



NÚMEROS DE CONDICIONAMENTO E PROPRIEDADES LIMITES DA
DIREÇÃO AFIM-ESCALA EM PROGRAMAÇÃO LINEAR

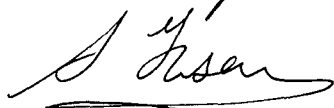
Hugo José Lara Urdaneta


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.

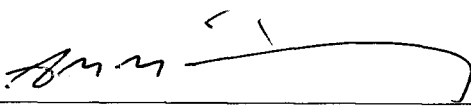
Aprovada por:


Prof. Clóvis Caesar Gonzaga, D. Sc.


Prof. Nelson Maculan Filho, D.Habil.


Prof. Alfredo N. Iusem, Ph. D.


Prof. Carlos Humes Jr., Ph.D.


Prof. José M. Martínez, D. Sc.

RIO DE JANEIRO, RJ - BRASIL

FEVEREIRO DE 1998

LARA URDANETA, HUGO JOSÉ

Números de Condicionamento e Propriedades

Limites da Direção Afim-Escala em Programação

Linear [Rio de Janeiro] 1998

VIII, 79p. 29.7 cm (COPPE/UFRJ, D.Sc.,

ENGENHARIA DE SISTEMAS E COMPUTAÇÃO, 1998)

TESE – Universidade Federal do Rio de
Janeiro, COPPE

1 – Algoritmos de Pontos Interiores

2 – Programação Linear

3 – Complexidade 4– Direção Afim-Escala

I. COPPE/UFRJ II. Título(Série).

*A Belkis
a Lupe
a Huguito
y a José*

Agradecimentos

Ao prof. Clovis C. Gonzaga, quem com paciência, sabedoria, competência e amizade soube orientar a realização do presente trabalho.

A meu colega e amigo Rómulo Castillo, quem compartilhou comigo os momentos bons e difíceis da realização acadêmica.

Aos professores Alfredo N. Iusen, Carlos Humes Jr., Nelson Maculan, e Jose M. Martinez por permitirem-me ter a mais brilhante e competente banca possível.

Aos meus professores da COPPE, especialmente Adilson E. Xavier e Susana S. de Makler.

Aos meus colegas de sala Marco Menezes, Marli Cardia, Elizabeth W. Karas, Luiz Matioli e Marcio Vilela.

Aos meus amigos e anfitriões da cidade de Florianópolis que me deram acolhida Maritza, Valdésio Marlene y Cesar.

Ao CONICIT pelo apoio financeiro.

A mi familia: Belkis, Carolina, Luis, Mariló, Lalita, Ana, Sobrinos, Tíos, amigos, Gracias por siempre estar ahí.

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

NÚMEROS DE CONDICIONAMENTO E PROPRIEDADES LIMITES DA DIREÇÃO AFIM-ESCALA EM PROGRAMAÇÃO LINEAR

Hugo José Lara Urdaneta

Fevereiro de 1998

Orientador: Clóvis Caesar Gonzaga

Programa: Engenharia de Sistemas e Computação

Dois números de condição descritos por Ye, Vavasis, Stewart, Todd e outros, têm sido usados para estudar a complexidade dos algoritmos de pontos interiores. O primeiro é dado pela “menor variável grande” na face ótima do problema de programação linear; e o outro pelo supremo das normas de todas as projeções oblíquas sobre o espaço imagem de A^T . O primeiro tem a desvantagem de depender do conhecimento da partição ótima na sua definição, enquanto que o segundo depende só dos dados, mas não usa a estrutura do problema de programação linear. A respeito do segundo número provamos que ele é igual ao supremo das normas das projeções oblíquas sobre o espaço nulo de A , e com relação ao primeiro mostramos uma caracterização que não depende de conhecer a partição ótima. Depois estudamos propriedades limites da direção afim-escala. Em particular calculamos um limitante para o ângulo entre todas as direções afim-escala e o vetor de custos. O limitante encontrado está relacionado com a caracterização do número de condição antes mencionado.

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

CONDITION NUMBERS AND LIMITING PROPERTIES OF THE AFFINE
SCALING DIRECTION IN LINEAR PROGRAMMING

Hugo Jose Lara Urdaneta

February/1998

Advisor: Clóvis Caesar Gonzaga

Department: Systems Engineering and Computer Science

Two condition numbers described by Ye, Vavasis, Stewart, Todd and others have been used to study the complexity of interior point algorithms. The first one is given by the “smallest large variable” in the optimal face of the linear programming problem, and the other by the supreme of the norms of all the oblique projection operators on the range space of A^T . The first one has the disadvantage of depending on the knowledge of the optimal partition, and the second one, depends only on the data, but does not use the structure of the linear programming problem. With respect to the second one, we show that it coincides with the supreme of the norms of all the oblique projections on the null space of A , and with respect to the first one, we give a characterization that does not depend on the knowledge of the optimal partition. Later we study limiting properties of the affine scaling direction. Particularly we compute a bound for the angle between all the affine scaling directions and the cost vector. The bound is related with the characterization of the condition number given above.

Índice

1	Introdução	1
2	Números de Condicionamento	7
2.1	Introdução	7
2.2	O problema de programação linear e algoritmos de pontos interiores	8
2.3	Número de condicionamento de Y_e e complexidade.	14
2.4	O número de condição $\bar{\chi}_A$	17
2.5	Propriedades duais de $\bar{\chi}_A$	22
2.6	Mais propriedades de $\bar{\chi}_A$	26
2.7	O número $\bar{\chi}_A$ e complexidade	30
2.8	Uma caracterização da menor variável grande	32
3	Propriedades limites da direção afim-escala	37
3.1	Introdução	37
3.2	O problema	37
3.3	O algoritmo afim-escala	38
3.4	A direção afim-escala primal	43
3.5	O ângulo limite	49
3.5.1	Preliminares	49
3.5.2	As arestas do conjunto viável	54
3.5.3	O cálculo de $\tau_{A,c}$	56
4	Comentários finais	71

Capítulo 1

Introdução

O problema de programação linear, que consiste em minimizar (ou maximizar) uma função linear em \mathbb{R}^n , sujeita a restrições também lineares de igualdade e desigualdade, é um modelo matemático para muitos problemas de economia, planejamento, engenharia, produção, etc. Em 1947, Dantzig [7] propôs o algoritmo simplex para resolver o problema de programação linear. Este método, que continua sendo usado até agora, constitui uma ferramenta poderosa para a resolução do problema e tem um desempenho computacional de bom nível. Ainda há muitos pesquisadores tentando melhorias, e é o método comercialmente padrão.

Dada uma classe de problemas, a análise dos algoritmos que resolvem as instâncias do problema busca provar se o algoritmo termina, e se nesse caso obtém-se uma solução do problema. Estuda também a eficiência, que é o desempenho do algoritmo em termos da quantidade de trabalho computacional requerido para achar uma solução. A eficiência é muitas vezes estudada pela complexidade computacional, quer dizer, a quantidade de trabalho computacional requerido para resolver o problema mais difícil da classe ou em outras palavras o “pior caso”. Nesse contexto existem classificações dos problemas como polinomiais e não polinomiais (ou exponenciais) segundo a existência de algoritmos que resolvem o “pior caso” com um número de operações aritméticas executadas limitadas por expressões polinomiais de parâmetros que representam o “tamanho dos dados”. Por exemplo, um algoritmo

é exponencial se o número de operações aritméticas é limitada por expressões como $O(Ln^m)$ ou $O(n!)$, $O(100^n)$, etc. onde L é o “tamanho em bits” do problema e m , n representam a dimensão das funções envolvidas. Um algoritmo é polinomial se o número de operações aritméticas é limitado por um polinômio dos parâmetros L , n e m (por exemplo $O(n^4L)$). (Ver por exemplo Campello e Maculan [5])

Por muito tempo foi desconhecido se o problema de programação linear seria polinomial ou não. Em 1970 Klee e Minty [23] criaram um exemplo (de “pior caso”) onde mostraram que o algoritmo simplex clássico requer um número exponencial de passos, assim mostraram a não polinomialidade do algoritmo de Dantzig. Em 1978 Khachian [22] desenvolveu um algoritmo polinomial para resolver o problema de programação linear chamado método elipsoidal, demonstrando que o problema de programação linear é polinomial. No entanto, o algoritmo de elipsóides não mantém na prática um desempenho satisfatório, aliás o pior caso está muito próximo ao caso médio, assim não consegue concorrer com o método simplex. Nesse momento se tinha dois algoritmos: O simplex, bom na prática, mas ruim na teoria, e o elipsoidal, bom na teoria mas fraco no desempenho prático. Em 1984, com seu algoritmo projetivo, Karmarkar [20] deu uma importante melhoria na teoria de programação linear com um algoritmo que tinha a propriedade teórica de polinomialidade, e que prometia desempenhos computacionais satisfatórios. Após o trabalho de Karmarkar, que obteve uma complexidade de $O(n^{3.5}L)$ operações aritméticas, aparecem vários trabalhos que melhoram o resultado. O algoritmo original de Karmarkar utilizava um formato especial para o problema, e era muito difícil. Em 1985 vários pesquisadores introduziram variantes para o problema em formato padrão (Anstreicher [1], Gay [12], De Ghellinck e Vial [13], Gonzaga [14], Todd e Burrell [36]), e ao mesmo tempo uma simplificação do algoritmo gerou o método afim-escala (Barnes[3], Vanderbei, Meketon e Freedman[44], Cavalier e Soyster [6]). O algoritmo afim-escala era conhecido havia vinte anos, publicado por Dikin [8] na União Soviética. Depois, em 1986 foram introduzidos resultados novos, onde o limitante da complexidade desce para $O(\sqrt{n}L)$ iterações, e $O(n^3L)$ para o número de operações aritméticas, em trabalhos

de Renegar [32], Gonzaga [14] e Vaidya [43]. Em março de 1997 Anstreicher [2] prova a menor complexidade conhecida até agora com $O(\sqrt{n}L)$ para o número de iterações, e no total $O(\frac{n^3}{\log n}L)$ operações aritméticas.

Em todos esses resultados, o modelo utilizado para a análise da complexidade foi o dado pelo “tamanho em bits” L do problema de programação linear, em poucas palavras, a quantidade de algarismos utilizados para armazenar os dados no computador sob a suposição de serem inteiros (ou racionais, considerando inteiros o numerador e denominador). Dito marco para estudar a complexidade é baseado num modelo de computação chamado “máquina de Turing”, que é próprio para problemas combinatórios. O problema de programação linear pode ser encaixado neste modelo, pois achar uma solução do problema de programação linear pode ser visto como identificar a partição ótima, a partição dos índices que identificam as variáveis que são positivas ou nulas no conjunto ótimo. No entanto, os algoritmos de pontos interiores têm como característica serem de natureza contínua, seguindo uma trajetória contínua bem definida (a trajetória central) que utiliza funções contínuas. Em cada iteração não “escolhem um vértice” como o método simplex mas encontram um ponto próximo à trajetória central. O problema de programação linear é desde este ponto de vista de natureza contínua, como as funções que o definem, e não discreta com é sugerido por procedimentos que tomam decisões, como o método simplex.

Existe outra maneira de estudar a complexidade num contexto contínuo: como no modelo de Turing, queremos rotular os problemas segundo alguma característica que de alguma forma meça quanto a instância que estamos estudando é difícil de resolver, como o “tamanho em bits” dos dados, e caracterize o “pior caso”. Queremos então definir outras medidas de dificuldade do problema mas supondo os dados como números reais, em lugar de inteiros como faz o modelo de Turing. Queremos também usar as operações aritméticas elementares de uma forma direta, e não operações complicadas como as propostas pelo modelo baseado no tamanho em bits. Esta teoria de complexidade é chamada complexidade analítica e foi formalizada por

Blum, Shub e Smale [4]. Existem vários trabalhos em algoritmos de pontos interiores que procuram estudar a complexidade usando modelos reais. Aparecem principalmente os trabalhos de Renegar [30], de Ye [50] e de Vavasis e Ye [45],[46],[47]. O primeiro define uma norma dos dados e utiliza a “distância aos problemas mal postos” para definir um modelo real de complexidade; o segundo define como número de condição a “menor variável grande” na face ótima, e estuda a complexidade com este número; por último, Vavasis e Ye definem um número como o supremo das normas dos operadores projeções oblíquas no espaço imagem de A^T e demonstram com ele alguns resultados de complexidade.

Com respeito aos números de condição damos uma caracterização da “menor variável grande” que não depende de conhecer a partição ótima, e provamos que o número definido como o supremo das normas das projeções oblíquas sobre a imagem de A^T coincide com o supremo das normas das projeções oblíquas sobre o espaço nulo de A .

O algoritmo afim-escala apareceu no contexto dos algoritmos de pontos interiores logo após do trabalho de Karmarkar [20] em 1984. Existia já desde 1967 quando Dikin [8] o publicou na União Soviética. Foi redescoberto por Barnes [3], Cavalier e Soyster [6], Karmarkar e Ramakrisnan [21], Kortanek e Shi [24], Vanderbei, Meketon e Freedman [44], sendo desenvolvido como uma simplificação do algoritmo de Karmarkar. A polinomialidade é uma pergunta aberta. Há duas versões principais do algoritmo; uma de passos curtos, onde o tamanho do passo é menor ou igual a um, e outra de passos longos, em que se caminha na direção afim-escala até a fronteira, para depois voltar uma fração. Existem resultados de convergência do algoritmo com hipótese de não degeneração das restrições, dadas pelo próprio Dikin [9] por Barnes [3], Vanderbei, Meketon e Freedman [44] para ambas versões do algoritmo, e sem hipótese de não degeneração tem-se provas para passos curtos, por Tseng e Luo [37] e Tsuchiya, e para passos longos até $\frac{1}{2}$ do passo completo, por Dikin [11] e $\frac{2}{3}$ por Tsuchiya e Maramatsu [40]. O trabalho de Tseng e Luo apresentou propriedades interessantes das direções afim-escala que foram usadas ativamente para provar a

convergência, como um limitante inferior para o cosseno do ângulo que as direções afim-escala fazem com o vetor de custos. Isto significa que o ângulo que faz o custo com todas as direções afim-escala não pode aproximar-se a $\pi/2$. A prova de Tseng e Luo foi feita por contradição, assim eles provaram a existência do limitante, mas não calcularam seu valor. No capítulo 3 calculamos dito limitante sob uma hipótese que será especificada nesse capítulo. Para isso achamos uma caracterização das direções afim-escala limite, e usamos ativamente os resultados limites de Megiddo e Shub [26].

No segundo capítulo descrevemos o problema de programação linear no formato padrão. Definimos trajetória central e medida de proximidade à trajetória. Depois mostramos o número de condição da “menor variável grande” com o teorema de complexidade de Ye. Também definiremos o número de condição definido pelas normas das projeções oblíquas sobre o espaço imagem de A^T , algumas propriedades e demonstraremos nossa relação de dualidade. Mais tarde estudaremos a complexidade no formato de Karmarkar, e por último descrevemos a caracterização da menor variável grande. No capítulo 3 mostramos o algoritmo afim-escala primal, estudamos as direções afim-escala, os teoremas de Tseng e Luo e de Megiddo e Shub; e ao final calculamos o limitante para o ângulo entre a direção afim-escala primal e o custo, que é nosso último resultado principal.

Notação: Em todo o trabalho o espaço considerado é \mathbb{R}^n , e a norma, se não especificamos no contexto é a norma euclideana. $\mathbb{R}_+^n = \{x \in \mathbb{R}^n \mid x_i \geq 0, i = 1, \dots, n\}$ e $\mathbb{R}_{++}^n = \{x \in \mathbb{R}_+^n \mid x_i > 0, i = 1, \dots, n\}$. Para uma matriz $E \in \mathbb{R}^{m \times n}$ denotamos como $\mathcal{N}(E)$ ao espaço nulo de E , e como $\mathcal{R}(E)$ o espaço imagem. E^T é a matriz transposta de E . O vetor e denota o vetor de dimensão dada pelo contexto e com cada componente igual a um. Os vetores são denotados por letras minúsculas. Se $d \in \mathbb{R}_{++}^n$ denotamos por D a matriz diagonal, cujas entradas da diagonal são as componentes de d . As operações com vetores uv , u/v e u^l onde $l \in \mathbb{R}$, denotam as operações usuais em \mathbb{R} componente por componente. Denotamos por P_E a matriz de projeção sobre o espaço $\mathcal{N}(E)$ e por \tilde{P}_E a matriz de projeção sobre $\mathcal{R}(E^T)$. Dado o

conjunto de índices $\{1, \dots, n\}$ e $J \subset \{1, \dots, n\}$ a cardinalidade de J denota-se como $|J|$; $J^c = \{1, \dots, n\} \setminus J$ denota o complemento de J ; v_J representa a restrição do vetor v indexada por J . Dado o vetor $d \in \mathbb{R}_+^n$, definimos $B(d) = \{i = 1, \dots, n \mid d_i > 0\}$, e $N(d) = B(d)^c$. A notação $g = O(r)$ significa que existe uma constante positiva M tal que $\|g\| \leq Mr$. Dado um conjunto $\mathcal{A} \subset \mathbb{R}^n$, a notação $v + \mathcal{A}$, para um vetor $v \in \mathbb{R}^n$ significa a soma de conjuntos $\{v\} + \mathcal{A}$. $cl(\mathcal{A})$ denota o fecho do conjunto \mathcal{A} , $relint(\mathcal{A})$ denota o interior relativo do conjunto \mathcal{A} , e \mathcal{A}^o é a interseção do conjunto \mathcal{A} com o ortante positivo. Uma seqüência no conjunto \mathcal{A} é denotada por (x^k) em \mathcal{A} , e subsequências por $(x^k)_K$ onde $K \subset \mathbb{N}$.

Capítulo 2

Números de Condicionamento

2.1 Introdução

No estudo da complexidade e algoritmos de pontos interiores para programação linear, é tradicionalmente usada uma regra de parada que envolve o tamanho em bits do problema de programação linear. Este modelo para estudar a complexidade, que é baseado na máquina de Turing, supõe que os dados de entrada do problema de programação linear são números inteiros, pois o computador não admite números reais como dados. O estudo da complexidade é uma estratégia para medir a eficiência dos algoritmos através da contagem das operações aritméticas executadas por cada iteração do algoritmo, e limitantes para o número de iterações no pior caso. É conhecido que o problema de programação linear é de natureza contínua, e não discreta como é sugerido pelo modelo de Turing. Parece então natural estabelecer um modelo real para definir a regra de parada. Blum, Shub e Smale [4] definem um modelo chamado de “máquina universal” em que se supõem os dados dos problemas como números reais e as operações aritméticas são definidas conforme a natureza real dos dados. Inspirado nestas idéias, aparecem trabalhos que estudam a complexidade dos algoritmos de pontos interiores para programação linear baseados em modelos reais, em características do problema que de alguma forma medem quanto uma instância do problema é difícil de resolver. Ditas características podem ser re-

presentadas por um número de condicionamento para o problema de programação linear. Nessa linha aparecem os trabalhos de Renegar [30],[31] que usam a distância ao mal condicionamento como número de condição, e os trabalhos de Ye [50], Vavasis e Ye [45],[46],[47] com números de condição baseados nas soluções ótimas e nas projeções escaladas sobre o conjunto ótimo.

Neste capítulo trataremos dois números de condição descritos por Ye, Vavasis, Stewart [33], Todd [35] e outros, que têm sido usados para estudar a complexidade dos algoritmos de pontos interiores. O primeiro é dado pela “menor variável grande” na face ótima do problema de programação linear; e o outro pelo supremo das normas de todas as projeções oblíquas sobre o espaço imagem de A^T . O primeiro tem a desvantagem de depender do conhecimento da partição ótima na sua definição, enquanto que o segundo, que depende só dos dados, não usa a estrutura do problema de programação linear. Com respeito ao segundo número provamos que ele é igual ao supremo das normas das projeções oblíquas sobre o espaço nulo de A , e com relação ao primeiro mostramos uma caracterização que não depende de conhecer a partição ótima.

2.2 O problema de programação linear e algoritmos de pontos interiores

Consideremos os problemas primal e dual de programação linear associados à instância \bar{A}, b, \bar{c} , onde $\bar{A} \in \mathbb{R}^{m \times n}$, $\bar{c} \in \mathbb{R}^n$ e $b \in \mathbb{R}^m$:

$$\begin{aligned} & \text{minimizar} && \bar{c}^T x \\ (P) \quad & \text{sujeito a} && \bar{A}x = b \\ & && x \geq 0 \end{aligned}$$

e

$$\begin{aligned} & \text{maximizar} && b^T y \\ (D) \quad & \text{sujeito a} && \bar{A}^T y + s = \bar{c} \\ & && s \geq 0 \end{aligned}$$

Definimos o conjunto viável primal-dual como

$$F = \{(x, s) \in \mathbb{R}^n \times \mathbb{R}^n \mid \bar{A}x = b, x \geq 0, s = \bar{c} - \bar{A}^T y, s \geq 0, y \in \mathbb{R}^m\},$$

e o conjunto de pontos interiores

$$F^o = \{(x, s) \in F \mid x > 0, s > 0\}.$$

Neste capítulo faremos as seguintes suposições padrões em algoritmos de pontos interiores:

- $F^o \neq \emptyset$
- \bar{A} é de posto máximo.
- $\bar{c} \notin \mathcal{R}(\bar{A}^T)$.

A segunda suposição é usada para simplificar a notação pois os resultados que se descrevem podem ser alcançados sem dita condição. A primeira e terceira suposições visam garantir a boa definição da trajetória central.

Pelas condições de otimalidade de (P) , (D) , um ponto (x^*, s^*) é ótimo se pertence a F e verifica a condição de complementaridade

$$x_i^* s_i^* = 0 \text{ para todo } i = 1, \dots, n. \tag{2.1}$$

Pelo teorema de Goldman-Tucker (ver [34]) existe uma solução ótima do problema de programação linear estritamente complementar, quer dizer existe um par (x^*, s^*) ótimo tal que

$$x_i^* s_i^* = 0 \text{ e } x_i^* + s_i^* > 0, \forall i \in \{1, \dots, n\}.$$

A condição de complementaridade põe em evidência a natureza combinatória do problema de programação linear, pois para obter uma solução ótima devemos decidir quais das variáveis x_i ou s_i serão nulas ou positivas na solução ótima. Dado que esse problema combinatório é difícil, os algoritmos de trajetória central relaxam

a condição de complementaridade da seguinte forma: Para um parâmetro $\mu > 0$ escolhemos pontos em F tais que

$$x_i s_i = \mu, \forall i \in \{1, \dots, n\},$$

Levando o parâmetro μ a zero e mantendo a viabilidade, devemos obter uma solução ótima do problema $(P)(D)$. É assim que se define a trajetória central:

$$\mathcal{CP} = \{(x, s) \in F \mid xs = \mu e, \mu > 0\}$$

Queremos estudar o condicionamento do problema $(P), (D)$, quer dizer quanto uma instância $\bar{A}, \bar{b}, \bar{c}$ dada define um problema “difícil” de resolver. Imaginemos então um procedimento iterativo para achar uma solução do problema $(P), (D)$. É necessário fornecer ao procedimento um ponto por onde começar a busca, e o número de iterações depende claramente de onde está o ponto inicial. Assim a instância não estará bem definida apenas com os dados $\bar{A}, \bar{b}, \bar{c}$, mas sim adicionando o ponto inicial aos dados, quer dizer $\bar{A}, \bar{b}, \bar{c}, x^o$. Quando estudamos a complexidade dos algoritmos estamos interessados em duas questões: A primeira é quantas contas executará o computador para fornecer resultados num problema considerado como o pior caso?, e a segunda é quais são os critérios de comparação entre as instâncias para serem consideradas “fáceis”, ou “difíceis”? Visando estabelecer um critério unificado para responder a segunda questão, definiremos um par de problemas primal-dual equivalentes a $(P), (D)$ quanto ao conjunto de soluções, mas incorporando informação adicional à instância, proveniente de sua estrutura, de maneira a fornecer um ponto inicial padrão.

