



**UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E COMPUTAÇÃO CIENTÍFICA**

MESTRADO PROFISSIONAL EM MATEMÁTICA

***ELEMENTOS DE PROGRAMAÇÃO LINEAR:
CONDIÇÕES DE OTIMALIDADE E LEMA DE FARKAS***

Ricardo Alexandre Alves Pereira

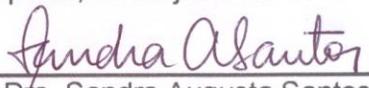
Orientadora: Prof^a Dra. Sandra Augusta Santos

CAMPINAS
JULHO – 2010

**ELEMENTOS DE PROGRAMAÇÃO LINEAR:
CONDIÇÕES DE OTIMALIDADE E LEMA DE FARKAS**

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por Ricardo Alexandre Alves Pereira e aprovada pela comissão julgadora.

Campinas, 29 de julho de 2010



Prof^ª. Dra. Sandra Augusta Santos
Orientadora

Banca Examinadora:

Prof^ª. Dra. Sandra Augusta Santos (UNICAMP)

Prof. Dr. Wesley Vagner Inês Shirabayashi (UEM)

Prof. Dr. Cristiano Torezzan (UNICAMP)

Dissertação apresentada ao Instituto de Matemática, Estatística e Computação Científica, UNICAMP, como requisito parcial para obtenção do Título de Mestre em Matemática.

**FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO IMECC DA UNICAMP**

Bibliotecária: Silvania Renata de Jesus Ribeiro Cirilo – CRB8 / 6592

Pereira, Ricardo Alexandre Alves

P414e Elementos de programação linear: condições de otimalidade e lema de Farkas/Ricardo Alexandre Alves Pereira-- Campinas, [S.P. : s.n.], 2010.

Orientador : Sandra Augusta Santos

Dissertação (mestrado) - Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

1. Otimização. 2. Lagrange, Multiplicadores de. 3. Dualidade (Matemática).. I. Santos, Sandra Augusta. II. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. III. Título.

Título em inglês: Elements of linear programming: optimality conditions and Farkas' lemma.

Palavras-chave em inglês (Keywords): 1. Optimization. 2. Lagrange multipliers. 3. Duality.

Área de concentração: Matemática Aplicada

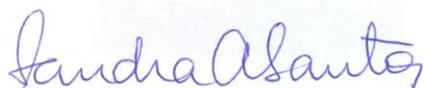
Titulação: Mestre em Matemática

Banca examinadora: Profa. Dra. Sandra Augusta Santos (IMECC-UNICAMP)
Prof. Dr. Wesley Vagner Ines Shirabayashi (UEM)
Prof. Dr. Cristiano Torezzan (FCA - UNICAMP)

Data da defesa: 29/07/2010

Programa de Pós-Graduação: Mestrado Profissional em Matemática

**Dissertação de Mestrado Profissional defendida em 29 de julho de 2010 e
aprovada pela Banca Examinadora composta pelos Profs. Drs.**



Prof. (a). Dr (a). SANDRA AUGUSTA SANTOS



Prof. (a). Dr (a). CRISTIANO TOREZZAN



Prof. (a). Dr (a). WESLEY VAGNER INES SHIRABAYASHI

Para meu pai
Paulo Alves Pereira (i.m.)

Agradecimentos

Agradeço em especial à minha esposa Gabriela que foi quem primeiro me incentivou e acreditou no meu potencial para fazer o mestrado, pelo seu carinho, compreensão e paciência durante todo o curso.

À minha mãe, meus irmãos e minha irmã que se fizeram presentes em vários momentos dessa caminhada.

Aos amigos Ricardo, Gabriel e Eduardo pela ajuda e incentivo que me dedicaram.

À professora Sandra por sua orientação, motivação e grande apoio para o desenvolvimento desse estudo.

Aos coordenadores, professores e monitores do Mestrado Profissional em Matemática pelos ensinamentos e oferecimento desse programa, o qual cumpre com uma das funções fundamentais da universidade pública que é a formação de cidadãos e profissionais qualificados e capacitados para a construção de uma sociedade melhor.

E a todos àqueles que de alguma forma colaboraram para esta conquista.

“Não se preocupe muito com as suas dificuldades em Matemática, posso assegurar-lhe que as minhas são ainda maiores.”

Albert Einstein

Resumo

Este trabalho teve por objetivo produzir um texto didático que auxiliasse no estudo e na compreensão dos Problemas de Programação Linear (PPL). Procuramos diminuir o “degrau” que existe entre o Cálculo, a Geometria e a Álgebra Linear no tratamento desses problemas, utilizando uma linguagem clara e objetiva. Dessa forma, fizemos apenas as demonstrações dos resultados que julgamos essenciais. Trabalhamos com os principais conceitos e definições que envolvem os PPL (otimização, vetor gradiente, derivada direcional, máximos e mínimos sobre conjunto compacto, Multiplicadores de Lagrange, espaço de exigência, solução ótima, dualidade entre outros) fazendo sempre que possível contextualizações através de diversas aplicações. Finalizamos este texto com o Lema de Farkas, utilizando argumentos simples e lógicos para a sua demonstração, com o uso de cálculo e da álgebra linear.

Palavras-chave: Problemas de Programação Linear; otimização; Multiplicadores de Lagrange; dualidade; Lema de Farkas.

Abstract

This study aimed to produce a didactic text which would help in the study and understanding of Linear Programming Problems (LPP). We seek to reduce the "gap" that exists between the Calculus, Geometry and Linear Algebra in the treatment of such problem using a clear and objective language. Thus, we have included only the proofs of the results that we consider essential. We work with key concepts and definitions involving PPL (optimization, gradient vector, directional derivative, maximum and minimum on a compact set, Lagrange multipliers, space requirement, optimal solution, duality, among others) including wherever possible a contextualization through various applications. We finish this text with the Farkas' Lemma, using simple and logical arguments for their demonstration with the use of calculus and linear algebra.

Keywords: Linear Programming Problems, Optimization, Lagrange multipliers, duality, Farkas' Lemma.

Sumário

Introdução	1
1. O problema de programação linear: ingredientes básicos	3
1.1 Representação	3
1.2 Estudo do conjunto viável	4
1.2.1 Análise do conjunto viável por linhas e por colunas	5
1.2.2 Classificação de um sistema linear	8
1.3 Uma visão geométrica da solução de problemas de programação linear	9
1.4 Conceitos de Cálculo relacionados ao PPL	14
1.4.1 O vetor gradiente	14
Exemplo numérico 1 - Cálculo e representação do vetor gradiente	14
1.4.2 Derivada direcional e gradiente	15
1.4.3 O Vetor gradiente e as curvas de nível	16
1.4.4 Máximos e Mínimos sobre conjunto compacto	17
1.4.5 Multiplicadores de Lagrange	18
1.5 Problemas contextualizados com solução geométrica	22
Exemplo numérico 2 – Maximizando o lucro de vendas	23
Exemplo numérico 3 - Questão 88 (Vestibular UNESP 2010)	29
Exemplo numérico 4 - Questão 89 (Vestibular UNESP 2010)	31
2. Aspectos algébricos do PPL: ampliando a análise	34
2.1 Problemas lineares sujeitos apenas a restrições de igualdade	34
2.2 Problemas lineares sujeitos a restrições gerais	37
2.3 Representação	37
2.4 O espaço de exigência	38
Exemplo numérico 5 – Ilustrando o espaço de exigência	39

2.5	Solução ótima	40
	Exemplo numérico 6 – Solução ótima finita.	41
	Exemplo numérico 7 – Problema ilimitado.	42
3.	Dualidade e o lema de Farkas	44
3.1	Dualidade	44
3.2	Representações do problema dual	49
	Exemplo numérico 8 – Dual de PPL na forma padrão.	49
	Exemplo numérico 9 – Dual de PPL na forma canônica.	51
	Exemplo numérico 10 – Dual do problema do exemplo 2	52
	Exemplo numérico 11 – O dual do dual.	53
3.3	O Lema de Farkas	55
3.4	O critério de otimalidade da programação linear: uma aplicação do Lema de Farkas	59
3.5	Uma prova do lema de Farkas	64
	Considerações finais e perspectivas futuras	77
	Referências Bibliográficas	78
	Apêndice A - Vértices (pontos extremos)	80
	Apêndice B - Conjunto Compacto	84
	Índice	86

Introdução

Muitas vezes, em indústrias, empresas e na ciência, nos deparamos com situações complexas, as quais necessitam de uma solução ótima. Por questões de simplicidade, em muitos casos, estas situações podem ser modeladas por variáveis linearmente relacionadas. Na busca de uma meta que também pode ser modelada por uma função linear, precisamos realizar uma investigação das condições existentes para a determinação da melhor solução. Nesse contexto, se inserem os Problemas de Programação Linear (PPL) que serão tratados neste texto.

