z \geq 0$ .

A hipótese de que  $A^k$  possui rank máximo implica na existência de uma correspondência biunívoca entre  $y$  e a folga dual  $s$  e outra correspondência biunívoca entre as variáveis primais  $x$  e  $z$ . Consequentemente, o conjunto equivalente a  $\mathcal{F}_d^k$  é o poliedro

$$\Omega^k = \{y \in \mathbb{R}_{++}^m : c^k - (A^k)^T y \geq 0\}.$$

Seja a quarta hipótese: *o conjunto  $\mathcal{F}_d^k$  é limitado.*

Então  $\Omega^k$  é um politopo no  $\mathbb{R}^m$ . Isto é, o conjunto viável do problema  $D^k$  é um politopo definido por  $p$  hiperplanos. Na figura (III.1) apresentamos um exemplo de um politopo definido por 7 cortes no  $\mathbb{R}^2$ .

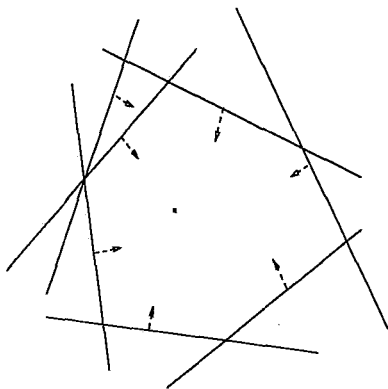


Figura III.1: Um polítopo no  $\mathbb{R}^2$  definido por 7 cortes

### Condições de Otimalidade

Sendo  $P^k$  um problema côncavo, temos que as condições de otimalidade de primeira ordem são necessárias e suficientes. Estas condições são apresentadas pelo Teorema de Karush-Kuhn-Tucker [23, 28] (aqui colocado de forma abreviada):  $(x, z)$  é um ponto de máximo se e somente se existe um vetor  $(y, s)$  satisfazendo o sistema:

$$\begin{aligned} (i) \quad & Sx = 0, \\ (ii) \quad & Yz = t, \\ (iii) \quad & A^k x - z = 0, \\ (iv) \quad & (A^k)^T y + s = ne, \\ & x \geq 0, \quad s \geq 0, \quad y > 0, \quad z > 0, \end{aligned}$$

onde  $\sum_{i=1}^m t_i = n$ . Tal sistema é muitas vezes chamado de sistema KKT, ele reúne as condições de viabilidade primal (iii) e dual (iv) e as condições de otimalidade dadas pela condição de complementaridade linear (i) e (ii). A equação (i) pode ser escrita por  $x^T s = 0$ , porque as componentes de ambos vetores são não-negativas. Outro fato importante diz respeito à relação combinatorial entre as variáveis  $x$  e  $s$ , pois para cada  $i$ ,  $i = 1, \dots, p$ ,  $x_i = 0$  ou  $s_i = 0$ . Observamos que a condição de viabilidade primal dada por  $e^T x = 1$ , não está explicitada acima, porém ela pode ser gerada a partir das equações do sistema KKT:

$$\begin{aligned} 0 = x^T s &= x^T (ne - (A^k)^T y) = (ne)^T x - (A^k x)^T y \\ &= ne^T x - y^T z = ne^T x - n = n(e^T x - 1). \end{aligned}$$

Como  $n \neq 0$ , temos  $e^T x = 1$ .



O problema primal-dual interno,  $(PD)^k$ , se coloca como sendo o problema que encontra uma solução ótima para o sistema acima. (Note que a existência de uma solução ótima é garantida pelas hipóteses feitas anteriormente.) Definimos o conjunto dos pontos primais-duais viáveis por

$$\mathcal{F}^k = \{(x, s) \in \mathbb{R}^{2p} : A^k x = z \text{ para algum } z \in \mathbb{R}_{++}^m, e^T x = 1, x \geq 0 \text{ e} \\ s = c^k - (A^k)^T y \text{ para algum } y \in \mathbb{R}_{++}^m, s \geq 0\},$$

deste modo, seu conjunto de pontos interiores é,

$$\overset{\circ}{\mathcal{F}}^k = \{(x, s) \in \mathcal{F}^k : x > 0 \text{ e } s > 0\}.$$

## III.2 O Algoritmo de Otimização

A seguir apresentamos um modelo do algoritmo de otimização que encontra uma solução ótima para o problema EMV. Cada iteração do algoritmo gera um ponto correspondente a um problema interno  $P^k$ .

### Algoritmo 1 – Algoritmo Principal

**Dados** uma discretização finita inicial de  $\Gamma$ ,  $P^0$ ,  $D^0$  e  $(x^0, s^0, y^0, z^0) > 0$ ;

**Faça:**  $k = 0$ ;

**Enquanto** não converge, **faça:**

•consulte o Oráculo, o qual responde:

–melhore o modelo introduzindo um novo ponto à discretização, atualize  $P^{k+1}$  e  $D^{k+1}$  e obtenha um novo ponto  $(x^{k+1}, s^{k+1}, y^{k+1}, z^{k+1})$  correspondente a  $P^{k+1}$  e  $D^{k+1}$ ;

–o modelo não pode ser melhorado, neste caso obtenha um novo ponto  $(x^{k+1}, s^{k+1}, y^{k+1}, z^{k+1})$  correspondente a  $P^k$  e  $D^k$ ;

•faça:  $k = k + 1$ ;

**Fim**

**Fim**

A seguir vamos estudar a definição de Oráculo, como introduzir um novo ponto à uma discretização existente e a implicação deste fato na definição de um novo polítopo  $\Omega^{k+1}$ .

No próximo capítulo, estudaremos um algoritmo que encontra um ponto primal-dual correspondente ao problema interno  $P^k$ . Vamos também estudar como obter um ponto primal-dual após a introdução de um ponto à discretização. O teste de convergência do algoritmo acima será apresentado formalmente no capítulo 6, quando tratamos da implementação do algoritmo.

### III.3 Introdução de um corte profundo: o Oráculo

#### Definição do Oráculo

Dado um conjunto e um ponto, toda vez que o Oráculo for consultado pelo ponto, ele responde com um corte se este for do tipo profundo. O corte ou hiperplano fornecido pelo Oráculo define um semi-espaço que contém o conjunto.

#### Definição de Corte Profundo

Considere  $\theta_1, \theta_2, \dots, \theta_p$  a discretização de  $\Gamma$  na iteração  $k$ . Dados  $P^k, D^k$  e um ponto  $(x^k, s^k, y^k, z^k) \in \overset{\circ}{\mathcal{F}}^k$ , é possível obter uma estimativa de  $F_{\Theta}$  na iteração  $k$  fazendo

$$F_{\Theta}^k(\cdot) = \sum_{j=1}^p x_j^k U(\theta_j)(\cdot),$$

onde  $U(\theta_j)(\cdot)$  é a função degrau,  $\sum_{j=1}^p x_j^k = 1$  e  $x_j^k \geq 0$ , com  $j = 1, \dots, p$ .

Nosso objetivo é adicionar mais um ponto à discretização existente, isto é, queremos adicionar a  $F_{\Theta}^k$  uma nova função degrau  $U(\theta_{p+1})$  tal que

$$F_{\Theta}^{k+1}(\cdot) = F_{\Theta}^k(\cdot) + x_{p+1} U(\theta_{p+1}).$$

Podemos interpretar isto como seguir uma direção de subida da função objetivo de EMV a partir de  $L(F_{\Theta}^k) \in \text{conv}(C)$  para um ponto  $l(\theta_{p+1}) \in C$ , como vimos no final do capítulo anterior.

Como antes (II.8), vamos denotar a direção de subida por  $g(\theta_{p+1}, F_{\Theta}^k)$  e desenvolver sua expressão:

$$\begin{aligned} g(\theta_{p+1}, F_{\Theta}^k) &= \sum_{i=1}^m t_i \left( \frac{l_i(\theta_{p+1})}{L_i(F_{\Theta}^k)} - 1 \right) \\ &= \sum_{i=1}^m t_i \left( \frac{l_i(\theta_{p+1})}{z_i^k} - 1 \right) \\ &= \sum_{i=1}^m \frac{t_i l_i(\theta_{p+1})}{z_i^k} - \sum_{i=1}^m t_i \\ &= \sum_{i=1}^m l_i(\theta_{p+1}) y_i^k - n \\ &= l(\theta_{p+1})^T y^k - n. \end{aligned}$$

Como  $g(\theta_{p+1}, F_{\Theta}^k) > 0$ , temos  $l(\theta_{p+1})^T y^k - n > 0$ . Para que esta desigualdade origine um corte profundo do tipo

$$l(\theta_{p+1})^T y \leq n, \quad (\text{III.5})$$

devemos escolher  $\theta_{p+1}$  tal que a diferença  $l(\theta_{p+1})^T y^k - n$  seja a maior possível. Então devemos escolher  $\theta_{p+1}$  tal que  $\theta_{p+1}$  seja um máximo local de  $g(\cdot, F_{\Theta}^k)$ .

Por outro lado, ao mesmo tempo que adicionamos um ponto  $\theta_{p+1}$  à discretização, de acordo com o nosso modelo (III.1), estamos adicionando uma variável primal  $x_{p+1}$  ao problema interno  $P^k$ , obtendo um novo problema interno  $P^{k+1}$ . Equivalentemente, quando introduzimos  $\theta_{p+1}$  à discretização existente de  $\Gamma$ , adicionamos um corte como (III.5) a  $\Omega^k$ , obtendo um novo polítopo  $\Omega^{k+1}$ .

Concluindo, quando o Oráculo for consultado pelo ponto  $y^k$ , ele responderá com um corte profundo caso exista um máximo local de  $g(\cdot, F_{\Theta}^k)$  tal que o valor de  $g(\cdot, F_{\Theta}^k)$  no máximo local seja positivo. Seja  $\theta_{p+1}$  um máximo local de  $g(\cdot, F_{\Theta}^k)$ . Então o corte profundo dado pelo Oráculo é um corte igual a (III.5), onde  $l(\theta_{p+1}) = (l_1(\theta_{p+1}), \dots, l_m(\theta_{p+1}))$ .

No caso de existir mais de um máximo local, então mais de um corte

será feito em  $\Omega^k$ . Considere que as funções núcleos (funções de densidade de probabilidade) sejam funções do tipo exponencial. Então o número de cortes fornecidos pelo Oráculo por iteração será finito, pois a função  $g(\cdot, F_{\Theta}^k)$  apresenta um número finito de máximos locais, Lema 8.2 em [31].

# Capítulo IV

## Otimização

Seja a formulação primal do problema interno de *EMV* na iteração  $k$  dada por

$$\begin{aligned} (P^k) \quad & \text{maximizar} \quad \sum_{i=1}^m t_i \log z_i \\ & \text{sujeito a} \quad A^k x = z \\ & \quad \quad \quad e^T x = 1 \\ & \quad \quad \quad x \geq 0 \\ & \quad \quad \quad z > 0. \end{aligned}$$

De acordo com as hipóteses do capítulo anterior, o problema de programação não-linear acima possui uma solução ótima.

Primeiramente neste capítulo, vamos estudar um algoritmo que resolve  $P^k$  a partir de um ponto inicial  $(x^k, s^k, y^k, z^k) > 0$  dado, pois queremos obter um ponto correspondente a  $P^k$ . Em nosso algoritmo não faremos uma resolução completa de  $P^k$ : obteremos somente uma solução aproximada do problema.

Mais adiante neste capítulo, conforme o Algoritmo Principal, estudaremos como obter um ponto primal-dual no novo politopo originado pela introdução de um ou mais cortes.

E no final do capítulo, vamos estudar como obter um ponto inicial apropriado para o Algoritmo Principal após a introdução de um ou mais cortes, visto que o ponto atual neste passo do Algoritmo se torna inviável após a introdução de um ou mais cortes.

## IV.1 Introdução a um Algoritmo de Pontos Interiores

Baseado no Teorema da Dualidade Forte [52], vamos encontrar a solução ótima de  $P^k$  resolvendo  $D^k$ .

O algoritmo escolhido para resolver  $D^k$  é um algoritmo de pontos interiores. Algoritmos de pontos interiores estão fundamentados em seguir aproximadamente uma trajetória no interior do conjunto viável do problema de otimização. A trajetória é obtida através de uma parametrização quando aplicamos a metodologia da barreira logarítmica ao problema de otimização.

### Função barreira

Define-se por função barreira logarítmica primal a função  $\rho : \mathbb{R}_{++}^p \rightarrow \mathbb{R}$  tal que para  $x \in \mathbb{R}_{++}^p$ ,  $\rho(x) = \sum_{j=1}^p \log x_j$ . Ela apresenta como propriedades ser estritamente côncava e decrescer indefinidamente nos pontos perto da fronteira de seu domínio.

Sejam o conjunto viável dual dado pelo polítopo

$$\Omega^k = \{y \in \mathbb{R}_{++}^m : c_j^k - (a_j^k)^T y \geq 0, j = 1, \dots, p\}, \quad (\text{IV.1})$$

onde  $c^k = ne$  e  $\overset{\circ}{\Omega}^k$  o seu conjunto interior. Considere a função barreira logarítmica dual  $\phi : \mathbb{R}_{++}^m \rightarrow \mathbb{R}$  assim definida: dado  $y \in \overset{\circ}{\Omega}^k$ ,

$$\phi(y) = \sum_{j=1}^p \log s_j,$$

onde  $s_j := c_j^k - (a_j^k)^T y$  é a folga da  $j$ -ésima desigualdade linear na iteração  $k$ . Note que  $\phi$  é uma função estritamente côncava em  $\overset{\circ}{\Omega}^k$  e também decresce indefinidamente nos pontos perto da fronteira do conjunto viável dual.

## Centro Analítico

O centro analítico de um politopo definido por um número finito de desigualdades lineares foi introduzido por Huard [20] e mais tarde por Sonnevend [47].

Seja  $\Omega^k$  limitado e com interior não-vazio, por hipótese, definimos

$$y^a := \operatorname{argmax}\{\phi(y) : y \in \overset{\circ}{\Omega}^k\} \quad (\text{IV.2})$$

e denotamos  $y^a$  como sendo o centro analítico do conjunto  $\Omega^k$ .

Veja a ilustração de  $y^a$  na figura (IV.1). Das condições necessárias e suficientes de otimalidade para  $y^a$  ser um centro analítico de  $\Omega^k$ , isto é  $\nabla\phi(y^a) = 0$ , temos que existe  $x^a > 0$  tal que

$$X^a s^a = e, \quad Ax^a = 0 \quad \text{e} \quad s^a = c - A^T y^a > 0,$$

onde  $c = ne$ .

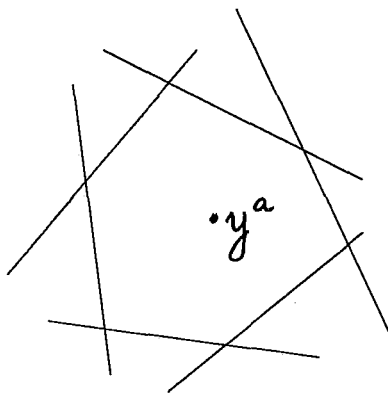


Figura IV.1:  $y^a$  centro analítico de  $\Omega^k$

## Função penalizada

Dado o problema  $D^k$ , o método da barreira logarítmica substitui as restrições de não-negatividade das variáveis por termos da barreira logarítmica na função-objetivo,

determinando um novo problema de otimização para cada  $\mu \in (0, +\infty)$ :

$$(D_\mu^k) \begin{array}{l} \text{maximizar} \quad \sum_{i=1}^m t_i \log y_i + \mu \sum_{j=1}^p \log s_j \\ \text{sujeito a} \quad (A^k)^T y + s = ne \\ \quad \quad \quad s \geq 0 \\ \quad \quad \quad y > 0, \end{array}$$

onde  $\mu$  é o parâmetro da barreira.

Denotamos por função penalizada dual a função-objetivo acima:

$$f_\mu^d(y, s) = \sum_{i=1}^m t_i \log y_i + \mu \sum_{j=1}^p \log s_j.$$

Analogamente, define-se por função penalizada primal a função

$$f_\mu^p(x, z) = \sum_{i=1}^m t_i \log z_i + \mu \sum_{j=1}^p \log x_j,$$

onde  $\sum_{i=1}^m t_i \log z_i$  é a função-objetivo do problema interno primal  $P^k$ .

A idéia geral do algoritmo que resolve  $D^k$  é resolver uma sequência de problemas  $D_\mu^k$  a cada parâmetro fixo  $\mu$ . Note que à medida que  $\mu$  tende a zero, a solução do problema  $D_\mu^k$  fica cada vez mais próxima da solução de  $D^k$ .

Para a nova formulação dual  $D_\mu^k$ , as condições necessárias e suficientes de otimalidade de primeira ordem para  $\mu$  fixo são apresentadas pelo sistema:

$$\begin{array}{rcl} Xs & = & \mu e, \\ Yz & = & t, \\ A^k x - z & = & 0, \\ (A^k)^T y + s & = & ne \\ x \geq 0, \quad s \geq 0, \quad y > 0, \quad z > 0. \end{array} \tag{IV.3}$$

Tal sistema é muitas vezes chamado de sistema KKT perturbado devido à presença do parâmetro da barreira logarítmica na primeira equação.

Observa-se que a restrição  $e^T x = 1 + \gamma\mu$ ,  $\gamma > 0$  fixo, nos é dada implicitamente a partir das equações acima:

$$\begin{aligned} p\mu &= x^T s = x^T (ne - (A^k)^T y) = (ne)^T x - (A^k x)^T y \\ &= ne^T x - y^T z = ne^T x - n = n(e^T x - 1). \end{aligned}$$



Como  $n \neq 0$  e  $p$  são dados, temos  $e^T x - 1 = (p/n)\mu$ . Seja  $\gamma = p/n$  fixo, com  $\gamma > 0$ , então  $e^T x = 1 + \gamma\mu$ . No nosso algoritmo,  $p$  é um número finito maior ou igual a  $m$ , como veremos no capítulo VI, onde abordamos a implementação do algoritmo.

Um fato importante surge quando aplicamos o método da barreira logarítmica ao problema interno de EMV na sua formulação dual: o conjunto viável dual é limitado enquanto que o conjunto viável primal é ilimitado, como podemos verificar pela equação  $e^T x = 1 + \gamma\mu$ , com  $\gamma > 0$  fixo, resultante do sistema KKT perturbado acima. Note que a restrição  $e^T x = 1$  não é satisfeita quando  $\mu > 0$ , o que nos leva a concluir que a variável primal  $x$  que satisfaz o sistema (IV.3) não é primal viável. Porém quando  $\mu \rightarrow 0$ , ela será satisfeita.

## IV.2 Trajetória Central

Seja  $c = ne$ . Considere o sistema formado pelas equações de igualdade do sistema KKT perturbado (IV.3):

$$\begin{cases} Xs = \mu c \\ Yz = t \\ Ax - z = 0 \\ s + A^T y = c \end{cases}$$

A matriz Jacobiana do sistema acima em relação ao ponto  $(x, s, y, z)$  é

$$J(x, s, y, z) = \begin{pmatrix} S & X & 0 & 0 \\ 0 & 0 & Z & Y \\ A & 0 & 0 & -I \\ 0 & I & A^T & 0 \end{pmatrix}.$$

**Proposição IV.1** *Seja  $A$  uma matriz com rank máximo e  $(x, s, y, z) > 0$ . A matriz  $J$  acima é não singular.*

*Demonstração.* Considere o ponto  $(p, q, u, v)$  no espaço nulo da matriz Jacobiana, isto é,

$$\begin{pmatrix} S & X & 0 & 0 \\ 0 & 0 & Z & Y \\ A & 0 & 0 & -I \\ 0 & I & A^T & 0 \end{pmatrix} \begin{pmatrix} p \\ q \\ u \\ v \end{pmatrix} = 0 \leftrightarrow \begin{cases} Sp + Xq = 0 & (1) \\ Zu + Yv = 0 & (2) \\ Ap - v = 0 & (3) \\ q + A^T u = 0 & (4) \end{cases}$$

De (3), temos  $Ap = v$ . Multiplicando por  $p^T$  em (4), temos

$$p^T q - p^T A^T u = p^T q - v^T u = 0. \quad (\text{IV.4})$$

De (1), tem-se  $p = -S^{-1}Xq$  ou  $q = -X^{-1}Sp$ . De (2), tem-se  $u = -Z^{-1}Yv$  ou  $v = -Y^{-1}Zu$ . Substituindo  $p$  ou  $q$  e  $u$  ou  $v$  em (IV.4), temos:

$$p^T X^{-1}Sp = v^T Z^{-1}Yv = u^T Y^{-1}Zu = q^T S^{-1}Xq. \quad (\text{IV.5})$$

Por outro lado, substituindo  $v$  e  $q$  dados pelas equações (3) e (4) nas equações (1) e (2), obtemos o sistema reduzido:

$$\begin{cases} Sp - XA^T u = 0 & (1') \\ Zu + YAp = 0 & (2') \end{cases}$$

De (1'),  $p = S^{-1}XA^T u$ . Substituindo  $p$  em (2'):  $Zu + YAS^{-1}XA^T u = Y^{-1}Zu + AS^{-1}XA^T u = (Y^{-1}Z + AS^{-1}XA^T)u = 0$ . Por hipótese,  $Z$  e  $Y$  são matrizes definidas positivas assim como  $AS^{-1}XA^T$ , pois  $A$  possui rank máximo. Logo a matriz  $Y^{-1}Z + AS^{-1}XA^T$  é definida positiva. Daí  $u^T(Y^{-1}Z + AS^{-1}XA^T)u = 0 \Leftrightarrow u = 0$ . Portanto, a igualdade em (IV.5) nos leva a concluir que  $(p, q, u, v) = 0$ . Isto é,  $J$  é não singular. ■

Dado um valor para  $\mu \in (0, +\infty)$ , definimos por ponto central dual associado a  $\mu$  o ponto

$$(y(\mu), s(\mu)) := \operatorname{argmax}\{f_\mu^d(y, s) : (y, s) \in \mathcal{F}_d^k\}.$$

A curva definida por  $\mu \in (0, +\infty) \mapsto (y(\mu), s(\mu))$  é chamada trajetória central dual. Veja na figura (IV.2) a ilustração da trajetória central dual definida num conjunto viável dual.

Analogamente, dado um valor para  $\mu \in (0, +\infty)$ , definimos por ponto central primal associado a  $\mu$  o ponto

$$(x(\mu), z(\mu)) := \operatorname{argmax}\{f_\mu^p(x, z) : (x, z) \in \mathcal{F}_{p, \mu}^k\}, \quad (\text{IV.6})$$

onde  $\mathcal{F}_{p, \mu}^k = \{(x, z) \in \mathbb{R}^{p+m} : A^k x = z, e^T x = 1 + \gamma\mu, x > 0, z > 0, \gamma > 0 \text{ fixo}\}$ . A curva definida por  $\mu \in (0, +\infty) \mapsto (x(\mu), z(\mu))$  é chamada trajetória central primal.

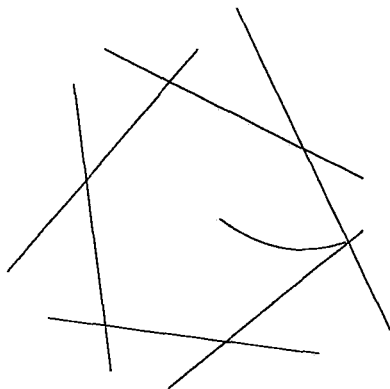


Figura IV.2: Exemplo de trajetória central dual.