Denotemos por $(x(\mu), s(\mu))$ os pontos de \mathcal{CP} . Tomemos o ponto $(\hat{x}, \hat{s}) = (x(1), s(1))$ e definamos a matriz de mudança de variáveis $\hat{D} = \hat{X}^{\frac{1}{2}} \hat{S}^{-\frac{1}{2}}$. Consideremos agora o problema escalado, quer dizer substituímos as condições de viabilidade primal e dual por $A\bar{x} = b$, onde $A = \bar{A}\hat{D}$, $\bar{x} = \hat{D}^{-1}x$, e $A^T y + \bar{s} = c$ com $\bar{s} \geq 0$ e $c = \hat{D}\bar{c}$. Como o vetor e é primal e dual viável, podemos escrever os problemas

como

$$\begin{aligned} & \text{minimizar} && c^T x \\ & \text{sujeito a} && Ax = b \\ & && x \geq 0 \end{aligned}$$

e

$$\begin{aligned} & \text{maximizar} && b^T y \\ & \text{sujeito a} && A^T y + s = c \\ & && s \geq 0. \end{aligned}$$

Após a mudança de coordenadas vemos que o vetor e é primal e dual viável: de fato, $Ae = \bar{A}\hat{D}e = \bar{A}\hat{D}(\hat{D})^{-1}\hat{x} = b$, e $A^T y + e = \hat{D}A^T y + \hat{D}\hat{s} = \hat{D}\bar{c} = c$. Por outra parte sabemos que $c + \mathcal{R}(A^T) = s^0 + \mathcal{R}(A^T)$ para todo $s^0 \in c + \mathcal{R}(A^T)$. Agora, como $e \in c + \mathcal{R}(A^T)$ podemos substituir c por e na definição do problema dual, e dado que $b = Ae$ a função objetivo dual é $e^T A^T y = e^T(e - s) = e^T e - e^T s$, assim os problemas primal e dual podem ser escritos como:

$$\begin{aligned} & \text{minimizar} && e^T x \\ (P_e) \quad & \text{sujeito a} && Ax = b \\ & && x \geq 0 \end{aligned}$$

e

$$\begin{aligned} & \text{maximizar} && e^T(e - s) \\ (D_e) \quad & \text{sujeito a} && A^T y + s = e \\ & && s \geq 0. \end{aligned}$$

Esta normalização nos proporciona uma interessante forma de escrever os problemas primal e dual: Denotemos o conjunto $\mathcal{N}(A)$ como S . Dado que $Ae = b$, a primeira restrição primal pode ser escrita como $A(x - e) = 0$ quer dizer $x \in e + S$. A primeira restrição dual pode ser escrita como $s \in e + S^\perp$ enquanto que a função objetivo dual é $b^T y = e^T A^T y = e^T(e - s)$. Maximizar esta função é equivalente a minimizar a função $e^T s$ com s dual viável. Desta maneira os problemas (P_e) e (D_e) podem ser descritos como:

$$\begin{aligned} & \text{minimizar} && e^T x \\ & \text{sujeito a} && x \in e + S \\ & && x \geq 0 \end{aligned}$$

e

$$\begin{aligned} & \text{minimizar} && e^T s \\ & \text{sujeito a} && s \in e + S^\perp \\ & && s \geq 0. \end{aligned}$$

Descrito assim, o par de problemas primal-dual é simétrico com relação a S e S^\perp . Temos mostrado então que existe uma mudança de variáveis tal que qualquer problema de programação linear, que verifique a condição de Slater, tem um problema equivalente que é simétrico com respeito ao subespaço que define as restrições lineares e seu complemento ortogonal.

Definimos o conjunto viável primal e dual como

$$F_e = \{(x, s) \in \mathbb{R}^n \times \mathbb{R}^n \mid Ax = b, x \geq 0, s = e - A^T y, s \geq 0, y \in \mathbb{R}^m\}$$

e a trajetória

$$\mathcal{CP}_e = \{(x, s) \in F_e \mid xs = \mu e, \mu > 0\}.$$

Podemos pensar nos problemas $(P_e)(D_e)$ como os problemas de programação linear normalizados, pois agora podemos estabelecer critérios para comparar as instâncias. Usaremos os problemas $(P_e)(D_e)$ no resto do presente capítulo.

Sabe-se que para o problema de programação linear, o conjunto ótimo é unicamente caracterizado por uma partição $\{B, N\}$ dos índices $\{1, \dots, n\}$ tal que $x_B^* > 0$, $x_N^* = 0$, $s_B^* = 0$ e $s_N^* > 0$ para todo x^*, s^* no interior relativo do conjunto ótimo primal-dual. Dada a partição ótima $\{B, N\}$, podemos particionar as colunas de A como $A = (A_B, A_N)$. Definamos então o conjunto ótimo:

$$\Omega := \{(x, s) \in F_e \mid x_B \geq 0, x_N = 0, s_B = 0, s_N \geq 0\}. \quad (2.2)$$

Sabe-se também que os algoritmos de trajetória central definem seqüências $(x^k, s^k) \in F_e^\circ$ que seguem a trajetória, quer dizer os pontos (x^k, s^k) verificam algum critério de proximidade com a trajetória. Definimos a medida de proximidade:

$$\mathbb{R}_{++}^n \times \mathbb{R}_{++}^n \times \mathbb{R}_+ \ni (x, s, \mu) \mapsto \delta(x, s, \mu) := \left\| \frac{xs}{\mu} - e \right\|.$$

Para alguns algoritmos, seguir a trajetória significa que para $\alpha \in (0, 1)$ fixo se satisfaz

$$\delta(x^k, s^k, \mu^k) \leq \alpha, \quad \forall k \in \mathbb{N}. \quad (2.3)$$

Dado um polítopo $PO \subset \mathbb{R}_{++}^n$, definimos o centro analítico de PO como

$$x_c = \operatorname{argmin} \left\{ - \sum_{i=1}^n \log x_i \mid x \in PO \right\}.$$

Sob a hipótese de Slater é possível caracterizar a trajetória central primal (similarmente a dual) através dos centros analíticos dos polítopos definidos por $Ax = b, x \geq 0$ e $e^T x \leq K$, onde K é um parâmetro que decresce até o valor ótimo $e^T x^*$ (ver Gonzaga [15]). Usando dita caracterização mostraremos um interessante lema relacionado com o tamanho das variáveis.

Lema 2.2.1 *Seja $x(\mu)$ um ponto primal centrado . Então*

$$x_i \leq n x_i(\mu)$$

$\forall i \in \{1, \dots, n\}$ e $\forall x$ viável tal que $e^T x \leq e^T x(\mu)$.

Demonstração. Associado a μ existe um parâmetro K tal que $x(\mu)$ é o centro analítico do polítopo definido por $Ax = b, x \geq 0$ e $e^T x \leq K$, segundo a caracterização no parágrafo anterior. Mudando a escala do problema, usando a matriz $\bar{D} = X(\mu)^{-1}$ temos $\bar{x} = \bar{D}x$ e $\bar{x}(\mu) = e$, que é o centro analítico do polítopo escalado. Sabemos que existe uma bola centrada em e e de raio n tal que todo ponto \bar{x} viável está dentro da bola. Isto é $\|\bar{x}\| \leq n$ e assim $\bar{x}_i < n$. Desta forma

$$\frac{x_i}{x_i(\mu)} \leq n.$$

■

Introduzimos sem demonstração o lema análogo relativo às variáveis duais:

Lema 2.2.2 *Seja $s(\mu)$ um dual ponto centrado . Então*

$$s_i \leq ns_i(\mu)$$

$\forall i \in \{1, \dots, n\}$ e $\forall s$ dual viável tal que $e^T s \leq e^T s(\mu)$.

Os lemas anteriores foram demonstrados por Vavasis e Ye [47] com uma prova mais complicada, e valendo a relação só para pontos $(x(\mu'), s(\mu'))$ da trajetória com $\mu' < \mu$. A nossa prova é mais simples e garante também a relação para todo ponto viável com custo menor ao dado por $e^T x(\mu)$.

Estabeleceremos sem demonstração um lema bem conhecido em algoritmos de pontos interiores. (ver Gonzaga [15]) relativo aos pontos cercanos á trajetória central:

Lema 2.2.3 *Se $\delta(x, s, \mu) \leq \alpha < 1$ para $(x, s) \in F$ e $\mu > 0$ fixo, e $(x(\mu), s(\mu)) \in \mathcal{CP}_\gamma$ é o ponto sobre a trajetória associado a μ então para cada $i \in \{1, \dots, n\}$ verifica-se*

i)

$$\left(1 - \frac{\alpha}{1 - \alpha}\right) x_i \leq x_i(\mu) \leq \left(1 + \frac{\alpha}{1 - \alpha} x_i\right)$$

ii)

$$\left(1 - \frac{\alpha}{1 - \alpha}\right) s_i \leq s_i(\mu) \leq \left(1 + \frac{\alpha}{1 - \alpha} s_i\right)$$

2.3 Número de condicionamento de Ye e complexidade.

Dada a face ótima Ω , e a partição ótima $\{B, N\}$ como vistas em (2.2) , definimos o número de condicionamento como segue:

$$\begin{aligned} \sigma_P &= \min_{j \in B} \{\max x_j \mid (x, s) \in \Omega\}, \\ \sigma_D &= \min_{j \in N} \{\max s_j \mid (x, s) \in \Omega\}, \\ \sigma &= \min\{\sigma_P, \sigma_D\}. \end{aligned}$$

Podemos interpretar este número de condição dos problemas $(P_e), (D_e)$ como a “menor variável grande na face ótima”. É intuitivo que σ atua como medida da dificuldade do problema, pois quanto menor for essa variável, maior será a dificuldade que enfrentarão os métodos para identificá-la como uma variável grande, e não confundí-la com as que serão nulas na face ótima.

Veremos como estudar a complexidade dos algoritmos de trajetória central usando σ na definição da regra de parada. Consideremos uma seqüência $(x^k, s^k, \mu^k) \in F^o \times \mathbb{R}_+$ gerada por um algoritmo de trajetória central, satisfazendo (2.4).

Dado que se um par de pontos $(x, s) \in \mathcal{CP}_e$ verifica $x_i s_i = \mu$ para todo i , e a condição (2.3) é satisfeita, uma estimativa razoável para μ^k é $\mu^k = \frac{x^{kT} s^k}{n}$. Definimos os conjuntos $B^k = \{j \in \{1, \dots, n\} \mid s_j^k \leq x_j^k\}$ e $N^k = \{j \in \{1, \dots, n\} \mid x_j^k < s_j^k\}$. O seguinte lema estabelece a regra de parada usando o número de condição de Ye:

Lema 2.3.1 *Consideremos uma seqüência $(x^k, s^k, \mu^k) \in F^o \times \mathbb{R}_+$ com $\mu^k \rightarrow 0$ satisfazendo (2.3), e os conjuntos B^k, B, N^k, N como definidos antes. Então $B^k = B$ e $N^k = N$ para todo k tal que*

$$\mu^k < \frac{\sigma^2}{n^2}(1 - \alpha).$$

Demonstração. Consideremos um par ótimo (x^*, s^*) que define a partição ótima $B = B(x^*), N = B(s^*)$ e satisfaz $x_i^* \geq \sigma, s_i^* \geq \sigma, \forall i$. Para todo k temos

$$(x^k - x^*)^T (s^k - s^*) = 0$$

assim $x^{kT} s^k - x^{kT} s^* - x^{*T} s^k + x^{*T} s^* = 0$. Dado que $x^{*T} s^* = 0$, e usando a partição $\{B, N\}$ temos

$$x^{kT} s^k = \sum_{i \in N} x_i^k s_i^* + \sum_{i \in B} x_i^* s_i^k.$$

Agora fixando $j \in B$ temos que $\forall k$ tal que $\mu^k < \sigma^2/n^2$,

$$s_j^k \leq \frac{x^{kT} s^k}{x_j^*} = \frac{n\mu^k}{x_j^*} \leq \frac{n\sigma^2(1 - \alpha)}{\sigma n^2} = \frac{\sigma(1 - \alpha)}{n}.$$

Por outro lado, pelos lemas (2.2.1) e (2.2.3) temos que

$$x_j^k \geq \frac{x_j^*}{n} \geq \frac{\sigma}{n}(1 - \alpha).$$

Isto significa que para todo $j \in B$ e para todo k tal que $\mu^k < \sigma^2/n^2$, obtemos $j \in B^k$.

Similarmente, tomamos $j \in N$ arbitrário e podemos mostrar que $\forall k$ tal que $\mu^k < \sigma^2/n^2$ obtemos $j \in N^k$. Assim $B^k = B$ e $N^k = N$ para todo k que satisfaz $\mu^k < \sigma^2/n^2$, completando a prova. ■

É também conhecido (ver Gonzaga [15]) para vários algoritmos de pontos interiores de trajetória central de passos curtos, que o parâmetro de centralização μ satisfaz uma redução em cada iteração dada por

$$\mu^k \leq \mu^0 \left(1 - \frac{\nu}{\sqrt{n}}\right)^k, \quad (2.4)$$

onde $\nu \in (0, 1)$. Podemos estimar limitantes para o número de iterações dos algoritmos de trajetória central usando (2.4). O modelo clássico para fazer a estimação supõe uma regra de parada dada por $\mu^k < 2^{-L}$, onde L é “o tamanho em bits” da instância A, b, e, x^0 , pressupondo os dados inteiros. No seguinte lema mostraremos o cálculo da limitante:

Lema 2.3.2 *Dada uma seqüência que satisfaz (2.3) e L como definido acima, então:*

$$k = O(\sqrt{n}L).$$

Demonstração. Por (2.4) temos

$$\frac{\mu^k}{\mu^0} \leq \left(1 - \frac{\nu}{\sqrt{n}}\right)^k.$$

Tomando logaritmos e usando a relação $\log(1 - t) \leq -t$, $\forall t \in (0, 1)$ obtemos

$$\log \frac{\mu^k}{\mu^0} \leq k \log\left(1 - \frac{\nu}{\sqrt{n}}\right) \leq -\frac{k\nu}{\sqrt{n}},$$

assim

$$k \leq -\frac{\sqrt{n}}{\nu} \log\left(\frac{\mu^k}{\mu^0}\right). \quad (2.5)$$

Agora substituímos a regra de parada $\mu^k < 2^{-L}$ e temos

$$k \leq \frac{\sqrt{n} \log\left(\frac{2}{\mu^0}\right)}{\nu} L,$$

portanto $k = O(\sqrt{n}L)$. ■

Com um raciocínio análogo, mas usando como regra de parada $\mu^k \leq \frac{\sigma^2}{n^2}$ estabelecemos o seguinte lema, que descreve a complexidade dos algoritmos de trajetória central usando o modelo real dado pelo número de condicionamento de Ye.

Lema 2.3.3 *Dada a seqüência (x^k, s^k, μ^k) que satisfaz (2.3), e a regra de parada definida no lema 2.3.1, verifica-se*

$$k = O(\sqrt{n}(|\log \sigma| + \log n)).$$

Demonstração. Substituímos $\mu^k \leq \frac{\sigma^2}{n^2}$ em (2.5) para obter

$$k \leq -\frac{\sqrt{n}}{\nu} \log\left(\frac{\sigma^2}{n^2 \mu^0}\right) = \frac{2\sqrt{n}}{\nu}(|\log \sigma| + \log n).$$

■

Os resultados acima descrevem os modelos para estudar a complexidade dos algoritmos de pontos interiores que seguem a trajetória central. O modelo clássico mantém a desvantagem de assumir os dados inteiros, e dribla a natureza real (contínua) do problema (P_e) , enquanto o outro modelo supõe conhecida a partição ótima do problema, que desde o ponto de vista de elaboração de algoritmos é incongruente. Embora tenha só importância teórica, o modelo proposto por Ye abre a possibilidade de estabelecer outros modelos reais que estudem a complexidade, possivelmente caracterizando o número de condição, ou achando limitantes inferiores que possam ser calculados num procedimento de inicialização. Na seção 2.8 mostramos uma caracterização.

2.4 O número de condição $\bar{\chi}_A$.

Definiremos nesta seção o número de condição estudado por Stewart [33], O’Leary [29], Todd [35] Vavasis e Ye [45],[46],[47],[48] e outros, que depende da matriz de entradas A do problema (P_e) (sem supor as componentes inteiras) e que têm sido usado ativamente em novos resultados em complexidade de algoritmos de trajetória central para programação linear, (ver [47]). Estudaremos também algumas propriedades.

Definiremos as projeções oblíquas de um ponto $v^o \in \mathbb{R}^n$ sobre os espaços $\mathcal{N}(A)$ e $\mathcal{R}(A^T)$, com respeito ao ponto $d \in \mathbb{R}_{++}^n$. A projeção oblíqua de $v^o \in \mathbb{R}^n$ sobre $\mathcal{R}(A^T)$ com respeito a $d \in \mathbb{R}_{++}^n$, e denotada por $\tilde{Q}_{A,d}v^o$ é a única solução do problema:

$$\begin{aligned} &\text{minimizar } \frac{1}{2}\|D(v - v^o)\|^2 \\ &\text{sujeito a } v = A^T w. \end{aligned} \quad (2.6)$$

Similarmente a projeção oblíqua do vetor v^o sobre $\mathcal{N}(A)$ com respeito a $d \in \mathbb{R}_{++}^n$, e denotada por $Q_{A,d}v^o$ é a única solução do problema:

$$\begin{aligned} &\text{minimizar } \frac{1}{2}\|D^{-1}(u - v^o)\|^2 \\ &\text{sujeito a } Au = 0. \end{aligned} \quad (2.7)$$

Pelas condições de otimalidade dos problemas acima, as projeções oblíquas $\tilde{Q}_{A,d}v^o$ e $Q_{A,d}v^o$ satisfazem as seguintes equações respectivamente:

$$\begin{aligned} D^2(v - v^o) + \lambda &= 0 \\ A\lambda &= 0 \\ A^T y &= v, \end{aligned} \quad (2.8)$$

e

$$\begin{aligned} D^{-2}(u - v^o) + A^T y &= 0 \\ Au &= 0. \end{aligned} \quad (2.9)$$

Se a matriz $A \in \mathbb{R}^{m \times n}$ é de posto máximo m , então usando as relações (2.8) e (2.9) podemos obter expressões explícitas para as matrizes de projeção $\tilde{Q}_{A,d}$ e $Q_{A,d}$:

$$\tilde{Q}_{A,d} = A^T(AD^2A^T)^{-1}AD^2 \quad (2.10)$$

e

$$Q_{A,d} = I - D^2A^T(AD^2A^T)^{-1}A, \quad (2.11)$$

mas no caso geral as matrizes de projeção são encontradas tomando colunas linearmente independentes de A^T , no caso de $\tilde{Q}_{A,d}$ ou linhas de A linearmente independentes para $Q_{A,d}$. Esta afirmação é possível pois as projeções não dependem da representação matricial A , mas sim dos espaços linha e coluna que define. Daremos outras representações das projeções oblíquas que serão úteis, e que se derivam diretamente das condições de otimalidade (2.8) e (2.9):

i) Existe $\tilde{y} \in \mathcal{R}(A^T)$ tal que

$$\tilde{Q}_{A,d}v^o = A^T\tilde{y} \quad (2.12)$$

e

ii) Existe $\hat{y} \in \mathcal{R}(A^T)$ tal que

$$Q_{A,d}v^o = D^2A^T\hat{y}. \quad (2.13)$$

ou também

$$\tilde{Q}_{A,d}v^o = (A_d^+A)^Tv^o = A^T(A_d^+)^Tv^o \quad (2.14)$$

e

$$Q_{A,d}v^o = (I - A_d^+A)v^o \quad (2.15)$$

onde A_d^+ denota a matriz pseudo inversa pesada com respeito a d , que satisfaz $AA_d^+ = I$.

A projeção oblíqua $\tilde{Q}_{A,d}$ permite decompor o espaço \mathbb{R}^n como soma direta dos subespaços $\mathcal{R}(A^T)$ e $\mathcal{N}(AD^2)$, enquanto que a projeção $Q_{A,d}$ da a decomposição como soma direta de $\mathcal{N}(A)$ e $\mathcal{R}(D^2A^T)$.

Definiremos agora os números de condição, χ_A e $\bar{\chi}_A$ que dependem só da matriz A como

$$\bar{\chi}_A := \sup\{\|\tilde{Q}_{A,d}v\| \mid \|v\| = 1, d \in \mathbb{R}_{++}^n\}$$

e

$$\chi_A := \sup\{\|A_d^+v\| \mid \|v\| = 1, d \in \mathbb{R}_{++}^n\}.$$

$\bar{\chi}_A$ pode ser interpretado como o supremo de todas as projeções oblíquas sobre $\mathcal{R}(A^T)$ de vetores de norma um. Lembremos que a norma de uma matriz M é definida como $\|M\| = \max\{Mx \mid \|x\| = 1\}$, assim

$$\|\tilde{Q}_{A,d}\| = \|A^T(AD^2A^T)^{-1}AD^2\| = \max\{\tilde{Q}_{A,d}v \mid \|v\| = 1\},$$

e o supremo pode ser escrito como

$$\sup\{\|A^T(AD^2A^T)^{-1}AD^2\| \mid d \in \mathbb{R}_{++}^n\}, \quad (2.16)$$

que é a forma usualmente encontrada na bibliografia. Numa primeira aproximação intuitiva poderíamos pensar que o supremo de (2.16) é infinito, mas Stewart [33] e O’Leary [29], provaram o contrário. Veremos esse resultado: Definimos os conjuntos

$$S = \{s \in \mathcal{R}(A^T) \mid \|s\| = 1\}$$

$$X = \{x \in \mathcal{N}(AD^2) \mid d \in \mathbb{R}_{++}^n\},$$

e a distância entre eles dada por

$$\rho_A = \inf\{\|s - x\| \mid s \in S, x \in X\}.$$

Lema 2.4.1 *Considere os conjuntos X e S definidos acima. Então*

$$\rho_A = \inf\{\|s - x\| \mid s \in S, x \in X\} \neq 0.$$

Demonstração. É suficiente mostrar que $(\text{cl}X) \cap S = \emptyset$. De fato, suponhamos por contradição que existe uma seqüência (x^k) em X tal que $\forall k \in \mathbb{N}$, $x^k \rightarrow x \in S$, então existe uma seqüência (d^k) em \mathbb{R}_{++}^n tal que $x^k \in \mathcal{N}(A(D^k)^2)$, de maneira que $(D^k)^2 x^k \perp x$, isto é,

$$0 = x^T D^{k^2} x^k = \sum_{i=1}^n x_i d_i^{k^2} x_i^k. \quad (2.17)$$

Por outro lado, se $x_i \neq 0$ então $x_i x_i^k > 0$, e assim o lado direito de (2.17) é positivo, o que é uma contradição. ■

O próximo teorema vincula a distância ρ_A com o supremo de (2.16), fornecendo um limitante superior para ele.