O estudo dos PPL é feito na disciplina de Programação Linear, a qual faz parte do currículo dos cursos de Matemática Aplicada e Computacional, sendo oferecida normalmente no 5º semestre. No entanto, as idéias básicas desse tipo de problema são muitas vezes vistas no final do curso de Álgebra Linear. A importância dos PPL é inegável sendo que o seu conhecimento está sendo cobrado dos alunos cada vez mais cedo. Isso foi exemplificado, nesse trabalho, por duas questões presentes na primeira fase do vestibular de 2010 da UNESP (Universidade Estadual Paulista).

Procuramos apresentar de forma clara e natural a relação entre os diversos conceitos da geometria, do cálculo e da álgebra linear, os quais são necessários para a compreensão dos PPL. Acreditamos que, dessa forma, o texto tornou-se didático e agradável para o estudo, sendo que, algumas demonstrações foram deixadas para o final, nos apêndices.

Inicialmente apresentamos os principais conceitos e definições que são necessários ao estudo dos PPL. No final do Capítulo 1 enunciamos problemas contextualizados com soluções geométricas.

No Capítulo 2 aprofundamos nossa análise dos PPL fazendo um estudo de problemas lineares sujeitos apenas a restrições de igualdades e a restrições gerais, o que nos fornece condições de estudarmos o espaço de exigência e buscarmos a solução ótima do problema.

No último capítulo tratamos do conceito de dualidade, apresentando vários exemplos, os quais evidenciam as relações entre um PPL e seu dual.

Finalizamos com a demonstração do famoso Lema de Farkas que, além de reunir os conceitos estudados, é de grande importância teórica para caracterizar a solução dos Problemas de Programação Linear.

Por exemplo, se tivermos

$$5x_1 + 3x_2 \leq 2,$$

a variável de folga x_3 ($x_3 \geq 0$) pode ser introduzida para conseguirmos

$$5x_1 + 3x_2 + x_3 = 2.$$

Podemos proceder de maneira análoga com desigualdades do tipo (\geq), só que neste caso as folgas não negativas devem ser subtraídas para levar a expressão à igualdade.

Em notação matricial, podemos escrever a nossa formulação inicial como:

$$\begin{array}{ll} \text{minimizar} & z = \mathbf{c}^T \mathbf{x} \\ \text{sujeito a} & \begin{cases} \mathbf{Ax} = \mathbf{b} \\ \mathbf{x} \geq \mathbf{0} \end{cases}, \end{array}$$

onde \mathbf{A} é uma matriz $m \times n$, \mathbf{b} um vetor coluna de dimensão m , \mathbf{c} e \mathbf{x} são vetores-coluna de dimensão n e \mathbf{c}^T indica o transposto de \mathbf{c} .

As restrições a que um problema de programação linear está sujeito formam um *conjunto viável*. Geometricamente temos que cada igualdade é representada por um hiperplano enquanto, cada inequação, por um semiespaço.

Em um problema de programação linear com duas variáveis (três ou mais variáveis) teremos as restrições dadas por retas (hiperplanos) e semiplanos (semiespaços) fechados, pois os conjuntos incluem sua fronteira.

1.2 Estudo do conjunto viável

Para estudarmos o conjunto viável vamos fazer uma análise do sistema linear que surge quando trabalhamos com um problema de programação linear.

Vimos que nosso problema estará sujeito a condições que podem ser representadas da seguinte maneira:

$$\begin{cases} \mathbf{Ax} = \mathbf{b} \\ \mathbf{x} \geq \mathbf{0} \end{cases}$$

onde $\mathbf{A} \in \mathfrak{R}^{m \times n}$, $\mathbf{b} \in \mathfrak{R}^m$ e $\mathbf{x} \in \mathfrak{R}^n$.

Sabemos que o conjunto de soluções viáveis é convexo cf.[3, p. 356]. Para um resumo dos elementos básicos sobre convexidade, sugerimos por exemplo [11, p.33].

1.2.1 Análise do conjunto viável por linhas e por colunas

Inicialmente vamos analisar apenas o sistema $\mathbf{Ax} = \mathbf{b}$, com $\mathbf{A} \in \mathfrak{R}^{m \times n}$.

Utilizaremos as seguintes notações cf.[13], em que a matriz é escrita tanto por um vetor com suas linhas quanto por um vetor com suas colunas:

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}_{1\bullet}^T \\ \mathbf{a}_{2\bullet}^T \\ \vdots \\ \mathbf{a}_{m\bullet}^T \end{pmatrix} = (\mathbf{a}_{\bullet 1} \ \mathbf{a}_{\bullet 2} \ \cdots \ \mathbf{a}_{\bullet n}), \quad \begin{cases} \mathbf{a}_{i\bullet} \in \mathfrak{R}^n, \ i = 1, \dots, m \\ \mathbf{a}_{\bullet j} \in \mathfrak{R}^m, \ j = 1, \dots, n. \end{cases}$$

- Por linhas temos:

$$\mathbf{Ax} = \mathbf{b} \Rightarrow \begin{pmatrix} \mathbf{a}_{11}x_1 + \mathbf{a}_{12}x_2 + \cdots + \mathbf{a}_{1n}x_n \\ \mathbf{a}_{21}x_1 + \mathbf{a}_{22}x_2 + \cdots + \mathbf{a}_{2n}x_n \\ \vdots \\ \mathbf{a}_{m1}x_1 + \mathbf{a}_{m2}x_2 + \cdots + \mathbf{a}_{mn}x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \Rightarrow \begin{pmatrix} \mathbf{a}_{11} & \mathbf{a}_{12} & \cdots & \mathbf{a}_{1n} \\ \mathbf{a}_{21} & \mathbf{a}_{22} & \cdots & \mathbf{a}_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{a}_{m1} & \mathbf{a}_{m2} & \cdots & \mathbf{a}_{mn} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \square$$

Se $\mathbf{a}_{1\bullet}, \mathbf{a}_{2\bullet}, \dots, \mathbf{a}_{m\bullet} \in \mathfrak{R}^n$ denotam os vetores-linha de \mathbf{A} , então o produto \mathbf{Ax} , com $\mathbf{x} \in \mathfrak{R}^n$, pode ser expresso como

$$\mathbf{Ax} = \begin{pmatrix} \mathbf{a}_{1\bullet}^T \mathbf{x} \\ \mathbf{a}_{2\bullet}^T \mathbf{x} \\ \vdots \\ \mathbf{a}_{m\bullet}^T \mathbf{x} \end{pmatrix}$$

Dessa forma, um sistema linear $\mathbf{Ax} = \mathbf{b}$ de m equações em n incógnitas pode ser escrito como m equações de hiperplanos, ou seja,

$$\mathbf{a}_{i\bullet}^T \mathbf{x} = b_i, \quad i = 1, \dots, m.$$

- Por colunas temos:

$$\mathbf{Ax} = \mathbf{b} \Rightarrow \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \Rightarrow \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} x_1 + \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix} x_2 + \dots + \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix} x_n = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

Se $\mathbf{a}_{\bullet 1}, \mathbf{a}_{\bullet 2}, \dots, \mathbf{a}_{\bullet n} \in \mathfrak{R}^m$ denotam os vetores-coluna de \mathbf{A} , então o produto \mathbf{Ax} pode ser expresso como uma combinação linear destes vetores-coluna tendo como coeficientes as componentes de \mathbf{x} cf.[1, p.184], ou seja,

$$\mathbf{Ax} = \mathbf{a}_{\bullet 1}x_1 + \mathbf{a}_{\bullet 2}x_2 + \dots + \mathbf{a}_{\bullet n}x_n.$$

Em forma de somatório, temos

$$\mathbf{Ax} = \sum_{j=1}^n \mathbf{a}_{\bullet j}x_j.$$

Assim, um sistema linear $\mathbf{Ax} = \mathbf{b}$ de m equações em n incógnitas pode ser escrito como

$$\sum_{j=1}^n \mathbf{a}_{\bullet j}x_j = \mathbf{b}.$$

Incluindo agora a condição $\mathbf{x} \geq 0$, presente no conjunto viável do PPL temos:

- Por linhas

$$\mathbf{a}_i^T \mathbf{x} = b_i, \quad i = 1, \dots, m$$

$$\mathbf{x} \geq 0$$

que correspondem a hiperplanos e semiespaços, os quais podem formar um *poliedro* que também pode ser um *politopo* cf.[2, p.67]. Vejamos a seguir as definições destes conjuntos:

Definição 1.1

Denominamos de *poliedro* à intersecção de um número finito de semiespaços.