**Lema IV.1** O vetor  $(x, z)$  é o ponto central primal associado a  $\mu$  na iteração  $k$  se e somente se  $(y = tZ^{-1}e, s = \mu X^{-1}e)$  é o ponto central dual associado a  $\mu$  na mesma iteração. Neste caso, o gap de complementaridade é dado por  $x^T s = p\mu$  e  $y^T z = n$ .

*Demonstração.* Considere  $(x, z) = (x(\mu), z(\mu))$  para um dado valor de  $\mu > 0$ . Então  $(x, z) \in \overset{\circ}{\mathcal{F}}_{p, \mu}^k$  e

$$\nabla f_{\mu}^p(x, z(x)) = \tilde{y}e,$$

onde  $\tilde{y}$  é o multiplicador de Lagrange associado a restrição  $e^T x = 1 + \gamma\mu$ ,  $\gamma > 0$ . Substituindo o gradiente na equação acima, temos  $(A^k)^T y + s = \tilde{y}e$ . Das condições de otimalidade para que  $(x, z)$  seja um ponto central primal na iteração  $k$ , obtém-se que  $\tilde{y} = n$ . Logo  $s = ne - A^T y$  com  $s > 0$  e  $y > 0$ , isto é,  $(y, s) \in \overset{\circ}{\mathcal{F}}_d^k$ .

No sentido contrário, considere agora  $(y, s) = (y(\mu), z(\mu))$  para um dado valor de  $\mu > 0$ . Então  $(y, s) \in \overset{\circ}{\mathcal{F}}_d^k$  e

$$\nabla f_{\mu}^d(y, s(y)) = Y^{-1}t - A\mu S^{-1}e = 0.$$

Substituindo  $z = Y^{-1}t$  e  $x = \mu S^{-1}e$  na equação acima, temos  $Ax - z = 0$ . E ainda das condições de otimalidade para que  $(y, s)$  seja um ponto central dual, obtemos  $e^T x = 1 + \gamma\mu$ ,  $\gamma > 0$ . Como  $z > 0$  e  $x > 0$ , então  $(x, z) \in \overset{\circ}{\mathcal{F}}_{p, \mu}^k$ .

Para  $(x, z)$  e  $(y, s)$  como acima, tomando as equações  $Xs = \mu e$  e  $Yz = t$  e somando as componentes das equações, obtemos respectivamente  $x^T s = p\mu$  e  $y^T z = n$ . ■

Dado  $\mu$ , diz-se que  $(x(\mu), s(\mu), y(\mu), z(\mu))$  é um ponto central primal-dual se satisfaz o sistema KKT perturbado. Ou seja, o ponto  $(x, s, y, z) \in \mathbb{R}^{2(p+m)}$  é o ponto central primal-dual associado a  $\mu > 0$  se e somente se  $(x, z) \in \mathcal{F}_{p,\mu}$ ,  $(s, y) \in \mathcal{F}_d$ ,  $Xs = \mu e$  e  $Yz = t$ .

A trajetória central primal-dual está definida pela parametrização:

$$\mu \in (0, +\infty) \mapsto (x(\mu), s(\mu), y(\mu), z(\mu)).$$

**Teorema IV.2** *Seja  $(x(\mu), s(\mu), y(\mu), z(\mu))$  na trajetória central.*

1. Dado  $\mu^0 > 0$ , o par central  $(x(\mu), s(\mu))$  é limitado para  $0 < \mu \leq \mu^0$ .
2. Para  $0 < \mu' < \mu$ ,  $\sum_{i=1}^m t_i \log z(\mu')_i > \sum_{i=1}^m t_i \log z(\mu)_i$  e  $\sum_{i=1}^m t_i \log y(\mu')_i > \sum_{i=1}^m t_i \log y(\mu)_i$ .
3.  $(x(\mu), s(\mu))$  tende para o par de solução ótima para  $P^k$  e  $D^k$  quando  $\mu$  tende para zero.

*Demonstração.* Dado  $\mu^0 > 0$ , pelas condições de otimalidade de primeira ordem apresentadas pelo sistema (IV.3), para a variável primal  $x$ ,  $0 \leq x_j \leq 1 + \gamma\mu^0$  para  $j = 1, \dots, p$  e  $0 < \gamma$  fixo. Em particular, para  $\mu < \mu^0$ ,  $0 \leq x_j(\mu) \leq 1 + \gamma\mu^0$ . Sabendo que os elementos da matriz  $A^k$  são não negativos e menores ou iguais a 1, e que  $y(\mu) > 0$  pois  $y(\mu)$  é dual viável para  $\mu < \mu^0$ . De  $s(\mu)_j = n - (a_j^k)^T y(\mu)$  onde  $a_j^k$  é a  $j$ -ésima coluna da matriz  $A^k$ , temos  $s(\mu)_j \leq n$  para todo  $\mu < \mu^0$ .

Quanto a função penalizada primal, dados  $\mu > \mu' > 0$ , sabemos que

$$\sum_{i=1}^m t_i \log z(\mu)_i + \mu \sum_{j=1}^p x(\mu)_j > \sum_{i=1}^m t_i \log z_i + \mu \sum_{j=1}^p x_j, \quad \forall (x, z) \in \mathcal{F}_{p,\mu}^{\circ}.$$

Em particular,

$$\sum_{i=1}^m t_i \log z(\mu)_i + \mu \sum_{j=1}^p x(\mu)_j > \sum_{i=1}^m t_i \log z(\mu')_i + \mu \sum_{j=1}^p x(\mu')_j. \quad (\text{IV.7})$$

Por outro lado,

$$\sum_{i=1}^m t_i \log z(\mu')_i + \mu' \sum_{j=1}^p x(\mu')_j > \sum_{i=1}^m t_i \log z_i + \mu' \sum_{j=1}^p x_j, \quad \forall (x, z) \in \mathcal{F}_{p, \mu}^{\circ}.$$

Em particular,

$$\sum_{i=1}^m t_i \log z(\mu')_i + \mu' \sum_{j=1}^p x(\mu')_j > \sum_{i=1}^m t_i \log z(\mu)_i + \mu' \sum_{j=1}^p x(\mu)_j. \quad (\text{IV.8})$$

Multiplicando por  $\mu'$  a equação (IV.7), multiplicando por  $\mu$  a equação (IV.8) e somando-as, temos  $(\mu - \mu') \sum_{i=1}^m t_i \log z(\mu')_i > (\mu - \mu') \sum_{i=1}^m t_i \log z(\mu)_i$ . Como  $(\mu - \mu') > 0$ , temos

$$\sum_{i=1}^m t_i \log z(\mu')_i > \sum_{i=1}^m t_i \log z(\mu)_i.$$

Considerando agora a função penalizada dual, temos

$$\sum_{i=1}^m t_i \log y(\mu)_i + \mu \sum_{j=1}^p s(\mu)_j > \sum_{i=1}^m t_i \log y_i + \mu \sum_{j=1}^p s_j, \quad \forall (y, s) \in \mathcal{F}_d^{\circ}.$$

Em particular,

$$\sum_{i=1}^m t_i \log y(\mu)_i + \mu \sum_{j=1}^p s(\mu)_j > \sum_{i=1}^m t_i \log y(\mu')_i + \mu \sum_{j=1}^p s(\mu')_j. \quad (\text{IV.9})$$

Por outro lado,

$$\sum_{i=1}^m t_i \log y(\mu')_i + \mu' \sum_{j=1}^p s(\mu')_j > \sum_{i=1}^m t_i \log y_i + \mu' \sum_{j=1}^p s_j, \quad \forall (y, s) \in \mathcal{F}_d^{\circ}.$$

Em particular,

$$\sum_{i=1}^m t_i \log y(\mu')_i + \mu' \sum_{j=1}^p s(\mu')_j > \sum_{i=1}^m t_i \log y(\mu)_i + \mu' \sum_{j=1}^p s(\mu)_j. \quad (\text{IV.10})$$

Multiplicando por  $\mu'$  a equação (IV.9), multiplicando por  $\mu$  a equação (IV.10) e somando-as, temos  $(\mu - \mu') \sum_{i=1}^m t_i \log y(\mu')_i > (\mu - \mu') \sum_{i=1}^m t_i \log y(\mu)_i$ . Como  $(\mu - \mu') > 0$ , concluímos

$$\sum_{i=1}^m t_i \log y(\mu')_i > \sum_{i=1}^m t_i \log y(\mu)_i.$$

Observando a primeira equação do sistema em (IV.3), se  $\mu$  tende a zero, então  $Xs$  tende a zero. Sobre a igualdade que resulta daquele sistema, sabemos

que  $e^T x$  tende a 1 quando  $\mu$  tende para zero. Logo o sistema (IV.3) tende para o sistema KKT correspondente aos problemas  $P^k$  e  $D^k$  apresentado no capítulo anterior. ■

Vários métodos foram criados para os algoritmos que seguem aproximadamente a trajetória central. Quanto à abordagem, os métodos podem ser primais, duais ou primais-duais. Cada método apresenta um modo específico para inicializar as variáveis e atualizar o parâmetro da barreira. A metodologia utilizada para atualizar o parâmetro da barreira classifica os algoritmos que seguem a trajetória central por algoritmos de passos curtos (os pontos gerados pelo algoritmo se encontram próximos da trajetória central) e algoritmos de passos longos (os pontos gerados pelo algoritmo não necessariamente se encontram próximos da trajetória central).

A seguir apresentamos um modelo de algoritmo que segue aproximadamente a trajetória central primal-dual.

### **Algoritmo 2 *Modelo de Algoritmo que Segue a Trajetória Central Primal-Dual***

**Dados :**  $\varepsilon > 0$  e  $\mu^0 > 0$ .

**Faça:**  $k := 0$ .

**Repita**

encontre aproximadamente  $(x(\mu^k), s(\mu^k), y(\mu^k), z(\mu^k))$ ,

escolha  $\mu^{k+1} < \mu^k$ ,

$k := k + 1$ ,

**Até que:**  $\mu^k < \varepsilon$ .

**Fim**

A seguir estudaremos o método escolhido para encontrar uma solução ótima aproximada de  $D^k$  denotado por Primal-Dual Predictor-Corretor. Obteremos aproximadamente um ponto central associado a  $\mu^{k+1}$  em  $\Omega^k$ , quando o modelo não puder ser melhorado, conforme o Algoritmo Principal.

### IV.3 O método predictor-corretor inviável

Tendo como referência resultados eficientes obtidos na prática pelos algoritmos que seguem a trajetória central primal-dual em programação linear [1], optou-se por utilizar um algoritmo primal-dual inviável para resolver o problema interno de encontrar a estimativa de máxima verossimilhança. Isso decorre do fato de que, apesar de o problema que queremos resolver ser não-linear, após a aplicação da barreira logarítmica, o problema interno de EMV apresenta condições de otimalidade de primeira ordem semelhantes às de um problema linear com barreira logarítmica.

O método primal-dual escolhido foi o método predictor-corretor, desenvolvido por Mehrotra [36] em 1990, que é uma variante do algoritmo de trajetória central primal-dual de passos longos. Utilizamos este método por ser o mais eficiente computacionalmente [1].

Como todo método primal-dual inviável, o método predictor-corretor inviável é iterativo. Cada ponto gerado pelo algoritmo é positivo (ficando caracterizado o fato do algoritmo ser de pontos interiores), mas não necessariamente primal e/ou dual viável. A viabilidade é obtida ao longo do processo à medida que a otimalidade é alcançada. O algoritmo termina quando o gap de complementaridade assim como a inviabilidade são reduzidos a uma tolerância pré-determinada.

#### Convergência

Mizuno, Todd e Ye [39] mostraram que o algoritmo primal-dual inviável segue aproximadamente uma superfície (vizinhança) de pontos centrais. Isto porque cada ponto

está associado a dois parâmetros independentes: o parâmetro da viabilidade e o da otimalidade. À medida que o algoritmo converge, os parâmetros tendem a zero e o algoritmo é capaz de indentificar a inviabilidade ou então fornece uma solução viável aproximada para o problema de otimização.

Conforme Mizuno [40], no caso do problema de otimização possuir uma solução ótima e se uma certa medida de viabilidade é pequena no ponto inicial, então o algoritmo preditor-corretor inviável termina em  $O(\sqrt{n}L)$  passos, onde  $L$  é o tamanho do problema de otimização com dados inteiros.

Lustig et. al [35], tendo com referência o trabalho de Kojima et al. [25] que provaram que os métodos primais-duais apresentam convergência global com algumas salva guardas, verificaram através dos resultados computacionais que o algoritmo preditor-corretor inviável apresenta convergência global quando algumas modificações de custo irrizório são inseridas no algoritmo.

Informações precisas a respeito do algoritmo primal-dual preditor-corretor inviável, sua convergência e referências correlatas podem ser obtidas na tese de doutorado do Marco A. F. Menezes [37].

## O método primal-dual inviável

De início, vamos estudar brevemente o método primal-dual inviável para o problema interno de EMV. E em seguida apresentamos a sua variante: o método primal-dual preditor-corretor inviável.

O método primal-dual inviável é derivado diretamente das condições de otimalidade de primeira ordem em (IV.3) quando aplicamos uma iteração do método de Newton. O método primal-dual encontra uma solução aproximada de (IV.3) para um valor de  $\mu$  dado e reduz este parâmetro. O processo continua até que o gap de complementaridade seja reduzido a uma tolerância pre-determinada.

Dados um ponto inicial  $(x^k, s^k, y^k, z^k) > 0$  e  $\mu^k$  na iteração  $k$ . Con-



sidere ainda que na iteração sejam conhecidos  $A^k$  e  $t$ . Uma iteração do algoritmo primal-dual inviável é um passo do método de Newton aplicado às equações do sistema (IV.3), o que equivale a resolver o seguinte sistema

$$\begin{aligned} S^k dx + X^k ds &= \mu^k e - X^k s^k, \\ Y^k dz + Z^k dy &= t - Y^k z^k, \\ A^k dx - dz &= z - A^k x^k, \\ (A^k)^T dy + ds &= ne - (A^k)^T y^k - s^k. \end{aligned}$$

E obtém o novo ponto

$$\begin{aligned} x^{k+1} &= x^k + \alpha_p dx, \\ s^{k+1} &= s^k + \alpha_d ds, \\ y^{k+1} &= y^k + \alpha_d dy, \\ z^{k+1} &= z^k + \alpha_p dz, \end{aligned}$$

onde  $\alpha_p$  e  $\alpha_d$  são respectivamente os tamanhos dos passos primal e dual calculados de maneira a garantir a não-negatividade destas variáveis. Em seguida o valor de  $\mu^{k+1}$  é obtido reduzindo  $\mu^k$ . Veja o modelo do algoritmo primal-dual inviável abaixo:

### Algoritmo 3 – Primal-Dual Inviável

**Dados**  $\varepsilon > 0$ ,  $\mu^0 > 0$  e  $(x^0, s^0, y^0, z^0) > 0$ ;

**Faça:**  $k:=0$ ;

**Repita**

*calcule a direção de Newton primal-dual,*

*encontre os tamanhos dos passos,*

*atualize o ponto,*

*reduza o valor de  $\mu^k$ ,*

*$k:=k+1$ ,*

**Até que:**  $\mu^k < \varepsilon$ ;

**Fim**

### O método preditor-corretor

Semelhante ao algoritmo primal-dual inviável, o método preditor-corretor inviável utiliza o sistema KKT perturbado (IV.3) substituindo diretamente o novo ponto nas

equações enquanto que os métodos primais-duais aplicam uma iteração do método de Newton.

Dados  $(x^k, s^k, y^k, z^k) > 0$ ,  $\mu^{k+1}$  e o sistema KKT perturbado como em (IV.3), quando substituimos diretamente o novo ponto nas equações, obtemos:

$$\begin{aligned} (X^k + Dx)(s^k + ds) &= \mu^{k+1}e, \\ (Y^k + Dy)(z^k + dz) &= t, \\ A^k(x^k + dx) - (z^k + dz) &= 0, \\ (s^k + ds) + (A^k)^T(y^k + dy) &= ne, \end{aligned}$$

onde  $Dx$  e  $Dy$  são matrizes diagonais cujos elementos da diagonal são as componentes dos vetores  $dx$  e  $dy$ , respectivamente. Podemos escrever o sistema acima como

$$\begin{aligned} S^k dx + X^k ds &= \mu^{k+1}e - X^k s^k - Dxds, \\ Y^k dz + Z^k dy &= t - Y^k z^k - Dydz, \\ A^k dx - dz &= z^k - A^k x^k, \\ (A^k)^T dy + ds &= ne - (A^k)^T y^k - s^k. \end{aligned} \tag{IV.11}$$

A presença dos termos não-lineares no lado direito do sistema acima diferencia o método preditor-corretor dos demais métodos primais-duais.

Para  $\mu^{k+1}$  dado, a solução do sistema acima é uma solução exata de (IV.3). Porém  $\mu^{k+1}$  não é conhecido, ele será calculado pelo método. Mehrotra encontra uma solução aproximada de (IV.3) passando por 2 estágios (os quais apresentamos adiante). Um novo ponto é obtido e o processo iterativo continua até que o gap de complementaridade seja reduzido a uma tolerância pré-estabelecida.

## Notação

Para efeito de simplificação, vamos denotar por  $r_p$  o resultado da expressão  $z^k - A^k x^k$  e chamar de resíduo primal na iteração  $k$ . O resultado de  $ne - (A^k)^T y^k - s^k$  chama-se resíduo dual na iteração  $k$  e denota-se por  $r_d$ .

## A direção afim-escala

Dado o ponto inicial  $(x^k, s^k, y^k, z^k) > 0$  na iteração  $k$ . Suponha conhecidos na iteração a matriz  $A^k$  e o vetor  $t$ . Primeiramente, Mehrotra sugere encontrar a direção afim-escala resolvendo o seguinte sistema

$$\begin{aligned} S^k dx^a + X^k ds^a &= -X^k s^k, \\ Y^k dz^a + Z^k dy^a &= t - Y^k z^k, \\ A^k dx^a - dz^a &= r_p, \\ (A^k)^T dy^a + ds^a &= r_d. \end{aligned} \tag{IV.12}$$

Assim, a direção  $(dx^a, ds^a, dy^a, dz^a)$  é utilizada para estimar o valor de  $\mu^{k+1}$  e corrigir a direção de centralização ou direção de Newton.

### Valor de $\mu^{k+1}$

Para escolher  $\mu^{k+1}$ , Mehrotra faz de início o teste da razão para as variáveis na direção afim-escala. Sejam

$$\begin{aligned} \delta_p &= \min \left\{ \min_j \left\{ \frac{x_j}{-dx_j^a}, dx_j^a < 0 \right\}, \min_j \left\{ \frac{z_j}{-dz_j^a}, dz_j^a < 0 \right\} \right\} \text{ e} \\ \delta_d &= \min \left\{ \min_j \left\{ \frac{s_j}{-ds_j^a}, ds_j^a < 0 \right\}, \min_j \left\{ \frac{y_j}{-dy_j^a}, dy_j^a < 0 \right\} \right\}. \end{aligned}$$

A fim de evitar que o novo ponto encoste na fronteira, faz-se

$$\begin{aligned} \delta_p &= 0.99\delta_p, \\ \delta_d &= 0.99\delta_d. \end{aligned}$$

Calcula-se o gap de complementaridade para a direção afim-escala

$$g^a = (x^k + \delta_p dx^a)^T (s^k + \delta_d ds^a).$$

Introduzindo a sugestão de Lustig et al. [34], que se mostrou na prática mais estável, o valor de  $\mu^{k+1}$  é determinado realizando o seguinte teste: seja  $gap = (x^k)^T s^k$ , então

$$\mu^{k+1} = \begin{cases} \left( \frac{g^a}{gap} \right)^2 \left( \frac{g^a}{p} \right), & gap \geq 1 \\ \frac{gap}{p^2}, & \text{caso contrário} \end{cases}.$$

Mehrotra calcula o valor de  $\mu^{k+1}$  avaliando somente a expressão do caso  $gap \geq 1$ .

A interpretação que se dá ao modo como  $\mu^{k+1}$  está sendo atualizado é a de que  $\mu^{k+1}$  deve ser pequeno quando a direção afim-escala é boa, neste caso o passo é longo. Ou  $\mu^{k+1}$  é grande quando existe a necessidade de centralizar pois a direção afim-escala não é boa.

### A direção de centralização

Finalmente, no segundo estágio a direção de centralização é encontrada para o valor de  $\mu^{k+1}$  que acabamos de calcular, substituindo os termos  $Dx ds$  por  $Dx^a ds^a$  e  $Dy dz$  por  $Dy^a dz^a$  e resolvendo

$$\begin{aligned} S^k dx + X^k ds &= \mu^{k+1} e - X^k s^k - Dx^a ds^a, \\ Y^k dz + Z^k dy &= t - Y^k z^k - Dy^a dz^a, \\ A^k dx - dz &= r_p, \\ (A^k)^T dy + ds &= r_d. \end{aligned} \tag{IV.13}$$

Novo teste da razão é realizado para o novo ponto na nova direção de centralização. Os tamanhos dos passos primal e dual,  $\alpha_p$  e  $\alpha_d$ , são calculados da mesma maneira apresentada acima e então utilizados na definição do novo ponto:

$$\begin{aligned} x^{k+1} &= x^k + \alpha_p dx, \\ s^{k+1} &= s^k + \alpha_d ds, \\ y^{k+1} &= y^k + \alpha_d dy, \\ z^{k+1} &= z^k + \alpha_p dz. \end{aligned}$$

Mehrotra levou em consideração o fato de que a fase de fatoração é mais cara que a fase de encontrar a solução na resolução de um sistema linear. Utilizou, portanto, para uma mesma fatoração duas soluções. Com isso, melhorou a direção de busca do algoritmo primal-dual quando introduziu termos não lineares para corrigir a direção a ser encontrada. O método primal-dual preditor-corretor apresenta um maior custo por iteração, porém, o custo total do algoritmo é melhor comparado com os demais métodos primais-duais [1].

Outros aspectos relacionados ao algoritmo serão abordados na seção referente à implementação apresentada no capítulo VI.

Sumariamente, apresentamos a seguir o algoritmo primal-dual preditor-corretor inviável. Ele será utilizado pelo Algoritmo Principal quando o modelo não puder ser melhorado: deve-se obter aproximadamente um novo ponto central primal-dual associado a  $\mu^{k+1}$  em  $\Omega^k$ , veja a figura (IV.3) em seguida.