Teorema 2.4.2 *Consideremos uma matriz $A \in \mathbb{R}^{m \times n}$ com posto m . Então*

$$\bar{\chi}_A = \sup\{\|A^T(AD^2A^T)^{-1}AD^2\| \mid d \in \mathbb{R}_{++}^n\} \leq \rho_A^{-1}$$

e

$$\chi_A = \sup\{\|(AD^2A^T)^{-1}AD^2\| \mid d \in \mathbb{R}_{++}^n\} \leq \rho_A^{-1}\|A^+\|.$$

Demonstração. Fixemos $d \in \mathbb{R}_{++}^n$ e $z \in \mathbb{R}^n$, tal que $\|z\| = 1$. Podemos decompor z como $z = \tilde{Q}_{A,d}z + (I - \tilde{Q}_{A,d})z$ onde $\tilde{Q}_{A,d}z \in \mathcal{R}(A^T)$ e $(I - \tilde{Q}_{A,d})z \in$

$\mathcal{N}(AD^2)$ são as projeções oblíquas discutidas acima. Suponhamos sem perda de generalidade que $\tilde{Q}_{A,d}z \neq 0$. Então

$$\|\tilde{Q}_{A,d}z\|^{-1} = \frac{\|z\|}{\|\tilde{Q}_{A,d}z\|} = \left\| \frac{\tilde{Q}_{A,d}z}{\|\tilde{Q}_{A,d}z\|} + \frac{(I - \tilde{Q}_{A,d})z}{\|\tilde{Q}_{A,d}z\|} \right\| \geq \rho_A$$

pois $\frac{\tilde{Q}_{A,d}z}{\|\tilde{Q}_{A,d}z\|} \in S$ e $\frac{(I - \tilde{Q}_{A,d})z}{\|\tilde{Q}_{A,d}z\|} \in X$. Isto é

$$\|\tilde{Q}_{A,d}z\| = \|A^T(AD^2A^T)^{-1}AD^2z\| \leq \rho_A^{-1}$$

para $d \in \mathbb{R}_{++}^n$ arbitrário. Assim tomando o supremo obtemos a primeira parte. A segunda segue-se já que $(AD^2A^T)^{-1}AD^2 = A^+A^T(AD^2A^T)^{-1}AD^2$. ■

O resultado afirma que as projeções oblíquas estão limitadas superiormente, quer dizer o supremo em (2.16) é sempre finito para qualquer matriz A , e portanto $\bar{\chi}_A$ esta bem definido.

Existe uma característica combinatória na definição de ρ_A . Para ver isto, definamos $\text{sig}(t)$ para $t \in \mathbb{R}$ como $\text{sig}(t) = 1$ se $t > 0$, $\text{sig}(t) = 0$ se $t = 0$ e $\text{sig}(t) = -1$ se $t < 0$. Para um vetor v , $\text{sig}(v)$ é definido componente a componente. Definimos para algum ponto $\hat{x} \in \mathbb{R}^n$ o cone

$$C(\hat{x}) = \{x \in \mathbb{R}^n \mid \exists d > 0 \text{ tal que } x = D\hat{x}\}.$$

é fácil ver que $C(\hat{x}) = \{x \in \mathbb{R}^n \mid \text{sig}(x) = \text{sig}(\hat{x})\}$. Podemos caracterizar o conjunto X pela união dos cones $C(\hat{x})$ com $\hat{x} \in \mathcal{N}(A)$:

Lema 2.4.3 *Para X e $C(x)$ definidos como antes temos:*

i)

$$X = \bigcup_{\hat{x} \in \mathcal{N}(A)} C(\hat{x}).$$

ii) *O número de cones $C(x)$ que caracterizam X em i) é finito.*

Demonstração. Para demonstrar i) consideremos $x \in X$, então existe um ponto $\bar{d} \in \mathbb{R}_{++}^n$ tal que $x \in \mathcal{N}(A\bar{D}^2)$. Definimos $\bar{x} = \bar{D}^2x \in \mathcal{N}(A)$, e assim $x = \bar{D}^{-2}\bar{x}$, portanto $x \in C(\bar{x}) \subset \bigcup_{\hat{x} \in \mathcal{N}(A)} C(\hat{x})$. Por outro lado, se $x \in \bigcup_{\hat{x} \in \mathcal{N}(A)} C(\hat{x})$ então

existe $\tilde{x} \in \mathcal{N}(A)$ tal que $x \in C(\tilde{x})$. Definimos \tilde{d} por $\tilde{d}_i = \sqrt{\frac{\tilde{x}_i}{x_i}}$ se $x_i \neq 0$ e $\tilde{d}_i = 1$ se $x_i = 0$. Temos então $A\tilde{D}^2x = A\tilde{x} = 0$, e assim $x \in X$. A parte ii) é trivial observando que o conjunto $\{v \in \mathbb{R}^n \mid v = \text{sig}(x) \text{ para algum } x \in \mathbb{R}^n\}$ é finito. ■

O lema (2.4.3) põe em evidência a natureza combinatória da distância ρ_A , pois a podemos calcular como a menor das distâncias entre o conjunto S e cada um dos cones $C(\hat{x})$ que caracterizam X . O seguinte resultado devido a O’Leary [29] explora esta relação para calcular ρ . Dada uma matriz $E \neq 0$, a notação $\text{inf}_+(E)$ significa o menor valor singular de E diferente de zero.

Lema 2.4.4 *Consideremos uma base ortonormal U do subespaço $\mathcal{R}(A^T)$. Então*

$$\rho_A = \min\{\text{inf}_+(U_J) \mid J \subset \{1, \dots, n\}\}. \quad (2.18)$$

O lema acima proporciona uma ferramenta para calcular ρ_A e assim $\bar{\chi}_A$; No entanto, o caráter combinatório da expressão (2.18) faz ao problema de calcular ρ_A tão ou mais difícil do que o problema de programação linear. Recentemente Tunçel [42] fez uma prova simplificada deste resultado.

2.5 Propriedades duais de $\bar{\chi}_A$

Nesta seção descrevemos o primeiro resultado principal do trabalho. Estudaremos o ponto de vista dual do número de condicionamento $\bar{\chi}_A$. Ele está definido como o supremo das normas das matrizes de projeção oblíqua sobre o conjunto $\mathcal{R}(A^T)$. Suponhamos que a matriz $A \in \mathbb{R}^{m \times n}$ tem posto máximo m (Esta suposição não é necessária, mas a fazemos para facilitar a notação). Dado $d \in \mathbb{R}_{++}^n$ a matriz de projeção oblíqua sobre o conjunto $\mathcal{N}(A)$ é dada por $Q_{A,d} = I - D^2A^T(AD^2A^T)^{-1}A$. Tomamos uma matriz Z cujas linhas definem uma base de $\mathcal{N}(A)$, assim $\mathcal{N}(A) = \mathcal{R}(Z^T)$ e podemos definir $\bar{\chi}_Z$ de maneira análoga a $\bar{\chi}_A$, quer dizer, como o supremo das projeções oblíquas sobre $\mathcal{R}(Z^T) = \mathcal{N}(A)$. Numa primeira inspeção vemos que

$$\|Q_{A,d}\| = \max_{\|x\|=1} \|I - D^2A^T(AD^2A^T)^{-1}Ax\| \leq 1 + \bar{\chi}_A,$$

com a desigualdade valendo pela desigualdade triangular e devido a que a norma de matrizes é preservada pela transposição, quer dizer

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

Isto nos levaria a pensar que possivelmente $\bar{\chi}_Z = \sup\{\|Q_{A,d}\| \mid d \in \mathbb{R}_{++}^n\} \neq \bar{\chi}_A$. Nós provamos em [18] que são iguais. Mostramos esse resultado a seguir: Consideremos os conjuntos

$$X' := \{x \in \mathcal{R}(Z^T) \mid \|x\| = 1\}$$

e

$$S' := \{s \in \mathcal{N}(ZD^2) \mid d \in \mathbb{R}_{++}^n\}.$$

Caracterizaremos esses conjuntos no seguinte fato:

Fato 2.5.1 *Dados os conjuntos X' e S' definidos como acima,*

$$X' = \{x \in \mathcal{N}(A) \mid \|x\| = 1\}$$

e

$$S' = \{s \in \mathcal{R}(D^{-2}A^T) \mid d \in \mathbb{R}_{++}^n\}.$$

Demonstração. A primeira igualdade é trivial. Vamos então à segunda: Se $\hat{s} \in S'$ então existe $\hat{d} \in \mathbb{R}_{++}^n$ tal que $Z\hat{D}^2\hat{s} = 0$. Isto é, $\hat{D}^2\hat{s} \in \mathcal{N}(Z)$ portanto $\hat{D}^2\hat{s} \perp \mathcal{R}(z^T) = \mathcal{N}(A)$; assim $\hat{D}^2\hat{s} \in \mathcal{R}(A^T)$ o que significa $\hat{s} \in \mathcal{R}(D^{-2}A^T)$ e finalmente $\hat{s} \in \{s \in \mathcal{R}(D^{-2}A^T) \mid d \in \mathbb{R}_{++}^n\}$. De forma similar podemos provar que se $\bar{s} \in \{s \in \mathcal{R}(D^{-2}A^T) \mid d \in \mathbb{R}_{++}^n\}$ então $\bar{s} \in S'$. Isto completa a prova. ■

Observemos que, devido ao lema acima, os conjuntos X, S, X', S' não dependem da matriz A mas sim dos subespaços $\mathcal{N}(A)$ e $\mathcal{R}(A^T)$. Desta forma poderíamos definir os números de condição com respeito a um subespaço em lugar da matriz que o representa. Definamos agora a distância ρ_Z :

$$\rho_Z := \inf\{\|x - s\| \mid x \in X', s \in S'\}.$$

Definimos o conjunto $C(\hat{s}) = \{s \in \mathbb{R}^n \mid \exists d \in \mathbb{R}_{++}^n, s = D\hat{s}\}$. Estabeleceremos sem demonstração o resultado análogo ao lema 2.4.3:

Lema 2.5.2 Para S' e $C(\hat{s})$ definidos como antes temos:

i)

$$S' = \bigcup_{\hat{s} \in \mathcal{R}(A^T)} C(\hat{s}).$$

ii) O número de cones $C(\hat{s})$ que caracterizam S' em i) é finito.

Consideremos para $d \in \mathbb{R}_{++}^n$ fixo, o problema

$$\begin{aligned} &\text{minimizar} && \frac{1}{2} \|x - s\|^2 \\ &\text{sujeito a} && Zs = 0 \\ &&& AD^2x = 0 \\ &&& \|s\| = 1. \end{aligned} \tag{2.19}$$

As condições de otimalidade para o problema dão para algum $y \in \mathbb{R}^m$, $w \in \mathbb{R}^{n-m}$ e $\alpha \in \mathbb{R}$

$$\bar{x} - \bar{s} + D^2A^T y = 0, \tag{2.20}$$

$$\bar{s} - \bar{x} + Z^T w - \alpha \bar{s} = 0. \tag{2.21}$$

Teorema 2.5.3 Nas condições acima,

$$\rho_A = \rho_Z.$$

Demonstração. Provaremos que $\rho_A \geq \rho_Z$. Isto é suficiente pela simetria da construção: Se começamos com a matriz Z , então as linhas de A proporcionam uma base para $\mathcal{N}(Z)$ e podemos concluir $\rho_Z \geq \rho_A$.

Consideremos vetores arbitrários $x \in X, s \in S$. Pela definição de X , existe $d \in \mathbb{R}_{++}^n$ tal que $AD^2x = 0$. Trabalharemos com este d fixo para provar que $\|x - s\| \geq \rho_Z$.

Sejam \bar{x} e \bar{s} soluções ótimas para o problema (2.19). Consideremos os vetores seguintes:

$$x' = \frac{\bar{x} - (1 - \alpha)\bar{s}}{\|\bar{x} - (1 - \alpha)\bar{s}\|}, s' = \beta(\bar{x} - \bar{s}),$$

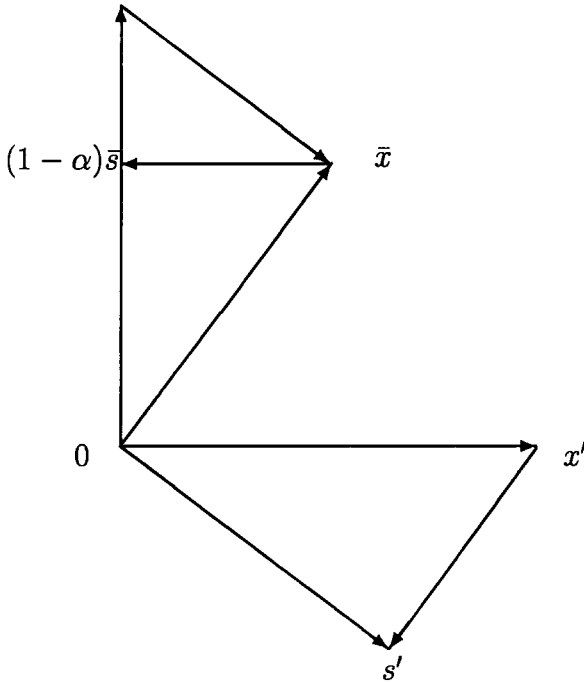
onde $\beta \in \mathbb{R}$ será especificado daqui a pouco. Então $x' \in X'$ por (2.21) e pelo fato que $\|x'\| = 1$, e $s' \in S'$ para qualquer $\beta \in \mathbb{R}$ por (2.20).

Toda a ação acontece no espaço bidimensional definido por \bar{x} e \bar{s} . Podemos usar a geometria em \mathbb{R}^2 , seguindo a figura. O triângulo $(0, \bar{s}, \bar{x})$ é retângulo, dado que por (2.21) $\bar{x} - \bar{s} \perp \bar{x} \in \mathcal{N}(AD^2)$. Seja θ o ângulo entre \bar{x} e \bar{s} . Então $\|\bar{x} - \bar{s}\| = \text{sen}\theta$.

Temos que $x' \perp \bar{s}$, devido a (2.21) e o fato que $\bar{s} \in \mathcal{R}(A^T)$. Como vimos no parágrafo anterior, $s' \perp \bar{x}$, para qualquer $\beta \in \mathbb{R}$. Segue-se que o sinal de β pode ser escolhido de maneira que o ângulo entre x' e s' seja θ , e seu valor absoluto pode ser escolhido tal que $\|s'\| = \|\bar{x}\|$. Assim o triângulo $(0, x', s')$ é idêntico ao triângulo $(0, \bar{s}, \bar{x})$, e segue-se que

$$\|x - s\| \geq \|\bar{x} - \bar{s}\| = \|x' - s'\| \geq \rho_Z.$$

Dado que $x \in X$ e $s \in S$ foram escolhidos arbitrariamente, a desigualdade permanece válida tomando ínfimo, e assim completamos a prova. ■



O resultado foi referenciado nos trabalhos de Ye [52], de Vavasis e Ye [49], e de Tunçel [42]. Em [49], Vavasis e Ye estudam o tamanho das variáveis ao longo da trajetória central \mathcal{CP} , separando em grandes e pequenas segundo um critério que

envolve o número de condição $\bar{\chi}_A$. Especificamente eles separam as variáveis x_i e x_j como grandes e pequenas se $\frac{x_i}{x_j} \geq g$ onde $g = K\sqrt{n}(\bar{\chi}_A + 1)$ para uma constante K . Segundo o nosso resultado, a expressão $(\bar{\chi}_A + 1)$ pode ser substituída por $\bar{\chi}_A$ na definição de g (Ver [49]).

2.6 Mais propriedades de $\bar{\chi}_A$

Agora mostraremos mais propriedades dos números de condição visando a relação com a complexidade:

Fato 2.6.1 *Consideremos a matriz $A \in \mathbb{R}^{m \times n}$. Então*

$$\bar{\chi}_A \geq 1.$$

Demonstração. Temos que $0 \in X$, então $1 = \|0 - x\| \geq \rho_A \geq \frac{1}{\bar{\chi}_A}$. ■

Lema 2.6.2 *Consideremos a matriz $A \in \mathbb{R}^{m \times n}$ não nula, e a partição $\{B, N\}$ das colunas da matriz A tal que $A = [A_B, A_N]$. Então*

$$\bar{\chi}_{A_B} \leq \bar{\chi}_A.$$

Demonstração. Suponhamos sem perda de generalidade que A_B é tomada como as primeiras colunas de A . Definimos

$$\rho_{A_B} = \rho_{Z_B} = \inf\{\|u - v\| \mid u \in U, v \in V\},$$

onde $U = \{u \in \mathcal{R}(D_B^2 A_B^T) \mid d \in \mathbb{R}_{++}^n\}$ e $V = \{v \in \mathcal{N}(A_B) \mid \|v\| = 1\}$. Consideremos $u \in U$ e $v \in V$ escolhidos arbitrariamente. Queremos provar que $\|u - v\| \geq \rho_Z$. Como $u \in U$, existe $d \in \mathbb{R}_{++}^n$ tal que $u = D_B^2 A_B^T q$ para algum q , e $v \in V$ implica que $\|v\| = 1$ e $A_B v = 0$. Definimos o vetor $u' = D'^2 A^T q$ onde $d' = (d_B^T, \lambda e_N^T)^T$; isto é, a parte de u' indexada por B é u , e a outra parte vai para zero quando $\lambda \rightarrow 0$. Assim $u' \in X'$ e

$$\left\| \begin{pmatrix} u \\ 0 \end{pmatrix} - u' \right\| \rightarrow 0, \lambda \rightarrow 0.$$

Por outro lado, consideremos o vetor $v' = (v^T, 0)^T \in \mathbb{R}^n$. Temos $Av' = 0$, e $\|v'\| = \|v\| = 1$; Isto é, $v' \in S'$. Dado que $u \in U$ e $v \in V$ temos

$$\|u - v\| \geq \rho_{AB},$$

mas $\|u' - v'\| \rightarrow \|u - v\|$ pois as últimas $n - |B|$ componentes de $u' - v'$ são arbitrariamente pequenas; assim

$$\|u - v\| \geq \rho_A.$$

■

O seguinte lema, devido a Vavasis e Ye [47] demonstra uma propriedade de separação entre variáveis grandes e pequenas:

Lema 2.6.3 *Consideremos $z \in \mathcal{N}(A)$, $w \in \mathcal{R}(A^T)$ e $d \in \mathbb{R}_{++}^n$ escolhido arbitrariamente. Então*

$$\begin{aligned} \|w\| &\leq \bar{\chi}_A \|w + D^2 z\|, \\ \|z\| &\leq \bar{\chi}_A \|z + D^2 w\|. \end{aligned}$$

Demonstração. Dado que $z \in \mathcal{N}(A)$ e $w \in \mathcal{R}(A^T)$, temos $D^2 w \in \mathcal{N}(AD^2)$ e a projeção oblíqua de $w + D^2 z$ sobre $\mathcal{R}(A^T)$ denotada por $\tilde{Q}_{A,d}(w + D^2 z)$, é igual a w , e $(I - \tilde{Q}_{A,d})(w + D^2 z) = D^2 z$. Assim

$$w = \tilde{Q}_{A,d}(w + D^2 z) = A^T (AD^2 A^T)^{-1} (w + D^2 z)$$

portanto $\|w\| \leq \bar{\chi}_A \|w + D^2 z\|$.

Por outro lado temos que $z = \tilde{Q}_{A,d}(z + D^2 w)$ e pelo teorema 2.5.3 obtemos que $\|z\| \leq \bar{\chi}_A \|z + D^2 w\|$. ■

Demonstraremos agora três lemas relacionados com o número $\bar{\chi}_A$ e o conjunto $T_{A,b} = \{x \in \mathbb{R}^n \mid Ax = b\}$ que serão usados no capítulo 3. Definimos, para $d \in \mathbb{R}_{++}^n$, como $t(d)$ a solução de mínima norma de $Ax = b$ usando a norma $\mathbb{R}^n \ni x \mapsto \|x\|_{d^{-1}} = \|D^{-1}x\|$, quer dizer

$$t(d) := \operatorname{argmin}\{\|D^{-1}t\| \mid At = b\}.$$

Podemos caracterizar $t(d)$ usando as condições de otimalidade, como a única solução que satisfaz

$$At = b \text{ e } t = D^2 A^T y, \quad (2.22)$$

assim, para cada $d \in \mathbb{R}_{++}^n$, temos $t(d) \in T \cap S'$, com S' definida na seção anterior. Denotemos por A^+ à matriz pseudo-inversa de A .

Lema 2.6.4 *Dada a matriz $A \in \mathbb{R}^{m \times n}$ e o vetor $b \in \mathcal{R}(A)$, verifica-se para todo $d \in \mathbb{R}_{++}^n$ que*

$$\|t(d)\| \leq \bar{\chi}_A \|A^+ b\|, \forall d \in \mathbb{R}_{++}^n$$

onde $t(d) := \operatorname{argmin}\{\|D^{-1}t\| \mid At = b\}$.

Demonstração. Pelas condições de otimalidade (2.22) obtemos para $d \in \mathbb{R}_{++}^n$ fixo mas arbitrário que $t(d) = D^2 A^T (AD^2 A^T)^{-1} b$. Agora

$$\begin{aligned} \|t(d)\| &= \|D^2 A^T (AD^2 A^T)^{-1} A A^+ b\| \\ &\leq \|D^2 A^T (AD^2 A^T)^{-1} A\| \|A^+ b\| \\ &= \|A^T (AD^2 A^T)^{-1} AD^2\| \|A^+ b\| \\ &\leq \bar{\chi}_A \|A^+ b\|. \end{aligned}$$

■

É fácil ver que

$$\{t \in \mathbb{R}^n \mid t = \operatorname{argmin}\{\|D^{-1}t\| \mid At = b\}, d \in \mathbb{R}_{++}^n\} = T_{A,b} \cap S',$$

assim $t(d) \in T_{A,b} \cap S'$. Pelo lema (2.4.3) caracterizamos o conjunto S' por $S' = \bigcup_{\hat{s} \in \mathcal{R}(A^T)} C(\hat{s})$ onde $C(\hat{s}) = \{s \mid \exists d \in \mathbb{R}_{++}^n \text{ tal que } s = D^2 \hat{s}\}$. O lema (2.22) afirma que $T_{A,b} \cap S'$ é limitado. Podemos então definir o vetor de máxima norma em $\operatorname{cl}(T_{A,b} \cap S') = T_{A,b} \cap \operatorname{cl}(S')$.