Definição 1.2

Quando o conjunto que forma o poliedro é uma região limitada dizemos que esse poliedro é um *politopo*.

- Por colunas

$$\sum_{j=1}^n \mathbf{a}_j x_j = \mathbf{b}$$

$$x_j \geq 0, \quad j = 1, \dots, n.$$

Os vetores coluna de \mathbf{A} geram um *cone convexo* ao qual \mathbf{b} pertence.

Definição 1.3

Um conjunto $D \subset \mathfrak{R}^n$ é um *cone convexo* quando $d \in D \Rightarrow \alpha d \in D, \forall \alpha \in \mathfrak{R}_+$.

Em outras palavras, a propriedade que caracteriza um *cone convexo* é que este tipo de conjunto contém todos os múltiplos não nulos e não negativos de qualquer de seus elementos.

1.2.2 Classificação de um sistema linear

Podemos classificar um sistema linear levando em consideração o seu número de incógnitas, o posto de sua matriz de coeficientes e da sua matriz ampliada cf.[3, p.45].

Tomando um sistema de m equações e n incógnitas e sua representação matricial $(\mathbf{A} | \mathbf{b})$, para acharmos o seu posto (p) devemos encontrar primeiro sua matriz-linha reduzida à forma escada, e depois contar suas linhas não nulas. Este número é o posto (p) de \mathbf{A} .

Chamando de p_c o posto da matriz dos coeficientes e p_a o da matriz ampliada $(\mathbf{A} | \mathbf{b})$, teremos os seguintes casos:

- $p_c = p_a \Rightarrow$ *Sistema Possível*
 - se $p_c = p_a = n \Rightarrow$ *Determinado (SPD)*
 - se $p_c = p_a < n \Rightarrow$ *Indeterminado (SPI)*
- $p_c \neq p_a \Rightarrow$ *Sistema Impossível (SI)*

Assim,

- (i) para um *Sistema Possível e Determinado (SPD)* \Rightarrow solução única;
- (ii) para um *Sistema Possível e Indeterminado (SPI)* \Rightarrow infinitas soluções e
- (iii) para um *Sistema Impossível (SI)* \Rightarrow nenhuma solução.

Para os problemas de programação linear é razoável considerar que o sistema linear $\mathbf{Ax} = \mathbf{b}$ de m equações em n incógnitas possui $p_c = p_a = m < n$, que corresponde à situação (ii), ou seja, teremos infinitas soluções, o que significa que o conjunto viável contém infinitos pontos. A tarefa de resolver o PPL consiste em encontrar qual, desses infinitos pontos, otimiza a função objetivo.

Vamos estudar com mais detalhes o caso (ii) logo em seguida.

1.3 Uma visão geométrica da solução de problemas de programação linear

No caso (ii), quando $p_c = p_a = m < n$ o sistema linear $\mathbf{Ax} = \mathbf{b}$ pode ser reescrito equivalentemente da seguinte maneira cf.[11, p.44]:

$$\bar{\mathbf{A}}\mathbf{x} = \bar{\mathbf{b}} \Rightarrow \begin{pmatrix} x_1 & & + \bar{a}_{1,m+1}x_{m+1} + \cdots + \bar{a}_{1,n}x_n \\ & x_2 & + \bar{a}_{2,m+1}x_{m+1} + \cdots + \bar{a}_{2,n}x_n \\ & & \ddots & \vdots & \vdots \\ & & & x_m & + \bar{a}_{m,m+1}x_{m+1} + \cdots + \bar{a}_{m,n}x_n \end{pmatrix} = \begin{pmatrix} \bar{b}_1 \\ \bar{b}_2 \\ \vdots \\ \bar{b}_m \end{pmatrix} \square$$

As variáveis x_1, \dots, x_m são chamadas básicas e as variáveis x_{m+1}, \dots, x_n são chamadas de não básicas.

O fato de considerarmos que as m primeiras variáveis são básicas não afeta a generalidade, pois é sempre possível reordenar as colunas de A desta maneira.

As variáveis básicas podem assumir diferentes valores em função das variáveis não básicas. Isso faz com que o conjunto solução do sistema que representamos por essa matriz tenha infinitos valores.

Acrescentando as restrições $x_1, \dots, x_n \geq 0$ teremos o conjunto viável de nosso problema, dentro do qual procuramos otimizar a função objetivo dada por:

$$f(\mathbf{x}) = \mathbf{c}^T \mathbf{x}.$$

Essa função representa a busca da situação ótima do problema. Essa ocorrerá, se existir, em um dos vértices do conjunto viável (estabelecemos esse fato na sequência).

Apresentamos a seguir a definição de *região poliedral convexa fechada*, a definição de *vértices* cf.[3, p.358] e enunciamos o *Teorema Fundamental da Programação Linear* cf.[3, p.368], que é o principal resultado da Programação Linear.

Definição 1.4

Uma *região poliedral convexa fechada* em \mathfrak{R}^n é a interseção de uma quantidade finita de semiespaços fechados do \mathfrak{R}^n .

Definição 1.5

Dada uma região poliedral convexa fechada do \mathfrak{R}^n (determinada por um sistema de inequações lineares), os *vértices* dessa região são os pontos da região que satisfazem um dos possíveis sistemas de n equações lineares independentes, obtidas substituindo-se as desigualdades por igualdades.

Observação: Depois de resolver um sistema, a fim de verificar se o ponto está na região, testamos para ver se ele satisfaz todas as desigualdades. Um estudo completo deste tipo de problema pode ser encontrado na dissertação de *Monticelli* (ver [14]).

Teorema 1.1

Seja $f(x_1, \dots, x_n) = c_1 x_1 + \dots + c_n x_n$ definida numa região poliedral convexa G do \mathfrak{R}^n . Suponha que f assuma um valor máximo (mínimo) nesta região. Então, se G possui vértice(s), este valor máximo (mínimo) será assumido num vértice.

Demonstração: [ver Apêndice A](#).

Vamos ilustrar essas ideias cf.[2, p.14], trabalhando com um sistema de inequações de duas variáveis em que o conjunto solução é um *conjunto compacto* ([ver Apêndice B](#)), que forma uma *região poliedral convexa fechada*.

Seja o problema de programação linear:

$$\text{minimizar } z = c_1x_1 + c_2x_2$$

$$\text{sujeito a } \begin{cases} \mathbf{Ax} = \mathbf{b} \\ \mathbf{x} \geq \mathbf{0}. \end{cases}$$

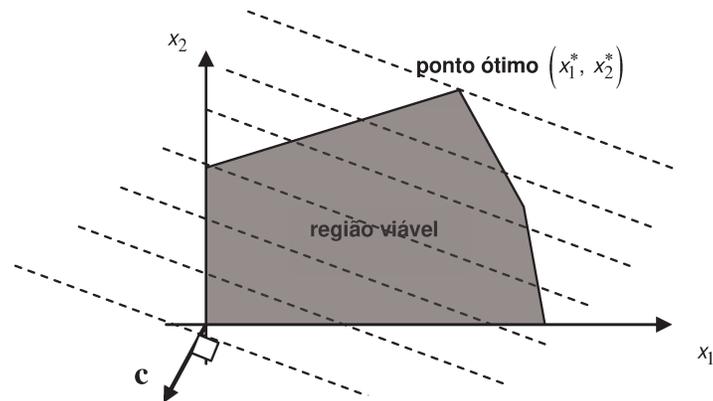


Figura 1 - Solução Geométrica

Temos na representação gráfica da Figura 1 a região viável que consiste em todos os pontos que satisfazem $\mathbf{Ax} = \mathbf{b}$ e $\mathbf{x} \geq \mathbf{0}$. Para a solução de nosso problema, desejamos o ponto que minimiza o valor da função objetivo z .