**Algoritmo 4 – Primal-Dual Preditor-Corretor Inviável**

**Dados**  $\varepsilon > 0$ ,  $\mu^0 > 0$  e  $(x^0, s^0, y^0, z^0) > 0$ ;

**Faça:**  $k:=0$ ;

**Repita**

calcule a direção afim-escala,

encontre o valor de  $\mu^{k+1}$ ,

calcule a direção de centralização,

encontre os tamanhos dos passos,

atualize o ponto,

$k:=k+1$ ;

**Até que:**  $\mu^k < \varepsilon$

**Fim**

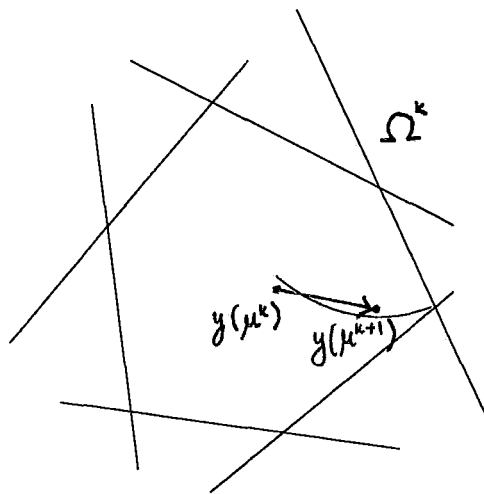


Figura IV.3: Um novo ponto central aproximado associado a  $\mu^{k+1}$

## IV.4 Um ponto central aproximado após a introdução de um ou mais cortes

Sejam  $k$  a iteração atual do Algoritmo Principal,  $\Omega^k$  o conjunto viável dual atual e  $\mu^k$  o parâmetro da barreira atual. Considere o passo do Algoritmo Principal em que o Oráculo tenha introduzido um ou mais pontos à discretização existente de tamanho  $p$ . Como vimos no final do capítulo III, isto equivale a dizer que o Oráculo fez um ou mais cortes no politopo atual obtendo  $\Omega^{k+1}$ .

Nosso objetivo, nesta seção, é obter um ponto  $(x^{k+1}, s^{k+1}, y^{k+1}, z^{k+1})$  correspondente a  $P^{k+1}$  e  $D^{k+1}$ . Para isto vamos utilizar um algoritmo primal-dual para encontrar aproximadamente um ponto central dual associado a  $\mu^k$  no politopo  $\Omega^{k+1}$ . Portanto, iremos encontrar aproximadamente um ponto central primal-dual associado a  $\mu^k$ .

Para o Algoritmo Principal, estamos encontrando um ponto aproximadamente central primal-dual associado a  $\mu^k$  no conjunto viável definido após a introdução de um ou mais cortes. Observe que este mesmo parâmetro foi utilizado para encontrar um ponto aproximadamente central no conjunto viável atual. No próximo capítulo introduziremos o conteúdo necessário para entender este passo do Algoritmo Principal. Veja a figura (IV.4) abaixo.

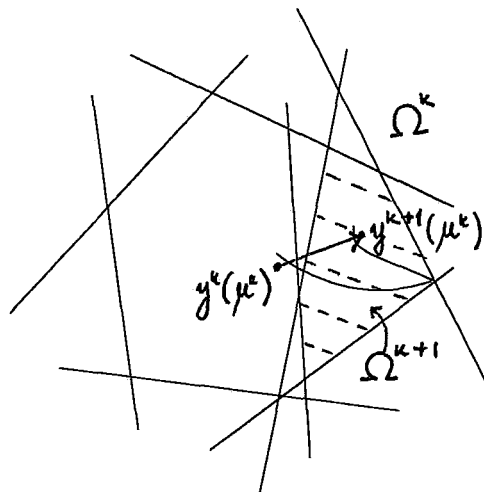


Figura IV.4: Um novo ponto central aproximado associado a  $\mu^k$  após os cortes

## Notação

A partir desta seção introduzimos uma nova notação que diz respeito aos novos cortes realizados no polítopo atual. As variáveis relativas aos novos cortes representamos por  $x_o$ ,  $s_o$ , e as variáveis do novo problema interno denotamos por  $\tilde{x} = (x, x_o)$  e  $\tilde{s} = (s, s_o)$ , por exemplo.

Seja  $\eta$  o número de cortes realizados na iteração  $k$ . Os cortes realizados em  $\Omega^k$  são do tipo

$$A_o^T y \leq ne,$$

onde  $A_o \in \mathbb{R}^{m \times \eta}$ . Seja

$$\Omega^{k+1} = \{y \in \mathbb{R}_{++}^m : \tilde{s} = ne - \tilde{A}^T y \geq 0\},$$

onde  $\tilde{A} = [A^k \ A_o] \in \mathbb{R}^{m \times (p+\eta)}$  e  $\tilde{s} \in \mathbb{R}^{p+\eta}$ .

Dado  $\mu^k$ , considere  $(y(\mu^k), \tilde{s}(\mu^k))$  o ponto central dual associado a  $\mu^k$  em  $\Omega^{k+1}$ . Por definição de ponto central dual (IV.6), temos que  $(y(\mu^k), \tilde{s}(\mu^k))$  é a solução do seguinte problema

$$\begin{aligned} (D_\mu^{k+1}) \quad & \text{maximizar} \quad \sum_{i=1}^m t_i \log y_i + \mu^k \sum_{j=1}^{p+\eta} \log \tilde{s}_j \\ & \text{sujeito a} \quad \tilde{s} = ne - \tilde{A}^T y \\ & \quad \tilde{s} \geq 0 \\ & \quad y > 0. \end{aligned}$$

Semelhante a (IV.3), as condições de otimalidade de primeira ordem para o problema acima são:

$$\begin{aligned} \tilde{X} \tilde{s} &= \mu^k e, \\ Yz &= t, \\ \tilde{A} \tilde{x} - z &= 0, \\ \tilde{A}^T y + \tilde{s} &= ne, \\ \tilde{x} \geq 0, \quad \tilde{s} \geq 0, \quad y > 0, \quad z > 0. \end{aligned} \tag{IV.14}$$

Logo, o ponto que satisfaz o sistema acima é o ponto central primal-dual associado a  $\mu^k$ .

Como não existe um algoritmo que em tempo finito encontre a solução exata do problema acima, para um dado  $t$  e  $\mu > 0$ , vamos definir a medida de pro-

ximidade primal-dual de um ponto  $(x, s, y, z)$  ao ponto  $(x(\mu), s(\mu), y(\mu), z(\mu))$  por

$$\delta(x, s, y, z, t, \mu) = \|Xs - \mu e\| + \|Yz - t\|.$$

Um algoritmo de centralização é construído aplicando o método de Newton às condições de otimalidade. No caso da abordagem ser primal-dual, aplicamos o método de Newton diretamente às equações do sistema KKT acima para obter um novo ponto. O algoritmo pára quando a medida de proximidade é menor que uma pré-determinada tolerância. Para um ponto inicial positivo, o algoritmo de centralização é polinomial, Mizuno em [41].

A seguir vamos apresentar o modelo do algoritmo de centralização primal-dual inviável e depois passamos a descrever uma iteração do algoritmo.

**Algoritmo 5** *Modelo do Algoritmo de Centralização Primal-Dual Inviável*

**Dados**  $\varepsilon > 0$ ,  $\mu^0 > 0$  e  $(\tilde{x}^0, \tilde{s}^0, y^0, z^0) > 0$ ;

**Faça:**  $k := 0$ ;

**Repita**

*encontre a direção de Newton primal-dual,*

*encontre os tamanhos dos passos,*

*atualize o ponto,*

$k := k + 1$ ;

**Até que:**  $\delta(\cdot) < \varepsilon$ ;

**Fim**

Uma iteração do algoritmo de centralização primal-dual segue os seguintes passos: dados  $(\tilde{x}, \tilde{s}, y, z) > 0$ ,  $\mu^k$ ,  $\tilde{A}$  e  $t$ ,



- resolva o sistema para encontrar a direção de Newton primal-dual:

$$\begin{aligned}
 \tilde{S}d\tilde{x} + \tilde{X}d\tilde{s} &= \mu^k e - \tilde{X}\tilde{s} \\
 Zdy + Ydz &= t - Yz \\
 \tilde{A}d\tilde{x} - dz &= z - \tilde{A}\tilde{x} \\
 \tilde{A}^T dy + d\tilde{s} &= ne - \tilde{A}^T y - \tilde{s}
 \end{aligned}
 \tag{IV.15}$$

- encontre os tamanhos dos passos  $\alpha_p$  e  $\alpha_d$  de modo a garantir a não negatividade das variáveis,

- atualize os pontos:

$$\begin{aligned}
 \tilde{x}^{k+1} &= \tilde{x} + \alpha_p d\tilde{x}, \\
 \tilde{s}^{k+1} &= \tilde{s} + \alpha_d d\tilde{s}, \\
 y^{k+1} &= y + \alpha_d dy, \\
 z^{k+1} &= z + \alpha_p dz.
 \end{aligned}$$

Decidimos realizar no nosso algoritmo somente uma iteração do algoritmo de centralização primal-dual inviável para obter um ponto aproximadamente central primal-dual associado a  $\mu^k$ . Isto decorre da nossa escolha do ponto inicial  $(\tilde{x}, \tilde{s}, y, z) > 0$ , como veremos a seguir.

## IV.5 O ponto inicial após a introdução de um ou mais cortes

Como vimos na seção anterior, precisamos de um ponto inicial adequado para dar início ao algoritmo de centralização.

A questão de utilizar uma solução ótima de um problema interno como ponto inicial de um novo problema interno foi considerada por muito tempo como sendo um problema na prática para os métodos de planos de corte. A dificuldade surgia quando a solução do primeiro problema se encontrava perto da fronteira ou mesmo na fronteira da região viável. Pois nestes casos, acontece de os algoritmos apresentarem problemas de convergência ou não conhecerem um ponto viável inicial para o novo problema.

Ainda assim, esta questão continuou intratável quando os métodos de planos de corte empregaram algoritmos de pontos interiores na resolução dos

problemas internos. O “gargalo” dos algoritmos de planos de corte somente foi eliminado quando os problemas internos deixaram de ser resolvidos “até o fim” pelos algoritmos de pontos interiores (no sentido de  $\mu$  ser menor que uma pre-determinada tolerância), possibilitando um grande desempenho na prática para algumas classes de problemas como: viabilidade convexa e multi-fluxo, e outros da programação estocástica e combinatória.

A seguir apresentamos os procedimentos adotados a fim de que a partir de uma solução ótima aproximada de um problema interno obtenhamos um ponto inicial para um novo problema quando acrescentamos uma ou mais restrições.

## Preliminares

Considere  $k$  a iteração atual. Seja  $p$  o tamanho atual da discretização de  $\Gamma$ . Sejam  $\eta$  o número de cortes e  $A_o \in \mathbb{R}^{m \times \eta}$  a matriz originada pelos novos cortes dados pelo Oráculo. Sejam  $P^k$  e  $D^k$  os problemas internos atuais de EMV. Considere ainda  $(x^k, s^k, y^k, z^k)$  o ponto atual do Algoritmo Principal e que  $\mu^k$ ,  $A^k$  e  $t$  sejam conhecidos. Observe que, de acordo como os  $\eta$  cortes foram gerados pelo Oráculo,  $y^k$  não é mais dual viável para  $\Omega^{k+1}$ :  $ne - A_o^T y^k < 0$ .

Queremos encontrar aproximadamente um ponto central primal-dual associado a  $\mu^k$ . Para isto precisamos de um ponto  $(\tilde{x}, \tilde{s}, y, z) > 0$  para dar início ao algoritmo de centralização primal-dual inviável. De acordo com a seção anterior,  $(\tilde{x}, \tilde{s}, y, z)$  é tal que

$$\tilde{x} = \begin{bmatrix} x \\ x_o \end{bmatrix}, \quad \tilde{s} = \begin{bmatrix} s \\ s_o \end{bmatrix} \in \mathbb{R}^{(p+\eta)}, \quad y, z \in \mathbb{R}^m,$$

onde  $x_o$  e  $s_o$  são as variáveis relativas aos novos cortes.

A seguir apresentamos diferentes procedimentos para atualizar as variáveis duais e primais, nesta ordem, quando um ou mais cortes são introduzidos pelo Oráculo. No final da seção apresentamos resumidamente dois métodos diferentes para a inicialização.

### IV.5.1 Atualização da Variável Dual

A seguir apresentamos dois procedimentos distintos para a atualização da variável dual quando uma ou mais restrições são adicionadas ao politopo atual.

Restaurar a viabilidade dual em um problema interno da programação linear é um problema difícil especialmente quando na iteração existem mais de um corte.

Mesmo no caso de apenas um corte realizado em  $\Omega^k$ , tem-se que a viabilidade dual depende de quão profundo ele é. Este estudo foi feito por Goffin e Vial em [12]. Eles nos mostram que somente com um corte do tipo pouco profundo, ou seja, um corte que intercepta o “elipsóide de Dikin”, a viabilidade dual é restaurada. Na figura (IV.5), damos uma idéia do que são os cortes do tipo pouco profundos, os quais estão representados pelos planos  $a, b, c$ , e os cortes profundos, os quais estão representados pelos planos  $d, e, f$ .

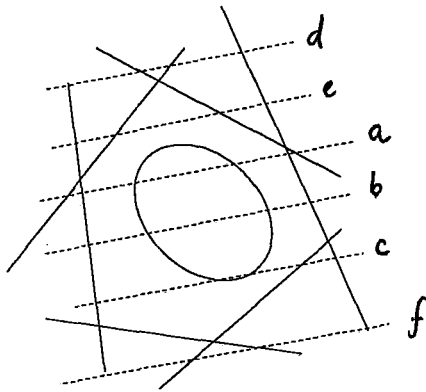


Figura IV.5: Exemplos de cortes pouco profundos ( $a, b, c$ ) e profundos ( $d, e, f$ )

Recentemente, na tese de doutorado de Marli Cardia [4], um novo algoritmo primal-dual para restaurar a viabilidade dual após um corte profundo foi apresentado. Porém, não o utilizamos nesta tese.

Como nossos cortes são do tipo profundos, resolvemos seguir alguns dos procedimentos adotados por Gondzio em [16], já que estamos utilizando um algoritmo primal-dual inviável para encontrar aproximadamente uma solução ótima

de  $D_\mu^{k+1}$ .

### Técnica de Gondzio

Baseado na própria experiência computacional, de que inviabilidades pequenas são facilmente absorvidas pelos algoritmos primais-duais inviáveis, Gondzio define um vetor  $\eta$ -dimensional das violações geradas pelos  $\eta$  cortes na iteração

$$v = ne - A_o^T y^k.$$

Dado  $\mu^k$ , ele estabelece que cortes profundos são do tipo  $v_j < -\sqrt{\mu^k}$  e cortes rasos são do tipo  $-\sqrt{\mu^k} < v_j < 0$ . Veja a figura (IV.6).

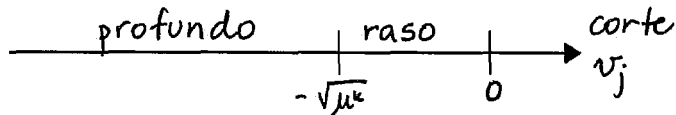


Figura IV.6: Redefinição dos cortes

Deste modo, faz-se, para  $j = 1, \dots, \eta$ ,

$$s_{oj} = \max\{\sqrt{\mu^k}, |v_j|\}.$$

Logo a atualização do ponto dual inicial será

$$y = y^k \text{ e } \tilde{s} = \begin{pmatrix} s^k \\ s_o \end{pmatrix} \in \mathbb{R}^{p+\eta},$$

não dual viável, porém positivo.

Gondzio afirma que tais procedimentos não garantem a viabilidade dual, porém, reduzem-na de maneira a ser absorvida mais facilmente pelo algoritmo.

### Técnica Natural no EMV

Entretanto, explorando a natureza do problema EMV, podemos facilmente obter um ponto dual viável. É sabido que a matriz  $A^k$  apresenta elementos não-negativos, assim como qualquer coluna da matriz  $A_o$  a ser adicionada a matriz  $A^k$ . Veja na

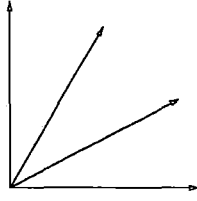


Figura IV.7: Exemplos de direções possíveis para os novos cortes.

figura (IV.7) exemplos de direções possíveis para os novos hiperplanos que cortam o politopo  $\Omega^k$ .

Para um politopo  $\Omega^{k+1}$  e um ponto  $y^k$  exterior como na figura (IV.8), podemos fazer uma busca no intervalo  $[0, y^k]$  até obtermos um ponto  $\lambda y^k$ , com  $\lambda \in (0, 1)$ , dentro do politopo.

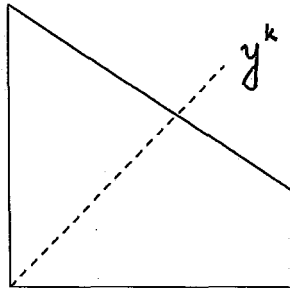


Figura IV.8: Busca para tornar  $y^k$  viável.

Logo, atualizamos as variáveis duais fazendo

dado  $\lambda \in (0, 1)$ ,  $y = \lambda y^k$  e  $\tilde{s} = ne - \tilde{A}^T y^k$ , onde  $\tilde{A} = [A^k \ A_o] \in \mathbb{R}^{m \times (p+\eta)}$ .

### IV.5.2 Atualização da Variável Primal

A seguir apresentamos um procedimento para atualizar a variável primal: a técnica da direção de Mitchell e Todd [38] aplicada ao problema interno de EMV quando um ou mais cortes são realizados na iteração.

### Um ponto primal interior para $P^{k+1}$ no caso de 1 corte

Seja o problema interno simplificado de EMV atual dado por

$$(D^k) \quad \begin{array}{l} \text{maximizar} \quad \sum_{i=1}^m t_i \log y_i \\ \text{sujeito a} \quad (A^k)^T y \leq ne \\ y > 0 \end{array}$$

O problema primal equivalente se apresenta no formato

$$(P^k) \quad \begin{array}{l} \text{maximizar} \quad \sum_{i=1}^m t_i \log z_i \\ \text{sujeito a} \quad A^k x - z = 0 \\ e^T x = 1 \\ x \geq 0 \\ z > 0 \end{array}$$

Fazer um corte na região viável dual equivale a adicionar uma restrição em  $D^k$ , com isto, o problema passa a se apresentar como

$$(D^{k+1}) \quad \begin{array}{l} \text{maximizar} \quad \sum_{i=1}^m t_i \log y_i \\ \text{sujeito a} \quad (A^k)^T y \leq ne \\ a_o^T y \leq n \\ y > 0, \end{array}$$

onde  $a_o \in \mathbb{R}^m$ . Como cada restrição em  $D^k$  está associada a uma variável em  $P^k$ , o formato primal correspondente a  $D^{k+1}$  é

$$(P^{k+1}) \quad \begin{array}{l} \text{maximizar} \quad \sum_{i=1}^m t_i \log z_i \\ \text{sujeito a} \quad A^k x - z + a_o x_o = 0 \\ e^T x + x_o = 1 \\ x \geq 0, x_o \geq 0 \\ z > 0, \end{array}$$

onde  $x_o \in \mathbb{R}$ .

O objetivo da técnica da direção de Mitchell e Todd é encontrar um ponto viável e interior de um problema linear no formato primal após a realização de um corte no dual. Por hipótese, considera-se um ponto primal viável e interior inicial do problema anterior ao corte. Igualando a zero a nova componente do ponto primal originada pelo corte, o novo ponto após o corte é viável porém não é interior. Esta técnica desloca o ponto viável de maneira que o torne também um ponto interior.

Veremos que a utilização desta técnica permite que obtenhamos um ponto inicial interior para o novo problema interno primal originado com a adição

de um corte no dual, porém não primal viável, visto que no nosso problema interno não podemos considerar  $(x^k, z^k)$  um ponto viável.

A fim de desenvolvermos a técnica para o problema interno do EMV no formato primal vamos considerar por hipótese que seja  $(x^k, z^k)$  um ponto viável e interior de  $P^k$ , isto é,  $A^k x^k - z^k = 0$ ,  $e^T x^k = 1$ ,  $x^k > 0$  e  $z^k > 0$ . Tomando  $(x^k, z^k, 0) \in \mathbb{R}^{p+m+1}$ , temos que este ponto é viável em  $P^{k+1}$ , porém não é interior. Considere ainda  $(y^k, s^k)$  um ponto viável interior de  $D^k$ , de modo que  $(A^k)^T y^k \leq ne$ ,  $y^k > 0$  e  $s^k = ne - (A^k)^T y^k > 0$ . Defina a matriz diagonal definida positiva

$$D = X^{k1/2} S^{k-1/2}.$$

Seja a mudança de escala primal-dual na variável  $x$  em  $P^k$  apresentada por  $\hat{P}^k$ , cujo formato é

$$\begin{aligned} (\hat{P}^k) \quad & \text{maximizar} \quad \sum_{i=1}^m t_i \log z_i \\ & \text{sujeito a} \quad \hat{A}\hat{x} - z = 0 \\ & \hat{e}^T \hat{x} = 1 \\ & \hat{x} \geq 0 \\ & z > 0, \end{aligned}$$

onde  $\hat{A} = A^k D$ ,  $\hat{e} = De$  e  $\hat{x} = D^{-1}x$ . Note que  $(\sqrt{S^k X^k} e, z^k)$  é um ponto viável e interior de  $\hat{P}^k$ . Agora, a versão escalada correspondente ao problema aumentado de  $P^{k+1}$  é

$$\begin{aligned} (\hat{P}^{k+1}) \quad & \text{maximizar} \quad \sum_{i=1}^m t_i \log z_i \\ & \text{sujeito a} \quad \hat{A}\hat{x} - z + a_o x_o = 0 \\ & \hat{e}^T \hat{x} + x_o = 1 \\ & \hat{x} \geq 0, \quad x_o \geq 0 \\ & z > 0 \end{aligned}$$

e o ponto  $(\sqrt{S^k X^k} e, z^k, 0) \in \mathbb{R}^{p+m+1}$  é viável porém não é interior.

Vamos utilizar a abordagem direta da técnica da direção de Mitchell e Todd que trabalha com o problema escalado e aumentado  $\hat{P}^{k+1}$ . Seja  $d = (d_{\hat{x}}, d_z, d_o)$  a direção de subida máxima do seguinte problema

$$\begin{aligned} & \text{maximizar} \quad x_o \\ & \text{sujeito a} \quad \hat{A}\hat{x} - z + a_o x_o = 0 \\ & \hat{x} > 0, \quad x_o > 0, \quad z > 0. \end{aligned}$$

Fazendo uma mudança de variáveis, reescrevemos o problema acima como

$$\begin{aligned} & \text{maximizar } x_o \\ & \text{sujeito a } \bar{A}\bar{x} + a_o x_o = 0 \\ & \bar{x} > 0, x_o > 0, \end{aligned}$$

onde

$$\bar{A} = [\hat{A} \ -I] \in \mathbb{R}^{m \times (p+m)} \text{ e } \bar{x} = \begin{pmatrix} \hat{x} \\ z \end{pmatrix} \in \mathbb{R}^{p+m},$$

cuja solução é

$$\begin{pmatrix} \bar{d} \\ d_o \end{pmatrix} := P_{[\bar{A} \ a_o]} \begin{pmatrix} \bar{0} \\ 1 \end{pmatrix},$$

onde  $P_M$  denota a matriz projeção no espaço nulo de  $M$  e  $\bar{0} \in \mathbb{R}^{p+m}$ . Utilizando a fórmula de Sherman-Morrison-Woodbury, temos

$$\begin{aligned} \begin{pmatrix} \bar{d} \\ d_o \end{pmatrix} &= \begin{pmatrix} \bar{0} \\ 1 \end{pmatrix} - \begin{bmatrix} \bar{A}^T \\ a_o^T \end{bmatrix} \left( [\bar{A} \ a_o] \begin{bmatrix} \bar{A}^T \\ a_o^T \end{bmatrix} \right)^{-1} [\bar{A} \ a_o] \begin{pmatrix} \bar{0} \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} \bar{0} \\ 1 \end{pmatrix} - \begin{bmatrix} \bar{A}^T \\ a_o^T \end{bmatrix} (\bar{A}\bar{A}^T + a_o a_o^T)^{-1} a_o \\ &= \begin{pmatrix} \bar{0} \\ 1 \end{pmatrix} - \begin{bmatrix} \bar{A}^T \\ a_o^T \end{bmatrix} \left[ (\bar{A}\bar{A}^T)^{-1} - \frac{1}{1+\bar{\theta}} (\bar{A}\bar{A}^T)^{-1} a_o a_o^T (\bar{A}\bar{A}^T)^{-1} \right] a_o \\ &= \begin{pmatrix} -\bar{A}^T (\bar{A}\bar{A}^T)^{-1} a_o + \frac{1}{1+\bar{\theta}} \bar{A}^T (\bar{A}\bar{A}^T)^{-1} a_o a_o^T (\bar{A}\bar{A}^T)^{-1} a_o \\ 1 - a_o^T (\bar{A}\bar{A}^T)^{-1} a_o + \frac{1}{1+\bar{\theta}} a_o^T (\bar{A}\bar{A}^T)^{-1} a_o a_o^T (\bar{A}\bar{A}^T)^{-1} a_o \end{pmatrix} \\ &= \begin{pmatrix} -\bar{A}^T (\bar{A}\bar{A}^T)^{-1} a_o + \frac{\bar{\theta}}{1+\bar{\theta}} \bar{A}^T (\bar{A}\bar{A}^T)^{-1} a_o \\ 1 - \bar{\theta} + \frac{1}{1+\bar{\theta}} \bar{\theta}^2 \end{pmatrix} \\ &= \frac{1}{1+\bar{\theta}} \begin{pmatrix} -\bar{A}^T (\bar{A}\bar{A}^T)^{-1} a_o \\ 1 \end{pmatrix}, \end{aligned}$$

onde  $\bar{\theta} = a_o^T (\bar{A}\bar{A}^T)^{-1} a_o$ .