Lema 2.6.5 *O vetor \hat{t} de máxima norma de $\operatorname{cl}(T_{A,b} \cap S')$ pertence à fronteira de $\operatorname{cl}(T_{A,b} \cap S')$.*

Demonstração. O conjunto $T_{A,b} \cap S'$ pode ser caracterizado como

$$T_{A,b} \cap \left(\bigcup_{\hat{s} \in \mathcal{R}(A^T)} C(\hat{s}) \right).$$

Pelo lema 2.6.4 este conjunto é limitado, portanto para qualquer $\hat{s} \in \mathcal{R}(A^T)$ fixo, o conjunto $T_{A,b} \cap C(\hat{s})$ também é limitado. Como

$$\hat{t} \in \text{cl}(T_{A,b} \cap (\bigcup_{\hat{s} \in \mathcal{R}(A^T)} C(\hat{s}))),$$

existe um ponto $\bar{s} \in \mathcal{R}(A^T)$ tal que $\hat{t} \in T_{A,d} \cap \text{cl}C(\bar{s})$. Cada conjunto $T_{A,d} \cap \text{cl}C(\hat{s})$ é um politopo, pois é a interseção de um cone poliedral convexo com uma variedade afim, e é limitado. Assim \hat{t} é o vetor de máxima norma do politopo $T_{A,d} \cap \text{cl}C(\bar{s})$. É bem conhecido que o vetor de máxima norma de um politopo está num vértice, quer dizer, na fronteira do politopo. Assim terminamos a prova. ■

Lema 2.6.6 *Se \hat{x} é um vértice de $\text{cl}(T_{A,b} \cap S')$ então existe um vetor $\bar{s} \in \mathcal{R}(A^T)$ tal que para a partição $\{B, N\} = \{B(\hat{x}), N(\hat{x})\}$ dos índices $\{1, \dots, n\}$, temos $A_B x_B = 0$, $x_N = 0$, $x_i \bar{s}_i > 0$ para todo $i \in B$, e as colunas de A_B são linearmente independentes.*

Demonstração. Seja \hat{s} um vértice de $T := \text{cl}(T_{A,b} \cap S')$. Pelo lema (2.5.2) escrevemos T como

$$T = \text{cl}((\bigcup_{\hat{s} \in \mathcal{R}(A^T)} C(\hat{s})) \cap T_{A,b}) = \bigcup_{\hat{s} \in \mathcal{R}(A^T)} \text{cl}(C(\hat{s}) \cap T_{A,b}).$$

Cada conjunto $\text{cl}(C(\hat{s}) \cap T_{A,b})$ é um politopo (pelo lema 2.6.4), então existe $\bar{s} \in$ tal que \hat{x} é um vértice do politopo $\text{cl}(C(\bar{s}) \cap T_{A,b})$; isto é existe uma seqüência (d^k) em \mathbb{R}_{++}^n tal que $\hat{x} = \lim_{k \rightarrow \infty} D^{k^2} A^T \bar{y}$ onde $\bar{s} = A^T \bar{y}$. Podemos descrever esse politopo pelas relações $Ax = b$, $x_i \bar{s}_i \geq 0$ para todo $i \in \{1, \dots, n\}$. Mudamos as variáveis da seguinte maneira: para cada $i \in \{1, \dots, n\}$, $z_i = x_i$ se $\bar{s}_i \geq 0$; e $z_i = -x_i$ se $\bar{s}_i < 0$. A equação $Ax = b$ é equivalente a $Az = b$, enquanto as relações $x_i \bar{s}_i \geq 0$ para todo $i \in \{1, \dots, n\}$ equivalem a $z_i \geq 0$ para todo $i \in \{1, \dots, n\}$. Tomamos \hat{z} definido desde \hat{x} mudando as variáveis. Então \hat{z} é um vértice do politopo definido por

$$\begin{aligned} Az &= b \\ z &\geq 0. \end{aligned}$$

assim $A_B \hat{z}_B = b$, $\hat{z}_N = 0$, e as colunas de A_B são linearmente independentes, pois \bar{z} é um vértice. Restaurando as variáveis obtemos a tese. ■

2.7 O número $\bar{\chi}_A$ e complexidade

Mostraremos agora uma interessante relação entre os números $\bar{\chi}_A$ e σ_A , dada por Vavasis e Ye [46] no contexto do problema da viabilidade no formato de Karmarkar ou formato homogêneo. Consideremos o poliedro da forma

$$HP = \{x \in \mathbb{R}^n \mid Ax = 0, e^T x = 1, x \geq 0\},$$

onde $A \in \mathbb{R}^{m \times n}$ é de posto m . Estabelecemos as seguintes suposições para o conjunto HP :

- Suposição 1: $HP \neq \emptyset$.
- Suposição 2: Existe $\bar{y} \neq 0$ tal que $A^T \bar{y} \geq 0$.

Dado que estamos interessados só no poliedro no formato de Karmarkar, e no problema de viabilidade associado, podemos redefinir o problema como

$$\begin{aligned} &\text{maximizar} && 0^T x \\ &\text{sujeito a} && Ax = 0 \\ &&& e^T x = 1 \\ &&& x \geq 0, \end{aligned}$$

com problema dual

$$\begin{aligned} &\text{minimizar} && z \\ &\text{sujeito a} && A^T y + ze \geq 0. \end{aligned}$$

A suposição 1 implica que $z = 0$ para toda solução ótima do dual. Definimos o conjunto viável primal-dual como

$$F_{PH} = \{(x, s) \mid Ax = 0, x \geq 0, e^T x = 1, s + ze \geq 0, A^T y = s, y \in \mathbb{R}^m\}.$$

Nossas suposições implicam que ambos os problemas são viáveis e que existe uma partição própria $\{B, N\}$ dos índices $\{1, \dots, n\}$ que definem o conjunto com interior relativo não vazio:

$$\Omega_{PH} = \{(x, s) \in F_{PH} \mid x_B > 0, x_N = 0, s_B = 0, s_N > 0\}.$$

(Usando o lema de Farkas vemos que a suposição 1 é verificada se e somente se $N \neq \{1, \dots, n\}$ e a suposição 2 é certa se e somente se $B \neq \{1, \dots, n\}$) Mais ainda, todas as soluções ótimas dos problemas primal-dual acima, satisfazem $x_N = 0$ e $s_B = 0$.

Definimos agora o número de condição σ_H para este problema:

$$\begin{aligned}\sigma_{HP} &= \min_{j \in B} \{\max x_j \mid (x, s) \in \Omega_{HP}\}, \\ \sigma_{HD} &= \min_{j \in N} \{\max s_j \mid (x, s) \in \Omega_{HP}\}, \\ \sigma_H &= \min\{\sigma_{HP}, \sigma_{HD}\}.\end{aligned}$$

Teorema 2.7.1 *Nas condições acima,*

$$\sigma_H \geq \frac{1}{\bar{\chi}_A}.$$

Demonstração. Provaremos primeiro que $\sigma_{HD} \geq \frac{1}{\bar{\chi}_A}$: Definimos $n1 = |N|$. Para cada $i \in N$ e $\mu > 0$ consideremos o seguinte problema:

$$\begin{aligned}\text{maximizar} \quad & s_i + \mu \sum_{i \in N} \log s_i \\ \text{sujeito a} \quad & (x, s) \in \Omega_{PH}.\end{aligned}$$

Este problema possui uma única solução que satisfaz $s_N(-e_i - x_N - \lambda e) = \mu e$, $s_N + A_N^T y = 0$, $s_B = 0$ e $A_B x_B + A_N x_N = 0$, onde x_B, x_N são os multiplicadores de Lagrange e e_i denota o i -ésimo vetor canônico. Como $s_N^T x_N = s^T x = 0$ temos $-\lambda = n1\mu + s_i$, e assim

$$s_N \bar{x}_N + s_N = s_N \bar{x}_N - A_N^T y = \frac{\mu}{n1\mu + s_i} e + \frac{1}{n1\mu + s_i} s_N e_i,$$

onde $\bar{x} = \frac{x}{n1\mu + s_i}$. Notemos também que como $\mu \rightarrow 0$, então

$$s_i \rightarrow s_i^* = \operatorname{argmax}\{s_i \mid (x, s) \in \Omega_{PH}\}.$$

Escolhemos a matriz D tal que $D_B = \mu I$ e D_N tem μ na i -ésima componente da diagonal, e 1 em outro caso. Então

$$\left\| \begin{pmatrix} D_B & 0 \\ 0 & D_N \end{pmatrix} \bar{x} - A^T y \right\| \rightarrow s_i, (\mu \rightarrow 0)$$

assim $s_i^* \geq s_i \geq \frac{1}{\bar{\chi}_A}$.

Observemos que a equação $Ax = 0$ pode ser escrita como $x = Z^T q$, para algum $q \in \mathbb{R}^m$. Substituindo A por Z em todo lugar, obtemos com um raciocínio similar que $\sigma_{HP} \geq \frac{1}{\bar{\chi}_Z}$, assim pelo teorema 2.5.3 obtemos $\sigma_{HP} \leq \frac{1}{\bar{\chi}_A}$ e concluímos a prova.

■

Como estudamos no lema 2.3.3, se a regra de parada é $\mu^k \leq \frac{\sigma_H^2}{n^2}$, então os algoritmos de trajetória central páram após $O(\sqrt{n}(|\log \sigma| + \log n))$ iterações. Assim no caso do problema no formato de Karmarkar, um algoritmo de trajetória central pára após $O(\sqrt{n}(\log \bar{\chi}_A + \log n))$ iterações.

Embora este resultado seja significativo pois definimos uma regra de parada que não precisa do conhecimento da face ótima, recentemente Tunçel [42] estabeleceu um exemplo onde o número σ_H é grande (da ordem de $\frac{1}{2}$) para uma família de instâncias A_ϵ que definem polígonos HP_ϵ e o número $\bar{\chi}_\epsilon$ depende de ϵ e é tão pequeno quanto quisermos. Isto demonstra que o número $\bar{\chi}_A$ não carrega informação acerca de σ_H , assim poderíamos dar o rótulo de mal condicionado a um problema que é essencialmente fácil. Em conclusão, existem ainda questões para ser respondidas acerca dos números de condicionamento e complexidade. Isto justifica a tentativa de definição de novos números de condição que usem mais estrutura do problema de programação linear, e não só da matriz A . Na próxima seção definiremos um novo número de condição.

2.8 Uma caracterização da menor variável grande

Nesta seção consideramos os problemas de programação linear no formato normalizado (P_ϵ) e (D_ϵ) apresentados na seção 2.2. Queremos estabelecer um limitante inferior para o número de condicionamento da menor variável grande na face ótima σ dado por Ye [50],[51] e que descrevemos na seção 2.3. O número σ representa uma medida do condicionamento do problema de programação linear, dado que quanto menor for, mais difícil será identificar a partição ótima, dada a dificuldade de distinguir se dita variável sera nula ou positiva no interior relativo da face ótima. A

principal desvantagem do uso deste número para estudar os algoritmos e sua complexidade é que sua definição depende de conhecer a partição ótima. Na seção 2.7 vimos uma tentativa de caracterizar dito número através de $\bar{\chi}_A$ mas como demonstrou Tunçel [42] o limitante dado por este número é muito grosseiro. Além disso o estudo de Vavasis e Ye foi feito no contexto do problema de programação linear no formato de Karmarkar, e no problema de viabilidade, portanto usa pouca estrutura do problema de programação linear. O resultado desta seção representa uma caracterização do número σ em programação linear no formato padrão normalizado. Usaremos a teoria de dualidade aplicada aos problemas que definem o maior valor de uma variável na face ótima, pois o menor destes valores define σ .

Para definir uma caracterização da menor variável grande na face ótima, precisamos uma característica do problema (P_e) que possa ser relacionada com a partição ótima, mas que só dependa dos dados. Definiremos os seguintes conjuntos visando encontrar dita característica.

$$\Gamma = \{J \subset \{1, \dots, n\} \mid \dim \mathcal{N}(A_J) = 1\},$$

$$W = \{w \in \mathbb{R}^n \mid J \in \Gamma, w_J \in \mathcal{N}(A_J), w_{J^c} = 0, \|w\| = 1, e^T w \geq 0\}.$$

e

$$\hat{W} = \{w \in W \mid e^T w > 0\}.$$

E para Z tal que $\mathcal{R}(Z^T) = \mathcal{N}(A)$ definimos

$$\Gamma' = \{J \subset \{1, \dots, n\} \mid \dim \mathcal{N}(Z_J) = 1\},$$

$$W' = \{v \in \mathbb{R}^n \mid J \in \Gamma', v_J \in \mathcal{N}(Z_J), v_{J^c} = 0, \|v\| = 1, e^T v \geq 0\}.$$

e

$$\hat{W}' = \{v \in W' \mid e^T v > 0\}.$$

Os vetores de W são os vetores de $\mathcal{N}(A)$ paralelos às arestas do conjunto definido por $Ax = b$ e $x \geq 0$. Similarmente, os elementos de W' pertencem a $\mathcal{R}(A^T)$ e são paralelos às arestas do conjunto definido por $s \in c + \mathcal{R}(A^T)$, e $s \geq 0$.

Para um vetor $t \in \mathbb{R}^n$, definimos $t^+ = \max\{t_i \mid i = 1, \dots, n\}$. Definimos então os números:

$$\xi_P = \min\left\{\frac{e^T w}{w^+} \mid w \in \hat{W}\right\},$$

$$\xi_D = \min\left\{\frac{e^T v}{v^+} \mid v \in \hat{W}'\right\}$$

e

$$\xi = \min\{\xi_P, \xi_D\}.$$

O próximo teorema caracteriza o número de condição dual σ_D :

Teorema 2.8.1 *Dados σ_D e ξ_P definidos antes, verifica-se*

$$\sigma_D \geq \xi_P.$$

Demonstração. Pela definição de σ_D dada na seção 2.3, o número σ_D é caracterizado como a menor das soluções, para $j \in N$, e utilizando e_j , o j -ésimo vetor canônico de $\mathbb{R}^{|N|}$ de

$$\begin{aligned} & \text{maximizar} && e_j^T s_N \\ (P_j) \quad & \text{sujeito a} && A_B^T y = e_B \\ & && A_N^T y + s_N = e_N \\ & && s_N \geq 0. \end{aligned}$$

Fixando $j \in N$, o problema dual associado a (P_j) é

$$\begin{aligned} & \text{minimizar} && e^T w \\ (D_j) \quad & \text{sujeito a} && Aw = 0 \\ & && w_N - e_j \geq 0. \end{aligned}$$

As condições de otimalidade são as condições de viabilidade primal e dual, junto com a complementaridade

$$(w_N - e_j)s_N = 0.$$

O problema (P_j) tem valor ótimo positivo s_j^* pois as componentes em N representam as variáveis que podem ser positivas numa solução ótima dual. Segue-se pela dualidade que o problema (D_j) também tem valor ótimo positivo, quer dizer, $e^T w > 0$

para qualquer w dual viável. Queremos provar que existe um vetor dual ótimo que também pertence a \hat{W} . De fato, tomamos uma solução ótima s^* do problema (P_j) tal que o número de componentes nulas seja máximo (um vértice). Definimos a seguinte partição de N : $N_1 = \{i \in N \mid s_i^* = 0\}$, $N_2 = N \setminus N_1$ portanto $N = N_1 \cup N_2$. Pela construção temos que $s_{N_2}^* > 0$. Dado que $s_j^* > 0$ concluímos que $j \in N_2$. Estudamos uma solução ótima w' de (D_j) associada a s^* . Pela complementaridade temos que $w'_j = 1$ e $w'_i = 0$ para todo $i \in N_2 \setminus \{j\}$; e pela viabilidade dual temos que $w'_{B \cup N_1}$ satisfaz

$$\begin{aligned} A_B w_B + A_{N_1} w_{N_2} + A_j &= 0 \\ w_{N_1} &\geq 0 \end{aligned} \tag{2.23}$$

Definimos $J^1 = \{i \in B \cup N_1 \mid w_i \neq 0\}$. Se as colunas da matriz A_{J^1} são linearmente independentes, então $w'/\|w'\|$ proporciona um vetor de \hat{W} pois tem norma um, $\bar{J}^1 := J^1 \cup \{i\} \in \Gamma$, $w'_{\bar{J}^1} \in \mathcal{N}(A_{\bar{J}^1})$, $w'_{(\bar{J}^1)^c} = 0$ e $\frac{e^T w'}{\|w'\|} > 0$. Se as colunas de A_{J^1} não são linearmente independentes então podemos construir, partindo de w' , uma solução de (2.23) eliminando colunas de A_{J^1} linearmente dependentes das outras, usando o seguinte procedimento de purificação: Escolhemos $i \in J^1$ tal que A_i seja combinação linear das outras colunas de A_{J^1} , quer dizer $A_i = A_{J^1 \setminus \{i\}} r$ para algum $r \in \mathbb{R}^{|J^1|-1}$. Definimos a direção $h = P_{A_{J^1}} e_i$ se $w'_i > 0$ ou $h = -P_{A_{J^1}} e_i$ se $w'_i < 0$. Avançamos na direção h desde w' até que a primeira componentes de $w'_{J^1} + \lambda h$ se anule, quer dizer escolhemos λ_{max} de maneira que

$$\lambda_{max} = \min_{l \in J^1} \left\{ -\frac{w'_l}{h_l} \mid \frac{w'_l}{h_l} < 0 \right\}.$$

Assim, pelo menos uma das componentes de $w_{J^1}^2 := w'_{J^1} + \lambda_{max} h$ é nula, e o sinal das componentes não nulas de w^2 coincide com o sinal da componente correspondente em w' , e além disso $w_{J^1}^2$ também satisfaz (2.23). Definimos como J^2 as componentes de J^1 que não são nulas em $w_{J^1}^2$. Repetimos o procedimento acima até achar J^l subconjunto de $B \cup N_1$ tal que as colunas de A_{J^l} sejam linearmente independentes, e gerem A_j . Desta maneira obtemos um vetor $w^l \in \hat{W}$. Temos demonstrado que para cada variável $j \in N$ existe um vetor $w^{(j)} \in \hat{W}$ tal que $\frac{e^T w^{(j)}}{w_j^{(j)}} = s_j^*$ que é o maior

valor que a variável s_j atinge na face ótima. Estudemos agora a relação

$$s_j^* = \frac{e^T w^{(j)}}{w_j^{(j)}} \geq \frac{e^T w^{(j)}}{w^{(j)+}}$$

e portanto, tomando os mínimos obtemos a relação $\sigma_D \geq \xi_P$, completando a prova.

■

O resultado dual, que apresentaremos sem demonstrar, tem uma prova análoga à anterior pela simetria da apresentação.

Teorema 2.8.2 *Dados σ_P e ξ_D definidos antes, verifica-se*

$$\sigma_P \geq \xi_D.$$

De maneira imediata obtemos o resultado principal desta seção:

Teorema 2.8.3 *Dados σ e ξ definidos antes, verifica-se*

$$\sigma \geq \xi.$$

Este resultado, junto com o teorema 2.3.3 permite estabelecer um limitante para o número de iterações de um problema de programação linear em função de ξ cuja definição não depende do conhecimento da partição ótima:

Corolário 2.8.4 *Dada uma seqüência (x^k, s^k, μ^k) gerada por algum algoritmo de trajetória central, satisfazendo (2.3); e a regra de parada*

$$\mu^k \leq \frac{\xi^2}{n^2},$$

então $k = O(\sqrt{n}(|\log \xi| + \log n))$.

Demonstração. Imediata usando o teorema 2.8.3 e o teorema 2.3.3. ■

Em nosso conhecimento, esta caracterização é o primeiro número de condição baseado na “menor variável grande” para o problema de programação linear em formato padrão. O número ξ depende dos dados A, b, c, x^o cujas entradas são supostas reais. O número permite estabelecer um modelo real para estudar a complexidade de algoritmos de pontos interiores, que é melhor do que σ , pois poderia ser usado algoritmos, ação impossível com σ .

Capítulo 3

Propriedades limites da direção afim-escala

3.1 Introdução

Neste capítulo estudamos a direção afim-escala primal. Começamos estabelecendo o problema de programação linear, e definindo os conjuntos viável e ótimo. Mostramos o algoritmo afim-escala primal de passos longos. Depois estudamos as propriedades limites das direções afim-escala, mostrando os teoremas de Megiddo e Shub, e de Tseng e Luo. Na última seção calculamos, sob uma hipótese que será explicada depois, o ínfimo dos cossenos dos ângulos entre as direções afim-escala e o vetor de custos.

3.2 O problema

Consideremos o problema de programação linear no formato padrão e suponhamos que $c \notin \mathcal{R}(A^T)$:

$$\begin{aligned}
& \text{minimizar} && c^T x \\
(P) & \text{sujeito a} && Ax = b \\
& && x \geq 0, \\
& \text{maximizar} && b^T y \\
(D) & \text{sujeito a} && A^T y + s = c \\
& && s \geq 0.
\end{aligned}$$

onde $A \in \mathbb{R}^{m \times n}$, $m < n$ e b é de posto m . Consideremos o conjunto viável primal-dual

$$F = \{x \in \mathbb{R}^n \mid Ax = b, x \geq 0\},$$

e o conjunto de pontos interiores

$$F^\circ = \{x \in F \mid x > 0\}.$$

Suporemos que existe um ponto $x^0 \in F^\circ$ (condição de Slater), e que o problema é limitado, quer dizer existem soluções ótimas. No capítulo 2 nós supomos a condição de Slater, tanto para o problema primal quanto para o dual, e transformamos o problema de programação linear a um formato normalizado, para incorporar informação do ponto inicial à instância. Neste capítulo não fazemos dita transformação, nem precisamos a condição de Slater, em ambos problemas. Só nos referimos ao problema primal, e a existência de soluções ótimas.

3.3 O algoritmo afim-escala

O algoritmo afim-escala foi o primeiro algoritmo de pontos interiores para programação linear, proposto por Dikin em 1967 [8], e conhecido como um dos mais simples e eficientes algoritmos de pontos interiores. Foi redescoberto por vários autores, como Barnes [3], Cavalier e Soyster [6], Karmarkar e Ramakrishnan [21], Kortanek e Shi [24] e Vanderbei, Meketon e Freedman [44], logo depois que Karmarkar propôs seu famoso algoritmo projetivo em 1984 [20]. Cada passo do algoritmo é:

- i) Construir um elipsóide, chamado “elipsóide de Dikin”, e
- ii) Tomar um passo na direção que minimiza a função objetivo no elipsóide para obter o próximo ponto.

Este procedimento simples tornou-se um recurso muito eficiente para resolver problemas de programação linear de grande porte, em poucas dezenas de iterações. A eficiência prática do algoritmo em problemas grandes foi o primeiro sinal de que algum algoritmo de pontos interiores pudesse concorrer com a eficiência do algoritmo Simplex. Devido a sua importância teórica e prática tem sido assunto de vários artigos, revelando gradativamente as interessantes estruturas matemáticas do algoritmo. A polinomialidade do algoritmo é uma pergunta aberta.

Estabeleceremos o algoritmo afim-escala: Consideremos o problema de programação linear (P) .