Sendo z da forma $z = c_1x_1 + c_2x_2$ (o seu gráfico é um plano em \mathfrak{R}^3), procuramos o menor valor que z assume, tendo como domínio a região viável.

Para que a função objetivo assumo o seu valor mínimo, analisando a família de retas paralelas $z = k$, curvas de nível, cf. [10, p.13], devemos achar o ponto \mathbf{x}^* , pertencente ao conjunto viável, da curva de nível com menor valor z . Nesse caso, o ponto \mathbf{x}^* é um dos vértices do polígono formado pela região viável, que será alcançado quando deslocarmos o máximo possível na direção $-\mathbf{c}$ que é perpendicular às curvas de nível (veremos o porquê disso na seção 1.4).

Analogamente, para problemas de maximização, devemos buscar o ponto da reta $z = k$ (curva de nível) na direção \mathbf{c} em que z assume o maior valor dentro de seu domínio.

No exemplo ilustrado na Figura 1 temos uma única solução ótima. Existem outros casos possíveis que podem ocorrer, dependendo da estrutura do problema.

Para problemas de minimização, os casos são:

1. *Única Solução Ótima Finita.*

Quando a solução ótima finita é única, ela ocorrerá em um vértice (*ponto extremo*). Na Figura 2 temos em (a) um exemplo desse caso em uma região limitada e em (b) em uma região ilimitada.

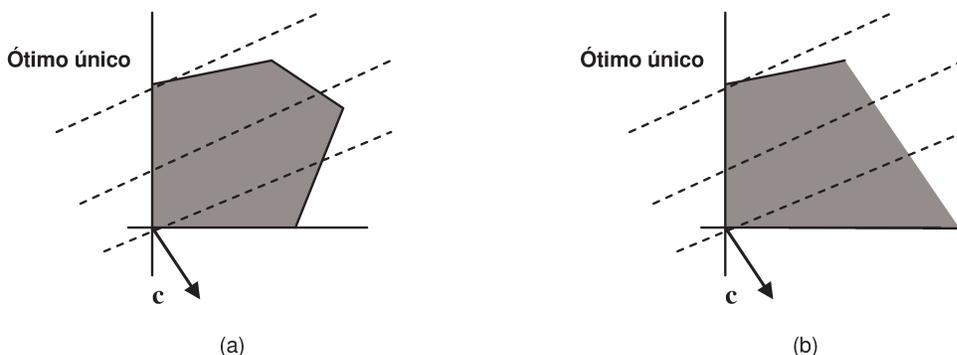


Figura 2 - Única solução ótima finita: (a) Região limitada. (b) Região ilimitada.

2. *Solução Ótima Alternativa.*

Nesse caso temos um conjunto infinito de soluções ótimas, sendo possível escolher alternativamente qualquer uma delas para que se obtenha a solução do problema. Na figura 3(a) está ilustrado este caso para uma região viável limitada. Nessa situação, temos como soluções os vértices x_1^* e x_2^* bem como qualquer ponto do segmento que os une ([ver Apêndice A](#)). Já na figura 3(b) temos uma região viável ilimitada, em que a solução pode ser qualquer ponto da semirreta, indicada na figura, com origem em x^* , e denominada direção extrema (cf.[2, p.66]).

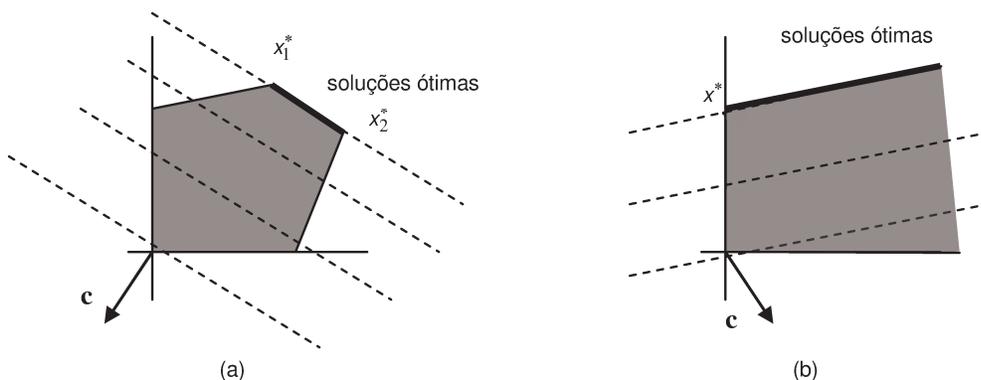


Figura 3 - Solução ótima alternativa: (a) Região limitada. (b) Região ilimitada.

3. Solução Ótima Inexistente.

Este caso, ilustrado na Figura 4, ocorre quando a região viável é ilimitada e a solução ótima é infinita. Para um problema de minimização, as curvas de nível $z = k$ podem ser obtidas na direção $-c$ ilimitadamente, sempre interceptando a região viável.

Neste caso o valor ótimo é menos infinito e não existe solução ótima.

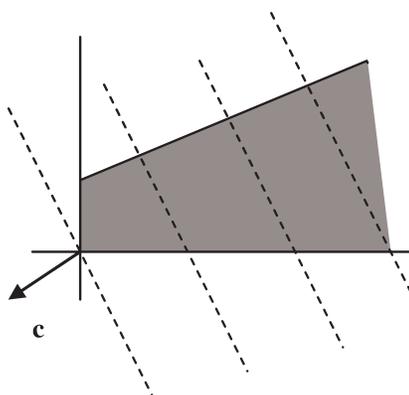


Figura 4 - Solução ótima inexistente

4. Região Viável Vazia.

Nesse caso, ilustrado na Figura 5, o sistema de equações ou inequações que definem a região viável é inconsistente, ou seja, não existem pontos que satisfazem simultaneamente essas equações ou inequações.

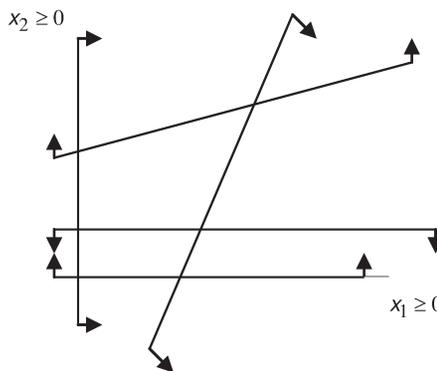


Figura 5 - Região viável vazia

Obs.: O vetor c corresponde ao vetor gradiente da função objetivo, como veremos a seguir (Seção 1.4.1), é sempre perpendicular às curvas de nível.

Para um melhor entendimento desse fato e em preparação à análise das condições de otimalidade do PPL, vamos retomar alguns conceitos do cálculo, como: gradiente, derivada direcional, máximos e mínimos sobre conjuntos compactos e multiplicadores de Lagrange.

Trabalharemos com funções de duas variáveis para que seja imediata a visualização geométrica. Isso, no entanto, não impede a generalização dos conceitos estudados para funções com três ou mais variáveis.

1.4 Conceitos de Cálculo relacionados ao PPL

1.4.1 O vetor gradiente

Vamos apresentar o vetor gradiente cf.[7, vol. 2, p.207].

Seja $z = f(x, y)$ uma função que admite derivadas parciais em (x_0, y_0) .

O vetor

$$\nabla f(x_0, y_0) = \left(\frac{\partial f}{\partial x}(x_0, y_0), \frac{\partial f}{\partial y}(x_0, y_0) \right)^T$$

denomina-se *gradiente* de f em (x_0, y_0) .

Geometricamente, interpretaremos $\nabla f(x_0, y_0)$ como um vetor aplicado no ponto (x_0, y_0) . O *gradiente* de f em (x_0, y_0) nada mais é que a derivada de $f(x, y)$ em (x_0, y_0) .

Assim, para $f(\mathbf{x}) = \mathbf{c}^T \mathbf{x}$, temos $\nabla f(\mathbf{x}) = \mathbf{c}$.

Exemplo numérico 1 - Cálculo e representação do vetor gradiente

Seja $f(x, y) = 2x^2 + y^3$. Calcule $\nabla f(1, 1)$ e represente-o geometricamente.

Solução

$$\nabla f(x, y) = \left(\frac{\partial f}{\partial x}(x, y), \frac{\partial f}{\partial y}(x, y) \right)^T = (4x, 3y^2)^T.$$

Portanto, $\nabla f(1, 1) = (4, 3)^T = 4\vec{i} + 3\vec{j}$, conforme ilustrado na Figura 6.