Logo a direção para as variáveis originais é

$$\begin{pmatrix} d_{\hat{x}} \\ d_z \\ d_o \end{pmatrix} = \frac{1}{1+\hat{\theta}} \begin{pmatrix} -\hat{A}^T (\hat{A}\hat{A}^T + I)^{-1} a_o \\ (\hat{A}\hat{A}^T + I)^{-1} a_o \\ 1 \end{pmatrix},$$

onde  $\hat{\theta} = a_o^T (\hat{A}\hat{A}^T + I)^{-1} a_o$ .

Sem perda de generalidade, desconsideraremos o fator da direção. Note que  $\hat{\theta}$  é um valor positivo: como  $\hat{A}$  possui rank máximo a matriz  $(\hat{A}\hat{A}^T + I)^{-1}$  é definida positiva.



Dado  $\alpha > 0$ . Seja o ponto  $(\hat{x}(\alpha), z(\alpha), x_o(\alpha))$  dado por

$$\begin{pmatrix} \hat{x}(\alpha) \\ z(\alpha) \\ x_o(\alpha) \end{pmatrix} = \begin{pmatrix} \sqrt{S^k X^k e} \\ z^k \\ 0 \end{pmatrix} + \alpha \begin{pmatrix} -\hat{A}^T(\hat{A}\hat{A}^T + I)^{-1} a_o \\ (\hat{A}\hat{A}^T + I)^{-1} a_o \\ 1 \end{pmatrix}.$$

Queremos encontrar o tamanho máximo do passo a ser dado ao longo da direção  $d$  a partir do ponto viável  $(\hat{x}(\alpha), z(\alpha), x_o(\alpha))$  de modo que o ponto resultante seja positivo. Para isto, realiza-se o teste da razão, fazendo

$$\alpha_{max} = \sup\{\alpha > 0 \mid (\hat{x}(\alpha), z(\alpha), x_o(\alpha)) + \alpha d > 0\}.$$

Substituindo o ponto e a direção

$$\alpha_{max} = \min \left\{ \min_{j=1, \dots, p} \frac{\sqrt{x_j^k s_j^k}}{-d_{\hat{x}_j}}, d_{\hat{x}_j} < 0; \min_{i=1, \dots, m} \frac{z_i^k}{-d_{z_i}}, d_{z_i} < 0 \right\}$$

ou

$$\alpha_{max} = \min \left\{ \min_{j=1, \dots, p} \frac{\sqrt{x_j^k s_j^k}}{\hat{a}_j^T (\hat{A}\hat{A}^T + I)^{-1} a_o}; \min_{i=1, \dots, m} \frac{z_i^k}{-\bar{a}_i^T a_o}, \bar{a}_i^T a_o < 0 \right\},$$

onde  $\hat{a}_j$  denota a  $j$ -ésima coluna da matriz  $\hat{A}$  e  $\bar{a}_i$  denota a  $i$ -ésima linha da matriz  $(\hat{A}\hat{A}^T + I)^{-1}$ .

Vamos definir

$$\xi(\alpha) := \hat{e}^T \hat{x}(\alpha) + x_o(\alpha).$$

Temos que  $\xi(\alpha) > 0$  pois  $\hat{e} > 0$ .

Logo o ponto

$$((1/\xi(\alpha))D\hat{x}(\alpha), (1/\xi(\alpha))z(\alpha), (1/\xi(\alpha))x_o(\alpha))$$

é viável e interior para o problema  $P^{k+1}$  desde que  $\alpha \in (0, \alpha_{max})$ . Verificando as

restrições de subespaço e normalização:

$$\begin{aligned}
& A^k \left( \frac{1}{\xi(\alpha)} D\hat{x}(\alpha) \right) - \frac{1}{\xi(\alpha)} z(\alpha) + a_o \left( \frac{1}{\xi(\alpha)} x_o(\alpha) \right) \\
&= \frac{1}{\xi(\alpha)} \left[ A^k D(D^{-1}x^k - \alpha \hat{A}^T(\hat{A}\hat{A}^T + I)^{-1}a_o) - (z^k + \alpha(\hat{A}\hat{A}^T + I)^{-1}a_o) + \alpha a_o \right] \\
&= \frac{1}{\xi(\alpha)} \left[ A^k x^k - \alpha \hat{A}\hat{A}^T(\hat{A}\hat{A}^T + I)^{-1}a_o - z^k - \alpha(\hat{A}\hat{A}^T + I)^{-1}a_o + \alpha a_o \right] \\
&= \frac{1}{\xi(\alpha)} \left[ -\alpha(\hat{A}\hat{A}^T + I)(\hat{A}\hat{A}^T + I)^{-1}a_o + \alpha a_o \right] \\
&= \frac{1}{\xi(\alpha)} (-\alpha a_o + \alpha a_o) \\
&= 0;
\end{aligned}$$

$$\begin{aligned}
e^T \left( \frac{1}{\xi(\alpha)} D\hat{x}(\alpha) \right) + \left( \frac{1}{\xi(\alpha)} x_o(\alpha) \right) &= \frac{1}{\xi(\alpha)} [e^T D\hat{x}(\alpha) + x_o(\alpha)] \\
&= \frac{1}{\xi(\alpha)} [\hat{e}^T \hat{x}(\alpha) + x_o(\alpha)] \\
&= \frac{1}{\xi(\alpha)} \xi(\alpha) = 1;
\end{aligned}$$

E finalmente verifica-se a condição de positividade das variáveis para  $\alpha \in (0, \alpha_{max})$ :

$$\frac{1}{\xi(\alpha)} D\hat{x}(\alpha) > 0, \quad \frac{1}{\xi(\alpha)} x_o(\alpha) > 0 \quad \text{e} \quad \frac{1}{\xi(\alpha)} z(\alpha) > 0.$$

Resumindo, o ponto  $(\tilde{x}, z)$ , onde

$$\tilde{x} = \begin{bmatrix} x(\alpha) \\ \psi \end{bmatrix} \in \mathbb{R}^{p+1}, \quad z = z(\alpha) \in \mathbb{R}^m,$$

e

$$\begin{aligned}
\psi &= 1/[\alpha^{-1} + 1 - e^T D^2(A^k)^T(A^k D^2(A^k)^T + I)^{-1}a_o], \\
x(\alpha) &= \psi \alpha^{-1} x^k - \psi D^2(A^k)^T(A^k D^2(A^k)^T + I)^{-1}a_o, \\
z(\alpha) &= \psi \alpha^{-1} z^k + \psi (A^k D^2(A^k)^T + I)^{-1}a_o,
\end{aligned}$$

é um ponto viável interior de  $P^{k+1}$ , com  $\alpha \in (0, \alpha_{max})$  sendo o tamanho do passo definido pelo teste da razão, e  $D^2 = X^k S^{k-1}$ .

Entretanto, para o nosso algoritmo, não podemos afirmar que  $(\tilde{x}, z)$  seja viável pois  $(x^k, z^k)$  não é viável visto que  $e^T x^k = 1 + \gamma \mu^k$ ,  $\gamma > 0$  fixo. A utilização deste procedimento em nosso algoritmo reduzirá a inviabilidade em geral, enquanto que a viabilidade será alcançada à medida que o algoritmo se aproxima da otimalidade.

### Um ponto primal interior para $P^{k+1}$ no caso de mais de 1 corte

Vamos estudar o caso em que adicionamos  $\eta$  cortes ao politopo, isto é, acrescentamos ao problema interno atual mais  $\eta$  restrições. Seja o novo problema interno com a adição de  $\eta$  cortes dado por

$$(D^{k+1}) \quad \begin{array}{l} \text{maximizar} \quad \sum_{i=1}^m t_i \log y_i \\ \text{sujeito a} \quad (A^k)^T y \leq ne \\ \quad \quad \quad A_o^T y \leq ne \\ \quad \quad \quad y > 0, \end{array}$$

onde  $A_o \in \mathbb{R}^{m \times \eta}$ . E o problema dual equivalente escrevemos no formato

$$(P^{k+1}) \quad \begin{array}{l} \text{maximizar} \quad \sum_{i=1}^m t_i \log z_i \\ \text{sujeito a} \quad A^k x - z + A_o x_o = 0 \\ \quad \quad \quad e^T x + e^T x_o = 1 \\ \quad \quad \quad x \geq 0, \quad x_o \geq 0 \\ \quad \quad \quad z > 0, \end{array}$$

onde  $x_o \in \mathbb{R}^\eta$ .

Suponha que  $(x^k, z^k)$  seja um ponto viável e interior de  $P^k$ . Temos que o ponto  $(x^k, z^k, 0) \in \mathbb{R}^{p+m+\eta}$  é viável em  $P^{k+1}$ , porém não é interior. Considere ainda  $(y^k, s^k)$  um ponto viável e interior de  $D^k$ .

De modo a contornar problemas provenientes do mal-condicionamento de matrizes quando utiliza-se diretamente a técnica da direção para  $\eta$  cortes, Mitchell e Todd abordaram o problema acima fazendo uma substituição de variáveis.

Definimos para um vetor  $w \in \mathbb{R}^\eta$  e  $w > 0$ :

$$\begin{aligned} \check{a} &= A_o w, \\ \check{e} &= e^T w, \end{aligned}$$

onde  $\check{a} \in \mathbb{R}^m$  e  $\check{e} \in \mathbb{R}$ . Seja  $\hat{P}^{k+1}$  o problema resultante da mudança de escala primal-dual na variável  $x$  em  $P^{k+1}$ . Podemos escrever  $\hat{P}^{k+1}$  num novo formato em que se tem apenas uma variável extra como em  $\hat{P}^{k+1}$ . Seja

$$(\check{P}) \quad \begin{array}{l} \text{maximizar} \quad \sum_{i=1}^m t_i \log z_i \\ \text{sujeito a} \quad \hat{A} \hat{x} - z + \check{a} \check{x} = 0 \\ \quad \quad \quad \hat{e}^T \hat{x} + \check{e} \check{x} = 1 \\ \quad \quad \quad \hat{x} \geq 0, \quad \check{x} \geq 0 \\ \quad \quad \quad z > 0, \end{array}$$

onde  $x_o = w\check{x} \in \mathbb{R}^n$  e  $\check{x} \in \mathbb{R}$ . O ponto  $(\sqrt{S^k X^k e}, z^k, 0) \in \mathbb{R}^{p+m+1}$  é viável porém não é interior de  $\check{P}$ . Portanto, com o novo formato de  $\check{P}$  podemos aplicar a metodologia anterior que encontra um ponto viável interior no caso de apenas um corte a partir do ponto viável que é

$$((1/\xi(\alpha))\hat{x}(\alpha), (1/\xi(\alpha))z(\alpha), (1/\xi(\alpha))\check{x}(\alpha)),$$

onde

$$\begin{pmatrix} \hat{x}(\alpha) \\ z(\alpha) \\ \check{x}(\alpha) \end{pmatrix} = \begin{pmatrix} \sqrt{S^k X^k e} \\ z^k \\ 0 \end{pmatrix} + \alpha \begin{pmatrix} -\hat{A}^T(\hat{A}\hat{A}^T + I)^{-1}\check{a} \\ (\hat{A}\hat{A}^T + I)^{-1}\check{a} \\ 1 \end{pmatrix}.$$

para  $\xi(\alpha) = \hat{e}^T \hat{x}(\alpha) + \check{e}\check{x}(\alpha) > 0$  e  $\alpha \in (0, \check{\alpha}_{max})$  onde

$$\check{\alpha}_{max} = \min \left\{ \min_{j=1, \dots, p} \frac{\sqrt{x_j^k s_j^k}}{\hat{a}_i^T (\hat{A}\hat{A}^T + I)^{-1} \check{a}}; \min_{i=1, \dots, m} \frac{z_i^k}{-\bar{a}_i^T \check{a}}, \bar{a}_i^T \check{a} < 0 \right\},$$

$\hat{a}_j$  denota a  $j$ -ésima coluna da matriz  $\hat{A}$  e  $\bar{a}_i$  denota a  $i$ -ésima linha da matriz  $(\hat{A}\hat{A}^T + I)^{-1}$ .

Logo, o ponto  $((1/\xi(\alpha))D\hat{x}(\alpha), (1/\xi(\alpha))z(\alpha), (1/\xi(\alpha))\check{x}(\alpha)w)$  é um ponto viável e interior de  $P^{k+1}$  desde que  $\alpha \in (0, \check{\alpha}_{max})$ .

Verificando:

$$\begin{aligned} & A^k \left( \frac{1}{\xi(\alpha)} D\hat{x}(\alpha) \right) - \frac{1}{\xi(\alpha)} z(\alpha) + A_o \left( \frac{1}{\xi(\alpha)} \check{x}(\alpha)w \right) \\ &= \frac{1}{\xi(\alpha)} \left[ A^k D(D^{-1}x^k - \alpha \hat{A}^T(\hat{A}\hat{A}^T + I)^{-1}A_o w) - (z^k + \alpha(\hat{A}\hat{A}^T + I)^{-1}A_o w) + \alpha A_o w \right] \\ &= \frac{1}{\xi(\alpha)} \left[ A^k x^k - \alpha \hat{A}\hat{A}^T(\hat{A}\hat{A}^T + I)^{-1}A_o w - z^k - \alpha(\hat{A}\hat{A}^T + I)^{-1}A_o w + \alpha A_o w \right] \\ &= \frac{1}{\xi(\alpha)} \left[ -\alpha(\hat{A}\hat{A}^T + I)(\hat{A}\hat{A}^T + I)^{-1}A_o w + \alpha A_o w \right] \\ &= \frac{1}{\xi(\alpha)} (-\alpha A_o w + \alpha A_o w) \\ &= 0; \end{aligned}$$

$$\begin{aligned} e^T \left( \frac{1}{\xi(\alpha)} D\hat{x}(\alpha) \right) + e^T \left( \frac{1}{\xi(\alpha)} \check{x}(\alpha)w \right) &= \frac{1}{\xi(\alpha)} \left[ e^T D\hat{x}(\alpha) + e^T \check{x}(\alpha)w \right] \\ &= \frac{1}{\xi(\alpha)} (\hat{e}^T \hat{x}(\alpha) + \check{e}\check{x}(\alpha)) \\ &= \frac{1}{\xi(\alpha)} \xi(\alpha) = 1; \end{aligned}$$

e

$$\frac{1}{\xi(\alpha)} D\hat{x}(\alpha) > 0, \quad \frac{1}{\xi(\alpha)} \tilde{x}(\alpha)w > 0 \quad \text{e} \quad \frac{1}{\xi(\alpha)} z(\alpha) > 0 \quad \forall \alpha \in (0, \check{\alpha}_{max}).$$

Resumindo, caso  $\eta$  cortes foram realizados na iteração e fazendo a escolha mais simples para  $w = e \in \mathbb{R}^\eta$ , o ponto primal  $(\tilde{x}, z)$ , onde

$$\tilde{x} = \begin{bmatrix} x(\alpha) \\ \psi e \end{bmatrix} \in \mathbb{R}^{p+\eta} \quad \text{e} \quad z = z(\alpha) \in \mathbb{R}^m$$

é interior em  $P^{k+1}$ , com  $\alpha \in (0, \alpha_{max})$  sendo o tamanho do passo definido pelo teste da razão e

$$\begin{aligned} \psi &= 1/[\alpha^{-1} + \eta - e^T D^2(A^k)^T (A^k D^2(A^k)^T + I)^{-1} a], \\ x(\alpha) &= \psi \alpha^{-1} x^k - \psi D^2(A^k)^T (A^k D^2(A^k)^T + I)^{-1} a, \\ z(\alpha) &= \psi \alpha^{-1} z^k + \psi (A^k D^2(A^k)^T + I)^{-1} a, \end{aligned}$$

com  $a = A_o e$  e  $D^2 = X^k S^{k-1}$ . Observa-se novamente que em nosso algoritmo a aplicação da técnica de Mitchell e Todd reduz a inviabilidade já que  $(x^k, z^k)$  não é um ponto viável.

A seguir apresentamos dois métodos diferentes, resultante da composição das técnicas apresentadas acima, para obter um ponto inicial positivo para o algoritmo de centralização primal-dual inviável. Dados  $(x^k, s^k, y^k, z^k) > 0$ ,  $\mu^k, \eta$ ,  $A^k$  e  $A_o$ , escolha um dos três métodos abaixo:

1. aplique a técnica da direção de Mitchell e Todd para atualizar as variáveis primais  $(\tilde{x}, z)$  e adote os procedimentos de Gondzio para atualizar as variáveis duais  $(y, \tilde{s})$ , ou
2. faça uma busca para viabilizar as variáveis duais  $(y, \tilde{s})$  aplicando a técnica natural do EMV e aplique a técnica da direção de Mitchell e Todd para atualizar as variáveis primais  $(\tilde{x}, z)$ .

O nosso algoritmo utiliza o segundo método de inicialização acima.

# Capítulo V

## O Algoritmo

Neste capítulo apresentamos o algoritmo de planos de corte baseado em centros para o problema de encontrar a estimativa de máxima verossimilhança.

Mais adiante, mostraremos uma relação importante da geometria de centros na teoria dos algoritmos de pontos interiores e apresentamos o algoritmo conceitual de planos de corte em centro analítico para resolver um problema geral de maximização de uma função côncava com infinitas restrições lineares. Pretendemos com isto entender cada passo do algoritmo apresentado abaixo.

Ainda neste capítulo, faremos uma introdução à análise de convergência e complexidade do algoritmo conceitual.

### **Algoritmo 6 – Algoritmo de Planos de Corte Baseado em Centros para EMV**

**Dados**  $u, t$ , a aproximação poliédrica inicial de  $\Omega$  :  $\Omega^0 = \{y \in \mathbb{R}_{++}^m : y \leq u\}$   
e o centro analítico de  $\Omega^0$  :  $y^0 = 0.5u \in \mathbb{R}^m$ ;

**Inicialize**

$$A^0 = I \in \mathbb{R}^{m \times m}, c^0 = u \in \mathbb{R}^m, s^0 = c^0 - (A^0)^T y^0 = 0.5u \in \mathbb{R}^m, \\ \mu^0 = 1, x^0 = 2u^{-1} \in \mathbb{R}^m, z^0 = 2tu^{-1} \in \mathbb{R}^m;$$

$$\begin{aligned}
 (P^0) \quad & \text{maximizar} \quad \sum_{i=1}^m t_i \log z_i \\
 & \text{sujeito a} \quad A^0 x = z \\
 & \quad \quad \quad e^T x = 1 \\
 & \quad \quad \quad x \geq 0 \\
 & \quad \quad \quad z > 0.
 \end{aligned}$$

$$\begin{aligned}
 (D^0) \quad & \text{maximizar} \quad \sum_{i=1}^m t_i \log y_i \\
 & \text{sujeito a} \quad y \in \Omega^0.
 \end{aligned}$$

**Faça:**  $k = 0$ ;

**Enquanto**  $\text{gap} > \varepsilon$ , **faça:**

• consulte o Oráculo em  $y^k$ , o qual responde:

– com um ou mais cortes:

adicione estes cortes a  $\Omega^k$  obtendo  $\Omega^{k+1}$ ;

atualize  $P^{k+1}$  e  $D^{k+1}$ ;

obtenha um ponto positivo escolhendo um dos métodos de inicialização;

realize uma ou mais iterações do Algoritmo de Centralização Primal-Dual Inviável para encontrar aproximadamente  $(x^{k+1}, s^{k+1}, y^{k+1}, z^{k+1})$ , um ponto central primal-dual associado a  $\mu^k$ ;

– com nenhum corte:

realize uma iteração do Algoritmo Primal-Dual Preditor-Corretor Inviável para obter o parâmetro  $\mu^{k+1}$  e o ponto  $(x^{k+1}, s^{k+1}, y^{k+1}, z^{k+1})$ , um ponto central primal-dual associado a  $\mu^{k+1}$ ;

• faça  $k = k + 1$ ;

**Fim**

**Fim**

A seguir apresentamos uma relação importante existente na teoria dos algoritmos de pontos interiores, que nos ajudará entender o algoritmo acima.

## V.1 Geometria

Nesta seção vamos estudar a relação existente entre o centro analítico de um conjunto de nível da função objetivo de  $D^k$  e um ponto sobre a trajetória central dual em  $\Omega^k$ . O lema que apresentaremos é o resultado da extensão da relação existente em programação linear.

**Lema V.1** *O centro analítico de um conjunto de nível da função objetivo do problema interno  $D^k$  é um ponto central dual em  $\Omega^k$ .*

*Prova.* Seja o problema interno dual de EMV na iteração  $k$  dado por

$$(D^k) \begin{array}{ll} \text{maximizar} & \sum_{i=1}^m t_i \log y_i \\ \text{sujeito a} & y \in \Omega^k, \end{array}$$

onde  $\Omega^k = \{y \in \mathbb{R}_{++}^m : c^k - (A^k)^T y \geq 0\}$  é não-vazio e limitado,  $c^k = ne$ ,  $s = c^k - (A^k)^T y$ , e  $A^k$  e  $t$  são conhecidos na iteração. Seja a função objetivo dual  $f : \mathbb{R}_{++}^m \rightarrow \mathbb{R}$  tal que para  $y \in \mathbb{R}_{++}^m$ ,

$$f(y) = \sum_{i=1}^m t_i \log y_i.$$

O gradiente de  $f$  em  $y \in \mathbb{R}_{++}^m$  é dado por

$$\nabla f(y) = Y^{-1}t.$$

Considere  $l$  um limite inferior para o valor ótimo da função  $f$ , isto é,  $l \leq l^* = f(y^*)$ , onde  $y^* \in \Omega^k$  é a solução ótima de  $D^k$ . Dado  $l$ , o conjunto de nível de  $f$  é definido por

$$\Omega^k(l) = \{y \in \mathbb{R}_{++}^m : c^k - (A^k)^T y \geq 0, f(y) - l \geq 0\}.$$

Observe que  $\Omega^k(l)$  é um conjunto convexo visto que  $f$  é côncava.