Definição 3.3.1 *Dado um ponto $\hat{x} \in F^\circ$ definimos a direção afim-escala $h(\hat{x})$ a partir de $\hat{x} \in \mathbb{R}_{++}^n$ como a única solução do problema*

$$\begin{aligned} &\text{minimizar } c^T h \\ &\text{sujeito a } Ah = 0 \\ &\|\hat{X}^{-1}h\| = 1. \end{aligned}$$

Estudaremos na próxima seção interessantes caracterizações desta direção. A segunda restrição do problema anterior sugere uma mudança de coordenadas do problema (P) , usando a matriz \hat{X} . Depois de mudar as coordenadas, pela equação $\bar{h} = \hat{X}^{-1}h$, o problema (3.3.1) é escrito como

$$\begin{aligned} &\text{minimizar } \bar{c}^T \bar{h} \\ &\text{sujeito a } \bar{A}\bar{h} = 0 \\ &\|\bar{h}\| = 1, \end{aligned} \tag{3.1}$$

onde $\bar{c} = \hat{X}c$, e $\bar{A} = A\hat{X}$. A solução deste problema é a direção de máxima descida normalizada da função objetivo $\bar{c}^T \bar{h}$, que é a direção oposta à projeção de \bar{c} em $\mathcal{N}(\bar{A})$ (de norma um). A segunda restrição de (3.1) define uma bola centrada na

origem e de raio um, enquanto que a restrição correspondente em (3.3.1) define um elipsóide centrado também na origem. O algoritmo que descreveremos é chamado afim-escala, pela mudança de escala feita no parágrafo anterior e pelo fato que cada ponto definido começando com um vetor x tal que $Ax = b$ e somando a direção afim-escala $h(x)$, o ponto $x^+ = x + h(x)$ também verifica $Ax^+ = b$. O elipsóide de Dikin é definido para \hat{x} viável como

$$E(\hat{x}) = \{x \in \mathbb{R}^n \mid x = \hat{x} + h, Ah = 0, \|\hat{X}^{-1}h\| = 1\}.$$

O conjunto correspondente ao elipsóide $E(\hat{x})$ no espaço escalado é a interseção do conjunto afim $\{x \in \mathbb{R}^n \mid \bar{A}x = b\}$ com a bola com centro em e e de raio um. Mostraremos o algoritmo a seguir:

Algoritmo 3.3.2 (*afim-escala*)

Dado $x^0 \in F^0$ e $\nu \in (0, 1)$

$k=0$

Repita:

Direção $h := h(x^k)$ como definida em (3.3.1)

Busca linear: $\lambda_{max} := \max\{\lambda > 0 \mid x^k + \lambda h \geq 0\}$

$x^{k+1} := x^k + \nu \lambda_{max} h$

$k := k + 1$

até convergência.

Dado que existem soluções ótimas, e h é uma direção de descida λ_{max} é finito, pois no caso contrário teríamos uma direção viável onde o custo decresceria indefinidamente. O algoritmo primeiro calcula o maior comprimento do passo possível no primeiro ortante, e depois toma um passo um pouco menor para garantir que o próximo ponto x^k seja estritamente positivo. O cálculo de λ_{max} é feito pelo teste da razão:

$$\lambda_{max} := \min\left\{-\frac{x_i^k}{h_i} \mid h_i < 0, i = 1, \dots, n\right\}.$$

O algoritmo afim-escala gera pontos só primais viáveis, quer dizer gera uma seqüência (x^k) tal que $Ax^k = b$ e $x^k > 0$. As condições de otimalidade para o problema (P) são dadas por estas relações de viabilidade primal, a viabilidade dual $s = c - A^T y \geq 0$ e a condição de complementaridade $x_i s_i = 0$ para todo $i = 1, \dots, n$. Para testar a otimalidade dos pontos limites da seqüência gerada, é importante ter uma seqüência de estimadores das variáveis duais (s^k, y^k) associada com a seqüência primal. Desejamos então uma seqüência que aproxime de alguma maneira a viabilidade dual e a complementaridade. Uma construção razoável de variáveis duais partindo da seqüência primal é a solução do seguinte problema de quadrados mínimos:

$$\begin{aligned} & \text{minimizar}_{y,s} && \frac{1}{2} \|X^k s\|^2 \\ & \text{sujeito a} && s = c - A^T y. \end{aligned}$$

A única solução do problema é

$$s(x^k) = (X^k)^{-1} P_{AX^k} X^k c.$$

Demostraremos na próxima seção que para qualquer $d \in \mathbb{R}_{++}^n$ existe $\alpha > 0$ tal que a direção afim-escala com respeito de d $h(d)$ pode ser escrita como $h(d) = -\alpha DP_{AD} Dc$. Se aceitamos esta relação podemos ver a conexão entre a direção afim-escala calculada em x^k e os estimadores duais acima:

$$(X^k)^{-1} h(x^k) = -P_{AX^k} X^k c = -X^k s(x^k).$$

Concluimos assim que o algoritmo afim-escala proporciona naturalmente estimadores das variáveis duais.

Para provar a convergência do algoritmo afim-escala é necessário provar que o limite da seqüência primal existe; e para provar que esse limite é um ponto ótimo é preciso mostrar que o limite de estimadores de multiplicadores existe e fornece uma solução dual viável complementar. Pela construção, a seqüência dual pertence ao conjunto afim $c + \mathcal{R}(A^T)$, assim que o maior trabalho é demonstrar que o limite s é não negativo e verifica a complementaridade com o limite da seqüência primal.

A versão original de Dikin usa a direção $h(x^k)$ de norma um (ou norma menor do que um em algumas referências) e não faz busca linear. Chamaremos a esta

versão método afim-escala de passos curtos. A versão apresentada no algoritmo 3.3.2 permite o tamanho do passo maior, constituindo um método de passos longos. Resumiremos os resultados de convergência do algoritmo afim-escala: Em 1974, Dikin [9] demonstrou a convergência global da seqüência primal e dos estimadores de dualidade, para a versão de passos curtos com hipótese de não degeneração primal. Desafortunadamente, o trabalho de Dikin não foi conhecido fora da Rússia até 1988 [10]. Logo depois de que Karmarkar apresentou seu algoritmo projetivo em 1984 [20], Barnes [3] e Vanderbei, Meketon e Freedman [44] acharam provas de convergência com hipótese de não degeneração primal e dual para as versões de passos curtos e passos longos respectivamente. Depois aparecem vários trabalhos tentando eliminar a hipótese de não degeneração como Tsuchiya 1989 [39], Tseng e Luo, 1989 [37], Tsuchiya, 1990 [38] para o algoritmo de passos curtos. Finalmente em 1992 Dikin [11], e Tsuchiya e Maramatsu [40] acharam provas para o caso de passos longos, sem hipótese de não degeneração, com $0 < \nu \leq \frac{1}{2}$ e $0 < \nu \leq \frac{2}{3}$ respectivamente. Existem outras referências com provas simplificadas, como a de Monteiro, Tsuchiya e Wang [27]; outras com propriedades interessantes, como por exemplo uma instância onde o algoritmo de passos longos com $\nu = 0.999$ converge para um ponto não ótimo (ver Mascarenhas [25]); versões do algoritmo para problemas convexos com restrições lineares como Gonzaga e Carlos [16] e Monteiro e Wang [28]; e vários artigos mostrando o aspecto computacional do algoritmo. O recente artigo de Tsuchiya [41], é um excelente survey do algoritmo afim-escala.

No artigo de Tseng e Luo foi mostrada uma propriedade interessante relativa ao ângulo entre a direção afim-escala primal e o vetor de custos que define o problema de programação linear. Eles provaram que o ínfimo dos cossenos dos ângulos entre o vetor de custos e qualquer direção afim-escala é positivo. Dita propriedade é fundamental na prova de convergência global do algoritmo afim-escala. A demonstração de Tseng e Luo foi feita por contradição, e a existência de dito limitante foi garantida mas não foi estimado. O resultado principal deste capítulo é o cálculo desse ínfimo.

3.4 A direção afim-escala primal

Consideremos para um vetor $d \in \mathbb{R}_{++}^n$ a única solução $u(d)$ do problema

$$\begin{aligned} &\text{minimizar} && -c^T u + \frac{1}{2} \|D^{-1}u\|^2 \\ &\text{sujeito a} && Au = 0. \end{aligned} \tag{3.2}$$

Usando as condições de otimalidade, concluímos que $u(d)$ satisfaz, para $y \in \mathbb{R}^m$

$$\begin{aligned} Au &= 0 \\ u &= D^2(c - A^T y). \end{aligned} \tag{3.3}$$

Agora lembremos a definição da direção afim-escala, dada na seção anterior:

Para $d \in \mathbb{R}_{++}^n$, a direção afim-escala primal $h(d)$ é a única solução do problema

$$\begin{aligned} &\text{minimizar} && c^T h \\ &\text{sujeito a} && Ah = 0 \\ &&& \|D^{-1}h\| = 1, \end{aligned}$$

quer dizer, $h(d)$ satisfaz as seguintes condições de otimalidade, para $w \in \mathbb{R}^m$, e $\beta \in \mathbb{R}$:

$$\begin{aligned} h &= \frac{1}{\beta} D^2(c + A^T w) \\ Ah &= 0 \\ \|D^{-1}h\| &= 1. \end{aligned} \tag{3.4}$$

Relacionamos as direções $u(d)$ e $h(d)$ no seguinte fato:

Fato 3.4.1 *Dado $d \in \mathbb{R}_{++}^n$. Existe $\alpha > 0$ tal que*

$$h(d) = -\alpha u(d).$$

Demonstração. Pelas condições (3.3) obtemos, multiplicando a segunda equação por A , e usando a primeira que

$$AD^2c - AD^2A^T y = 0.$$

Como $A \in \mathbb{R}^m$ é de posto máximo m , e a matriz AD^2A^T é inversível, temos

$$y = (AD^2A^T)^{-1} AD^2c,$$

e assim

$$\begin{aligned} u(d) &= D^2(I - A^T(AD^2A^T)^{-1}AD^2)c \\ &= D(I - DA^T(AD^2A^T)^{-1}AD)Dc \\ &= DP_{AD}Dc. \end{aligned}$$

Por outra parte, usando as equações (3.4) obtemos, multiplicando a primeira equação por A e usando a segunda que

$$\frac{1}{\beta}AD^2(c - A^T w) = 0,$$

assim

$$w = (AD^2A^T)^{-1}AD^2c,$$

e portanto

$$h = -\frac{1}{\beta}DP_{AD}Dc.$$

Multiplicando esta equação por D^{-1} e tomando normas obtemos por (3.4)

$$\beta = \|P_{AD}Dc\| > 0,$$

em consequência

$$h = -\frac{1}{\|P_{AD}Dc\|}DP_{AD}Dc = -\alpha u(d),$$

com $\alpha = 1/\|P_{AD}Dc\|$. ■

O resultado anterior vale também sem a hipótese de posto máximo para a matriz A , dado que a projeção não depende da representação matricial, mas sim dos espaços nulo e imagem definidos pela matriz.

Para facilitar a notação, chamaremos à direção dada por $u(d)$ como direção afim-escala primal. De imediato podemos ver algumas caracterizações da direção afim-escala $u(d)$: Das condições de otimalidade (3.3) temos que

$$u(d) = DP_{AD}Dc. \tag{3.5}$$

Sabemos que para qualquer vetor $s \in c + \mathcal{R}(A^T)$ tem-se que $s^T u = c^T u$ para todo $u \in \mathcal{N}(A)$, então podemos substituir c por qualquer $s \in c + \mathcal{R}(A^T)$ no problema

(3.2), obtendo a mesma solução $u(d)$. Isto é, para qualquer $s \in c + \mathcal{R}(A^T)$, tem-se que $u(d)$ é solução do problema

$$\begin{aligned} & \text{minimizar} && -s^T u + \frac{1}{2} \|D^{-1}u\|^2 \\ & \text{sujeito a} && Au = 0. \end{aligned}$$

Como consequência temos que para qualquer $s \in c + \mathcal{R}(A^T)$, $u(d)$ satisfaz, para $y \in \mathbb{R}^m$

$$\begin{aligned} Au &= 0 \\ u &= D^2(s - A^T y). \end{aligned} \tag{3.6}$$

ou equivalentemente

$$u(d) = DP_{AD}Ds. \tag{3.7}$$

Existem algumas escolhas interessantes de s na caracterização (3.6) como por exemplo $s = P_{AC}$, e uma que definiremos em seguida: consideremos uma partição $\{B, N\}$ dos índices $\{1, \dots, n\}$.

Fato 3.4.2 *Se $c_B \in \mathcal{R}(A_B^T)$ então existe um vetor $s \in c + \mathcal{R}(A^T)$ tal que $s_B = 0$ e $s_N \neq 0$.*

Demonstração. Tomamos \bar{y} tal que $c_B = A_B^T \bar{y}$ e fazemos $s = c - A^T \bar{y}$. Pela construção $s_B = 0$, e pelo fato que $c \notin \mathcal{R}(A^T)$, necessariamente $s_N \neq 0$. ■

Proposição 3.4.3 *Dada uma partição dos índices B, N tal que $c_B \in \mathcal{R}(A_B^T)$ então existe um vetor $s \in c + \mathcal{R}(A^T)$ tal que a direção $u(d)$ ($d \in \mathbb{R}_{++}^n$) pode ser caracterizada por*

$$A_B u_B + A_N u_N = 0 \tag{3.8}$$

$$u_B = -D_B^2 A_B^T y \tag{3.9}$$

$$u_N = D_N^2 (s_N - A_N^T y). \tag{3.10}$$

Demonstração. Trivial usando 3.6 e o fato 3.4.2. ■

Outra propriedade interessante da direção afim-escala, para $d \in \mathbb{R}_{++}^n$ fixo é:

$$u(\alpha d) = \alpha^2 u(d), \quad \forall \alpha > 0 \tag{3.11}$$

Comentário em notação: Quando nos referimos à direção afim-escala do problema com instância A, c usamos a notação $u(d)$; para as restrições desse vetor indexadas por um conjunto de índices J , a notação é $u_J(d)$, e se nos referimos à direção afim-escala da instância A_J, c_J , usamos a notação $u_J(d_J)$. Observemos que $u_J(d)$ não é o mesmo que $u_J(d_J)$. A primeira é a restrição indexada por J da solução $u(d)$ definida por 3.2, que satisfaz por exemplo as condições (3.3); enquanto $u_J(d_J)$ satisfaz as equações análogas para a instância A_J, c_J :

$$\begin{aligned} A_J u_J &= 0 \\ u_J &= D_J^2(c_J - A_J^T y). \end{aligned} \tag{3.12}$$

Queremos estudar a aplicação $u(d)$ quando $d \rightarrow \bar{d} \geq 0$ ($\bar{d} \neq 0$). Dado \bar{d} , definimos os conjuntos $B(\bar{d}) = \{i \in \{1, \dots, n\} \mid \bar{d}_i > 0\}$ e $N(\bar{d}) = \{1, \dots, n\} \setminus B(\bar{d})$. As variáveis com índices em $B(\bar{d})$ são chamadas variáveis grandes, e as outras variáveis pequenas. É difícil descrever o comportamento das variáveis pequenas $u_N(\bar{d})$. Megiddo e Shub [26] e Gonzaga e Tapia [17] estudaram o comportamento das variáveis grandes. Descreveremos esses resultados: Tomamos $B = B(\bar{d}), N = N(\bar{d})$, e escrevemos o problema (3.2) como

$$\begin{aligned} \text{minimizar} \quad & -c_B^T u_B - c_N^T u_N + \frac{1}{2} \|D_B^{-1} u_B\|^2 + \frac{1}{2} \|D_N^{-1} u_N\|^2 \\ \text{sujeito a} \quad & A_B u_B = -A_N u_N. \end{aligned} \tag{3.13}$$

Suporemos agora que $u_N(d)$ é conhecido. Então $u_B(d)$ resolve o problema

$$\begin{aligned} \text{minimizar} \quad & -c_B^T u_B + \frac{1}{2} \|D_B^{-1} u_B\|^2 \\ \text{sujeito a} \quad & A_B u_B = -A_N u_N(d). \end{aligned}$$

Dado que $d_N \rightarrow 0$ e $u_N(d) = D_N^2(c_N - A_N^T y)$, onde $(c_N - A_N^T y)$ é finito, então $u_N(\bar{d}) = 0$, e assim $u_B(\bar{d}) = \bar{D}_B P_{A_B \bar{D}_B} \bar{D}_B c_B$. Estudaremos a aplicação ponto conjunto θ definida para $d > 0$ por

$$\mathbb{R}_{++}^n \ni d \mapsto \theta(d) = \{u_B \in \mathbb{R}^{|B|} \mid A_B u_B = -A_N u_N(d)\}, \tag{3.14}$$

próximo ao ponto $\bar{d} \geq 0$ ($\bar{d} \neq 0$). Notemos que neste ponto $\theta(\bar{d}) = \mathcal{N}(A_B)$. Lembremos agora a definição de uma aplicação ponto-conjunto contínua.

Definição 3.4.4 Seja φ uma aplicação ponto conjunto $\mathbb{R}^n \ni x \mapsto \varphi(x) \subset \mathbb{R}^m$. A aplicação φ é contínua em \bar{x} se para qualquer seqüência (x^k) que converge para \bar{x} , verifica-se:

- i) Para qualquer seqüência convergente (z^k) , onde $z^k \in \varphi(x^k)$, necessariamente $\bar{y} = \lim z^k \in \varphi(\bar{x})$. (Semi-continuidade superior)
- ii) Para qualquer ponto $z' \in \varphi(\bar{x})$ existe uma seqüência (z^k) que converge para z' onde $z^k \in \varphi(x^k)$. (Semi-continuidade inferior)

Lema 3.4.5 A aplicação θ é contínua em $\bar{d} \in \mathbb{R}_+^n$.

Demonstração.

Consideremos uma partição $\{B, N\}$ de $\{1, \dots, n\}$ e um ponto $\bar{d} \in \mathbb{R}_+^n$ tal que $\bar{d}_B > 0$ e $\bar{d}_N = 0$.

- i) Semi-continuidade superior: Consideremos uma seqüência (d^k) em \mathbb{R}_{++}^n tal que $d^k \rightarrow \bar{d}$ e uma seqüência (u_B^k) tal que $A_B u_B^k = -A_N u_N(d^k)$ e $u_B^k \rightarrow \bar{u}_B = u_B(\bar{d})$. Devemos provar que $A_B \bar{u}_B = 0$. Basta observar que seqüência $u_N(d^k)$ vai para zero, pois $u(d^k) = D^k P_{AD^k} D^k c$, $P_{AD^k} D^k c$ é limitado, e $d_N^k \rightarrow 0$.
- ii) Semi-continuidade inferior: Consideremos um ponto arbitrário $\bar{u} \in \theta(\bar{d}) = \mathcal{N}(A_B)$. Dada uma seqüência arbitrária $(d^k) \subset \mathbb{R}_{++}^n$ tal que $d^k \rightarrow \bar{d}$, devemos construir (u_B^k) tal que $A_B u_B^k = -A_N u_N(d^k)$ e $u_B^k \rightarrow \bar{u}_B$. Consideremos (d^k) em \mathbb{R}_{++}^n tal que $d^k \rightarrow \bar{d}$. Definimos $u_N^k = u_N(d^k)$. Para cada k seja \tilde{u}_B^k a solução de mínima norma de $A_B u_B = -A_N u_N^k$. Se A_B^+ denota a matriz pseudo inversa de A_B , então podemos escrever $\tilde{u}_B^k = -A_B^+ A_N u_N^k$. Segue-se $\tilde{u}_B^k \rightarrow 0$, dado que $u_N^k \rightarrow 0$. Construimos $u_B^k = u_B(\bar{d}) + \tilde{u}_B^k$, então

$$A_B u_B^k = A_B u_B(\bar{d}) + A_B \tilde{u}_B^k = -A_N u_N^k.$$

Dado que $u_B(\bar{d}) \in \mathcal{N}(A_B)$, temos $u_B^k \in \theta(d^k)$; e como $\tilde{u}_B^k \rightarrow 0$ temos $u_B^k \rightarrow u_B(\bar{d})$.

■

Lema 3.4.6 Consideremos $u(d)$ como definida em (3.2), $d \rightarrow \bar{d}$ tal que $\bar{d}_B > 0$ e $\bar{d}_N = 0$. Então

$$i) u_B(d) \rightarrow u_B(\bar{d}_B) = \bar{D}_B P_{A_B \bar{D}_B} \bar{D}_B c_B.$$

$$ii) (P_{AD} Dc)_N \rightarrow 0.$$

Demonstração.

i) A aplicação $d \mapsto u_B(d) = \operatorname{argmin}\{-c_B^T u_B + \frac{1}{2} \|D_B^{-1} u_B\|^2 \mid u_B \in \theta(d)\}$ está bem definida já que a função $-c_B^T u_B + \frac{1}{2} \|D_B^{-1} u_B\|^2$ é estritamente convexa. É contínua em \bar{d} como consequência das projeções (ver por exemplo Hogan [19]). Dado que $d \rightarrow \bar{d}$ temos $u_B(\bar{d}) = \operatorname{argmin}\{-c_B^T u_B + \frac{1}{2} \|\bar{D}_B^{-1} u_B\|^2 \mid u_B \in \mathcal{N}(A_B)\}$ ou $u_B(\bar{d}_B) = \bar{D}_B P_{A_B \bar{D}_B} \bar{D}_B c_B$, estabelecendo a parte (i).

ii) Sabemos que $\bar{D}_N c_N = 0$. Denotemos por $y(d) = P_{AD} Dc$. Suponhamos para alguma seqüência (d^k) em \mathbb{R}_{++}^n com $d^k \rightarrow \bar{d}$ que temos $y(d^k)_N \rightarrow \bar{y}_N \neq 0$. Definimos $\epsilon = \|\bar{y}_N\|^2 > 0$. Assim

$$\|y(d^k) - D^k c\|^2 = \|y_B(d^k) - D_B^k c_B\|^2 + \|y_N(d^k) - D_N^k c_N\|^2.$$

Pela parte (i), $u_B(d^k) \rightarrow \bar{u}_B = \bar{D}_B P_{A_B \bar{D}_B} \bar{D}_B c_B$, assim $y_B(d^k) \rightarrow \bar{y}_B = P_{A_B \bar{D}_B} \bar{D}_B c_B$. Para k suficientemente grande,

$$\|y(d^k) - D^k c\|^2 > \|\bar{y}_B - \bar{D}_B c_B\|^2 - \frac{\epsilon}{2}. \quad (3.15)$$

Agora construímos a seqüência:

$$\tilde{y}_B^k = (D_B^k)^{-1} \bar{D}_B \bar{y}_B, \quad \tilde{y}_N^k = 0,$$

segue-se que $\tilde{y}_B^k \rightarrow \bar{y}_B$, e $\tilde{y}^k \in \mathcal{N}(AD^k)$, pois

$$AD^k \tilde{y}^k = A_B D_B^k (D_B^k)^{-1} \bar{D}_B \bar{y}_B = 0.$$

Comparando com (3.15) temos para k suficientemente grande $\|\tilde{y}^k - D^k c\| < \|y(d^k) - D^k c\|$ and $\tilde{y}^k \in \mathcal{N}(AD^k)$, contradizendo a definição de $y(d^k) = P_{AD^k} D^k c$ e completando a prova. ■

3.5 O ângulo limite

3.5.1 Preliminares

Queremos estudar limitantes inferiores para o cosseno do ângulo que fazem a direção afim-escala $u(d)$ e o vetor de custos c para todo $d \in \mathbb{R}_{++}^n$. Tseng e Luo [37] mostraram que o ínfimo é positivo, mas não deram uma forma de calculá-lo. Nosso objetivo nesta seção é estimar dito ínfimo, que faremos com uma hipótese que explicaremos depois. O lema de Tseng e Luo foi também descrito e simplificado por Monteiro e Tsuchiya [27], e permite uma prova de convergência para o algoritmo afim-escala primal sem hipótese de degeneração. Nossa prova é uma simplificação da demonstração feita por Monteiro e Tsuchiya:

Lema 3.5.1 (Tseng e Luo) *Consideremos o problema (P) com $c \notin \mathcal{R}(A^T)$. Existe uma constante $L = L(A, c) > 1$ tal que $\forall d > 0$,*

$$\frac{c^T u(d)}{\|u(d)\|} \geq \frac{1}{L}. \quad (3.16)$$

Antes de provar o lema lembremos um resultado relacionado ao tamanho das soluções de um sistema linear, conhecido como lema de Hoffman:

Lema 3.5.2 *Consideremos uma matriz $E \in \mathbb{R}^{m \times n}$ fixa. Para todo vetor $g \in \mathcal{R}(E)$, existe uma solução \hat{x} do sistema de equações lineares $Ex = g$ tal que $\hat{x} = O(\|g\|)$.*

Demonstração. Dado $g \in \mathcal{R}(E)$, consideremos a solução de mínima norma de $Ex = g$, denotada por $\hat{x} = E^+g$, onde E^+ é a matriz pseudo-inversa de E . Tomando normas obtemos

$$\|\hat{x}\| = \|E^+g\| \leq \|E^+\| \|g\|,$$

completando a prova. ■

Usamos o resultado para provar o lema 3.5.1:

Demonstração. (Lema 3.5.1) Suponhamos por contradição que para alguma seqüência (d^k) em \mathbb{R}_{++}^n , $\frac{c^T u(d^k)}{\|u(d^k)\|} \rightarrow 0$. Pela propriedade (3.11) podemos supor sem

perda de generalidade que para todo k , $c^T u(d^k) = 1$, (basta fazer $\alpha_k = \frac{1}{c^T u(d^k)}$ e $d^k := \alpha_k d^k$). Agora, $\frac{1}{\|u(d^k)\|} \rightarrow 0$ e portanto $\|u(d^k)\| \rightarrow \infty$.