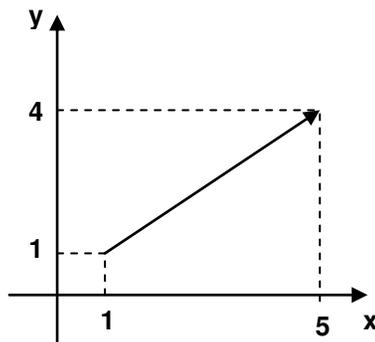


Figura 6 - Gradiente da função f no ponto $(1, 1)$.

Verificamos que o vetor gradiente é representado no domínio da função dada.

■

1.4.2 Derivada direcional e gradiente

Se f for diferenciável em (x_0, y_0) , então f admitirá *derivada direcional* em todas as direções, no ponto (x_0, y_0) e cf.[7, vol. 2, p.262]:

$$\frac{\partial f}{\partial \vec{u}}(x_0, y_0) = \nabla f(x_0, y_0) \cdot \vec{u}$$

com $\vec{u} = (a, b)$ um vetor unitário.

Como, $\frac{\partial f}{\partial \vec{u}}(x_0, y_0) = \nabla f(x_0, y_0) \cdot \vec{u} = \|\nabla f(x_0, y_0)\| \cdot \|\vec{u}\| \cos \theta$, com $\vec{u} = (a, b)$ um vetor unitário e θ o ângulo entre os vetores ∇f e \vec{u} , a direção e sentido que \vec{u} deve tomar para que f cresça mais rapidamente é a do vetor gradiente, pois $\|\vec{u}\| = 1$ e o valor

máximo que $\cos \theta$ pode assumir também é igual a 1. Analogamente, o sentido que se deve tomar para que f decresça mais rapidamente é o oposto do vetor gradiente.

Essa propriedade é de grande importância para compreendermos o porquê de caminharmos na direção do vetor gradiente quando procuramos o valor máximo (mesmo sentido) ou o valor mínimo (sentido oposto) da função objetivo de um problema de programação linear.

Portanto, se considerarmos uma função de duas variáveis f e um ponto $P(x_0, y_0)$ em seu domínio, vimos que o vetor gradiente $\nabla f(x_0, y_0)$ indica a direção e sentido de maior crescimento da função f a partir de (x_0, y_0) . Quando nos movemos a partir de P permanecendo na superfície de nível $f(P)$, o valor da função f não se altera. É o movimento na direção ortogonal à superfície de nível que nos dará o maior aumento, ou a maior diminuição, para f .

1.4.3 O Vetor gradiente e as curvas de nível

A derivada direcional $\frac{\partial f}{\partial \vec{u}}(x_0, y_0) = \nabla f(x_0, y_0) \cdot \vec{u} = \|\nabla f(x_0, y_0)\| \cdot \|\vec{u}\| \cos \theta$ é nula se, e somente se, $\cos \theta = 0$, ou seja, se $\theta = 90^\circ$. Isto é, se \vec{u} é perpendicular ao vetor $\nabla f(x_0, y_0)$.

Como a derivada direcional na direção da tangente à curva de nível é nula, o vetor gradiente deve ser perpendicular à tangente à curva de nível em um ponto qualquer $P(x_0, y_0)$ (Figura 7).

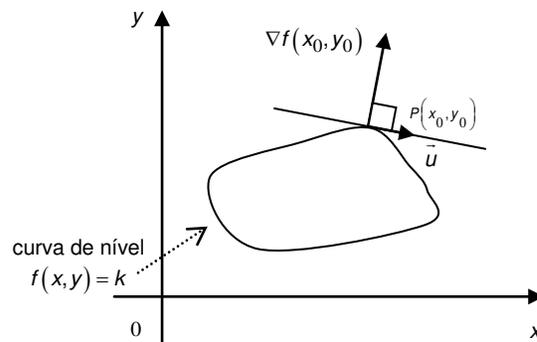


Figura 7 - O vetor gradiente é perpendicular à tangente à curva de nível.

Como para os problemas de programação linear formados por funções de duas variáveis as curvas de nível são retas, teremos o vetor gradiente sempre perpendicular a essas. Já nos casos de problemas com funções de mais de duas variáveis o gradiente será sempre ortogonal (normal) às superfícies de nível.

1.4.4 Máximos e Mínimos sobre conjunto compacto

Conforme [7, vol. 2, p.317] e [16, p.956] para uma função f de uma variável, o *Teorema de Weierstrass* diz que sendo $f(x)$ contínua em $[x_1, x_2]$, então f assume um valor mínimo absoluto e um valor máximo absoluto neste intervalo. Determinamos esses valores calculando f não somente nos pontos críticos, mas também nos extremos do intervalo $[x_1, x_2]$, escolhendo o menor e o maior valor dentre todos os candidatos.

Para funções de duas variáveis esta ideia se estende de forma parecida.

Teorema 1.2 (de Weierstrass)

Se $f(x, y)$ for contínua no compacto C , então existirão pontos (x_1, y_1) e (x_2, y_2) em C tais que, para todo (x, y) em C ,

$$f(x_1, y_1) \leq f(x, y) \leq f(x_2, y_2).$$

Como vimos na Seção 1.3, os problemas de programação linear podem ser divididos em diferentes casos. No caso do conjunto viável ser compacto os pontos de máximo absoluto e de mínimo absoluto existem e estão localizados na fronteira do domínio de f .

Assim, para encontrarmos um máximo ou mínimo absoluto de $f(x,y)$, nas condições acima, devemos:

- Estabelecer os valores extremos de f na fronteira de C , via multiplicadores de Lagrange (seção 1.4.5)
- O maior dos valores encontrados será o máximo absoluto e o menor, o mínimo absoluto.

1.4.5 Multiplicadores de Lagrange

Embora o *Método dos Multiplicadores de Lagrange* não permita resolver o problema de programação linear, vamos apresentá-lo na sequência como uma ferramenta preliminar da otimização restrita, para resolver problemas com restrições somente de igualdade. Além de enfatizar as interpretações geométricas e algébricas dos multiplicadores, esta seção visa complementar os aspectos teóricos do nosso texto.

Esse método consiste na identificação do valor máximo e do valor mínimo de uma função em uma região, determinada pelas restrições de igualdade impostas ao problema.

Para apresentar as ideias do método vamos trabalhar com curvas em \mathfrak{R}^2 e superfícies e curvas em \mathfrak{R}^3 . As regiões são, portanto, conjuntos do tipo:

$$\{(x,y) : g(x,y) = 0\}, \{(x,y,z) : g(x,y,z) = 0\} \text{ e} \\ \{(x,y,z) : g(x,y,z) = 0 \text{ e } h(x,y,z) = 0\},$$

com $g : \mathfrak{R}^2 \rightarrow \mathfrak{R}$ no primeiro caso, $g, h : \mathfrak{R}^3 \rightarrow \mathfrak{R}$ no segundo e terceiro casos cf.[7, vol. 2, p.322].

Vamos primeiro explicar a base geométrica do método de Lagrange para as funções de duas variáveis cf.[16, p.963]. Seja $f(x,y)$ sujeita à restrição da forma $g(x,y)=k$. Estamos interessados nos valores extremos de $f(x,y)$ quando o ponto $(x,y) \in g$ pertence à curva de nível $g(x,y)=k$. Na figura 8 representamos essa curva de nível com várias curvas de nível da função f . A família de curvas de nível é representada por $f(x,y)=t$, onde $t=6, 7, 8$ e 9 . Para maximizarmos $f(x,y)$ sujeita a $g(x,y)=k$ devemos achar qual o maior valor de t tal que a curva de nível $f(x,y)=t$ intercepte $g(x,y)=k$. Isso ocorre quando essas curvas se tocam, ou seja, quando essas curvas têm uma reta tangente em comum (caso contrário, poderíamos aumentar o valor de t). Isso significa que as retas normais ao ponto (x_0, y_0) onde as duas curvas se tocam devem ser as mesmas. Logo os vetores gradientes são paralelos, ou seja, se $\nabla g(x_0, y_0) \neq 0$ então para algum escalar λ temos:

$$\nabla f(x_0, y_0) = \lambda \nabla g(x_0, y_0).$$

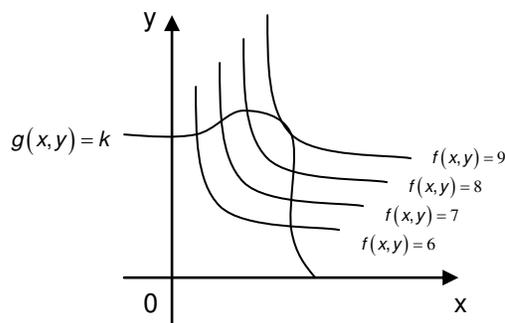


Figura 8 - Curvas de nível

Agora vamos utilizar esse argumento para o problema de achar os valores extremos de $f(x,y,z)$ sujeita à restrição $g(x,y,z)=k$. Assim, o ponto (x,y,z) está restrito a pertencer à superfície com equação $g(x,y,z)=k$. Em vez das *curvas de nível* da Figura 8, devemos considerar as *superfícies de nível* de $f(x,y,z)$ e argumentar que, se o valor máximo de f é $f(x_0, y_0, z_0) = t$, então a superfície de nível

$f(x, y, z) = t$ é tangente à superfície de nível $g(x, y, z) = k$, e então os correspondentes gradientes são paralelos.