Sendo  $\Omega^k(l) \subset \Omega^k$ , o centro analítico de  $\Omega^k(l)$  denotado por  $y^a(l)$  é, por definição,

$$y^a(l) := \operatorname{argmax} \left\{ \sum_{j=1}^p \log s_j + \log(f(y) - l) : y \in \overset{\circ}{\Omega}^k(l) \right\},$$



se e somente se  $y^a(l) \in \overset{\circ}{\Omega}^k(l)$  e satisfaz a condição de otimalidade

$$-A^k \frac{e}{s^a(l)} + \frac{1}{f(y^a(l)) - l} \nabla f(y^a(l)) = 0, \quad (\text{V.1})$$

onde  $s^a(l) = c^k - (A^k)^T y^a(l) > 0$  e  $s^a(l)_{p+1} = f(y^a(l)) - l > 0$ . Veja um exemplo ilustrado de um conjunto de nível  $\Omega^k(l)$  e seu centro analítico  $y^a(l)$  na figura (V.1).

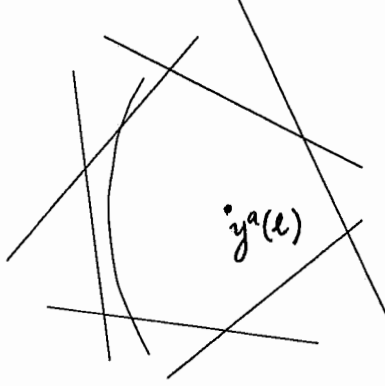


Figura V.1: Conjunto de nível  $\Omega^k(l)$  e seu centro analítico  $y^a(l)$

Fazendo  $x^a(l)' = s^a(l)^{-1} > 0$ ,  $x^a(l)' \in \mathbb{R}^p$  e  $x^a(l)'_{p+1} = 1/s^a(l)_{p+1} > 0$ ,  $x^a(l)'_{p+1} \in \mathbb{R}$ , a equação (V.1) pode ser substituída pelo seguinte sistema

$$\begin{bmatrix} X^a(l)' & 0 \\ 0 & x^a(l)'_{p+1} \end{bmatrix} \begin{bmatrix} s^a(l) \\ s^a(l)_{p+1} \end{bmatrix} = e \quad (\text{V.2})$$

$$[A^k \quad -\nabla f(y^a(l))] \begin{bmatrix} x^a(l)' \\ x^a(l)'_{p+1} \end{bmatrix} = 0 \quad (\text{V.3})$$

$$\begin{bmatrix} s^a(l) \\ s^a(l)_{p+1} \end{bmatrix} = \begin{bmatrix} c^k - (A^k)^T y^a(l) \\ f(y^a(l)) - l \end{bmatrix} > 0$$

Dividindo a equação (V.2) por  $x^a(l)'_{p+1}$ , obtemos

$$\begin{cases} \frac{1}{x^a(l)'_{p+1}} X^a(l)' s^a(l) = \frac{1}{x^a(l)'_{p+1}} e \\ s^a(l)_{p+1} = \frac{1}{x^a(l)'_{p+1}} \end{cases}$$

ou

$$X^a(l) s^a(l) = \mu e,$$

onde  $x^a(l)_j = x^a(l)'_j / x^a(l)'_{p+1}$  para todo  $j = 1, \dots, p$  e  $\mu = s^a(l)_{p+1} = f(y^a(l)) - l$ .

Do mesmo modo, dividindo a equação (V.3) por  $x^a(l)'_{p+1}$ ,

$$A^k \frac{x^a(l)'}{x^a(l)'_{p+1}} = \frac{x^a(l)'_{p+1}}{x^a(l)'_{p+1}} \nabla f(y^a(l))$$

ou

$$A^k x^a(l) = \nabla f(y^a(l)) = (Y^a(l))^{-1}t.$$

Fazendo  $z^a(l) = (Y^a(l))^{-1}t$ , então a equação (V.1) é equivalente ao sistema

$$\begin{cases} X^a(l)s^a(l) = \mu e \\ Y^a(l)z^a(l) = t \\ A^k x^a(l) - z^a(l) = 0 \\ (A^k)^T y^a(l) + s^a(l) = c^k \end{cases},$$

onde  $\mu = f(y^a(l)) - l$  e  $c^k = ne$ .

Note que o sistema acima representa as condições de otimalidade do seguinte problema

$$\begin{aligned} & \text{maximizar} && \sum_{i=1}^m t_i \log y_i + \mu \sum_{j=1}^p \log s_j \\ & \text{sujeito a} && y \in \Omega^k, \end{aligned}$$

cuja solução  $(y(\mu), s(\mu)) \in \overset{\circ}{\Omega}^k$  é o ponto central dual associado a  $\mu$  em  $\Omega^k$ .

Concluindo,  $(y^a(l), s^a(l))$  é o ponto central dual associado a  $\mu = f(y^a(l)) - l$  em  $\Omega^k$ . ■

Seja  $l^*$  o valor ótimo da função objetivo dual em  $\Omega^k$ . Observa-se que quando  $l \rightarrow l^*$ , temos  $\mu \rightarrow 0$  e quando  $l \rightarrow -\infty$  acontece de  $\mu \rightarrow \infty$ . Apesar disto, não podemos encontrar uma relação geral para os valores de  $l$  e  $\mu$ .

Observamos que os pontos  $y^a(l)$  e  $y(\mu)$  acima são idênticos, exceto a parametrização. Veja a figura (V.2).

## V.2 Algoritmo de Planos de Corte para Resolver o Problema de Maximização

Nesta seção apresentamos um algoritmo conceitual de planos de corte em centro analítico para resolver um problema geral de maximização de uma função côncava com infinitas restrições lineares. Na literatura existente encontramos um algoritmo

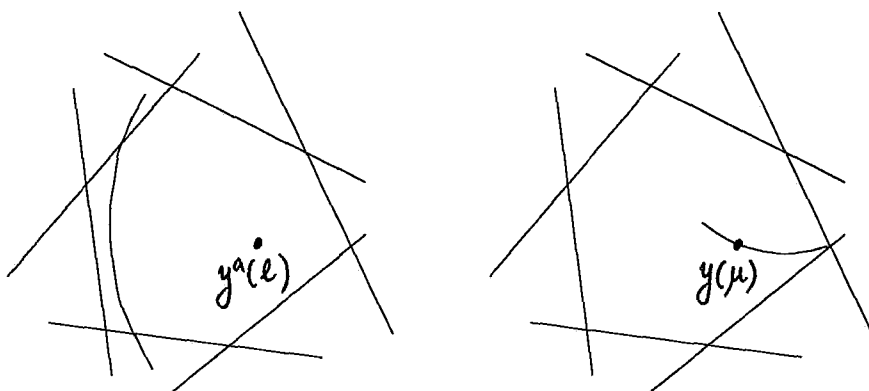


Figura V.2: Centro analítico de um conjunto de nível e o correspondente ponto central no politopo

conceitual de planos de corte em centro analítico que resolve uma classe de problemas da programação convexa [42] bem mais complexo do que iremos apresentar. Mais adiante faremos as comparações entre o algoritmo conceitual e o nosso algoritmo com o objetivo de entender o último.

### Algoritmo Conceitual: Preliminares

Considere  $\Omega \subset \mathbb{R}_{++}^m$  um conjunto convexo de interior (topológico) não-vazio e definido por um sistema de desigualdades lineares infinito. Seja o problema

$$\begin{array}{ll} \text{maximizar} & f(y) \\ \text{sujeito a} & y \in \Omega, \end{array}$$

onde  $f : \mathbb{R}^m \rightarrow \mathbb{R}$  é uma função côncava duas vezes diferenciável em  $\Omega$ .

Vamos supor que  $\Omega$  seja limitado de modo que

$$\Omega \subset \Omega^0 = \{y \in \mathbb{R}_{++}^m : y \leq u\},$$

onde  $u \in \mathbb{R}^m$  é um vetor dado.  $\Omega$  é definido implicitamente por um Oráculo, o qual para um  $y^0 \in \Omega^0$  responde se  $y^0 \in \Omega$  ou fornece um ou mais hiperplanos separadores

$$a_j^T y \leq a_j^T y^0, \quad j = 1, \dots, \eta,$$

onde  $a_j \in \mathbb{R}^m$  para  $j = 1, \dots, \eta$  e  $1 \leq \eta \leq m$ , tal que

$$\Omega \subset \Omega^0 \cap \{y \in \mathbb{R}^m : a_j^T y \leq a_j^T y^0, \quad j = 1, \dots, \eta\}.$$

Para encontrar  $y^* \in \Omega$  tal que  $l^* = f(y^*) \geq f(y)$ ,  $\forall y \in \Omega$ , vamos utilizar o algoritmo de planos de corte no centro analítico que resolve um problema de viabilidade convexa.

Um problema de viabilidade convexa é um problema do tipo: encontrar  $y \in \Omega$ , onde  $\Omega \subset \mathbb{R}_{++}^m$  é um conjunto convexo de interior (topológico) não-vazio.

Existe uma equivalência entre os dois problemas definidos acima:

dado  $\varepsilon \geq 0$ ,

$$\text{encontrar } y^* \in \Omega \text{ tal que } \max_{y \in \Omega} f(y) - f(y^*) \leq \varepsilon$$

equivale a

$$\text{encontrar } y^* \text{ no conjunto convexo } \{y \in \Omega : f(y) - \max_{y \in \Omega} f(y) + \varepsilon \geq 0\}.$$

O algoritmo de planos de corte em centro analítico constrói uma sequência de centros analíticos  $\{y^k\}$ . As respostas do Oráculo em  $y^k$  definem juntamente com o politopo inicial  $\Omega^0$  uma aproximação poliédrica  $\Omega^k$  de  $\Omega$  de modo que

$$\Omega^0 \supset \dots \supset \Omega^k \supset \Omega^{k+1} \supset \dots \supset \Omega.$$

Portanto, ao utilizar o algoritmo de planos de corte que resolve um problema de viabilidade convexa para resolver um problema de maximização, estamos incorporando a  $\Omega^0$  a restrição

$$f(y) - l^0 \geq 0,$$

onde  $l^0$  é um limite inferior conhecido para o valor ótimo da função  $f$ . E a partir do novo conjunto convexo

$$\Omega^0(l) = \{y \in \mathbb{R}_{++}^m : y \leq u, f(y) - l^0 \geq 0\} \subset \Omega^0,$$

damos início ao algoritmo de planos de corte em centro analítico. Veja o modelo do algoritmo a seguir.

**Algoritmo 7 – Algoritmo Conceitual de Planos de Corte em Centro Analítico para Maximizar uma Função Côncava com Infinitas Restrições Lineares**

**Dados**  $f(\cdot)$ ,  $l^0$ ,  $l^*$  e  $u$ .

**Inicialize**

$$A^0 = I \in \mathbb{R}^{m \times m}, \quad c^0 = u \in \mathbb{R}^m,$$

$$\Omega^0(l) = \{y \in \mathbb{R}_{++}^m : (A^0)^T y \leq c^0, f(y) - l^0 \geq 0\} \quad e$$

$y^0(l)$  o centro analítico de  $\Omega^0(l)$ .

**Faça:**  $k = 0$ ;

**Enquanto**  $l^k < l^*$ , **faça:**

• consulte o Oráculo em  $y^k(l)$ , o qual responde:

– se  $y^k(l) \notin \Omega$ , então adicione cortes a  $\Omega^k(l)$ .

Seja  $\eta$  o número de cortes do tipo  $(A_o)^T y \leq c_o$ .

Faça  $A^{k+1} = [A^k \quad A_o] \in \mathbb{R}^{m \times (p+\eta)}$  e  $c^{k+1} = \begin{pmatrix} c^k \\ c_o \end{pmatrix} \in \mathbb{R}^{p+\eta}$ .

Defina o novo conjunto de nível

$$\Omega^{k+1}(l) = \{y \in \mathbb{R}_{++}^m : (A^{k+1})^T y \leq c^{k+1}, f(y) - l^k \geq 0\} \quad e$$

encontre o seu centro analítico  $y^{k+1}(l)$ .

– se  $y^k(l) \in \Omega$ , então translade o corte convexo.

Seja  $l^{k+1} = f(y^k(l)) \geq l^k$ . Defina o novo conjunto de nível

$$\Omega^{k+1}(l) = \{y \in \mathbb{R}_{++}^m : (A^k)^T y \leq c^k, f(y) - l^{k+1} \geq 0\} \quad e$$

encontre o seu centro analítico  $y^{k+1}(l)$ .

• faça  $k = k + 1$ ;

**Fim**

**Fim**

No algoritmo apresentado acima, observe que  $l^k \leq l^*$ ,  $\forall k \geq 0$ , pois  $l^k = f(y^k(l))$  onde  $y^k(l) \in \Omega$ .

A seguir vamos entender o nosso algoritmo analisando o algoritmo acima quando este maximiza a função objetivo do problema interno dual num conjunto definido por infinitas restrições lineares.

De acordo com o lema (V.1), observamos que a parametrização utilizada no algoritmo acima é o limite inferior de  $f$ , enquanto que o nosso algoritmo utiliza  $\mu$  como parâmetro.

Quando fixamos  $\mu$  e encontramos aproximadamente um novo ponto central associado a  $\mu$  após a introdução de  $\eta$  cortes em nosso algoritmo, estamos mantendo o mesmo limite inferior no algoritmo conceitual e encontrando um novo centro analítico aproximado do conjunto de nível modificado após a adição de novos cortes.

Quando o algoritmo conceitual atualiza o limite inferior para o valor ótimo da função e acha um novo centro analítico do novo conjunto de nível resultante, significa que, na parametrização feita por  $\mu$ , o algoritmo encontra um novo ponto central reduzindo  $\mu$ . De acordo com o Teorema IV.2 no capítulo IV, este novo ponto obtido majora a função objetivo.

A seguir damos uma idéia de como seriam as iterações do algoritmo acima na figura (V.3).

### V.3 Introdução à Análise de Convergência

A seguir faremos uma introdução à análise de convergência e complexidade do algoritmo conceitual de planos de corte em centro analítico para resolver um problema geral de maximização de uma função côncava com infinitas restrições lineares.

Iniciamos a análise de convergência estudando a variação dos po-

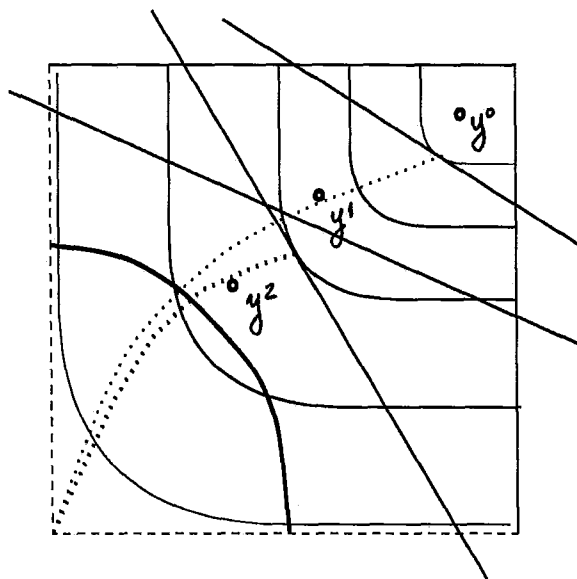


Figura V.3: Iterações do algoritmo

tenciais dos conjuntos de níveis após a atualização do limite inferior e após a introdução de um corte linear. Não fazemos a análise do número de passos de Newton necessários para a centralização. Nosso estudo segue utilizando o limite inferior do valor máximo da função objetivo como parâmetro na determinação de centros analíticos. Vale lembrar que a introdução à análise que fazemos é mais simples, comparando-a com aquela que encontramos em Mokhtarian e Goffin [42].

## Complexidade

Como vimos, existe uma equivalência entre os problemas de viabilidade convexa e maximizar uma função côncava em um conjunto convexo definido por uma infinidade de restrições lineares. Portanto, as complexidades dos algoritmos que resolvem estes problemas devem ser iguais.

A análise de complexidade feita por Ye em [51] quanto ao número total de consultas ao Oráculo do algoritmo de múltiplos planos de corte em centro analítico que resolve um problema de viabilidade convexa, no qual o conjunto é definido por um número infinito de desigualdades lineares, é de  $O^*\left(\frac{m^2}{\epsilon^2}\right)$ . A notação  $O^*$  indica que termos de ordem inferior foram ignorados e  $\epsilon > 0$  é o raio da bola

contida em  $\Omega$ .

Considerando que  $\eta \leq m$  é o número de cortes dados pelo Oráculo por iteração, o algoritmo pára após  $O^*\left(\frac{\eta^2 m^2}{\epsilon^2}\right)$  planos de corte gerados.

O número de passos de Newton para calcular um novo centro analítico aproximado de um politopo após  $\eta$  cortes é no máximo  $O^*\left(\frac{\eta^2 m^2}{\epsilon^2} \log \frac{1}{\epsilon}\right)$

Quando não existir mais cortes a fazer no conjunto convexo, então o algoritmo que utiliza o método primal-dual preditor-corretor inviável encontrará o ponto de máximo em  $O(\sqrt{n}L)$  passos para um problema de programação linear com dimensão  $n$ .

## O Potencial

Um dos critérios utilizados para avaliar o progresso de um algoritmo de otimização é medir o conjunto que contém o conjunto solução do problema de otimização. O algoritmo da bissecção que encontra uma raiz de uma função continua é um exemplo de algoritmo que se utiliza deste critério.

Nos algoritmos de planos de corte ou geração de colunas, os conjuntos contentores do conjunto solução são politopos. Para estes, foi criada uma medida analítica associada ao centro analítico do conjunto.

Seja  $\Omega$  um politopo no  $\mathbb{R}^m$  definido por  $n$  desigualdades lineares

$$\Omega = \{y \in \mathbb{R}_{++}^m : c_j - a_j^T y \geq 0, j = 1, \dots, n\}.$$

Podemos representar  $\Omega$  em termos da variável  $s$  como

$$S_\Omega = \{s \in \mathbb{R}^n : s = c - A^T y, s \geq 0, \text{ para algum } y \in \mathbb{R}_{++}^m\}.$$

O potencial de  $\Omega$  é definido por

$$P(\Omega) := \max_{y \in \Omega} \phi(y) = \sum_{j=1}^n \log(c_j - a_j^T y).$$



Isto é,  $P(\Omega) = \phi(y^a)$ , onde  $y^a$  é o centro analítico de  $\Omega$ .

Seja o elipsóide regular centrado em  $s \in S_\Omega$  dado por

$$E_s = \{t \in \mathbb{R}^n : \|S^{-1}(t - s)\| \leq 1\} \subset \mathbb{R}_+^n.$$

O volume do elipsóide é dado pela fórmula

$$\text{vol}(E_s) = \pi_n \det(S) = \pi_n \prod_{j=1}^n s_j,$$

onde  $\pi_n$  é o volume de uma bola unitária no  $\mathbb{R}^n$ . O volume de  $E_s$  será máximo em  $\mathbb{R}_+^n$  quando o seu centro for  $s^a = c - A^T y^a$ , visto que  $y^a$  é o ponto de máximo de  $\sum_{j=1}^n \log s_j$ .

Portanto, para medir  $\Omega$  podemos utilizar a função  $\prod_{j=1}^n s_j$  calculada no ponto  $y^a$ , ou, então, tomar como medida de  $\Omega$  o valor  $\exp[P(\Omega)]$  [52].

Dado  $t \in E_{s^a}$ . Seja  $t = c - A^T y$  para algum valor de  $y \in \mathbb{R}^m$ . Fazendo uma mudança de variável em  $E_{s^a}$ , obtemos o seguinte elipsóide de centro em  $y^a$  inteiramente contido em  $\Omega$

$$E_{y^a} = \{y \in \mathbb{R}^m : \|X^a A^T (y - y^a)\| \leq 1\},$$

denotado por elipsóide de Dikin. Veja a figura (V.4).

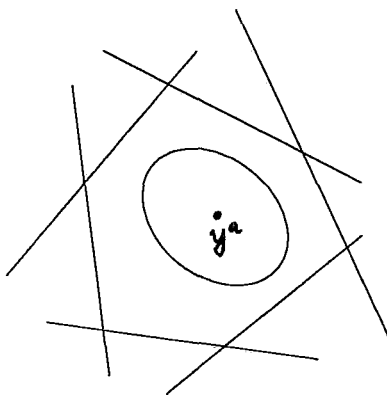


Figura V.4: Elipsóide de Dikin

## O conjunto de nível e sua medida

Seja  $f$  é uma função côncava duas vezes diferenciável  $f : \mathbb{R}_+^m \rightarrow \mathbb{R}$  e  $l$  é um limite inferior de  $f$ . Conseqüentemente, vale a proposição:

$$f(y) \leq f(y') + \nabla f(y')(y - y') \quad \forall y, y' \in \mathbb{R}_+^m.$$

Seja  $\Omega(l)$  o conjunto de nível, convexo e limitado no  $\mathbb{R}^m$ , definido por  $n$  desigualdades lineares e uma desigualdade convexa

$$\Omega(l) = \{y \in \mathbb{R}_{++}^m : c_j - a_j^T y \geq 0, j = 1, \dots, n, f(y) - l \geq 0\}, \quad (\text{V.4})$$

e seu conjunto interior denotado por  $\overset{\circ}{\Omega}(l)$ . Dado  $y \in \overset{\circ}{\Omega}(l)$ , define-se por função barreira  $\phi_l : \mathbb{R}_{++}^m \rightarrow \mathbb{R}$  tal que

$$\phi_l(y) := \sum_{j=1}^n \log s_j + \log(f(y) - l),$$

onde  $s_j := c_j - a_j^T y$  é a folga da  $j$ -ésima desigualdade linear.

Como  $\Omega(l) \subset \Omega$ , temos que  $\Omega(l)$  é limitado e visto que  $\Omega(l)$  é não-vazio, então podemos definir

$$y^a(l) := \operatorname{argmax}\{\phi_l(y) : y \in \overset{\circ}{\Omega}(l)\}$$

e denotar  $y^a(l)$  como sendo o centro analítico do conjunto  $\Omega(l)$ . Das condições de otimalidade para que  $y^a(l)$  seja o ponto de máximo, isto é,  $\nabla \phi_l(y^a(l)) = 0$ , temos que, existe  $s^a(l) > 0$ , tal que

$$Ax^a(l) = \frac{1}{f(y^a(l)) - l} \nabla f(y^a(l)), \quad s^a = c - A^T y^a(l) > 0 \quad \text{e} \quad X^a(l) s^a(l) = e. \quad (\text{V.5})$$

O potencial de  $\Omega(l)$  é definido por

$$P(\Omega(l)) := \max_{y \in \Omega(l)} \phi_l(y).$$

**Lema V.2** *Considere o elipsóide em  $\mathbb{R}_+^m$  com centro em  $y^a(l)$*

$$E_{y^a(l)} = \{y \in \mathbb{R}^m : \|\tilde{X}^a(l) \tilde{A}^T (y - y^a(l))\| \leq 1\}, \quad (\text{V.6})$$

onde

$$\tilde{x}^a(l) = \begin{pmatrix} x^a(l) \\ \frac{1}{f(y^a(l))-1} \end{pmatrix} \in \mathbb{R}^{n+1} \quad e \quad \tilde{A} = [A \quad -\nabla f(y^a(l))] \in \mathbb{R}^{m \times (n+1)}.$$

Então  $E_{y^a(l)} \subset \Omega$ .