Tomando subsequências se necessário podemos construir uma partição J, J^c dos índices tal que

$$\begin{aligned} |u_i(d^k)| &\rightarrow \infty && \text{para } i \in J \\ u_i(d^k) &\text{ limitado} && \text{para } i \in J^c. \end{aligned}$$

Usamos o lema de Hoffman no sistema:

$$\begin{aligned} c^T u &= 1 \\ Au &= 0 \\ u_i &= u_i(d^k), \quad \forall i \in J^c. \end{aligned}$$

Dado que o lado direito do sistema é limitado, pelo lema de Hoffman, para qualquer k existe uma solução \tilde{u}^k tal que $\|\tilde{u}^k\|$ é limitado. Segue-se então que para $i \in J$ e k suficientemente grande,

$$\|u_i(d^k)\| > \|\tilde{u}^k\| \geq |\tilde{u}_i^k|$$

então,

$$\begin{aligned} \|(D^k)^{-1} \tilde{u}^k\|^2 &= \|(D_J^k)^{-1} \tilde{u}_J^k\|^2 + \|(D_{J^c}^k)^{-1} \tilde{u}_{J^c}^k\|^2 \\ &< \|(D_J^k)^{-1} u_J(d^k)\|^2 + \|(D_{J^c}^k)^{-1} u_{J^c}(d^k)\|^2, \end{aligned}$$

porque $\tilde{u}_{J^c}^k = u_{J^c}(d^k)$ por construção. Portanto, para k suficientemente grande,

$$\|(D^k)^{-1} \tilde{u}\|^2 < \|(D^k)^{-1} u(d^k)\|^2$$

e como $c^T \tilde{u} = c^T u(d^k) = 1$,

$$c^T \tilde{u} + \frac{1}{2} \|(D^k)^{-1} \tilde{u}\|^2 < c^T u(d^k) + \frac{1}{2} \|(D^k)^{-1} u(d^k)\|^2,$$

contradizendo a definição de $u(d^k)$ e completando a prova. ■

O lema de Tseng e Luo diz que o coseno do ângulo entre $u(d)$ e c não pode aproximar-se de zero quando $u(d)$ é calculado para qualquer sequência no ortante positivo. A demonstração de Tseng e Luo foi feita por contradição. Isto é a existência do limitante é garantida mas seu valor não é calculado. É interessante calcular dito limitante pois poderia ser usado para medir o condicionamento do problema de

programação linear (P), pois quanto mais perto de $\frac{\pi}{2}$ estiver o ângulo entre $u(d)$ e c , mais lenta será a convergência dos métodos que usam a direção afim-escala como direção de busca. Nosso objetivo é calcular

$$\tau_{A,c} := \inf_{d>0} \frac{c^T u(d)}{\|u(d)\|}. \quad (3.17)$$

Afirmamos que existe uma seqüência $(d^k) \subset \mathbb{R}_{++}^n$ convergente para algum $\bar{d} \in \mathbb{R}_+^n$ tal que $\|d^k\| = 1$ e

$$\lim_{d^k \rightarrow \bar{d}} \frac{e^T u(d^k)}{\|u(d^k)\|} = \tau_{A,c}. \quad (3.18)$$

De fato, tomemos uma seqüência (d^k) tal que $\lim_{k \rightarrow \infty} \frac{c^T u(d^k)}{\|u(d^k)\|} = \tau_{A,c}$. Pela propriedade (3.11), tomando $\alpha_k = \|d^k\|^{-1}$ escolhemos a seqüência $d^k := \alpha_k d^k$ que é limitada, e tomando uma subsequência convergente obtemos a tese.

Denotamos por \mathcal{SE} ao conjunto cujos elementos são as seqüências em \mathbb{R}_{++}^n . Consideremos os conjuntos:

$$\bar{\mathcal{D}}_{A,c} = \{(d, (d^k)) \in \mathbb{R}_+^n \times \mathcal{SE} \mid \|d^k\| = 1, \lim_{d^k \rightarrow d} \frac{c^T u(d^k)}{\|u(d^k)\|} = \tau_{A,c}\}$$

e

$$\mathcal{D}_{A,c} = \{d \in \mathbb{R}^n \mid (d, (d^k)) \in \bar{\mathcal{D}}_{A,c}\}.$$

O seguinte lema, que é fundamental no nosso tratamento, afirma que quando existem direções afim-escala diferentes à trivial $c_p := P_A c$, então os vetores de $\mathcal{D}_{A,c}$ pertencem à fronteira de \mathbb{R}_{++}^n .

Lema 3.5.3 *Suponhamos que existe algum $\hat{d} \in \mathbb{R}_{++}^n$ tal que $u(\hat{d})$ não é colinear com c_p . Então para qualquer $\bar{d} \in \mathcal{D}_{A,c}$, a partição $B(\bar{d}), N(\bar{d})$ é própria.*

Demonstração. Tomamos $\bar{d} \in \mathcal{D}_{A,c}$. Como \bar{d} é o limite de uma seqüência de vetores de norma um, $B(\bar{d}) \neq \emptyset$. Falta provar que $N(\bar{d}) \neq \emptyset$. Suponhamos por contradição que $N(\bar{d}) = \emptyset$, quer dizer, $\bar{d} \in \mathbb{R}_{++}^n$. Desta forma \bar{d} é solução ótima do problema

$$\text{minimizar}_{d>0} \frac{c^T u(d)}{\|u(d)\|}.$$

Pela propriedade (3.11), para $\alpha > 0$ tem-se

$$\frac{c^T u(\alpha d)}{\|u(\alpha d)\|} = \frac{\alpha^2 c^T u(d)}{\alpha^2 \|u(d)\|} = \frac{c^T u(d)}{\|u(d)\|}.$$

Isto é a função objetivo é homogênea de grau zero. Para cada $d \in \mathbb{R}_{++}^n$, tomando $\alpha^2 = \frac{1}{c^T u(d)}$ e $\tilde{d} := \alpha d$ obtemos $c^T u(\tilde{d}) = 1$. Conclui-se que o valor ótimo do problema não muda se introduzimos a restrição $c^T u(d) = 1$. Portanto o problema acima é equivalente a

$$\begin{aligned} & \text{maximizar}_{d>0} \quad \|u(d)\| \\ & \text{sujeito a} \quad c^T u(d) = 1. \end{aligned}$$

Usando a caracterização 3.6 e escolhendo $s = c_p := P_{Ac}$, este problema pode ser escrito como

$$\begin{aligned} & \text{minimizar}_{d>0} \quad -\frac{1}{2} u^T u \\ & \text{sujeito a} \quad u = D^2 c_p - D^2 A^T w \end{aligned} \tag{3.19}$$

$$Au = 0 \tag{3.20}$$

$$c_p^T u = 1, \tag{3.21}$$

cujas condições de otimalidade são, para os multiplicadores $\lambda \in \mathbb{R}^n$, $\mu \in \mathbb{R}^m$, $\beta \in \mathbb{R}$, e $\Lambda = \text{diag}(\lambda)$

$$-u + \beta c_p + \lambda + A^T \mu = 0 \tag{3.22}$$

$$AD^2 \lambda = 0 \tag{3.23}$$

$$2D\Lambda(-c_p + A^T w) = 0. \tag{3.24}$$

Dado que a solução \bar{d} é positiva, então pelas equações (3.19) e (3.21) obtemos $\Lambda u = 0$, quer dizer

$$\lambda_i u_i = 0, \quad i = 1, \dots, n. \tag{3.25}$$

Multiplicamos (3.22) por u^T para obter

$$-u^T u + \beta c_p^T u + \lambda^T u + u^T A^T \mu = 0.$$

Como $\lambda^T u = 0$ por (3.25), $u^T A^T \mu = 0$ por (3.20) e $c_p^T u = 1$ por (3.21), esta equação reduz-se a

$$\beta = u^T u.$$

Multiplicando (3.22) por AD^2 , e usando (3.23) obtemos

$$AD^2 \mu = AD^2(u - \beta c_p),$$

assim

$$\mu = (AD^2 A^T)^{-1} AD^2(u - \beta c_p)$$

e portanto por (3.22) obtemos

$$\lambda = u - \beta c_p - A^T (AD^2 A^T)^{-1} AD^2(u - \beta c_p),$$

em consequência

$$D^2 \lambda = DP_{AD} D(u - \beta c_p). \quad (3.26)$$

Esta equação indica, por (3.5) que o vetor $\frac{D^2 \lambda}{\beta}$ é a única solução do problema

$$\begin{aligned} \text{minimizar} \quad & -\left(\frac{u}{\beta} - c_p\right)^T v + \frac{1}{2} \|D^{-1} v\|^2 \\ \text{sujeito a} \quad & Av = 0, \end{aligned} \quad (3.27)$$

mas por (3.25) sabemos que $\frac{D^2 \lambda}{\beta} u = 0$.

Por outro lado, multiplicando (3.26) por c_p^T obtemos

$$\begin{aligned} c_p^T D^2 \lambda &= c_p^T DP_{AD} D(u - \beta c_p) \\ &= u^T u - \beta u^T c_p = 0. \end{aligned}$$

A segunda igualdade é válida por (3.5), enquanto que a terceira é verdadeira dado que $\beta = u^T u$ e $c_p^T u = 1$. Dado que $\frac{D^2 \lambda}{\beta} u = 0$ e $\frac{D^2 \lambda}{\beta} c_p = 0$ concluímos que o problema (3.27) é equivalente a

$$\begin{aligned} \text{minimizar} \quad & \frac{1}{2} \|D^{-1} v\|^2 \\ \text{sujeito a} \quad & Av = 0 \\ & \left(\frac{u}{\beta} - c_p\right)^T v = 0, \end{aligned}$$

cujas soluções ótimas são $v = 0$. Portanto concluímos que $\frac{D^2\lambda}{\beta} = 0$ que significa $\lambda = 0$. Agora multiplicando a equação (3.22) por c_p obtemos

$$-c_p^T u + \beta c_p^T c_p = 0,$$

isto é como $c_p^T u = (c_p^T u)^2 = 1$ temos, dividindo a expressão acima por $\|u\|^2$ que $\frac{(c_p^T u)^2}{\|u\|^2} = \|c_p\|^2$, assim

$$\frac{c_p^T u}{\|u\|} = \|c_p\|.$$

Por hipótese existe uma direção afim-escala não colinear com c_p , que certamente verifica

$$\frac{c_p^T u(\hat{d})}{\|c_p\| \|u(\hat{d})\|} < 1,$$

portanto

$$\frac{c_p^T u(\hat{d})}{\|u(\hat{d})\|} < \|c_p\| = \frac{c_p^T u}{\|u\|}$$

contradizendo a definição de ínfimo. ■

3.5.2 As arestas do conjunto viável

Queremos calcular $\tau_{A,c}$ em função dos dados, quer dizer, em características da instância A, c . Para isto definiremos os seguintes conjuntos:

$$\Gamma = \{J \subset \{1, \dots, n\} \mid \dim \mathcal{N}(A_J) = 1\},$$

$$W = \{w \in \mathbb{R}^n \mid J \in \Gamma, w_J \in \mathcal{N}(A_J), w_{J^c} = 0, \|w\| = 1, c^T w \geq 0\}.$$

e

$$\hat{W} = \{w \in W \mid c^T w > 0\}.$$

Podemos interpretar os elementos de W como os vetores de $\mathcal{N}(A)$ que são paralelos com as arestas de algum polígono definido por $Ax = \tilde{b}$ e $x \geq 0$, com $b \in \mathcal{R}(A)$ (não necessariamente $\tilde{b} = b$). Na definição dos conjuntos Γ, W e \hat{W} é só necessário conhecer a matriz A e o vetor de custos c . Se $\dim \mathcal{N}(A) \geq 1$ então Γ e W são não vazios. O caso $\dim \mathcal{N}(A) = 0$ é trivial pois dado que $A \in \mathbb{R}^{m \times n}$ é de posto máximo m então o conjunto definido por $Ax = b$ e $x \geq 0$ tem um único elemento, assim

a única direção viável é zero. Dado que Γ é finito, podemos escrever os conjuntos por extenso como segue: $\Gamma = \{J_1, \dots, J_l\}$, onde $l = |\Gamma|$, e $W = \{w^{J_1}, \dots, w^{J_l}\}$. Se $\hat{W} \neq \emptyset$ (como provaremos adiante) então podemos calcular $\min\{c^T w \mid w \in \hat{W}\}$, número que depende só da instância A, c . No trabalho demonstraremos que este número caracteriza $\tau_{A,c}$. Vejamos alguns fatos relacionados com os conjuntos W e \hat{W} :

Fato 3.5.4 *Os elementos de W geram $\mathcal{N}(A)$.*

Demonstração. Dado que $A \in \mathbb{R}^{m \times n}$ é de posto máximo, e $m < n$ então podemos escolher m colunas de A linearmente independentes, que indexaremos por B . Assim $\dim \mathcal{N}(A_B) = 0$ e as colunas de A_B geram $\mathcal{R}(A)$. Denotamos por N aos índices das restantes $n - m$ colunas. Cada coluna A_j indexada por $j \in N$, pode-se escrever como $A_B y_B^i = A_i$, com $y_B^i \neq 0$. Definimos para cada $i \in N$ o vetor $w^i = \beta^i \begin{pmatrix} y_B^i \\ -1 \\ 0_{N \setminus \{i\}} \end{pmatrix}$ onde β^i é escolhido de maneira que $\|w^i\| = 1$ e $c^T w^i \geq 0$. Por construção $w^i \in W$ para todo $i \in N$, e como $\dim \mathcal{N}(A) = n - m$, $|N| = n - m$ e os vetores w^i , $i \in N$ são linearmente independentes, concluímos que esses vetores geram $\mathcal{N}(A)$. ■

Fato 3.5.5 *$\hat{W} \neq \emptyset$.*

Demonstração. Suponhamos por contradição que $\hat{W} = \emptyset$, quer dizer, para todo $w \in W$, $c_p^T w = 0$. Isto é, $c_p \perp \mathcal{N}(A)$ porque W gera $\mathcal{N}(A)$. Dado que $c_p \in \mathcal{N}(A)$, c_p tem que ser nulo o que contradiz $c \notin \mathcal{R}(A^T)$. ■

Fato 3.5.6 *Cada elemento de \hat{W} é o limite de direções afim-escala.*

Demonstração. Consideremos $w^J \in \hat{W}$ arbitrário, temos que $A_J w_J^J = 0$ e $J \in \Gamma$. Tomamos a seqüência $(d^k) \subset \mathbb{R}_{++}^n$ definida como $d^k = \begin{pmatrix} e_J \\ \lambda_k e_{J^c} \end{pmatrix}$ onde

$\lambda_k \downarrow 0$, que converge para $\bar{d} = \begin{pmatrix} e_J \\ 0_{J^c} \end{pmatrix}$. A seqüência $u(d^k) = D^k P_{AD^k} D^k c$ converge para $P_{A_J} c_J$ pelo lema 3.4.6, e dado que $\mathcal{N}(A_J)$ é gerado por w^J , e $c_J^T P_{A_J} c_J > 0$ concluímos que $w^J = \frac{u(\bar{d})}{\|u(\bar{d})\|}$, completando a prova. ■

Fato 3.5.7 *Se para todo $d \in \mathbb{R}_{++}^n$ se verifica que $u(d)$ e c_p são colineares, então \hat{W} tem como único elemento o vetor c_p .*

Demonstração. Dado que $\hat{W} \neq \emptyset$, pelo fato 3.5.5, existe um vetor $w \in \hat{W}$ que pelo fato 3.5.6 é limite de direções afim-escala; mas todas as direções $u(d)$ são escritas como $u(d) = \alpha_d c_p$ para algum $\alpha_d > 0$, portanto $w = \frac{c_p}{\|c_p\|}$ é o único elemento de \hat{W} . ■

Dado um subconjunto J dos índices, denotamos por W_J (respectivamente \hat{W}_J) ao conjunto definido como W (respectivamente \hat{W}) mas com a instância dada por A_J, c_J .

Fato 3.5.8 *Dado $J \subset \{1, \dots, n\}$ Se $t \in W_J$ (respectivamente $t \in \hat{W}_J$) então $\begin{pmatrix} t \\ 0_{J^c} \end{pmatrix} \in W$ (respectivamente $\begin{pmatrix} t \\ 0_{J^c} \end{pmatrix} \in \hat{W}$).*

Demonstração. Se $t \in W_J$ (respectivamente $t \in \hat{W}_J$) então existe um subconjunto de J digamos I tal que $\|t_I\| = 1$, $I \in \Gamma$, $t_I \in \mathcal{N}(A_I)$ e $c_J^T t \geq 0$ (respectivamente $c^T t > 0$). Completando com zeros o vetor t até completar $\{1, \dots, n\}$ e usando as relações acima obtemos a tese. ■

3.5.3 O cálculo de $\tau_{A,c}$

Escreveremos agora uma seqüência de lemas visando caracterizar $\tau_{A,c}$.

Lema 3.5.9 *Consideremos $J \subset \{1, \dots, n\}$ tal que $\mathcal{N}(A_J) \neq \{0\}$ e $c_J \notin \mathcal{R}(A_J^T)$. Então*

$$\tau_{A_J, c_J} \geq \tau_{A,c}.$$

Demonstração. Tomamos $d_J \in \overline{\mathbb{R}_{++}^{|J|}}$ arbitrário. Então, como $c_J \notin \mathcal{R}(A_J^T)$ segue-se que

$$u_J(d_J) = D_J P_{A_J D_J} D_J c_J \neq 0.$$

Agora, construímos uma seqüência (d^k) em \mathbb{R}_{++}^n tal que $d_J^k = d_J$, $d_{J^c}^k = \lambda_k e$, onde (λ_k) é uma seqüência em \mathbb{R}_{++} tal que $\lambda_k \rightarrow 0$. Dado que $d_J^k \rightarrow d_J$ e $d_{J^c}^k \rightarrow 0$, temos pelo lema 3.4.6 que $u_J(d^k) \rightarrow u_J(d_J) \neq 0$ e $u_{J^c}(d^k) \rightarrow 0$; portanto $\|u(d^k)\| \|u_J(d_J)\| \neq 0$. Por definição,

$$\tau_{A,c} \leq \frac{c^T u(d^k)}{\|u(d^k)\|},$$

e tomando limites

$$\tau_{A,c} \leq \frac{c_J^T u_J(d_J)}{\|u_J(d_J)\|}.$$

Como esta relação vale para qualquer $d_J \in \mathbb{R}_{++}^{|J|}$,

$$\tau_{A,c} \leq \inf_{d_J \in \mathbb{R}_{++}^{|J|}} \frac{c_J^T u_J(d_J)}{\|u_J(d_J)\|} = \tau_{A_J, c_J},$$

completando a prova. ■

Lema 3.5.10 *Consideremos um par $(d, (d^k)) \in \bar{\mathcal{D}}_{A,c}$. Se para alguma partição $\{J, J^c\}$ dos índices $\{1, \dots, n\}$, temos que*

$$\frac{u_{J^c}(d^k)}{\|u(d^k)\|} \rightarrow 0,$$

então

$$\tau_{A,c} \geq \tau_{A_J, c_J}.$$

Demonstração. Consideremos um par $(d, (d^k)) \in \mathcal{D}_{A,c}$, e uma partição $\{J, J^c\}$ como acima. Definamos para $k \in \mathbb{N}$ a seqüência (r^k) por

$$r^k := \frac{u(d^k)}{\|u(d^k)\|}.$$

A seqüência (r^k) é limitada, e por hipótese, $r_{J^c}^k \rightarrow 0$. Portanto $\|r_J^k\| \rightarrow 1$. Tomando subsequências se necessário, suponhamos que

$$r^k \rightarrow r \in \mathbb{R}^n, \quad \|r^k\| \rightarrow 1, \quad \text{e } \|r_{J^c}^k\| > 0.$$

Temos então para $k \in \mathbb{N}$,

$$r_J^k = \frac{u_J(d^k)}{\|u_J(d^k)\|} \frac{\|u_J(d^k)\|}{\|u(d^k)\|},$$

e tomando limites,

$$r_J = \lim_{k \rightarrow \infty} \frac{u_J(d^k)}{\|u_J(d^k)\|}. \quad (3.28)$$

Por definição de $u(\cdot)$, temos para $k \in \mathbb{N}$ que $Au(d^k) = 0$, e portanto

$$A_J u_J(d^k) = -A_{J^c} u_{J^c}(d^k),$$

ou equivalentemente, dividindo por $\|u(d^k)\|$

$$A_J r_J^k = -A_{J^c} r_{J^c}^k;$$

e tomando limites,

$$A_J r_J = 0.$$

Temos ainda

$$0 < \tau_{A,c} = \lim_{k \rightarrow \infty} \frac{c^T u(d^k)}{\|u(d^k)\|} = c^T r = c_J^T r_J. \quad (3.29)$$

Mostraremos que $c_J^T r_J \geq \tau_{A_J, c_J}$, construindo uma seqüência de direções afim-escala $\hat{u}_J(d_J^k)$ no problema reduzido, com $\frac{\hat{u}_J(d_J^k)}{\|\hat{u}_J(d_J^k)\|} \rightarrow r_J$. Fixando $k \in \mathbb{N}$, tomamos $d = d^k$, $u = u(d^k)$ e estudamos o subespaço afim definido por $A_J u_J = -A_{J^c} u_{J^c}$. Vamos calcular Δu_J tal que

$$\hat{u}_J = u_J - \Delta u_J \in \mathcal{N}(A_J). \quad (3.30)$$

Escolhemos um ponto \tilde{u}_J no conjunto afim, o ponto de mínima norma

$$\hat{u}_J = -A_J^+ A_{J^c} u_{J^c}. \quad (3.31)$$

Calculamos a projeção oblíqua de \tilde{u}_J sobre $\mathcal{N}(A_J)$ e fazemos

$$\begin{aligned} \Delta u_J &= \tilde{u}_J - Q_{A_J, d_J} \tilde{u}_J \\ &= (I - Q_{A_J, d_J}) \tilde{u}_J \\ &= -(I - Q_{A_J, d_J}) A_J^+ A_{J^c} u_{J^c}. \end{aligned} \quad (3.32)$$

Pela expressão (2.13) temos que existe $\tilde{y} \in \mathbb{R}^m$ tal que

$$Q_{A_J, d_J} \tilde{u}_J = \tilde{u}_J - D_J^2 A_J^T \tilde{y}$$

e portanto

$$\Delta u = D_J^2 A_J^T \tilde{y}. \quad (3.33)$$

Finalmente podemos concluir a demonstração, com os seguintes fatos:

i) A direção \hat{u}_J é uma direção afim-escala. De (3.33), (3.29) e (3.30)

$$\hat{u}_J = D_J^2(c_J - A_J^T y) - D_J^2 A_J^T \tilde{y} = D_J^2(c_J - A_J^T(y + \tilde{y})),$$

e $\hat{u}_J \in \mathcal{N}(A_J)$, pois $A_J \Delta u_J = -A_{J^c} u_{J^c}$ (por (3.32)), assim $A_J \hat{u}_J = A_J u_J + A_{J^c} u_{J^c} = 0$. Essas duas expressões significam por (2.9) que $\hat{u}_J = u_J(d_J)$.

ii) De (3.30) e (3.32), para $k \in \mathbb{N}$

$$\frac{\hat{u}_J^k}{\|u(d^k)\|} = \frac{u_J^k}{\|u(d^k)\|} + (I - Q_{A_J, d_J}) A_J^+ A_{J^c} \frac{u_{J^c}(d^k)}{\|u(d^k)\|}.$$

Tomando limites, e observando que pelo lema (2.4.2), a matriz $(I - Q_{A_J, d_J})$ é limitada, $r_{J^c}^k \rightarrow 0$, e $r_J^k \rightarrow r_J$, temos

$$\frac{\hat{u}_J^k}{\|u(d^k)\|} \rightarrow r_J.$$

Tomando normas,

$$\frac{\|\hat{u}_J^k\|}{\|u(d^k)\|} \rightarrow 1$$

de onde conclui-se que

$$\frac{\hat{u}_J^k}{\|u(d_J^k)\|} \rightarrow r_J.$$

iii) Como $\hat{u}_J^k = u_J(d_J^k)$, esta última expressão implica que

$$\tau_{A_J, c_J} \leq \frac{c^T \hat{u}_J^k}{\|\hat{u}_J^k\|}.$$

Tomando limites e usando (3.29),

$$\tau_{A_J, c_J} \leq c_J^T r_J = \tau_{A, c},$$

completando a demonstração.