Portanto, se $\nabla g(x_0, y_0, z_0) \neq 0$, existe um escalar λ tal que:

$$\nabla f(x_0, y_0, z_0) = \lambda \nabla g(x_0, y_0, z_0).$$

Resumindo

- Os *candidatos a extremantes locais* de $f(x, y)$ sujeita à restrição da forma $g(x, y) = k$ são os pontos (x, y) que tornam compatível o sistema

$$\begin{cases} \nabla f(x, y) = \lambda \nabla g(x, y) \\ g(x, y) = k. \end{cases}$$

- Os *candidatos a extremantes locais* de $f(x, y, z)$ sujeita à restrição da forma $g(x, y, z) = k$ são os pontos (x, y, z) que tornam compatível o sistema

$$\begin{cases} \nabla f(x, y, z) = \lambda \nabla g(x, y, z) \\ g(x, y, z) = k. \end{cases}$$

Após calcular todos os pontos e seus respectivos valores de f , em cada caso, o maior desses valores será o valor máximo de f , e o menor será o valor mínimo de f .

Até o momento trabalhamos com apenas uma restrição, veremos como fica se a função $f(x, y, z)$ estiver sujeita a duas restrições.

Nesse caso cf.[16, p.967] vamos supor que as condições são da forma $g(x, y, z) = k$ e $h(x, y, z) = w$, ou seja, duas superfícies em \mathfrak{R}^3 que interceptadas formam uma curva em \mathfrak{R}^3 .

Analisando geometricamente, estamos procurando pelos valores extremos de f quando (x, y, z) está restrita a pertencer à curva E , obtida pela interseção das superfícies de nível $g(x, y, z) = k$ e $h(x, y, z) = w$. Se f tiver um valor extremo no ponto $P(x_0, y_0, z_0)$, sabemos que ∇f é ortogonal a E , o vetor ∇g é ortogonal a $g(x, y, z) = k$ e o vetor ∇h é ortogonal a $h(x, y, z) = w$, ou seja, ∇g e ∇h são ambos ortogonais a E ,

sendo todos os gradientes calculados em (x_0, y_0, z_0) . Portanto, o vetor gradiente $\nabla f(x_0, y_0, z_0)$ pertence ao plano determinado por $\nabla g(x_0, y_0, z_0)$ e $\nabla h(x_0, y_0, z_0)$ (Figura 9).

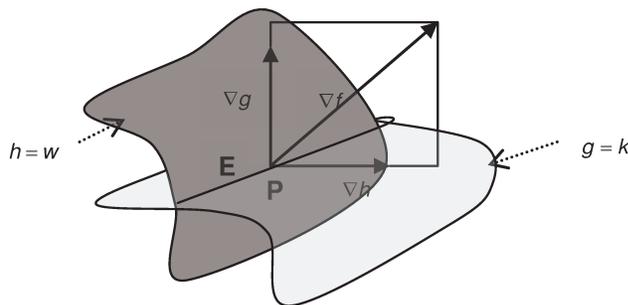


Figura 9 - Curva E obtida pela interseção das superfícies de nível $g(x, y, z) = k$ e $h(x, y, z) = w$ e os vetores ∇g , ∇h e ∇f .

Supondo que os vetores gradientes ∇g e ∇h em (x_0, y_0, z_0) não são paralelos nem nulos, existem escalares λ e μ (multiplicadores de Lagrange), tais que

$$\nabla f(x_0, y_0, z_0) = \lambda \nabla g(x_0, y_0, z_0) + \mu \nabla h(x_0, y_0, z_0).$$

Assim, queremos resolver:

$$(1) \quad \begin{cases} \nabla f(x, y, z) = \lambda \nabla g(x, y, z) + \mu \nabla h(x, y, z) \\ g(x, y, z) = k \\ h(x, y, z) = w. \end{cases}$$

Do conjunto solução desse sistema calcula-se os respectivos valores de f . O maior desses valores será o valor máximo de f , e o menor será o valor mínimo de f .

Embora o processo visto tenha sido apresentado apenas para problemas com funções de duas ou três variáveis, para problemas com funções de mais variáveis, a extensão perde em termos da visualização geométrica, mas é praticamente imediata do ponto de vista algébrico. No caso geral o sistema não linear (1) passa a ter tantos multiplicadores quantas forem as restrições de igualdade. Suas soluções são pontos em que o gradiente da função objetivo possa ser escrito como combinação linear dos gradientes das restrições, tendo os multiplicadores de Lagrange como coeficientes. Além disso, a viabilidade precisa ser satisfeita, portanto, os pontos desejados precisam

cumprir as equações que definem as restrições do problema. Para garantir que o sistema (1) possua solução em problemas não lineares gerais são necessárias hipóteses sobre o conjunto viável. Uma hipótese que aparece com frequência na literatura é a independência linear dos vetores gradientes das restrições.

1.5 Problemas contextualizados com solução geométrica

Uma ferramenta muito utilizada para a resolução de PPL é o *Método Simplex*. Desenvolvido por George Dantzig em 1947, esse é um procedimento algébrico e iterativo que fornece a solução exata do PPL não degenerado em um número finito de iterações. É também capaz de indicar se o problema tem solução ilimitada, se não tem solução ou se possui infinitas soluções.

Estas características do simplex permitiram sua codificação em programas extremamente rápidos e eficientes, possibilitando a solução de sistemas com milhares de variáveis. Extensões posteriores, como o simplex revisado e o princípio da decomposição, aumentaram sua capacidade para dezenas de milhares e, finalmente, centenas de milhares de variáveis. Além disto, o simplex possui uma interpretação geométrica bastante simples. De qualquer maneira, a apresentação e a análise de tal método fogem do escopo deste trabalho. Para tanto, sugerimos a leitura de [2, 9, 10].

Apresentamos a seguir um exemplo numérico extraído de [15, p.41]. Por possuir duas variáveis, permite a obtenção da solução geometricamente. A situação problema é apresentada de maneira a permitir ainda um breve exercício de modelagem para formular o problema a ser resolvido.

Observação: Por uma questão de notação, como estamos chamando de x às variáveis dos problemas de minimização, identificaremos por y às variáveis dos problemas de maximização.

Exemplo numérico 2 – Maximizando o lucro de vendas

Uma fábrica produz dois produtos diferentes (P_1 e P_2). Cada um deles deve passar por três tipos de máquinas (M_1 , M_2 e M_3). Para fazer o produto P_1 , as máquinas do tipo M_1 devem trabalhar 2 horas e as máquinas dos tipos M_2 e M_3 , 1 hora cada. Para fazer o produto P_2 , as máquinas do tipo M_1 e M_2 devem trabalhar 1 hora cada e as máquinas do tipo M_3 , 3 horas. Sabe-se que a fábrica tem um lucro de 40 reais por unidade do produto P_1 e 60 reais por unidade do produto P_2 . Considerando que a fábrica vende toda a sua produção e que as máquinas dos tipos M_1 , M_2 e M_3 podem trabalhar no máximo 70, 40 e 90 horas por semana, respectivamente, qual quantidade de cada tipo de produto deve ser produzida por semana, para que seu lucro seja máximo?

Solução

Inicialmente vamos organizar os dados do problema na Tabela 1.