*Prova.* Dado  $\bar{y} \in E_{y^a(l)}$ , então

$$\begin{aligned} \|\tilde{X}^a(l)\tilde{A}^T(\bar{y} - y^a(l))\| &= \left\| \begin{bmatrix} X^a(l) & 0 \\ 0 & \frac{1}{f(y^a(l))-1} \end{bmatrix} \begin{pmatrix} A^T \\ -\nabla f(y^a(l))^T \end{pmatrix} (\bar{y} - y^a(l)) \right\| \\ &= \left\| \begin{pmatrix} X^a(l)A^T(\bar{y} - y^a(l)) \\ -\frac{\nabla f(y^a(l))^T(\bar{y} - y^a(l))}{f(y^a(l))-1} \end{pmatrix} \right\| \leq 1. \end{aligned}$$

Da propriedade de  $\|x\|_\infty \leq \|x\|$ , a desigualdade acima equivale a

$$-e \leq \begin{pmatrix} X^a(l)A^T(\bar{y} - y^a(l)) \\ \frac{\nabla f(y^a(l))^T(\bar{y} - y^a(l))}{f(y^a(l))-1} \end{pmatrix} \leq e.$$

Para as primeiras  $n$  componentes do vetor, temos

$$\begin{aligned} X^a(l)A^T(\bar{y} - y^a(l)) \leq e &\Leftrightarrow X^a(l)(A^T\bar{y} - A^Ty^a(l)) \leq e \\ &\Leftrightarrow X^a(l)(-c + A^T\bar{y} + c - A^Ty^a(l)) \leq e \\ &\Leftrightarrow -X^a(l)c + X^a(l)A^T\bar{y} + e \leq e. \end{aligned}$$

Como  $X^a(l)$  é não singular, temos  $-c + A^T\bar{y} \leq 0$  ou  $c - A^T\bar{y} \geq 0$ . ■

Para ilustrar, apresentamos  $E_{y^a(l)}$  na figura (V.5).

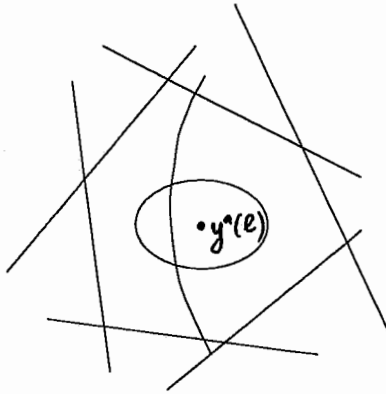


Figura V.5: Elipsóide com centro em  $y^a(l)$

A seguir vamos analisar o “estreitamento” natural dos conjuntos contenedores determinado pelo algoritmo de planos de corte em centro analítico quando

este resolve um problema geral de maximização. Este estudo diz respeito ao centro analítico de fato e utiliza a parametrização dada por  $l$ , conforme visto na seção (V.1). Nosso intuito é iniciar a análise do decréscimo do potencial.

### $\beta$ -translação de um corte convexo

Vamos agora analisar como o potencial  $P(\Omega(l))$  se comportará quando  $\Omega(l)$  é modificado trasladando a desigualdade convexa existente. Fazemos a  $\beta$ -translação de um corte convexo de maneira análoga à  $\beta$ -translação de um corte linear. O novo conjunto resultante de uma  $\beta$ -translação de um corte convexo de  $\Omega(l)$  é

$$\bar{\Omega}(l) = \{y \in \mathbb{R}_{++}^m : c_j - a_j^T y \geq 0, j = 1, \dots, n, f(y) - l - (1 - \beta)(f(y^a(l)) - l) \geq 0\},$$

onde  $\beta \in [0, 1)$  é fixo. Quando  $\beta = 0$ , a última desigualdade passa pelo centro analítico  $y^a(l)$  de  $\Omega(l)$ . Com  $\beta \in (0, 1)$ , o corte é feito em algum lugar entre o centro analítico e o antigo corte convexo (dado por  $\beta = 1$ ). Também podemos escrever o corte convexo como  $f(y) - \beta l - (1 - \beta)f(y^a(l)) \geq 0$ . Veja a figura (V.6). Considere

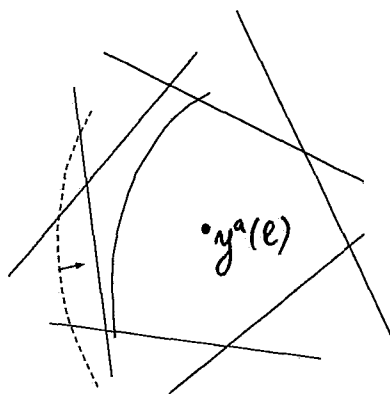


Figura V.6:  $\beta$ -translação de um corte convexo: caso  $0 < \beta < 1$

$\bar{y}^a(l)$  o centro analítico de  $\bar{\Omega}(l)$ .

**Lema V.3** *Sejam  $\Omega(l)$  (V.4) e  $\bar{\Omega}(l)$  dado como acima. Seja  $\beta \in [0, 1)$  fixo . Então*

$$P(\bar{\Omega}(l)) \leq P(\Omega(l)) - (1 - \beta).$$

*Prova.* Considerando que  $y^a(l)$  é o centro analítico de  $\Omega(l)$ , temos por (V.5) que

$$e^T X^a(l) s^a(l) = (x^a(l))^T (c - A^T y^a(l)) = c^T x^a(l) - (Ax^a(l))^T y^a(l).$$

Por outro lado,

$$e^T X^a(l) s^a(l) = \sum_{j=1}^n x^a(l)_j s^a(l)_j = n.$$

Então,

$$c^T x^a(l) = n + (Ax^a(l))^T y^a(l).$$

Repare que sendo  $f$  uma função côncava, temos que

$$f(y^a(l)) \geq f(\bar{y}^a(l)) + \nabla f(y^a(l))^T (y^a(l) - \bar{y}^a(l)).$$

Agora, considere o seguinte

$$\begin{aligned} e^T X^a(l) \bar{s}^a(l) &= \sum_{j=1}^n x^a(l)_j \bar{s}^a(l)_j \\ &= (x^a(l))^T (c - A^T \bar{y}^a(l)) \\ &= c^T x^a(l) - (Ax^a(l))^T \bar{y}^a(l) \\ &= n + (Ax^a(l))^T y^a(l) - (Ax^a(l))^T \bar{y}^a(l) \\ &= n + (Ax^a(l))^T (y^a(l) - \bar{y}^a(l)) \\ &= n + \frac{\nabla f(y^a(l))^T (y^a(l) - \bar{y}^a(l))}{f(y^a(l)) - l}. \end{aligned}$$

Daí, temos

$$\begin{aligned}
\frac{\exp[P(\bar{\Omega}(l))]}{\exp[P(\Omega(l))]} &= \frac{(f(\bar{y}^a(l)) - l - (1 - \beta)(f(y^a(l)) - l)) \prod_{j=1}^n \bar{s}^a(l)_j}{(f(y^a(l)) - l) \prod_{j=1}^n s^a(l)_j} \\
&= \frac{f(\bar{y}^a(l)) - l - (1 - \beta)(f(y^a(l)) - l)}{f(y^a(l)) - l} \prod_{j=1}^n x^a(l)_j \bar{s}^a(l)_j \\
&\leq \left( \frac{\frac{f(\bar{y}^a(l)) - l - (1 - \beta)(f(y^a(l)) - l)}{f(y^a(l)) - l} + \sum_{j=1}^n x^a(l)_j \bar{s}^a(l)_j}{n + 1} \right)^{n+1} \\
&= \left( \frac{\frac{f(\bar{y}^a(l)) - l}{f(y^a(l)) - l} - (1 - \beta) + n + \frac{\nabla f(y^a(l))^T (y^a(l) - \bar{y}^a(l))}{f(y^a(l)) - l}}{n + 1} \right)^{n+1} \\
&= \left( \frac{n + \frac{f(\bar{y}^a(l)) + \nabla f(y^a(l))^T (y^a(l) - \bar{y}^a(l)) - l}{f(y^a(l)) - l} - (1 - \beta)}{n + 1} \right)^{n+1} \\
&\leq \left( \frac{n + \frac{f(y^a(l)) - l}{f(y^a(l)) - l} - (1 - \beta)}{n + 1} \right)^{n+1} \\
&= \left( \frac{n + 1 - (1 - \beta)}{n + 1} \right)^{n+1} \\
&= \left( 1 - \frac{1 - \beta}{n + 1} \right)^{n+1} \\
&\leq \exp(-(1 - \beta)).
\end{aligned}$$

Logo

$$\exp[P(\bar{\Omega}(l))] \leq \exp[P(\Omega(l))] \exp(-(1 - \beta)). \blacksquare$$

Este resultado é igual ao obtido para a variação do potencial após um corte linear ser  $\beta$ -transladado num politopo. Aquele caso representa um passo do algoritmo de planos de corte em centro analítico para resolver um problema de viabilidade, [52].

### Adição de um corte linear

Considere agora o caso em que um corte linear é adicionado ao conjunto convexo  $\Omega(l)$ . O novo conjunto convexo é

$$\bar{\Omega}(l) = \{y \in \mathbb{R}_{++}^m : c_j - a_j^T y \geq 0, j = 1, \dots, n, c_o - a_o^T y \geq 0, f(y) - l \geq 0\},$$

onde  $c_o = a_o^T y^a(l)$ , com seu centro analítico denotado por  $\bar{y}^a(l)$ . Note que este caso refere-se à adição de um hiperplano em  $\Omega(l)$  passando pelo seu centro analítico  $y^a(l)$ . Veja a figura (V.7)

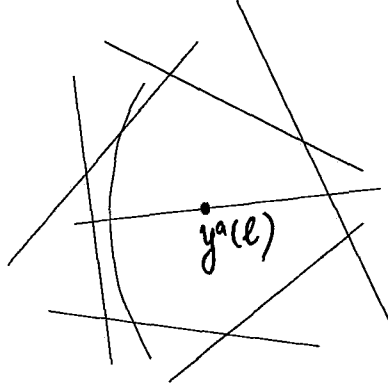


Figura V.7: Adição de um corte linear a  $\Omega(l)$  em  $y^a(l)$

Das propriedades de  $\bar{y}^a(l)$  ser o centro analítico de  $\bar{\Omega}(l)$ , temos que  $\exists \bar{x}^a(l) > 0 (\in \mathbb{R}^n)$  e  $\bar{x}_o^a(l) (\in \mathbb{R})$  tal que

$$\begin{aligned} \bar{X}^a(l)(c - A^T \bar{y}^a(l)) &= e, & \bar{x}_o^a(l)(c_o - a_o^T \bar{y}^a(l)) &= 1, \\ A \bar{x}^a(l) + a_o \bar{x}_o^a(l) &= \frac{\nabla f(\bar{y}^a(l))}{f(\bar{y}^a(l)) - l}. \end{aligned}$$

**Lema V.4** Seja  $r(\Omega(l)) = \sqrt{a_o^T (\tilde{A}(\tilde{X}^a(l))^2 \tilde{A}^T)^{-1} a_o}$ , onde

$$\tilde{x}^a(l) = \begin{pmatrix} x^a(l) \\ \frac{1}{f(y^a(l)) - l} \end{pmatrix} \in \mathbb{R}^{n+1} \quad \text{e} \quad \tilde{A} = [A \quad -\nabla f(y^a(l))] \in \mathbb{R}^{m \times (n+1)}.$$

E sejam  $\Omega(l)$  (V.4) e  $\bar{\Omega}(l)$  dado como acima. Então

$$P(\bar{\Omega}(l)) \leq P(\Omega(l)) + \log(r(\Omega(l))) + 2 \log 2 - 1.5.$$

*Prova.* Sendo  $y^a(l)$  o centro analítico de  $\Omega(l)$ , temos por (V.5) que

$$e^T X^a(l) s^a(l) = (x^a(l))^T (c - A^T y^a(l)) = c^T x^a(l) - (A x^a(l))^T y^a(l),$$

e por outro lado, temos

$$e^T X^a(l) s^a(l) = \sum_{j=1}^n x^a(l)_j s^a(l)_j = n,$$

logo

$$c^T x^a(l) = n + (Ax^a(l))^T y^a(l).$$

Da concavidade da função  $f$ , temos

$$f(y^a(l)) \geq f(\bar{y}^a(l)) + \nabla f(y^a(l))^T (y^a(l) - \bar{y}^a(l)).$$

Sendo  $\bar{y}^a(l)$  o centro analítico de  $\bar{\Omega}(l)$ , considere o desenvolvimento abaixo para as  $n$  primeiras componentes de  $x^a(l)$  e  $\bar{s}^a(l)$  mais o produto de  $1/(f(y^a(l)) - l)$  por  $f(\bar{y}^a(l)) - l$ :

$$\begin{aligned} e^T X^a(l) \bar{s}^a(l) + \frac{f(\bar{y}^a(l)) - l}{f(y^a(l)) - l} &= (x^a(l))^T \bar{s}^a(l) + \frac{f(\bar{y}^a(l)) - l}{f(y^a(l)) - l} \\ &= (x^a(l))^T (c - A^T \bar{y}^a(l)) + \frac{f(\bar{y}^a(l)) - l}{f(y^a(l)) - l} \\ &= c^T x^a(l) - (Ax^a(l))^T \bar{y}^a(l) + \frac{f(\bar{y}^a(l)) - l}{f(y^a(l)) - l} \\ &= n + (Ax^a(l))^T y^a(l) - (Ax^a(l))^T \bar{y}^a(l) + \frac{f(\bar{y}^a(l)) - l}{f(y^a(l)) - l} \\ &= n + (Ax^a(l))^T (y^a(l) - \bar{y}^a(l)) + \frac{f(\bar{y}^a(l)) - l}{f(y^a(l)) - l} \\ &= n + \frac{\nabla f(y^a(l))^T (y^a(l) - \bar{y}^a(l))}{f(y^a(l)) - l} + \frac{f(\bar{y}^a(l)) - l}{f(y^a(l)) - l} \\ &= n + \frac{f(\bar{y}^a(l)) + \nabla f(y^a(l))^T (y^a(l) - \bar{y}^a(l)) - l}{f(y^a(l)) - l} \\ &\leq n + \frac{f(y^a(l)) - l}{f(y^a(l)) - l} \\ &= n + 1. \end{aligned}$$



Da nova folga linear associada a  $\bar{y}^a(l)$ , obtemos

$$\begin{aligned}
\bar{s}^a(l)_o &= a_o^T(y^a(l) - \bar{y}^a(l)) \\
&= a_o^T(\tilde{A}(\tilde{X}^a(l))^2 \tilde{A}^T)^{-1}(\tilde{A}(\tilde{X}^a(l))^2 \tilde{A}^T)(y^a(l) - \bar{y}^a(l)) \\
&= a_o^T(\tilde{A}(\tilde{X}^a(l))^2 \tilde{A}^T)^{-1} \tilde{A}(\tilde{X}^a(l))^2 \begin{pmatrix} A^T \\ -\nabla f(y^a(l))^T \end{pmatrix} (y^a(l) - \bar{y}^a(l)) \\
&= a_o^T(\tilde{A}(\tilde{X}^a(l))^2 \tilde{A}^T)^{-1} \tilde{A}(\tilde{X}^a(l)) \begin{bmatrix} X^a(l) & 0 \\ 0 & \frac{1}{f(y^a(l))-l} \end{bmatrix} \begin{pmatrix} A^T(y^a(l) - \bar{y}^a(l)) \\ -\nabla f(y^a(l))^T(y^a(l) - \bar{y}^a(l)) \end{pmatrix} \\
&= a_o^T(\tilde{A}(\tilde{X}^a(l))^2 \tilde{A}^T)^{-1} \tilde{A}(\tilde{X}^a(l)) \begin{pmatrix} X^a(l)(A^T y^a(l) - A^T \bar{y}^a(l)) \\ \frac{-\nabla f(y^a(l))^T(y^a(l) - \bar{y}^a(l))}{f(y^a(l))-l} \end{pmatrix} \\
&= a_o^T(\tilde{A}(\tilde{X}^a(l))^2 \tilde{A}^T)^{-1} \tilde{A}(\tilde{X}^a(l)) \begin{pmatrix} X^a(l)(-c + A^T y^a(l) + c - A^T \bar{y}^a(l)) \\ \frac{-\nabla f(y^a(l))^T(y^a(l) - \bar{y}^a(l))}{f(y^a(l))-l} \end{pmatrix} \\
&= a_o^T(\tilde{A}(\tilde{X}^a(l))^2 \tilde{A}^T)^{-1} \tilde{A}(\tilde{X}^a(l)) \begin{pmatrix} X^a(l)(\bar{s}^a(l) - s^a(l)) \\ \frac{-\nabla f(y^a(l))^T(y^a(l) - \bar{y}^a(l))}{f(y^a(l))-l} \end{pmatrix} \\
&= a_o^T(\tilde{A}(\tilde{X}^a(l))^2 \tilde{A}^T)^{-1} \tilde{A}(\tilde{X}^a(l)) \begin{pmatrix} X^a(l)\bar{s}^a(l) - e \\ \frac{-\nabla f(y^a(l))^T(y^a(l) - \bar{y}^a(l))}{f(y^a(l))-l} \end{pmatrix} \\
&\leq \|\tilde{X}^a(l)\tilde{A}^T(\tilde{A}(\tilde{X}^a(l))^2 \tilde{A}^T)^{-1} a_o\| \left\| \begin{pmatrix} X^a(l)\bar{s}^a(l) - e \\ \frac{-\nabla f(y^a(l))^T(y^a(l) - \bar{y}^a(l))}{f(y^a(l))-l} \end{pmatrix} \right\| \\
&= r(\Omega(l)) \left\| \begin{pmatrix} X^a(l)\bar{s}^a(l) - e \\ \frac{-\nabla f(y^a(l))^T(y^a(l) - \bar{y}^a(l))}{f(y^a(l))-l} \end{pmatrix} \right\| \\
&= r(\Omega(l)) \sqrt{\sum_{j=1}^n (x^a(l)_j \bar{s}^a(l)_j - 1)^2 + \left( \frac{\nabla f(y^a(l))^T(y^a(l) - \bar{y}^a(l))}{f(y^a(l)) - l} \right)^2} \\
&\leq r(\Omega(l)) \sqrt{\sum_{j=1}^n (x^a(l)_j \bar{s}^a(l)_j - 1)^2 + \left( \frac{f(y^a(l)) - f(\bar{y}^a(l))}{f(y^a(l)) - l} \right)^2} \\
&= r(\Omega(l)) \sqrt{\sum_{j=1}^n (x^a(l)_j \bar{s}^a(l)_j - 1)^2 + \left( \frac{f(y^a(l)) - l - f(\bar{y}^a(l)) + l}{f(y^a(l)) - l} \right)^2} \\
&= r(\Omega(l)) \sqrt{\sum_{j=1}^n (x^a(l)_j \bar{s}^a(l)_j - 1)^2 + \left( \frac{(f(y^a(l)) - l) - (f(\bar{y}^a(l)) - l)}{f(y^a(l)) - l} \right)^2} \\
&= r(\Omega(l)) \sqrt{\sum_{j=1}^n (x^a(l)_j \bar{s}^a(l)_j - 1)^2 + \left( 1 - \frac{f(\bar{y}^a(l)) - l}{f(y^a(l)) - l} \right)^2} \\
&= r(\Omega(l)) \left\| \begin{pmatrix} X^a(l)\bar{s}^a(l) - e \\ \frac{f(\bar{y}^a(l))-l}{f(y^a(l))-l} - 1 \end{pmatrix} \right\|,
\end{aligned}$$

onde  $r(\Omega(l))$  é o valor máximo do seguinte problema de região de confiança (V.6):

$$\text{maximizar } s_o(l) = a_o^T(y^a(l) - y)$$

$$\text{sujeito a } y \in E_{y^a(l)} = \{y \in \mathbb{R}_+^m : (y - y^a(l))^T \tilde{A}(\tilde{X}^a(l))^2 \tilde{A}^T(y - y^a(l)) \leq 1\},$$

quando o resolvemos utilizando o método afim-escala [12]. Daí, temos

$$\begin{aligned} \frac{\exp[P(\bar{\Omega}(l))]}{r(\Omega(l)) \exp[P(\Omega(l))]} &= \frac{\bar{s}^a(l)_o(f(\bar{y}^a(l)) - l) \prod_{j=1}^n \bar{s}^a(l)_j}{r(\Omega(l))(f(y^a(l)) - l) \prod_{j=1}^n s^a(l)_j} \\ &= \frac{\bar{s}^a(l)_o(f(\bar{y}^a(l)) - l) \prod_{j=1}^n \bar{s}^a(l)_j}{r(\Omega(l))(f(y^a(l)) - l) \prod_{j=1}^n s^a(l)_j} \\ &\leq \left\| \begin{pmatrix} X^a(l)\bar{s}^a(l) - e \\ \frac{f(\bar{y}^a(l))-l}{f(y^a(l))-l} - 1 \end{pmatrix} \right\| \frac{(f(\bar{y}^a(l)) - l) \prod_{j=1}^n \bar{s}^a(l)_j}{(f(y^a(l)) - l) \prod_{j=1}^n s^a(l)_j}. \end{aligned}$$

Considere o problema de maximização a seguir:

$$\begin{aligned} \text{maximizar } f(v) &= \|v - e\| \prod_{j=1}^{n+1} v_j \\ \text{sujeito a } e^T v &\leq n + 1, v > 0, \end{aligned}$$

onde  $v = \begin{pmatrix} X^a(l)\bar{s}^a(l) \\ \frac{f(\bar{y}^a(l))-l}{f(y^a(l))-l} \end{pmatrix} \in \mathbb{R}^{n+1}$ . Queremos uma cota superior para o valor da função-objetivo no ponto de máximo. Para o ponto de máximo [52], sem perda de generalidade, escolhemos para  $v_1 = \gamma > 1$  e  $v_2 = \dots = v_{n+1} = (n + 1 - \gamma)/n > 0$ , então

$$\begin{aligned} f(v) &\leq \sqrt{(\gamma - 1)^2 + n \left( \frac{n + 1 - \gamma}{n} - 1 \right)^2} \gamma \left( \frac{n + 1 - \gamma}{n} \right)^n \\ &= (\gamma - 1) \sqrt{\frac{n + 1}{n}} \gamma \left( \frac{n + 1 - \gamma}{n} \right)^n \\ &= 4 \sqrt{\frac{n + 1}{n}} \frac{\gamma(\gamma - 1)}{2} \left( \frac{n + 1 - \gamma}{n} \right)^n \\ &\leq 4 \sqrt{\frac{n + 1}{n}} \left( \frac{\frac{\gamma}{2} + \frac{\gamma - 1}{2} + \sum_{i=1}^n \frac{n + 1 - \gamma}{n}}{n + 2} \right)^{n+2} \\ &= 4 \sqrt{\frac{n + 1}{n}} \left( \frac{\gamma - 0.5 + \frac{n(n + 1 - \gamma)}{n}}{n + 2} \right)^{n+2} \\ &= 4 \sqrt{\frac{n + 1}{n}} \left( \frac{n + 2 - 1.5}{n + 2} \right)^{n+2} \\ &\leq 4 \exp(-1.5). \end{aligned}$$

Logo

$$\exp[P(\bar{\Omega}(l))] \leq 4r(\Omega(l)) \exp[P(\Omega(l))] \exp(-1.5). \blacksquare$$

Este resultado é igual ao obtido para a variação do potencial após a adição de um corte linear num politopo passando pelo centro analítico. Aquele caso representa um passo do algoritmo de planos de corte em centro analítico para resolver um problema de viabilidade, [52].