Lema 3.5.11 Consideremos um par $(d, (d^k)) \in \bar{\mathcal{D}}_{A,c}$ e $u^k = u(d^k)$. Se existe uma partição própria $\{J, J^c\}$ dos índices $\{1, \dots, n\}$ tal que $c_J \notin \mathcal{R}(A_J^T)$ e

$$\tau_{J^c} = \frac{u_{J^c}^k}{\|u^k\|} \rightarrow 0,$$

então

$$\tau_{A,c} = \tau_{A_J, c_J}.$$

Demonstração. Pelo lema 3.5.9, dado que $c_J \notin \mathcal{R}(A_J^T)$, concluímos que

$$\tau_{A,c} \leq \tau_{A_J, c_J};$$

e pelo lema 3.5.10, dado que $r_{J^c}^k \rightarrow 0$ temos

$$\tau_{A,c} \geq \tau_{A_J, c_J};$$

isto é, $\tau_{A,c} = \tau_{A_J, c_J}$. ■

Lema 3.5.12 Suponhamos que para uma partição própria $\{J, J^c\}$ dos índices $\{1, \dots, n\}$ temos $\tau_{A,c} = \tau_{A_J, c_J}$. Se $d_J \in \mathcal{D}_{A_J, c_J}$ então o vetor não nulo $\bar{d} \in \mathbb{R}_+^n$ definido por $\bar{d}_J = d_J$ e $\bar{d}_{J^c} = 0$ pertence a $\mathcal{D}_{A,c}$.

Demonstração. Tomamos $d_J \in \mathcal{D}_{A_J, c_J}$ e a seqüência (d_J^k) em $\mathbb{R}_+^{|J|}$ tal que $d_J^k \rightarrow d_J$. Construimos a seqüência (\hat{d}^k) em \mathbb{R}_+^n da seguinte forma: $\hat{d}_J^k = d_J^k$ e $\hat{d}_{J^c}^k = \lambda_k e$ para uma seqüência (λ_k) em \mathbb{R}_+ convergente a zero. Pelo lema 3.4.6 temos que $u_J(\hat{d}^k) \rightarrow u_J(d_J)$ e $u_{J^c}(\hat{d}^k) \rightarrow 0$. Assim,

$$\frac{c^T u(\hat{d}^k)}{\|u(\hat{d}^k)\|} \rightarrow \frac{c_J^T u_J(d_J)}{\|u(d_J)\|} = \tau_{A,c},$$

completando a prova. ■

Chamaremos a uma instância A, c redutível se existe uma partição dos índices $\{J, J^c\}$ tal que $\tau_{A,c} = \tau_{A_J, c_J}$.

Pelo lema 3.5.11 e o lema 3.5.12 concluímos que podemos estudar o número $\tau_{A,c}$ começando com instâncias reduzidas, sempre que a instância original seja redutível. Assim basta com estudar as instâncias não redutíveis

Lema 3.5.13 Consideremos o par $(d, (d^k)) \in \bar{\mathcal{D}}_{A,c}$, a partição dos índices $\{B, N\} = \{B(d), N(d)\}$, e suponhamos que a seqüência $r^k = u(d^k)/\|d(d^k)\|$ é convergente, com limite r . Suponhamos também que u^k não são colineares com $c_p := P_{Ac}$ e que a instância A, c é não redutível. Então

i) $r_i \neq 0$ para todo $i \in \{1, \dots, n\}$.

ii) $u^k \rightarrow 0$.

iii) $c_B \in \mathcal{R}(A_B^T)$.

Demonstração.

i) Suponhamos que existe $i \in \{1, \dots, n\}$ tal que $r_i \rightarrow 0$. Como a instância é não redutível, devemos ter pelo lema 3.5.11 que $c_J \in \mathcal{R}(A_J^T)$ com $J = \{1, \dots, n\} \setminus \{i\}$. Então pelo lema 3.4.2 podemos substituir o vetor c na definição de $u(d)$ por s tal que $s_J = 0$ e $s_i \neq 0$, obtendo

$$\begin{aligned} A_J u_J + A_i u_i &= 0 \\ u_J &= -D_J^2 A_J^T y \\ u_i &= d_i^2 (s_i - A_i^T y). \end{aligned} \tag{3.34}$$

Agora, para as seqüências (d^k) , (u^k) e (r^k) definidas acima, temos

$$r^k = \frac{u^k}{\|u^k\|} \rightarrow \frac{u}{\|u_J\|},$$

com $r_i^k \rightarrow 0$ e $\|r_J^k\| \rightarrow 1$, já que por hipótese $r_i^k \rightarrow 0$. Pela primeira equação de (3.34) obtemos

$$A_J r_J^k + A_i r_i^k = 0,$$

assim $A_J r_J^k \rightarrow 0$ e portanto $r_J \in \mathcal{N}(A_J)$ com $\|r_J\| = 1$. Isto é o vetor r_J pertence ao conjunto $X' = \{x \in \mathcal{N}(A_J) \mid \|x\| = 1\}$. Também pela segunda equação de (3.34) obtemos

$$r_J^k \in \mathcal{R}(D_J^{k^2} A_J^T) \subset S' = \{s \in \mathcal{R}(D_J^2 A_J^T) \mid d_J \in \mathbb{R}_{++}^n\}$$

Pelo lema 2.4.1 a distância entre esses conjuntos é positiva, que contradiz o fato de que $\|r_J^k - r_J\| \rightarrow 0$.

ii) Pelo lema 3.4.6 temos que $u_N^k \rightarrow 0$. Suponhamos que para algum $i \in B$, $|u_i^k| \geq \epsilon > 0$, então $\|u^k\| \geq \epsilon$, isto é,

$$\frac{u_N^k}{\|u^k\|} \rightarrow 0$$

contradizendo a parte i).

iii) Pelo lema 3.4.6, $u_B = u_B(d_B)$. Se $c_B \notin \mathcal{R}(A_B^T)$ então $u_B(d_B) \neq 0$ contradizendo ii).

■

O próximo lema estuda $\tau_{A,c}$ no caso de instâncias não redutíveis. Este é o caso mais difícil pois o teorema de Megiddo e Shub (lema 3.4.6) não pode ser aplicado diretamente. Sabemos pelo lema 3.5.13 que $c_B \in \mathcal{R}(A_B)$ e portanto pelo lema 3.4.6 temos que $u_B(d) = 0$ e também $u_N(d) = 0$, assim o valor de $\tau_{A,c}$ não pode ser calculado diretamente da direção limite. Devemos estudar então a seqüência de direções afim-escala. Apresentamos um resultado parcial, já que faremos a suposição de que todas as componentes da seqüência de variáveis pequenas (d_N^k) vão para zero com a mesma velocidade.

Lema 3.5.14 *Consideremos o conjunto \hat{W} definido antes, o par $(d, (d^k)) \in \bar{\mathcal{D}}_{A,c}$, e a partição $\{B, N\} = \{B(d), N(d)\}$ dos índices $\{1, \dots, n\}$. Suponhamos que existe um $\epsilon > 0$ tal que a seqüência de variáveis pequenas (d_N^k) satisfaz $d_i^k / \|d_N^k\| > \epsilon$ para todo $i \in N$; para algum $k \in \mathbb{N}$, $u(d^k)$ é não colinear com c_p , e que a instância A, c é não redutível. Então*

$$\tau_{A,c} = c^T w^{J^*} = \min\{c^T w^J \mid w^J \in \hat{W}\}$$

Demonstração. Faremos a demonstração em três partes:

Primeira parte: Estabeleceremos propriedades da seqüência de direções $(u(d^k))$. Tomamos $(d, (d^k)) \in \bar{\mathcal{D}}_{A,c}$, $\{B, N\}$ e u^k definidos como acima. Dado que a instância é não redutível, então pelo fato 3.5.13 temos que $c_B \in \mathcal{R}(A_B^T)$. Pelo lema

3.4.2 existe um vetor $\tilde{s} \in c + \mathcal{R}(A^T)$ tal que $\tilde{s}_B = 0$, $\tilde{s}_N \neq 0$ e as direções afim-escala satisfazem, para $d^k \in \mathbb{R}_{++}^n$

$$\begin{aligned} A_B u_B^k + A_N u_N^k &= 0 \\ u_B^k &= -(D_B^k)^2 A_B^T \tilde{y}^k \\ u_N^k &= (D_N^k)^2 (\tilde{s}_N - A_N^T \tilde{y}^k). \end{aligned} \tag{3.35}$$

A seqüência (\tilde{y}^k) é limitada, pois por (2.9) obtemos

$$A(D^k)^2 A^T \tilde{y}^k = A(D^k)^2 \tilde{s},$$

assim,

$$\begin{aligned} \|\tilde{y}^k\| &= \|(A(D^k)^2 A^T)^{-1} A(D^k)^2 \tilde{s}\| \\ &\leq \|(A(D^k)^2 A^T)^{-1} A(D^k)^2\| \|\tilde{s}\| \leq \chi_A \|\tilde{s}\|. \end{aligned}$$

A última desigualdade vale pelo lema (2.4.2). Tomando subsequências podemos supor que $\tilde{y}^k \rightarrow \bar{y}$. Definimos os conjuntos

$$J = \{i \in N \mid \tilde{s}_i - A_i^T \bar{y} = 0\}, \text{ e } J^c = N \setminus J.$$

(Depois provaremos que $J = \emptyset$, portanto $J^c = N$.) $J^c \neq \emptyset$ pois $\tilde{s} \notin \mathcal{R}(A^T)$. Temos pela segunda equação de (3.35) e a parte *ii*) do lema 3.5.13 que $A_B^T \bar{y} = 0$. Pela construção de J temos $A_J^T \bar{y} = \tilde{s}_J$. Agora definimos o vetor $s = \tilde{s} - A^T \bar{y}$. Verifica-se então que $s_B = 0$, $s_J = 0$ e $s_i = \tilde{s}_i - A_i^T \bar{y} \neq 0$ para todo $i \in J^c$. Definimos a seqüência (y^k) como $y^k = \tilde{y}^k - \bar{y}$. Então

$$\begin{aligned} u_B^k &= -D_B^{k^2} A_B^T \tilde{y}^k = -D_B^{k^2} A_B^T (\tilde{y}^k - \bar{y}) = -D_B^{k^2} A_B^T y^k \\ u_J^k &= D_J^{k^2} (\tilde{s}_J - A_J^T \tilde{y}^k) = D_J^{k^2} (\tilde{s}_J - A_J^T \bar{y} - A_J^T (\tilde{y}^k - \bar{y})) = -D_J^{k^2} A_J^T y^k \\ u_{J^c}^k &= D_{J^c}^{k^2} (\tilde{s}_{J^c} - A_{J^c}^T \tilde{y}^k) = D_{J^c}^{k^2} (s_{J^c} - A_{J^c}^T y^k). \end{aligned}$$

assim podemos escrever u^k como a solução de

$$\begin{aligned} u_B^k &= -(D_B^k)^2 A_B^T y^k \\ u_J^k &= -(D_J^k)^2 A_J^T y^k \\ u_{J^c}^k &= (D_{J^c}^k)^2 (s_{J^c} - A_{J^c}^T y^k), \end{aligned} \tag{3.36}$$

com $y^k \rightarrow 0$. Definimos a seqüência (h^k) como $h^k = d_N^k / \|d_N^k\|$. Temos que $\|h^k\| = 1$. Tomando subseqüências podemos supor que $h^k \rightarrow h$ e que $\|h\| = 1$. Por hipótese $h_i > 0$ para todo $i \in N$. Escrevemos agora a seqüência (d_N^k) como $d_N^k = \lambda_k h^k$, onde $\lambda_k = \|d_N^k\|$. Assim temos

$$u_{J^c}^k = \lambda_k^2 H_{J^c}^{k^2} (s_{J^c} - A_{J^c}^T y^k). \quad (3.37)$$

Estudemos agora $r_{J^c}^k$: Tomamos um i em J^c . Sabemos por hipótese que $h_i^k \rightarrow h_i > 0$; Vejamos r_i^k . Pela definição de r_i^k e a expressão $d_N^k = \lambda_k h^k$ obtemos

$$r_i^k = \frac{\lambda_k^2}{\|u^k\|} h_i^k (s_i - A_i^T y^k).$$

Sabemos pelo lema 3.5.13 que $r_i^k \rightarrow r_i \neq 0$; pela construção que $y^k \rightarrow 0$ e $h_i s_i \neq 0$, assim tomando limite

$$r_i = \lim_{k \rightarrow \infty} \frac{\lambda_k^2}{\|u^k\|} h_i^2 s_i \in \mathbb{R} \setminus \{0\},$$

portanto

$$\delta := \lim_{k \rightarrow \infty} \frac{\lambda_k^2}{\|u^k\|} \in \mathbb{R}_{++}. \quad (3.38)$$

Essa relação nos permite mostrar que $J = \emptyset$: De fato, como $y^k \rightarrow 0$, e pelo lema 3.5.13 temos que $r_i \neq 0$ para todo $i \in \{1, \dots, n\}$. Também sabemos pela terceira equação de (3.35) que $u_J^k = \lambda_k^2 H_J^{k^2} (s_J - A_J^T y^k)$ então

$$r_i = \delta h_i s_i \neq 0, \text{ para todo } i \in J.$$

isto é, $s_i \neq 0$ para todo $i \in J$, que é contradição da definição de J , portanto $J^c = N$.

Temos agora:

$$u_B^k = -(D_B^k)^2 A_B^T y^k,$$

$$u_N^k = (D_N^k)^2 (s_N - A_N^T y^k),$$

$y^k \rightarrow 0$, $s_i \neq 0$ para todo $i \in N$, $d_N^k = \lambda_k h^k$ com $h^k \rightarrow h$, $\|h\| = 1$ e $h_i > 0$ para todo $i \in N$.

Segunda parte: Mostraremos agora que $\mathcal{N}(A_B) = \{0\}$. Para isso estudaremos a seqüência (t^k) em \mathbb{R}^n definida por $t^k = u^k / \lambda_k^2$. Por (3.38) e o lema 3.5.11 temos que

$$t^k = \frac{u^k \|u^k\|}{\lambda_k^2 \|u^k\|} = r^k \frac{\|u^k\|}{\lambda_k^2} \rightarrow \frac{r}{\delta} =: t.$$

portanto $t_i \neq 0$ para todo $i \in \{1, \dots, n\}$.

Suponhamos conhecida a seqüência u_N^k . Então u_B^k pode-se escrever como a solução de mínima norma do conjunto $A_B u_B = -A_N u_N^k$, com relação à norma definida por $d_B^k \in \mathbb{R}_{++}^{|B|}$. De fato, o problema que calcula u_B^k , conhecida u_N^k é

$$\begin{aligned} & \text{minimizar} \quad -s_B^T u_B + \frac{1}{2} \|D_B^{k^{-1}} u_B\| \\ & \text{sujeito a} \quad A_B u_B = -A_N u_N^k. \end{aligned}$$

Como $s_B = 0$, o problema acima coincide com o problema de projeção oblíqua da origem no conjunto afim definido por $A_B u_B = -A_N u_N^k$, com respeito a d_B^k . Isto é u_B^k pode-se escrever como

$$u_B^k = -A_{B,d_B^k}^+ A_N u_N^k \quad (3.39)$$

onde $A_{B,d_B^k}^+$ é a matriz pseudo inversa pesada de A_B com respeito a d_B^k . Dado que $N = J^c$, e pela expressão (3.37) obtemos

$$u_N^k = \lambda_k^2 H^{k^2} (s_N - A_N^T y^k), \quad (3.40)$$

e pela equação (3.39) obtemos

$$u_B^k = -\lambda_k^2 A_{B,d_B^k}^+ A_N H^{k^2} (s_N - A_N^T y^k). \quad (3.41)$$

Assim dividindo por λ_k^2 estas duas expressões obtemos

$$t_N^k = H^{k^2} (s_N - A_N^T y^k) \rightarrow H^2 s_N.$$

e

$$t_B^k = -A_{B,d_B^k}^+ A_N H^{k^2} (s_N - A_N^T y^k) \rightarrow t_B.$$

Para cada $k \in \mathbb{N}$, o vetor t_B^k é a solução de mínima norma com respeito a d_B^k , do sistema linear $A_B \rho_B = -A_N H^{k^2} (s_N - A_N^T y^k)$. Também sabemos que $t_B^k = -D_B^{k^2} A_B^T y^k / \lambda_k^2$, então o vetor limite t_B pertence ao fecho do conjunto

$$T = \{\rho_B \in \mathbb{R}^{|B|} \mid A_B \rho_B = -A_N H^2 s_N, \rho_B \in \mathcal{R}(\tilde{D}_B^2 A_B^T), \tilde{d}_B \in \mathbb{R}_{++}^{|B|}\}.$$

Notemos que $d_B > 0$ e $h > 0$. Pelo lema 2.6.4, o fecho de T é caracterizado como a união de politopos. O vetor de máxima norma de um politopo está num vértice.

Vejam agora

$$r^k = \frac{u^k}{\|u^k\|} = \frac{\lambda_k^2 t^k}{\|\lambda_k^2 t^k\|} = \frac{t^k}{\|t^k\|} \rightarrow \frac{t}{\|t\|}.$$

Sabemos que

$$\tau_{A,c} = c^T r = \frac{s^T t}{\|t\|} = \frac{s_N^T t_N}{\|t\|}, \quad (3.42)$$

portanto

$$\tau_{A,c} = \frac{s_N^T H^2 s_N}{\sqrt{\|t_B\|^2 + \|H^2 s_N\|^2}}. \quad (3.43)$$

Afirmamos que t_B é o vetor de máxima norma do conjunto $cl(T)$. Suponhamos por contradição que o vetor de máxima norma \hat{t}_B de $cl(T)$ é diferente de t_B . Pela definição do conjunto T , o vetor \hat{t}_B é caracterizado por uma seqüência \hat{d}_B^k em $\mathbb{R}_{++}^{|B|}$ tal que $\hat{t}_B = \lim_{k \rightarrow \infty} (\hat{D}_B^k)^2 A_B^T w^k$, com $w^k \in \mathbb{R}^m$. Construiremos uma seqüência de direções afim-escala (\hat{u}^k) tal que $\hat{u}_N^k / \lambda_k^2 \rightarrow H^2 s_N$ e $\hat{u}_B^k / \lambda_k^2 \rightarrow \hat{t}_B$. Consideremos a seqüência (\hat{d}^k) definida como $\hat{d}_B^k := \hat{d}_B^k$ e $\hat{d}_N^k := \lambda_k h^k$, e a seqüência $u(\hat{d}^k)$. De fato, dada a seqüência u_N^k podemos escrever \hat{u}_B^k como a única solução do problema

$$\begin{aligned} &\text{minimizar} && -s_B^T u_B + \frac{1}{2} \|(\hat{D}_B^k)^{-1} u_B\|^2 \\ &\text{sujeito a} && A_B u_B = -A_N u_N^k. \end{aligned}$$

Observemos que $s_B = 0$ e que $u_N^k = \lambda_k^2 H^2 (s_N - A_N^T y^k)$. As condições de otimalidade do problema são:

$$\begin{aligned} u_B + (\hat{D}_B^k)^2 A_B^T w &= 0 \\ A_B u_B &= -A_N u_N^k. \end{aligned} \quad (3.44)$$

Para mostrar que \hat{u}_B^k definido como acima, junto com u_N^k formam a direção afim-escala associada a \hat{d}^k , devemos provar que para w^k tal que

$$\hat{u}^k = (\hat{D}^k)^2 (s - A^T w^k),$$

se satisfaz

$$\hat{u}_N^k = u_N^k = (\hat{D}_N^k)^2 (s_N - A_N^T y^k).$$

De fato, o vetor w^k satisfaz $A(\hat{D}^k)^2 A^T w^k = A(\hat{D}^k)^2 s$. Usando a partição $\{B, N\}$, $\hat{d}_N^k = \lambda_k^2 h^k$ e o fato que $s_B = 0$ obtemos

$$A_B (\hat{D}_B^k)^2 A_B^T w^k + \lambda_k^2 A_N (H_N^k)^2 A_N^T w^k = \lambda_k^2 A_N (H^k)^2 s_N. \quad (3.45)$$

Pelas equações (3.44) obtemos, multiplicando a primeira por A_B e substituindo na segunda que

$$A_B(\hat{D}_B^k)^2 A_B^T w^k = \lambda_k^2 A_N H^{k^2} (s_N - A_N^T y^k),$$

agora, substituindo esta última em (3.45) temos

$$A_N H^{k^2} A_N^T w^k = A_N H^{k^2} A_N^T y^k$$

assim,

$$\hat{u}_N^k = -\lambda_k^2 A_N H_N^{k^2} (s_N - A_N^T w^k) = -\lambda_k^2 A_N H_N^{k^2} (s_N - A_N^T y^k) = u_N^k,$$

portanto $A_B \hat{u}_B^k + A_N u_N^k = 0$, $\hat{u}_B^k = -(\hat{D}_B^k)^2 A_B^T w^k$ e $\hat{u}_N^k = u_N^k = (\hat{D}_B^k)^2 (s_N - A_N^T w^k)$.