Tipo de máquina	Horas necessárias para produzir uma unidade de P_1	Horas necessárias para produzir uma unidade de P_2	Total de horas disponíveis
M_1	2	1	70
M_2	1	1	40
M_3	1	3	90
Lucro por unidade de $P_1 = 40$ reais		Lucro por unidade de $P_2 = 60$ reais	

Tabela 1 - Horas necessárias para a produção de P_1 e P_2 pelas máquinas dos tipos M_1 , M_2 e M_3 ; total de horas disponíveis por semana para cada tipo de máquina e lucro por unidade de P_1 e P_2 .

Agora, vamos formular este problema matematicamente.

Chamemos de y_1 a quantidade produzida de P_1 e y_2 a quantidade produzida de P_2 , por semana. Como o lucro por unidade produzida de P_1 é 40 reais e por unidade produzida de P_2 é 60 reais, o lucro total, em reais, é dado pela função

$$f(y_1, y_2) = 40y_1 + 60y_2.$$

Se forem fabricadas y_1 unidades do produto P_1 , o tempo gasto pela máquina do tipo M_1 é $2y_1$ e se são fabricadas y_2 unidades do produto P_2 , o tempo gasto pela máquina do tipo M_1 é y_2 , ou seja, o tempo total usado por esse tipo de máquina é

$$2y_1 + y_2.$$

Analogamente, temos o seguinte tempo gasto pelas máquinas dos tipos M_2 e M_3 , respectivamente

$$y_1 + y_2 \text{ e } y_1 + 3y_2.$$

Já que cada tipo de máquina pode ser usada até 70, 40 e 90 horas por semana, respectivamente, nós devemos ter

$$2y_1 + y_2 \leq 70$$

$$y_1 + y_2 \leq 40$$

$$y_1 + 3y_2 \leq 90.$$

Além disso, como y_1 e y_2 não podem assumir valores negativos, temos que

$$y_1 \geq 0 \text{ e } y_2 \geq 0.$$

Dessa forma, chegamos à seguinte formulação matemática:

Encontrar valores de y_1 e y_2 que maximizam

$$f(y_1, y_2) = 40y_1 + 60y_2$$

sujeito às cinco restrições

$$(A) 2y_1 + y_2 \leq 70$$

$$(B) y_1 + y_2 \leq 40$$

$$(C) y_1 + 3y_2 \leq 90$$

$$(D) y_1 \geq 0$$

$$(E) y_2 \geq 0$$

que estão representadas separadamente na Figura 10.

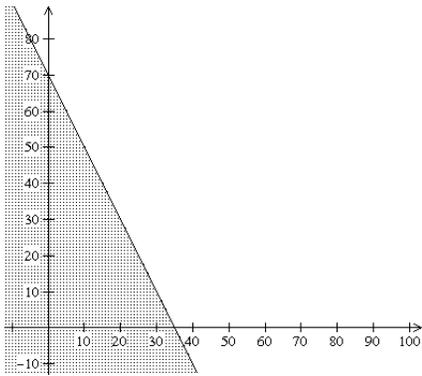


Gráfico A: $2y_1 + y_2 \leq 70$

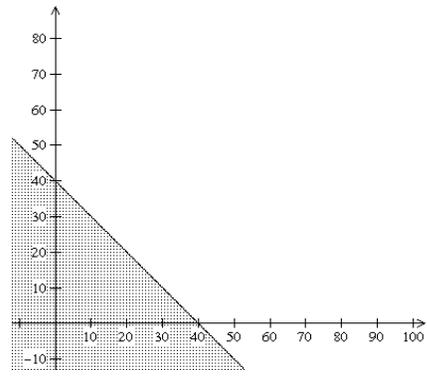


Gráfico B: $y_1 + y_2 \leq 40$

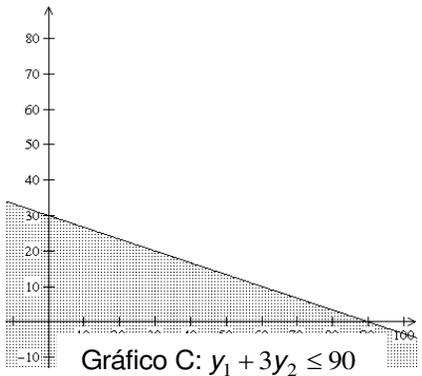


Gráfico C: $y_1 + 3y_2 \leq 90$

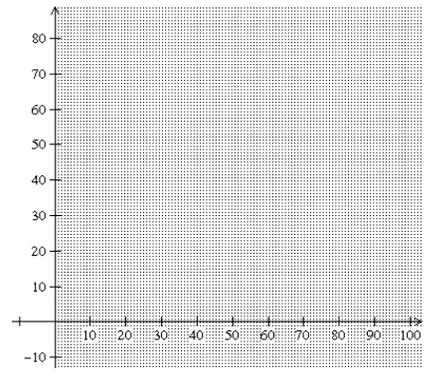


Gráfico D: $y_1 \geq 0$

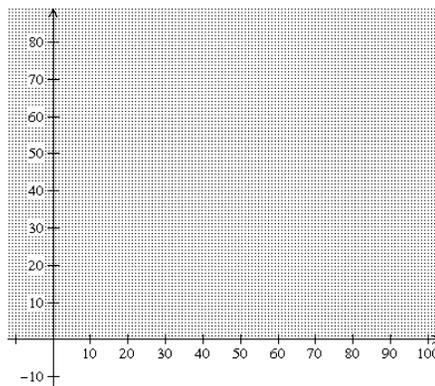


Gráfico E: $y_2 \geq 0$

Figura 10 - Gráficos A, B, C, D e E representando as regiões correspondentes às restrições do problema.

Nesses gráficos podemos verificar que a fronteira da região é determinada por apenas uma equação linear (reta). Cada reta (hiperplano) divide o

plano em dois semiplanos (semiespaços). O semiplano (semiespaço) é fechado quando contém a reta que o determina (hiperplano) e aberto quando não a (o) contém.

Podemos observar nas representações acima que as desigualdades determinam semiplanos fechados (semiespaços fechados).

A região viável de nosso exemplo consiste na intersecção das inequações (A)–(E) representada pela região da Figura 11.

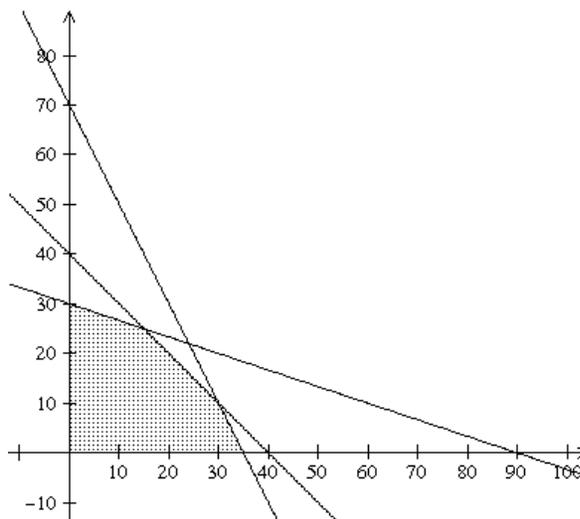


Figura 11 - Representação gráfica da região determinada pela intersecção dos gráficos A, B, C, D e E.

Observe que o conjunto viável é determinado pela intersecção de semiplanos fechados, formando uma região poliedral convexa fechada (cf. definição 1.4).

Como buscamos que o lucro dado pela função

$$f(y_1, y_2) = 40y_1 + 60y_2$$

seja máximo, estamos buscando na região destacada qual é o ponto (y_1, y_2) em que $f(y_1, y_2)$ atinge seu valor máximo.

O gráfico dessa função linear de y_1 e y_2 consiste em um plano em \mathfrak{R}^3 e suas curvas de nível são retas. À medida que nos deslocamos perpendicularmente a estas retas (direção do vetor gradiente), a função objetivo cresce (mesmo sentido do vetor gradiente) ou decresce (sentido oposto ao do vetor gradiente). Já vimos

(Teorema 1.1) que dentro de uma região viável limitada, o valor máximo e mínimo de $f(y_1, y_2)$ ocorre nos pontos extremos, ou seja, nos vértices da região poliedral formada.

Observe a região viável do problema com algumas curvas de nível e a solução ótima na Figura 12.