# Capítulo VI

## Experiência Computacional

### VI.1 Interpretação do problema estatístico

#### Iniciando APCC

De posse dos dados de um problema EMV, o Algoritmo de Planos de Corte Baseado em Centros para EMV, de agora em diante APCC, considera o tamanho da amostra como sendo a dimensão da variável dual  $y$  quando os valores dos dados são distintos. Neste caso, cada função núcleo possui multiplicidade 1. Caso contrário, a dimensão de  $y$  será o número de agrupamentos de valores distintos dos dados.

Prosseguindo, definimos para cada observação distinta da amostra uma função núcleo do estimador que é uma função de densidade de probabilidade condicionada.

Uma vez que todos os pontos suportes  $\theta_j$ , com  $j = 1, \dots, p$ , estão em  $\Gamma \subset \mathbb{R}$ , primeiro, APCC encontra o intervalo ao qual pertencem todos os pontos suportes ótimos da distribuição  $F_\Theta$  ótima. Para isto, uma busca linear é realizada em cada função núcleo do estimador  $f_{V_i|\Theta}$ ,  $i = 1, \dots, m$ , para encontrar a única moda em  $\tilde{\theta}_i$ . O intervalo é definido como

$$I = [\min_i \tilde{\theta}_i, \max_i \tilde{\theta}_i],$$

de acordo com Corolário no final do capítulo II. Podemos notar que o intervalo  $I$

está somente relacionado aos dados do problema. Ele é calculado a partir de limites inferior e superior razoavelmente escolhidos para a realização da busca linear nas funções núcleos para cada observação da amostra.

### Escolha de $u$

No começo do algoritmo APCC, a única restrição considerada é

$$y \leq u.$$

Essa restrição é exigida pelo algoritmo de planos de corte. A escolha de  $u$  é feita baseada nos dados e de modo que a restrição seja fraca, garantindo que a interseção dos hiperplanos definidos pela restrição acima contenha o conjunto solução do EMV.

Dada a formulação dual do problema interno de EMV na iteração  $k$  e sendo  $\sum_{i=1}^m t_i = n$ , temos

$$(A^k)^T y \leq ne, \quad \text{ou} \quad (a_j^k)^T y \leq n, \quad \forall j = 1, \dots, p,$$

onde  $a_j^k$  é a  $j$ -ésima coluna de  $A^k$ . Portanto, fixando  $i = 1, \dots, m$  e  $j = 1, \dots, p$ , temos

$$f_{V_i|\Theta}(v_i, \theta_j) y_i \leq n - \sum_{l=1, l \neq i}^m f_{V_l|\Theta}(v_l, \theta_j) y_l, \quad l = 1, \dots, m \text{ e } l \neq i.$$

Suponha que para cada  $i = 1, \dots, m$  fixo, escolhamos um valor para  $\theta_j$  tal que  $f_{V_i|\Theta}(v_i, \theta_j) = 1$ . Logo  $n$  é um limite superior para cada  $y_i$ , devido ao fato de que os elementos de  $A^k$  e  $y$  são não-negativos. Note que os elementos de  $A^k$  são medidas de probabilidade.

### Oráculo

Seguindo a estratégia do método ISDM em [30], em cada iteração, APCC encontra todos os máximos locais da derivada direcional da função do estimador da verossimilhança no intervalo  $I$ . Isto é feito dividindo  $I$  em subintervalos e aplicando o

método da bisseção para o subintervalo  $[a, b]$ , onde a derivada segunda da função do estimador é positiva em  $a$  e negativa em  $b$ .

Os máximos locais somente serão considerados quando a derivada direcional for positiva de acordo com a definição de corte profundo no final do capítulo III.

As funções núcleos avaliadas nesses máximos locais fornecem colunas. Logo, APCC adiciona todas as colunas tais que, para um centro analítico aproximado, este centro não é dual viável.

Verificamos que APCC pode adicionar mais que uma coluna por iteração, caracterizando o algoritmo de múltiplos planos de corte. Sabemos pelo capítulo III, que o número de cortes dados pelo Oráculo é finito.

### **Conjunto Suporte Inicial e Ótimo**

De acordo como  $u$  foi calculado, os pontos suportes iniciais são os dados distintos da amostra. Portanto a cardinalidade inicial do conjunto suporte é  $m$ .

O conjunto suporte ótimo será formado por valores de  $\theta$ , para os quais a derivada direcional se anula e o gap de complementaridade é menor que uma pre-determinada tolerância. Verificamos que os pontos suportes ótimos são aqueles que determinam as restrições ativas na última iteração do algoritmo APCC.

### **Vetor de Massa Inicial e Ótimo**

O vetor de massa inicial do algoritmo APCC,  $x^0$ , é o inverso do vetor dual inicial  $y^0$ , visto que  $\mu^0 = 1$ ,  $c^0 = u$  e  $p = m$ . Pela teoria, sabemos que no início do algoritmo a soma das massas não é 1, porém ela tende para 1 à medida que o algoritmo converge, como visto na seção (IV.1).

A variável primal  $x$ , o vetor de massas, é atualizada a cada iteração do

algoritmo APCC. Quando APCC converge, isto é, quando não for possível encontrar mais uma direção de subida, o gap de otimalidade é menor que uma pré-determinada tolerância assim como a variável dual é viável,  $x$  é exatamente o vetor de massas dos pontos suportes de  $F_{\Theta}$  ótimo. Para cada elemento não-nulo de  $x$  existe uma correspondência com um dos pontos suportes.

## VI.2 Implementação

Num algoritmo de pontos interiores o maior custo computacional de uma iteração ocorre na obtenção da direção de Newton que nada mais é do que resolver um sistema de equações lineares. No nosso algoritmo, a tarefa de resolver os sistemas do método de Mehrotra (IV.12, IV.13) e o sistema da direção do algoritmo de centralização (IV.15) demanda o maior esforço computacional. Na implementação do algoritmo, optou-se por resolver tais sistemas utilizando um método direto tendo como abordagem as equações normais. O método direto escolhido aplica a fatoração de Cholesky juntamente com as rotinas de substituição “forward” e “backward”.

Um motivo reforça a utilização da abordagem das equações normais para encontrar as direções do método de Mehrotra e de centralização para o nosso algoritmo: não há variáveis livres na formulação primal do nosso problema de otimização [1].

### Equações normais

A seguir apresentamos como exemplo o esquema da abordagem das equações normais para resolver o sistema linear correspondente a encontrar a direção de centralização do método de Mehrotra. Para o sistema posto em (IV.13), após algumas manipulações algébricas, temos:

$$\begin{aligned} (AS^{-1}XA^T + Y^{-1}Z)dy &= -\mu AS^{-1}e + AXe + AS^{-1}Xr_d + AS^{-1}Dx^a ds^a \\ &+ Y^{-1}t - Ze - Y^{-1}Dy^a dz^a + r_p. \end{aligned}$$

A direção  $dy$  é portanto a solução de um sistema de equações lineares normais. O termo normal surge de um abuso de linguagem pois a matriz  $M = AS^{-1}XA^T + Y^{-1}Z$  apresenta um termo com o produto  $A \cdot A^T$ .

Considerando que  $A$  contém a matriz identidade associada aos coeficientes dos hiperplanos definidos pelas restrições iniciais,  $A$  apresenta rank máximo em linhas. Portanto a matriz  $AS^{-1}XA^T$  é definida positiva. A matriz  $Y^{-1}Z$  é diagonal positiva, todos os elementos da diagonal são positivos. De acordo com a proposição (IV.1) no capítulo IV, a matriz  $M$  é simétrica e definida positiva. Tais características da matriz  $M$  aliadas ao fato de que os elementos da matriz são reais garantem uma única solução real ao sistema de equações lineares.

Logo, podemos resolver o sistema pela decomposição de Cholesky já que  $M$  satisfaz as condições necessárias para ser fatorada. As demais direções,  $dx$ ,  $ds$  e  $dz$ , são encontradas quando substituímos a direção  $dy$  nas expressões:

$$\begin{aligned} dx &= \mu S^{-1}e - Xe - S^{-1}Xr_d + S^{-1}XA^T dy - S^{-1}Dx^a ds^a, \\ ds &= r_d - A^T dy, \\ dz &= Y^{-1}t - Ze - Y^{-1}Zdy - Y^{-1}Dy^a dz^a. \end{aligned}$$

### Critério de Parada

Como o nosso algoritmo gera uma sequência infinita de pontos, precisamos definir como terminá-lo. O critério de parada adotado está baseado no gap de complementaridade (III.4) no capítulo III. Escolhida uma tolerância  $\varepsilon = 10^{-8}$  e dado  $(x, s, y, z) \in \overset{\circ}{\mathcal{F}}$ , o critério de parada se expressa como

$$\frac{x^T s}{1 + \|x\|_1 + \|s\|_1} \leq \varepsilon.$$

## VI.3 Resultados Numéricos

O código computacional foi escrito na linguagem C e todos os testes computacionais foram executados em uma “workstation” Sun SPARC-20 Solaris. Um dos testes computacionais é proveniente de dados reais da literatura e os demais gerados aleatoria-



mente. Para obter os resultados computacionais a seguir, utilizamos a metodologia de inicialização que obtém um ponto primal-dual viável interior para o algoritmo de centralização em APCC.

### EMV unidimensional

Executou-se o algoritmo APCC para resolver o problema EMV unidimensional em [30], cujos dados  $v_i$  e  $\sigma_i^2$  são apresentados na Tabela 1 abaixo num total de 13 observações relativas à diferença nas taxas de cura entre terapias tradicional e inovadora. Utilizamos este exemplo porque ele serviu de comparação em [30] entre os algoritmos existentes, dos quais ISDM foi o mais eficiente.

Para essa amostra o modelo considerado é o de que as variáveis aleatórias  $V_i$ ,  $i = 1, \dots, 13$ , sejam independentes e normalmente distribuídas, com variâncias fixas e conhecidas. Uma variável aleatória  $V_i$  possui distribuição normal se a função densidade de probabilidade é

$$f_{V_i|\Theta}(v_i|\theta_i) = \frac{1}{\sigma_i\sqrt{2\pi}} \exp\left[-\frac{(v_i - \theta_i)^2}{2\sigma_i^2}\right], \quad (\text{VI.1})$$

e é denotada por  $N(\theta_i, \sigma_i)$ , onde  $\theta_i$  é a média das observações e  $\sigma_i^2$  é a sua variância. Supondo que  $\Theta = (\Theta_1, \dots, \Theta_{13})$  é uma variável aleatória independente cuja função distribuição acumulada  $F_\Theta$  queremos estimar e sendo todos os dados da amostra distintos, a função-objetivo do problema EMV é

$$\mathcal{L}(F_\Theta) = \sum_{i=1}^{13} \log \int_{\Gamma} \frac{1}{\sigma_i\sqrt{2\pi}} \exp\left[-\frac{(v_i - \theta_i)^2}{2\sigma_i^2}\right] dF_\Theta(\theta).$$

Tabela 1 – Diferença nas taxas de cura entre terapias tradicional e inovadora

$v_i$	$\sigma_i$
-0.18	0.017
-0.14	0.028
-0.09	0.006
-0.07	0.001
-0.06	0.003
-0.04	0.011
0.00	0.003
0.02	0.001
0.06	0.008
0.07	0.008
0.16	0.067
0.19	0.017
0.25	0.013

Para a tolerância  $\varepsilon = 10^{-8}$ , APCC pára na iteração 14, com uma solução dual viável e o parâmetro da barreira  $\mu = 6.1809 \times 10^{-10}$ . APCC adicionou 14 colunas para encontrar a solução ótima aproximada. A estimativa do suporte ótimo é  $(-0.0557, 0.0149, 0.2103)$  e a estimativa do vetor de massa ótimo é  $(0.4156, 0.5355, 0.0491)$ . O valor ótimo da função do logaritmo do estimador da verossimilhança encontrado por APCC é 23.3398. E o valor máximo da derivada direcional na estimativa ótima é  $6.02 \times 10^{-5}$ . Tempo de execução: 0.26s.

ISDM encontra para a estimativa do suporte ótimo  $(-0.0557, 0.0150, 0.2099)$  e para a estimativa do vetor de massa ótimo  $(0.4154, 0.5355, 0.0491)$ . O valor ótimo da função do logaritmo do estimador da verossimilhança encontrado por ISDM é 23.3396. E o valor máximo da derivada direcional na estimativa ótima é  $1.44 \times 10^{-8}$ . Tempo de execução de ISDM: 3min.

Observamos que o valor ótimo da função do logaritmo do estimador da verossimilhança encontrado por APCC é maior do que o encontrado por ISDM, Uma possível explicação para isto é que utilizamos o método da bisseção para encontrar os máximos locais das derivadas direcionais, enquanto que ISDM encontra utilizando o método de Newton. De acordo com o comentário em [30], o algoritmo ISDM é sensível à escolha da metodologia para encontrar estes máximos locais. Acre-

ditamos que APCC também possua a mesma sensibilidade.

De acordo com o Teorema II.1 no final do capítulo II, os valores da derivada direcional do logaritmo do estimador da verossimilhança estão próximos de zero quando avaliada nos pontos das estimativas dos suportes ótimos dados por APCC, como podemos verificar pelo gráfico abaixo.

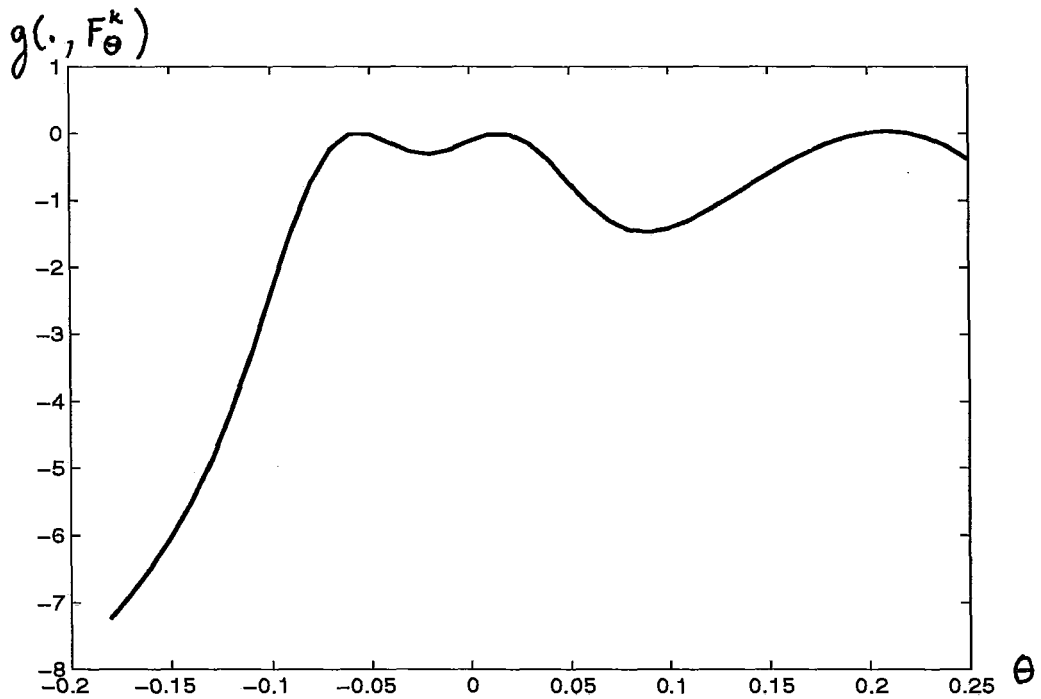


Figura VI.1: Função derivada direcional do logaritmo do estimador da verossimilhança.

### Outros testes

Para a realização de mais testes computacionais do algoritmo APCC geramos a partir de distribuições normais alguns conjuntos de dados aleatórios seguindo os procedimentos básicos encontrados na literatura, como por exemplo em [45].

Primeiramente escolhemos o número total de observações  $n$  a serem geradas em um conjunto de dados. Fixamos  $\hat{p}$  valores do parâmetro populacional  $\hat{\theta}_i$  que dão origem à distribuição mista das observações assim como as massas associadas,  $\hat{x}_i$ , e suas variâncias,  $\sigma_i^2$ .

Para cada  $i, i = 1, \dots, \hat{p}$ , conhecidos  $n$  e  $\hat{x}_i$ , passamos a gerar  $n_i = n\hat{x}_i$  valores aleatórios independentes e identicamente distribuídos conforme uma distribuição normal com média  $\hat{\theta}_i$  e desvio-padrão  $\sigma_i$  utilizando os seguintes procedimentos. Para cada par  $(\hat{\theta}_i, \sigma_i)$  dado, geramos de cada vez 2 valores aleatórios, independentes e uniformemente distribuídos em  $(0, 1)$ ,  $u_1$  e  $u_2$ . Uma variável aleatória  $U$  possui distribuição uniforme se a função densidade de probabilidade de  $U$  é dada por

$$f_U(u) = \begin{cases} 1, & 0 \leq u \leq 1 \\ 0, & \text{caso contrário} \end{cases}, u \in \mathbb{R}.$$

A partir destes valores obtemos outros 2 valores aleatórios e independentes  $z_1$  e  $z_2$  de uma distribuição normal-padrão, fazendo

$$\begin{aligned} z_1 &= \sqrt{-2 \ln u_1} \cos(2\pi u_2) \\ z_2 &= \sqrt{-2 \ln u_1} \sin(2\pi u_2) \end{aligned},$$

cuja função densidade de probabilidade é dada por

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right), \quad z \in \mathbb{R}$$

Como  $Z = (V - \theta)/\sigma$ , fixados  $\hat{\theta}_i$  e  $\sigma_i$ , temos que, para  $j = 1, 2$ ,  $v_j = \hat{\theta}_i + \sigma_i z_j$  e com isso conseguimos dois valores para as observações  $v_1$  e  $v_2$ , cuja função densidade de probabilidade é exatamente (VI.1) e denotada por  $N(\hat{\theta}_i, \sigma_i)$ . Repetimos o procedimento acima até que tenhamos  $n_i$  observações relativas ao ponto suporte  $\hat{\theta}_i$ .

Para a definição dos valores dos desvios-padrões relativos aos pontos suportes, podemos considerar um valor fixo e conhecido para todos os pontos suportes ou ainda gerar aleatoriamente seus valores dentro de um certo intervalo, por exemplo.

## Validação

Seja  $p$  o número de pontos suportes estimados. Com os valores das estimativas dos pontos suportes juntamente com as estimativas de massas podemos estimar a função

distribuição acumulada  $F_{\Theta}$  fazendo

$$\begin{aligned} \theta \in (-\infty, \theta_1) &\Rightarrow F_{\Theta}(\theta) = 0, \\ \theta \in [\theta_1, \theta_2) &\Rightarrow F_{\Theta}(\theta) = x_1, \\ \theta \in [\theta_2, \theta_3) &\Rightarrow F_{\Theta}(\theta) = x_1 + x_2, \\ &\vdots \\ \theta \in [\theta_i, \theta_{i+1}) &\Rightarrow F_{\Theta}(\theta) = x_1 + \dots + x_i, \\ &\vdots \\ \theta \in [\theta_p, \infty) &\Rightarrow F_{\Theta}(\theta) = x_1 + \dots + x_p \approx 1 \end{aligned}$$

É ilustrar a estimativa de  $F_{\Theta}$  pelo gráfico abaixo

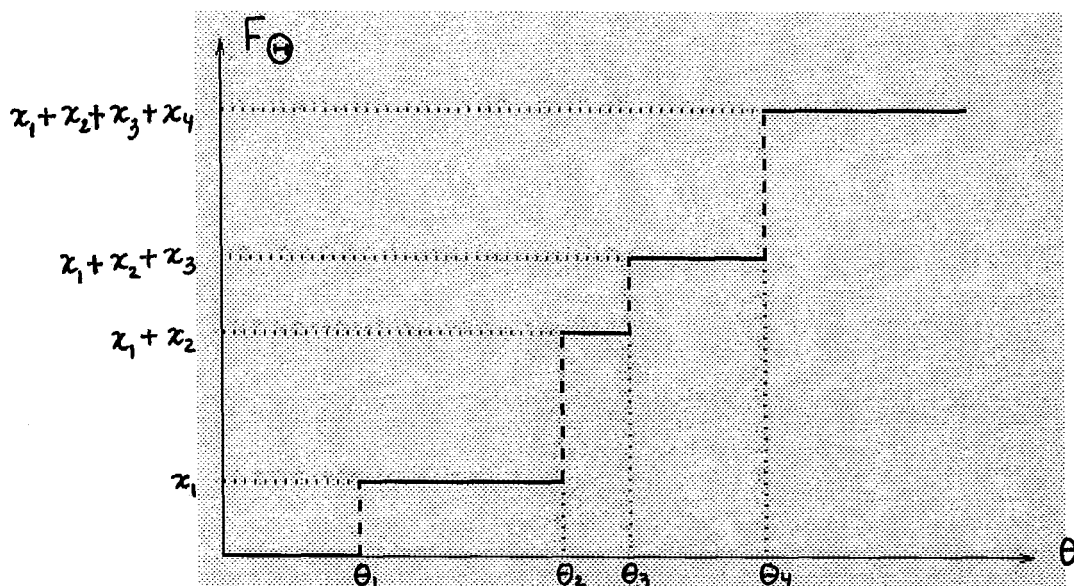


Figura VI.2: Estimativa de  $F_{\Theta}$

A partir dos pontos suportes e suas massas associados utilizados na geração de dados aleatórios, encontramos a função distribuição acumulada  $\hat{F}_{\Theta}$  teórica utilizando o mesmo processo acima. Com isso, validamos nossos resultados comparando a distribuição relativa aos dados gerados aleatoriamente,  $\hat{F}_{\Theta}$ , com a distribuição determinada pela otimização,  $F_{\Theta}$ . Agora, seja  $\hat{p}$  o número de pontos suportes usados para a geração aleatória dos dados e  $p$  o número de estimativas de pontos suportes ótimos provenientes do resultado da otimização. Essa comparação é feita calculando o erro utilizando o Erro Quadrático Médio:

$$Erro = \left( \frac{1}{|\lambda_M - \lambda_m|} \int_{\lambda_m}^{\lambda_M} (\hat{F}_{\Theta}(\lambda) - F_{\Theta}(\lambda))^2 d\lambda \right)^{1/2},$$

onde  $\lambda_m = \min\{\hat{\theta}_i, \theta_j\}$  e  $\lambda_M = \max\{\hat{\theta}_i, \theta_j\}$  para  $i = 1, \dots, \hat{p}$  e  $j = 1, \dots, p$ .