Isto é, \hat{u}^k é uma seqüência de direções afim-escala tal que $\frac{\hat{u}_N^k}{\lambda_k^2} \rightarrow H^2 s_N$ e $\frac{\hat{u}_B^k}{\lambda_k^2} \rightarrow \hat{t}_B$.

Dado que $\|\hat{t}_B\| > \|t_B\|$ obtemos

$$\frac{s^T u(\hat{d}^k)}{\|u(\hat{d}^k)\|} \rightarrow \frac{s_N^T H^2 s_N}{\sqrt{\|\hat{t}_B\|^2 + \|H^2 s_N\|^2}} < \frac{s_N^T H^2 s_N}{\sqrt{\|t_B\|^2 + \|H^2 s_N\|^2}} = \tau_{A,c},$$

que é uma contradição pela infimalidade de $\tau_{A,c}$. Isto significa que t_B é o vetor de máxima norma de $cl(T)$; portanto t_B pertence a um vértice de $cl(T)$. Pelo lema 2.6.6 temos que para a partição $\{B(t_B), N(t_B)\}$ de B , verifica-se $\mathcal{N}(A_{B(t_B)}) = \{0\}$, mas como todas as componentes de t_B são não nulas temos que $B(t_B) = B$ e portanto $\dim \mathcal{N}(A_B) = 0$.

Terceira parte: Mostraremos agora que o limite $t/\|t\|$ pertence a \hat{W} : Como t^k é solução de mínima norma com respeito a d_B^k do sistema $A_B \rho_B = -A_N H^{k^2} (s_N - A_N^T y^k)$, e o lado direito converge para $-A_N H^2 s_N$ que tem as componentes todas positivas, concluímos que t_B é o vetor de mínima norma do sistema linear $A_B \rho_B = -A_N H^2 s_N$ com respeito à norma definida por d_B , assim podemos escrever t como:

$$\begin{pmatrix} t_B \\ t_N \end{pmatrix} = \begin{pmatrix} -A_{B,d_B}^+ A_N H^2 s_N \\ H^2 s_N \end{pmatrix}.$$

onde A_{B,d_B}^+ é a matriz pseudoinversa pesada de A_B com respeito a d_B .

Agora fazemos uma partição das componentes de N da seguinte maneira: Dado que A é de posto m então existem m colunas de A linearmente independentes.

Como as colunas de A_B são linearmente independentes (já que $\mathcal{N}(A_B) = \{0\}$) então escolhemos $m - |B|$ colunas de A_N linearmente independentes entre elas, e com as colunas de A_B para completar uma base de $\mathcal{R}(A)$. Denotemos por N_0 este conjunto de índices. Assim as colunas de $A_{B \cup N_0}$ são linearmente independentes, então a dimensão de $\mathcal{N}(A_{B \cup N_0}) = 0$. Agora com as restantes $l := n - m$ colunas de A_N construímos os conjuntos $J_i = B \cup N_0 \cup \{i\}$ $i = 1, \dots, l$ que verificam $\dim \mathcal{N}(A_{J_i}) = 1$, portanto $J_i \in \Gamma$ para todo $i = 1, \dots, l$. Para cada $i = 1, \dots, l$ denotamos $N_i = N_0 \cup \{i\}$. Definimos os vetores $z^i, i = 1, \dots, l$ como

$$\begin{pmatrix} -A_{B,d_B} A_{N_1} H_{N_1}^2 s_{N_1} \\ H_{N_1}^2 s_{N_1} \\ 0 \end{pmatrix}, \begin{pmatrix} -A_{B,d_B} A_{N_2} H_{N_2}^2 s_{N_2} \\ 0 \\ H_{N_2}^2 s_{N_2} \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} -A_{B,d_B} A_{N_l} H_{N_l}^2 s_{N_l} \\ 0 \\ H_{N_l}^2 s_{N_l} \end{pmatrix}.$$

Sabemos que

$$s^T z^i = s_{N_i}^T H_{N_i}^2 s_{N_i} > 0, \text{ para todo } i = 1, \dots, l. \quad (3.46)$$

Por outra parte temos

$$A_{J_i} z^i = (A_B, A_{N_i}) \begin{pmatrix} -A_{B,d_B} A_{N_i} H_{N_i}^2 s_{N_i} \\ H_{N_i}^2 s_{N_i} \end{pmatrix} = 0, \quad (3.47)$$

Para cada $i = 1, \dots, l$. Temos por (3.46) que $c^T z^i > 0$ e por (3.47) $z^i \in \mathcal{N}(A_{J_i})$ assim como $\dim(A_{J_i}) = 1$ concluímos que existem constantes positivas α_i tais que $z^i = \alpha_i' w^{J_i}$, para todo $i = 1, \dots, l$, onde $w^{J_i} \in \hat{W}$. Desta forma podemos escrever t como

$$t = \sum_{i=1}^l z^i = \sum_{i=1}^l \alpha_i w^{J_i}$$

Agora substituindo em (3.42) obtemos

$$\begin{aligned} \tau_{A,c} &= \frac{c^T t}{\|t\|} \\ &= \sum_{i=1}^l \frac{\alpha_i}{\|\sum_{i=1}^l \alpha_i w^{J_i}\|} c^T w^{J_i} \\ &\geq \sum_{i=1}^l \frac{\alpha_i}{\sum_{i=1}^l \alpha_i} c^T w^{J_i} \\ &\geq \sum_{i=1}^l \frac{\alpha_i}{\sum_{i=1}^l \alpha_i} c^T w^{J_*} \\ &= c^T w^{J_*} \end{aligned}$$

onde $c^T w^{J^*} := \min\{c^T w^J \mid w^J \in \tilde{W}\}$. A primeira desigualdade segue-se pela desigualdade triangular e $\|w^{J_i}\| = 1, \forall i$. A segunda pela definição de w^{J^*} e

$$\sum_{i=1}^l \frac{\alpha_i}{\sum_{i=1}^l \alpha_i} = 1.$$

Por outra parte, sabemos pelo fato 3.5.6 que w^{J^*} é o limite de direções afim-escala, Isto é $c^T w^{J^*} \geq \tau_{A,c}$ e assim

$$\tau_{A,c} = c^T w^{J^*} = \min\{c^T w^J \mid w^J \in \tilde{W}\}$$

completando a prova. ■

O teorema final resume os resultados anteriores:

Teorema 3.5.15 *Consideremos o conjunto \hat{W} e o número $\tau_{A,c}$ como definidos antes. Se existe um par $(d, (d^k)) \in \bar{\mathcal{D}}_{A,c}$ e $\epsilon > 0$ tal que $\forall k \in \mathbb{N}$, $d_i^k / \|d_N^k(d)\| \geq \epsilon$ e $i \in N$, então*

$$\tau_{A,c} = \min\{c^T w \mid w \in \hat{W}\}.$$

Demonstração. Tomamos $(d, (d^k)) \in \bar{\mathcal{D}}_{A,c}$ como na hipótese. Se $u(d)$ é colinear com c_p para todo $d \in \mathbb{R}_{++}^n$ então pelo fato 3.5.7 c_p é o único elemento de \hat{W} , assim

$$\tau_{A,c} = \frac{c^T c_p}{\|c_p\|} = \min\{c^T w \mid w \in \hat{W}\}.$$

Suponhamos agora que existem direções afim-escala não colineares com c_p . Se a instância A, c é redutível então existe uma partição $\{I, I^c\}$ dos índices $\{1, \dots, n\}$ tal que $\tau_{A,c} = \tau_{A_I, c_I}$. Pelo lema 3.5.14 temos que

$$\tau_{A_I, c_I} = \min\{c_I^T w_I \mid w_I \in \hat{W}_I\}.$$

Tomamos $\bar{w}_I^* \in \operatorname{argmin}\{c_I^T w_I \mid w_I \in \hat{W}_I\}$. Pelo fato 3.5.8 concluímos que o vetor w^* definido como $w_I^* := \bar{w}_I^*$ e $w_{I^c}^* = 0$ pertence a \hat{W} . Afirmamos que $w^* \in \operatorname{argmin}\{c^T w \mid w \in \hat{W}\}$ pois no caso contrário existiria $\tilde{w} \in \hat{W}$ tal que $c^T \tilde{w} < c^T w^*$, mas dado que \hat{W} é limite de direções afim-escala, pelo fato 3.5.6, obtemos uma contradição da definição de $\tau_{A,c}$, assim completamos a prova. ■

O teorema acima precisa da suposição que as componentes pequenas (d_N^k) converjam para zero com a mesma taxa, mas precisamos de tal hipótese só nas variáveis pequenas da instância que resta depois de fazer todas as reduções possíveis. Se variáveis pequenas d_J^k ($J \subset N$) que podem ser reduzidas não se aproximam a zero com a mesma velocidade, elas são eliminadas no proceso de redução.

Capítulo 4

Comentários finais

Nós apresentamos três resultados principais. Descreveremos a importância, o alcance e as expectativas de cada um:

Vavasis e Ye apresentaram na seqüência de artigos [45], [46], [47] e [48] resultados de complexidade envolvendo o número de condicionamento $\bar{\chi}_A$. Utilizaram $\bar{\chi}_A$ por exemplo para separar as variáveis x_i e x_j como grandes e pequenas se $\frac{x_i}{x_j} \geq g$, onde $g = K\sqrt{n}(\bar{\chi}_A + 1)$ para uma constante K . Segundo o nosso resultado, (teorema 2.5.3) a expressão $(\bar{\chi}_A + 1)$ pode ser substituída por $\bar{\chi}_A$ na definição de g (Ver [49]). A melhoria conseguida por nosso teorema não é tanto no refinamento do limitante, já que a diferença de uma unidade nos números $\bar{\chi}_A$ e $(\bar{\chi}_A + 1)$ é insignificante dada a magnitude deles, mas sim na compreensão dos conceitos e em fechar a diferença entre os números de condição primal e dual que a definição de Vavasis e Ye mantinha. Foi escrito e publicado já um artigo com este resultado (ver [18]), e referenciado em [48], [52] e [42].

Desde 1990, quando Ye [50] apresentou o teorema de complexidade baseado na “menor variável grande” σ (lema 2.3.1), ele tentou descrever dito número com outras características do problema que não fossem a partição ótima. Mais tarde em 1993, Ye junto com Vavasis [45] conseguiu um limitante inferior para σ no contexto do problema de viabilidade de restrições lineares no formato de Karmarkar (ou formato homogêneo), que mostramos na seção 2.8. O limitante envolve o inverso

de $\bar{\chi}_A$. Dita caracterização, segundo mostrou Tunçel [42] é muito rigorosa: Tunçel apresentou uma família de instâncias (parametrizadas por ϵ) com menor variável grande da ordem de $1/2$, enquanto o número $(\bar{\chi}_{A_\epsilon})^{-1}$ é arbitrariamente pequeno. Desejamos caracterizar σ pois com ele podemos estabelecer um modelo real para a análise da complexidade de alguns algoritmos de pontos interiores em programação linear. Na prática, desde o ponto de vista da construção de algoritmos, não podemos usar diretamente σ , pois para defini-lo precisamos conhecer a partição ótima, que é equivalente a resolver o problema antes de usar o algoritmo. Nosso resultado de caracterização de σ (teorema 2.8.3) permite usar o teorema de complexidade de Ye na análise de alguns algoritmos de trajetória central. O número de condição ξ que propomos é, em nosso conhecimento, a primeira caracterização de σ definida para o problema de programação linear em formato padrão. A chave de nossa análise é estudar o problema de programação linear normalizado de maneira tal que os problemas primal e dual sejam simétricos. A teoria de dualidade joga um papel importante. Estudamos os problemas de otimização que definem a “menor variável grande” primal; depois estabelecemos os problemas duais associados com estes problemas primais e demonstramos que para cada um dos sub-problemas, algumas soluções duais podem ser definidas diretamente dos dados do problema. De fato, mostramos que existem soluções ótimas dos subproblemas duais que coincidem com os vetores do espaço dual original que são paralelos às arestas do conjunto dual viável original. Acreditamos que nosso resultado possa ser base para projetos de algoritmos que utilizem informação do tamanho das variáveis ótimas na sua definição.

Em [47] Vavasis e Ye apresentaram uma caracterização da trajetória central quando o parâmetro da trajetória varia em $(0, +\infty)$. Eles mostraram que a trajetória está formada por trechos (quasi) retos e curvos, e estimaram um limitante para o número de mudanças que acontecem (da ordem $O(n^2)$) baseados no conceito de “crossover events”, que estabelece um modelo para medir quando uma variável é grande ou pequena, e conta o número total de vezes que as variáveis passam de ser

grandes a pequenas ou vice-versa. É conhecido que a direção afim-escala calculada num ponto sobre a trajetória central é tangente à trajetória. O resultado central do capítulo 3 (teorema 3.5.15) descreve as direções afim-escala dentro de um cone gerado pelos elementos de \hat{W} que são vetores de $\mathcal{N}(A)$ paralelos às arestas do conjunto viável. Nós acreditamos que as parcelas retas da trajetória poden estar relacionadas com as direções de \hat{W} , assim poderíamos estimar o número de retas e curvas da trajetória segundo a cardinalidade de \hat{W} .

Sabemos que o algoritmo afim-escala primal de passos longos tem prova de convergência global sem hipótese de não degeneração para o tamanho do passo de $2/3$ do tamanho máximo viável. No entanto, existe uma instância dada por Mascarenhas [25] onde o algoritmo afim-escala converge para um ponto não ótimo quando é utilizado um comprimento do passo de 0.9999 do tamanho máximo viável. Esperamos que nossa análise do capítulo 3 ajude a esclarecer o vazio entre eses dois números. Esperamos também em breve demonstrar o teorema 3.5.15 sem hipótese nenhuma.

Bibliografia

- [1] K. Anstreicher. “A combined ‘phase I-phase II’ projective algorithm for linear programming”, Manuscript, Yale School of Organization and Management, New Havenm CT, to appear in *Mathematical Programming*
- [2] K. Anstreicher. “Linear programming in $O(\frac{n^3}{\log n}L)$ operations” Dept. of Management Sciences, University of Iowa, Iowa City, IA 52242, USA, March, 31 (1997).
- [3] E. Barnes. “A variation on Karmarkar’s algorithm for solving linear programming problems”, *Mathematical Programming*, Vol. 36,pp 174-182 (1986).
- [4] L. Blum, M. Shub and S. Smale. “On a theory of computation and complexity over the real numbers: NP-completeness, recursive functions and universal machines”, *AMS Bull. (New series)*, 21, pp 1-46 (1989).
- [5] R. E. Campello and N. Maculan. Algoritmos e Heurísticas: Desenvolvimento e Avaliação de performance. Editora da Universidade Federal Fluminense, EDUFF. (1994).
- [6] T.M Cavalier and A.L Soyster. “Some computational experience and a modification of the Karmarkar Algoritm”, The Pensilvania Sate University, ISME Working Paper, pp 85-105 (1985).
- [7] G. Danzing. “Programming in linear structure”, Report of the September 9, 1948 meeting in Madison, *Econometrica*, 17, pp 73-74 (1948).

- [8] I. Dikin. "Iterative solution of problems of linear and quadratic programming." *Soviet Mathematics Doklady*, Vol. 8, pp 674-675 (1967).
- [9] I. I Dikin. "O Skhodimosti Odnogo iteratsionnogo protsessa" (in Russian), *Upravlyaemye sistemi*, Vol 2 pp 54-60 (1986).
- [10] I. I Dikin "Letter to the editor" *Mathematical Programming*, Vol.41, pp 393-394 (1988).
- [11] I. Dikin. "The convergence of the dual variables". Technical Report, Siberian Energy Institute, Irkutsic, Rusia, December, (1991).
- [12] D. Gay. "A variant of the Karmarkar's linear programming algorithm for problems in standard form", *Mathematical Programming*, 37. pp 81-89, (1987).
- [13] G. De Ghellinck and J. P. Vial. "A oikynomial newton method for linear programming", *Algorithmica*, 1, pp 425-453 (1986).
- [14] C. C. Gonzaga. "An algorithm for solving linear programming problems in $O(n^3L)$ operations", Memorandum UCB/ERL M87/10, Ekelectronics Research Laboratory, University of California, Berkeley, CA, March (1987).
- [15] C. C. Gonzaga. "Path-following Methods for linear programming." *SIAM Review*. Vol. 34, No. 2, pp 167-224, June (1992).
- [16] C. Gonzaga and A. Carlos. "A primal affine-scaling algorithm for linearly constrained convex programs". Technical Report ES-238/90, Departament of Systems Engeneering anf Computer Science, COPPE-Federal University of Rio de Janeiro. Brazil, December (1990).
- [17] C. C. Gonzaga and R. A. Tapia. On the convergence of the Mizuno-Todd-Ye algorithm to the analytic center of the solution set. *SIAM J. Opt*, Vol 7 No 1: 47-65, 1997.
- [18] C. C. Gonzaga and H. J. Lara. A note on properties of condition numbers. *Linear Algebra and its Applications*, 261:269-273, 1997.

- [19] W. W. Hogan. Point-to-set maps in mathematical programming. *SIAM Rev*, 15:591–603, 1973.
- [20] N. Karmarkar. “A new polynomial-time algorithm for linear programming”. *Combinatorica*, Vol. 4, No. 4, pp. 373-395 (1984).
- [21] N. Karmarkar and K. Ramakrisnan. “Further developments in the new polynomial-time algorithm for linear programming”, Talk given at ORSA/TIMS National Meeting, Boston, MA, USA April (1985).
- [22] L. G. Khachian. “A polynomial algorithm for linear programming,” *Doklady, Akademia Nauk URSS*, 244, pp 1093-1096 (1979).
- [23] V. L. Klee and G. J. Minty. “How good is the simplex algorithm?”, in O. Shisha (ed), *Inequalities III*, Academic Press, London and New York, pp 159-175 (1972).
- [24] K.O Kortanek and M. Shi. “Convergence results and numerical experiments on a linear programming hibrid algorithm”, *European Journal of Operations Research*, Vol.32, pp 47-61, (1987).
- [25] W. Mascarenhas. “The affine scaling algorithm fails for $\lambda = 0.999$.” Technical Report, Universidade Estadual de Campinas, Campinas. SP. Brazil (1993).
- [26] N. Megiddo and M. Shub. Boundary Behavior of interior point algorithms in linear programming. *Mathematics of Operation Reserch*, vol 14. No 1:97–146, 1989.
- [27] R. Monteiro and T. Tsuchiya. “A simplifued global convergence proof of the affine scaling algorithm”, *Annals of Operation Research*, Vol. 47, pp: 443-482 (1993)
- [28] R. Monteiro and Y. Wang. “Trust region affin scaling algorithms for linearly constrained convex and concave programs”, Manuscript, School of Industrial and Systems Engeneering, Georgia Institute of Technology. Atlanta. USA. (1995).

- [29] D. P. O’Leary. “On bounds for scaled projections and pseudoinverses”, *Linear algebra and its applications*, 132, pp 115-117, (1990).
- [30] J. Renegar. “Incorporating condition measures into the complexity theory of linear programming,” *SIAM Journal on Optimization*, Vol. 5, No 3, pp 506-524, August (1995)
- [31] J. Renegar. “Some perturbation Theory for linear programming ”, *Mathematical Programming*, No 65, (1), pp 73-91, (1994).
- [32] J. Renegar. “A polynomial-time algorithm based on Newton’s method for linear programming. Report MSRI 07 118-86. Mathematical Sciences Research Institute, Berkeley, California, June (1986).
- [33] G. W. Stewart. On scaled projections and pseudoinverses. *Linear Algebra and its Applications*, 112:189–193, 1989.
- [34] A. Schrijver. *Theory of Linear and Integer Programming*. (Wiley. New York, 1986)
- [35] M. J. Todd. A Dantzig-Wolfe-like variant of Karmarkar’s interior-point linear programming algorithm. *Operations Research*, 38:1006–1018, 1990.
- [36] M. Todd and B. Burrell. “An extension of Karmarkar’s algorithm for linear programming using dual variables”, *Algorithmica*, 1, pp: 409-424, (1986).
- [37] P. Tseng and Z. Q. Luo. On the convergence of the affine scaling algorithm. Technical Report CICS-P-169. Center for Intelligent Control Systems, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. 1989.
- [38] T. Tsuchiya “Global convergence of the affine scaling method for degenerated linear programming problems”, *Mathematical programming*, Vol. 52, pp 377-404, (1991).

- [39] T. Tsuchiya “Global convergence property of the affine scaling method for degenerated linear programming problems”, *Mathematics of Operation Research*, Vol. 17, No 3, pp527-557 (1992).
- [40] T. Tsuchiya and M. Maramatsu. “Global convergence of a long-step affine scaling algorithm for degenerated linear programming problems” *Siam Journal on Optimization*, Vol. 5, No. 3, pp 525-551 (1995)
- [41] T. Tsuchiya. “Affine Scaling Algorithm.” T. Terlaky(Ed). Interior Point Methods of Mathematical Programming, pp 35-82. Kluwer Academic Publishers. (1996).
- [42] L. Tunçel. “On the condition numbers for polyedra in Karmarkar’s form”, Research Report CORR 97-24. Departament of Combinatorics and Optimization. Faculty of Mathematics. University of Waterloo. Waterloo Ontario. Canada. November (1997).
- [43] P. Vaidja. “An algorithm for linear programming which requires $O(((m+n)n^2 + (m+n)^{1.5}n)L)$ arithmetic operations”, Preprint, AT&T Bell Laboratories, Murray Hill, NJ,(1987)
- [44] R. J Vanderbei, M. S Meketon and B. A Freedman. “A modification of Karmarkar’s linear programming algorithm”, *Algoritmica*, Vol.1 pp 395-407 (1986)
- [45] S. A. Vavasis. Stable numerical algorithms for equilibrium systems. *SIAM J. Matrix Analysis*, 15:1108–1131, 1994.
- [46] S. A. Vavasis and Y. Ye. Condition numbers for polyhedra with real number data. Technical Report 93-1398, Dept. of Computer Sciences, Cornell University, Ithaca, NY, 1993.
- [47] S. A. Vavasis and Y. Ye. Identifying an optimal basis in linear programming. Technical report, Dept. of Computer Sciences, Cornell University, Ithaca, NY, 1993.

- [48] S. A. Vavasis and Y. Ye. An accelerated interior point method whose running time depends only on A . Technical Report 93-1391, Dept. of Computer Sciences, Cornell University, Ithaca, NY, 1993.
- [49] S. Vavasis and Y. Y Ye. "A simplification to 'A primal-dual interior point method whose runing time depends only on the constraint matrix' " To appear. (1997)
- [50] Y. Ye. On the finite convergence of interior-point algorithms for linear programming. *Mathematical Programming*, 57(1992) 325-335.
- [51] Y. Ye. "Toward probabilistic analysis of interior point algorithms for linear programming". *Mathematics of Operations Research*, Vol 19. No 1, February 1994. 38-52.
- [52] Y. Y Ye. "How partial knowledge help to solve linear programs", Department of Management Sciences, The University of Iowa, Iowa City IA 52242, (1996).