## Experimento 1

Com objetivo de mostrar a evolução da estimativa de  $\hat{F}_\Theta$  ao longo do algoritmo APCC, geramos um conjunto de 12 dados independentes e normalmente distribuídos segundo uma “mistura” de 3 distribuições normais (veja os dados na Tabela 2). Isto é, a verdadeira distribuição  $\hat{F}_\Theta$  é uma função discreta, cujos valores para o conjunto suporte verdadeiro são  $(-2., 0., 2.)$  e o vetor de massas correspondente é  $(0.5, 0.333, 0.167)$ . A variância foi fixada para todas as distribuições:  $\sigma^2 = 0.04$ .

Tabela 2

$v_i$
-1.89616
-1.79389
-2.29948
-2.06602
-1.96902
-1.94217
0.16808
0.16745
-0.34241
0.28461
1.96437
1.99195

Queremos estimar  $\hat{F}_\Theta$  baseado nos dados, portanto executamos o algoritmo APCC, o qual encontrou para a tolerância  $\varepsilon = 10^{-8}$  uma estimativa de  $\hat{F}_\Theta$  em 8 iterações com  $\mu = 1.399616 \times 10^{-9}$ . APCC fez 6 cortes para encontrar a solução ótima aproximada. A estimativa do suporte ótimo é  $(-1.9563, -0.3346, 0.1988, 1.9781)$  e a estimativa do vetor de massas ótimo é  $(0.5000, 0.0794, 0.2539, 0.1667)$ . O valor máximo da derivada direcional na estimativa ótima é  $2.96 \times 10^{-6}$  e o tempo de execução foi 0.17s. Quanto a validação do nosso resultado, o erro é de 0.000064.

Nas próximas figuras (VI.3), (VI.4), (VI.5) apresentamos as estimativas de  $\hat{F}_\Theta$  ao longo das iterações  $k = 0, 4$  e  $8$  de APCC.

Observe que  $F_\Theta$  na iteração  $k = 0$  é tal que  $Pr(\Theta \in \Gamma) = 2$ . A probabilidade diminui até que na última iteração,  $k = 8$ ,  $F_\Theta$  é tal que  $Pr(\Theta \in \Gamma) = 1$ ,

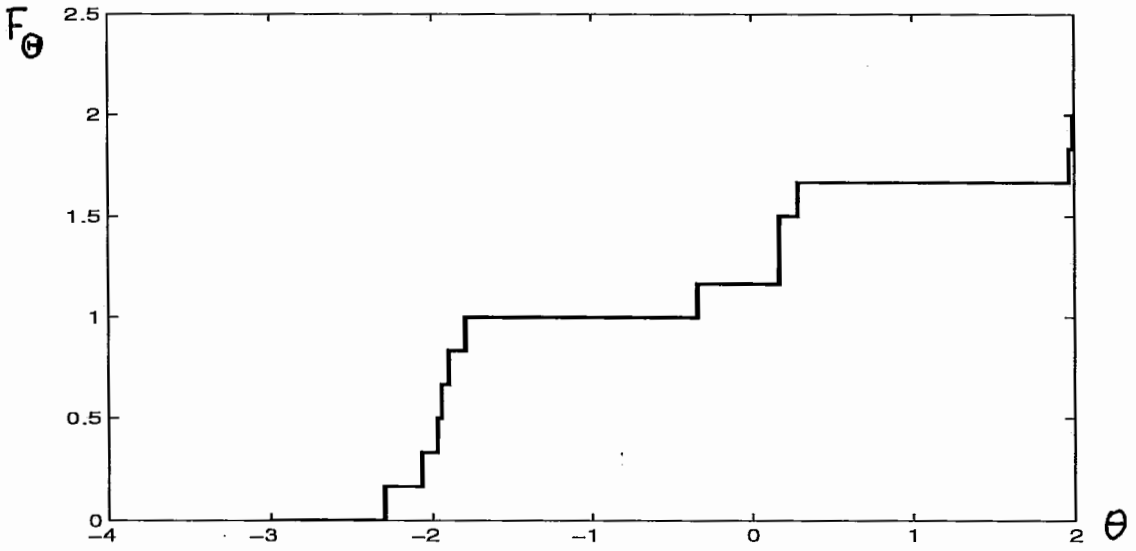


Figura VI.3: Estimativa de  $\hat{F}_\Theta$  na iteração  $k = 0$  :  $e^T x = 2$ .

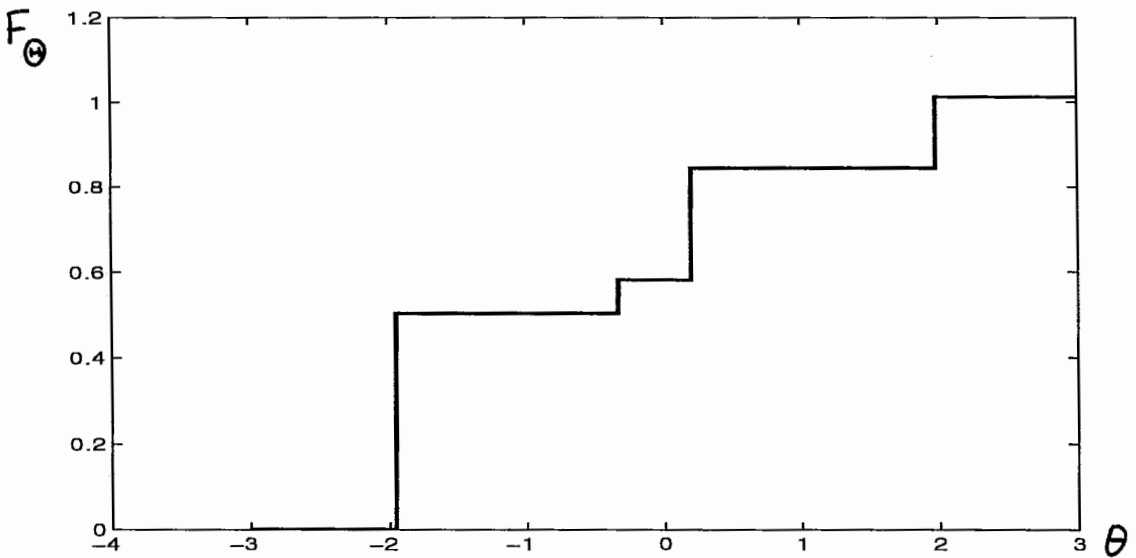


Figura VI.4: Estimativa de  $\hat{F}_\Theta$  na iteração  $k = 4$  :  $1 < e^T x < 2$ .

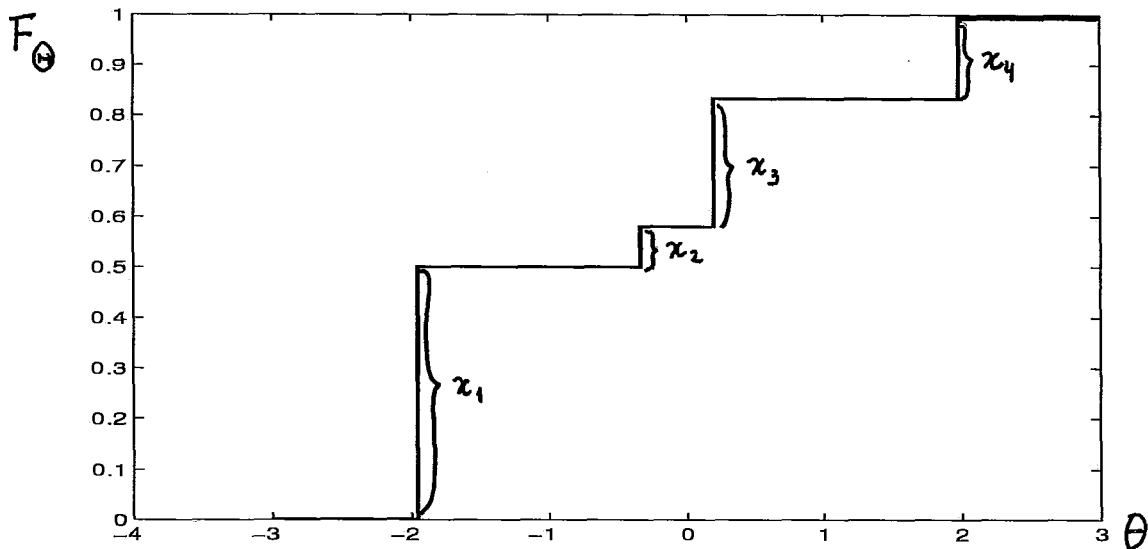


Figura VI.5: Estimativa de  $\hat{F}_\Theta$  na iteração  $k = 8$ :  $e^T x \approx 1$ .

de acordo com a teoria do algoritmo APCC. Observe também que na iteração  $k = 0$ , o conjunto suporte é exatamente igual ao conjunto de dados.

## Experimento 2

Nossa segunda experiência computacional apresenta o desempenho do algoritmo APCC para tamanhos de amostras variados quando conhecemos a verdadeira distribuição  $\hat{F}_\Theta$ . Os dados da amostra também são independentes e normalmente distribuídos. A tolerância para o critério de parada de APCC é  $\varepsilon = 10^{-8}$ .

Seja  $\hat{F}_\Theta$  tal que o suporte ótimo é  $(-2., 0., 1.)$  e o vetor de massa ótimo é  $(0.40, 0.533, 0.0667)$  com variância fixa  $\sigma^2 = 0.04$ . Na Tabela 3 apresentamos os resultados e comparações de APCC quando variamos o tamanho da amostra.

Tabela 3

tamanho	cortes	tempo (s)	erro
30	3	0.25	0.000333
60	5	0.72	0.000086
120	3	1.07	0.000059
240	4	5.87	0.000054



Tabela 3 – Continuação

tamanho	suporte	massas
30	(-1.8967,-0.0217,0.9260)	(0.4000,0.5333,0.0667)
60	(-1.9645,-0.0558,1.1318)	(0.4000,0.5329,0.0671)
120	(-1.9980, 0.0324, 1.0088)	(0.4000,0.5372,0.0628)
240	(-1.9836,-0.0256,0.9022,1.4270)	(0.4000,0.5331,0.0501,0.0168)

### Experimento 3

Nosso objetivo é executar APCC num amostra grande. Primeiro geramos a amostra com 400 dados aleatórios. Para este conjunto de dados 50% dos dados são gerados a partir de uma distribuição  $N(-3., 0.066016)$ , 30% dos dados são gerados a partir de  $N(0., 0.00772)$  e o restante a partir de  $N(2., 0.023814)$ . De acordo com a notação em (VI.1), o primeiro número do par é o valor do parâmetro  $\theta$ . O segundo número do par é o valor da variância. Os valores das variâncias foram obtidos aleatoriamente do intervalo  $(0, 0.25)$ .

Em 5 iterações APCC obtive num total de 3 cortes realizados para  $\varepsilon = 10^{-8}$ : estimativa do suporte ótimo  $(-3.0165, 0.0180, 1.9784)$ , estimativa do vetor de massas ótimo  $(0.5000, 0.3000, 0.2000)$ , erro resultante da validação 0.000060, tempo de execução 9.73s, parâmetro da barreira no final da execução  $\mu = 5.4826 \times 10^{-9}$ .

# Capítulo VII

## Comentários Finais

Acreditamos que o algoritmo que propomos para encontrar a estimativa de máxima verossimilhança de uma distribuição acumulada com suporte finito seja o mais competitivo entre os existentes, especialmente para as amostras grandes.

Podemos justificar a falha dos algoritmos de Terlaky e Vial [49] que resolvem o mesmo problema utilizando algoritmos de pontos interiores com abordagem primal e primal-dual, por não terem considerado o problema de otimização como sendo um problema onde não podemos explicitar as restrições e por isso a necessidade de um algoritmo de planos de corte.

Quando montamos o algoritmo de planos de corte em centro analítico com abordagem primal-dual, não havia na literatura nada parecido. Durante a implementação de nosso algoritmo, um artigo de Gondzio [16] chegava ao “site” dos Métodos de Pontos Interiores na INTERNET apresentando o primeiro algoritmo de planos de corte em centro analítico para resolver um problema da programação linear com abordagem primal-dual. Portanto, a otimização desenvolvida nesta tese para o problema não-linear originado pelo problema estatístico é inovadora, assim como o algoritmo que apresentamos, o qual apresenta algumas diferenças em relação ao de Gondzio.

Nosso algoritmo conceitual assim como a introdução à análise de convergência são mais simples do que aqueles que encontramos em Mokhtarian e

Goffin [42], onde o problema de maximização é mais geral: a função objetivo é convexa não linear e as restrições são definidas por infinitas funções pseudo-convexas. Este artigo também foi colocado no “site” logo depois que obtivemos os primeiros resultados teóricos e práticos.

Resolvemos o problema de encontrar a estimativa de máxima verossimilhança de uma distribuição acumulada com suporte finito, considerando apenas um parâmetro desconhecido. Podemos agora pensar em resolver este problema considerando mais de um parâmetro desconhecido.

Resumindo, a contribuição desta tese foi reformular o problema estatístico de modo a aplicar dois algoritmos eficientes: algoritmo de planos de corte e algoritmo de pontos interiores.

# Referências Bibliográficas

- [1] E. D. Andersen, J. Gondzio, C. Mészáros and X. Xu. Implementation of Interior Point Methods for Large Scale Linear Programming. in *Interior Point Methods in Mathematical Programming*, chapter 6, T. Terkaly Editor, Kluwer Academic Publishers, 1996.
- [2] O. Bahn, O. du Merle, J.-L. Goffin e J.-P. Vial. *A Cutting Plane Method from Analytic Centers for Stochastic Programming*. *Mathematical Programming*, 69 (1995), 45-73.
- [3] D. Böhning. *A Vertex-Exchange-Method in D-Optimal Design Theory*. *Metrika* 33, 337-347, 1986.
- [4] M. Cardia. *Usando a Trajetória Central para Calcular o Centro Analítico de um Polítopo após a Adição de um Plano de Corte Profundo*. Dissertação da Tese de Doutorado no Programa de Engenharia de Sistemas e Computação na COPPE/UFRJ, abril de 1998.
- [5] I. D. Coope e G. A. Watson. *A Projected Lagrangian Algorithms for Semi-Infinite Programming*. *Mathematical Programming*, 32 (1985), 337-356.
- [6] G. B. Dantzig. *Maximization of a linear function of variables subject to linear inequalities*. in TjC. Koopmans, ed., *Activity Analysis of Production and Allocation*, (Willey, New York, 1951) pp 339-347.
- [7] G. B. Dantzig e P. Wolfe. *The decomposition algorithm for linear programming*. *Econometrica*, 29 (1961), 767-778.

- [8] A. V. Fiacco e G. P. McCormick. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*. John Wiley and Sons, New York, 1968; reprinted as Classics in Applied Mathematics 4, SIAM, 1990.
- [9] D. den Hertog. *Interior Point Approach to Linear, Quadratic and Convex Programming, Algorithms and Complexity*. PhD Thesis, Faculty of Mathematics and Informatics, TU Delft, NL-2628 BL delft, The Netherlands, September 1992.
- [10] J.-L. Goffin, Z.-Q. Luo e Y. Ye. *Complexity Analysis of an Interior Cutting Plane Method for Convex Feasibility Problems*. SIAM Journal on Optimization 6 (1996), 638-652.
- [11] J.-L. Goffin e J.-P. Vial. *A two-cut approach in the analytic center cutting plane method*. Logilab Technical Report 97.6, University of Geneva, 1997.
- [12] J.-L. Goffin e J.-P. Vial. *Shallow, deep and very deep cuts in the analytic center cutting plane method*. Logilab Technical Report 96.1, University of Geneva, 1996.
- [13] J.-L. Goffin e J.-P. Vial. *Cutting Planes and Column Generation Techniques with the Projective Algorithm*. Journal of Optimization Theory and its Applications, No. 65, pages 409-429, 1989.
- [14] J.-L. Goffin, J. Gondzio, R. Sarkissian e J.-P. Vial. *Solving Nonlinear Multicommodity Flow Problems by the Analytic Center Cutting Plane Method*. Logilab Technical Report 94.21, University of Geneva, 1994, revised november 1995.
- [15] J.-L. Goffin e J.-P. Vial. *Multiple cuts in the analytic center cutting plane method*. Logilab Technical Report 98.10, University of Geneva, 1998.
- [16] J. Gondzio. *Warm Start of the Primal-Dual Method Applied in the Cutting Plane Scheme*. Logilab Technical Report 96.3, University of Geneva, May 1996, resived April 1997.

- [17] J. Gondzio and R. Sarkissian. *Column Generation with a Primal-Dual Method*. Logilab Technical Report 96.6, University of Geneva, June 1996.
- [18] C. C. Gonzaga. *Path following methods for linear programming*. SIAM Review Vol. 34, No. 2, pp. 167-224.
- [19] C. C. Gonzaga. *Interior Point Methods: Algorithms and Complexity*. Notas de aula de um minicurso, maio de 1997.
- [20] P. Huard. *Resolution of mathematical programming with nonlinear constraints by the method of centers*. In Nonlinear Programming, pages 207-219, J. Abadie, editor. NorthHolland, Amsterdam, The Netherlands, 1967.
- [21] J. Kaliski, D. Haglin, C. Roos and T. Terlaky. *Logarithmic barrier decomposition methods for semi-infinite programming*. Technical Report 96-51, TUDelft, Faculty of Technical Mathematics and Informatics, Abril de 1996.
- [22] N. K. Karmarkar. *A new polynomial-time algorithm for linear programming*. Combinatorica 4 (1984) 373-395.
- [23] W. Karush. *Minima of Functions of Several Variables with Inequalities as Side Constraints*. M.S. Thesis, Department of Mathematics, University of Chicago, 1939.
- [24] J. E. Kelley. *The Cutting Plane Method for Solving Convex Programs*. SIAM Journal of Applied Mathematics, Vol. 8, No. 4, 1960.
- [25] M. Kojima, N. Megiddo e S. Mizuno. *A primal-dual infeasible -interior point algorithm for linear programming*. Mathematical Programming 61 (1993) 263-280.
- [26] M. Kojima, S. Mizuno e A. Yoshida. *A primal-dual interior point algorithm for linear programming*. In N. Megiddo, editor, Progress in Mathematical Programming: Interior Point and Related Methods, pages 29-47. Springer Verlag, New York, 1989.

- [27] R. Hettich and K. O. Kortanek *Semi-Infinite Programming: Theory, Methods, and Applications*. SIAM Review, vol. 35, No. 3, pp. 380-429, Setembro de 1993.
- [28] H. W. Kuhn and A. W. Tucker. *Nonlinear Programming*, in J. Neyman (ed.). Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, Calif., pp. 481-492, 1950.
- [29] N. Laird. *Nonparametric Maximum Likelihood Estimation of a Mixing Distribution*. Journal of the American Statistical Association, 73, 805-810, 1978.
- [30] M. L. Lesperance and J. D. Kalbfleisch. *An Algorithm for Computing the Nonparametric MLE of a Mixing Distribution*. Journal of the American Statistical Association, Vol. 87, No. 417, pp. 120-126, March 1992.
- [31] B. G. Lindsay. *The Geometry of Mixture Likelihoods: A General Theory*. The Annals of Statistics, 1983, Vol. 11, No. 1, 86-94.
- [32] B. G. Lindsay. *Mixture Models: Theory, Geometry and Applications*. Institute of Mathematical Statistics, Hayward, Calif., and American Statistical Association, Alexandria, Va., Publishers, 1995.
- [33] D. G. Luenberger. *Introduction to Linear and Nonlinear Programming*. Addison-Wesley Publishing Company, 1973.
- [34] I. J. Lustig, R. E. Marsten e D. F. Shanno. *On implementing Mehrotra's predictor-corrector interior-point method for linear programming*. SIAM Journal on Optimization, 2(3):435-449, 1992.
- [35] I. J. Lustig, R. E. Marsten e D. F. Shanno. *Computational experience with a globally convergent primal-dual predictor-corrector algorithm for linear programming*. Mathematical Programming 66(1994) 123-135.
- [36] S. Mehrotra. *On the implementation of a (primal-dual) interior point method*. SIAM Journal on Optimization 2 (1992), 575-601

- [37] M. A. F. de Menezes. *Um Algoritmo de Ponto-Interior-Inviável com Complexidade  $O(\sqrt{n}L)$  Iterações para Programação Linear*. Dissertação da Tese de Doutorado no Programa de Engenharia de Sistemas e Computação na COPPE/UFRJ, agosto de 1998.
- [38] J. E. Mitchell e J. Todd. *Solving combinatorial optimization problems using Karmarkar's algorithm*. *Mathematical Programming* 56 (1992) 245-284.
- [39] S. Mizuno, M. J. Todd e Y. Ye. *A Surface of Analytic Centers and Primal-Dual Infeasible-Interior-Point Algorithms for Linear Programming*. *Mathematics of Operations Research* Vol. 20, No. 1, 1995.
- [40] S. Mizuno. *Polynomiality of infeasible-interior-point algorithms for linear programming*. *Mathematical Programming* 67 (1994) 109-119.
- [41] S. Mizuno. *In Interior Point Methods of Mathematical Programming*. Applied Optimization chapter 5, T. Terlaky editor, KLUWER Academic Publishers, 1996.
- [42] F. S. Mokhtarian e J.-L. Goffin. *A Nonlinear Analytic Center Cutting Plane Method for a Class of Convex Programming*. GERAD Technical Report G-96-53, november 1996.
- [43] Y. Nesterov. *Cutting plane algorithms from analytic centers: efficiency estimates*. *Mathematical programming, series B, Nondifferentiable and Large Scale Optimization*, Vol. 69, No. 1, pp. 149-176, 1995.
- [44] S. M. Ross. *Introduction to Probability and Statistics for Engineers and Scientists*. Wiley Series in Probability and mathematical Statistics, John Wiley & Sons, 1987.
- [45] R. Y. Rubinstein. *Simulation and the Monte Carlo Method*. Wiley Series in Probability and mathematical Statistics, John Wiley & Sons, 1981.
- [46] S. D. Silvey. *Optimal Design*. Chapman & Hall, New York, 1980.



- [47] G. Sonnevend. *An analytic center for polyhedrons and new classes of global algorithms for liner (smooth,convex) programming.* in Lecture Notes Control Inform. Sci., 866-876. Springer-Verlag, New York, NY, 1985.
- [48] M. J. Todd. *Potencial-reduction methods in mathematical programming.* Technical Report No. 1112, Scholl of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY, 1995.
- [49] T. Terlaky e J.-Ph. Vial. *Computing maximum likelihood estimators of convex density functions.* Technical Report 95-49, TUDelft, Faculty of Technical Mathematics and Informatics, Maio de 1995.
- [50] Y. Ye. *A Potential Reduction Algorithm Allowing Column Generation.* SIAM J. Optimization, Vol 2, No 1, pp 7-20, Fevereiro 1992.
- [51] Y. Ye. *Complexity Analysis of the Analytic Center Cutting Plane Method That Uses Multiple Cuts.* Mathematical Programming 78, 85-104, 1997.
- [52] Y. Ye. *Interior Point Algorithm: Theory and Analysis.* John Wiley & Sons, New York, 1997.
- [53] P. Whittle. *Some general points in the theory of optimal experimental design.* J. Roy. Statist. Soc. Ser. B, 35, 123-130, 1973.
- [54] R. J. Wonnacott and T. H. Wonnacott. *Introductory Statistics.* Wiley Series in Probability and mathematical Statistics, John Wiley & Sons, Fourth Edition, 1985.