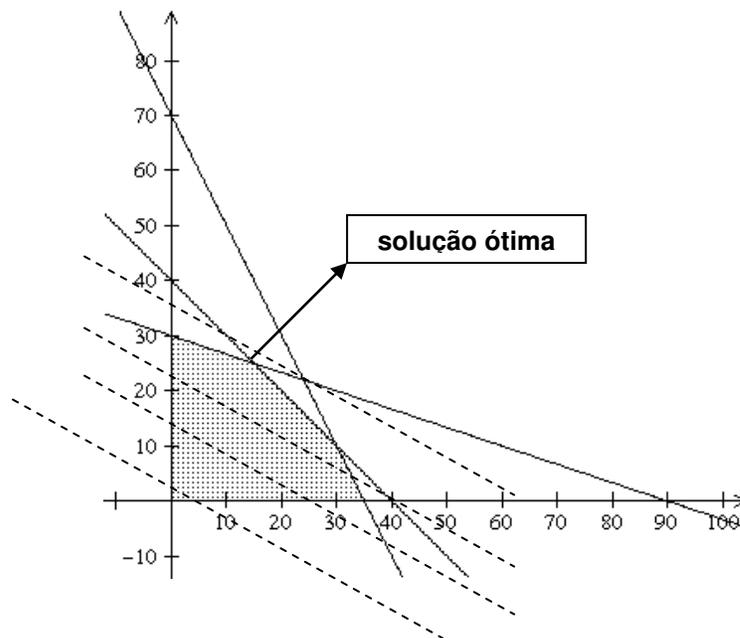


Figura 12 - Representação gráfica da região viável do problema com algumas curvas de nível.

Podemos verificar graficamente que a solução ótima realmente corresponde a um dos vértices da região poliedral que forma a região viável do problema

Agora, vamos identificar as coordenadas de cada um dos pontos extremos (vértices) e os valores correspondentes para $f(y_1, y_2)$ na Tabela 2.

Ponto Extremo (y_1, y_2)	Valor de $f(y_1, y_2)$
(0,0)	0
(0,30)	1800
(15,25)	2100
(30,10)	1800
(35,0)	1400

Tabela 2 - Coordenadas de cada um dos pontos extremos (vértices) e os valores correspondentes para $f(y_1, y_2)$.

Notamos que o maior valor de $f(y_1, y_2)$ é 2100 e a correspondente solução ótima é (15,25).

Observe que as máquinas do tipo M_2 e M_3 são usadas com capacidade plena, mas somente são usadas 55 horas do tempo disponível nas máquinas M_1 , 15 horas a menos do que sua capacidade. Dessa forma, poderiam ser vendidas algumas máquinas do tipo M_1 , sem que isso afetasse o lucro da fábrica.

Assim, o fabricante desses produtos atinge um lucro máximo de R\$ 2.100,00 quando fabrica 15 produtos P_1 e 25 produtos P_2 por semana, podendo se desfazer de algumas máquinas M_1 .

■

O conhecimento da otimização está sendo esperado do estudante cada vez mais cedo. Isso pode ser exemplificado pelo vestibular da primeira fase da UNESP (Universidade Estadual Paulista) de 2010, que apresentou duas questões envolvendo conhecimentos de otimização dentre um total de sete questões, ou seja, cerca de 30% das questões de matemática destinadas a avaliar o conhecimento adquirido pelo aluno durante a sua formação no ensino fundamental e médio envolviam este tema como elemento contextualizador.

O enunciado a seguir é comum para as duas questões mencionadas [<http://www.unesp.br/vestibular/vestibular-2010/> (06/03/2010)], que são propostas e resolvidas na sequência do texto .

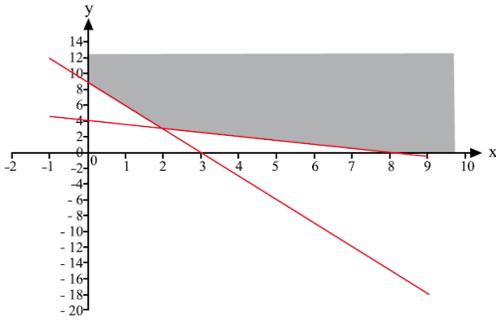
Uma fábrica utiliza dois tipos de processos, P_1 e P_2 , para produzir dois tipos de chocolates, C_1 e C_2 . Para produzir 1.000 unidades de C_1 são exigidas 3 horas de trabalho no processo P_1 e 3 horas em P_2 . Para produzir 1.000 unidades de C_2 são necessárias 1 hora de trabalho no processo P_1 e 6 horas em P_2 . Representando por x a quantidade diária de lotes de 1.000 unidades de chocolates produzidas pelo processo P_1 e por y a quantidade diária de lotes de 1.000 unidades de chocolates produzidas pelo processo P_2 , sabe-se que o número de horas trabalhadas em um dia no processo P_1 é $3x+y$, e que o número de horas trabalhadas em um dia no processo P_2 é $3x+6y$ ^{*}.

Exemplo numérico 3 - Questão 88 (Vestibular UNESP 2010)

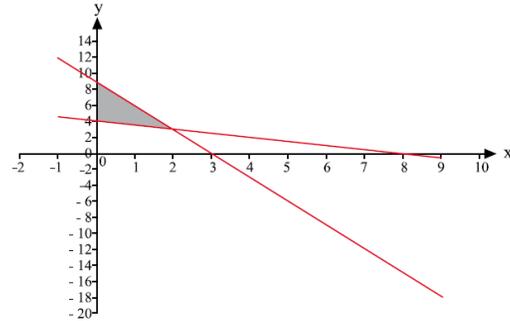
Dado que no processo P_1 pode-se trabalhar no máximo 9 horas por dia e no processo P_2 pode-se trabalhar no máximo 24 horas por dia, a representação no plano cartesiano do conjunto dos pontos (x,y) que satisfazem, simultaneamente, às duas restrições de número de horas possíveis de serem trabalhadas nos processos P_1 e P_2 , em um dia, é (Figura 13):

* Observe que o enunciado geral dos exemplos numéricos 3 e 4 apresenta dois equívocos no significado das variáveis do problema, pois x deve representar a quantidade diária produzida de lotes de 1000 unidades de *chocolates* C_1 e, y a quantidade diária produzida de lotes de 1000 unidades de *chocolates* C_2 , sem as referências aos processos.

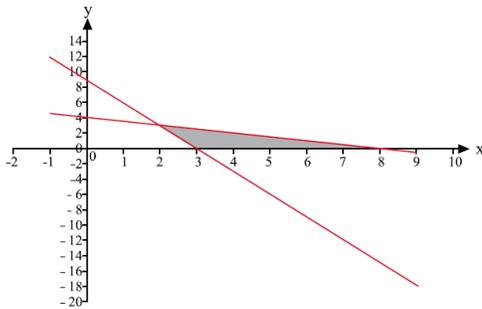
a)



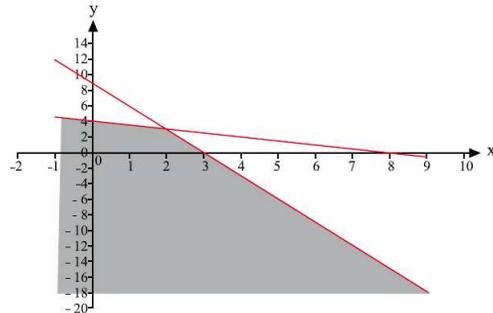
b)



c)



d)



e)

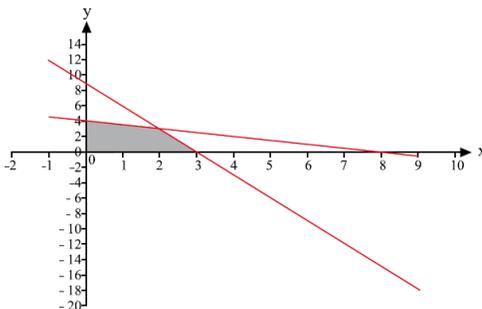


Figura 13 - Alternativas para a escolha da representação no plano cartesiano do conjunto dos pontos (x,y) que satisfazem, simultaneamente, às duas restrições.

Solução

Com as devidas correções na interpretação das variáveis, vamos organizar os dados do problema na Tabela 3.

Processos	Horas necessárias para produzir 1.000 unidades de C_1	Horas necessárias para produzir 1.000 unidades de C_2	Total de horas disponíveis
P_1	3	1	9
P_2	3	6	24

Tabela 3 - Horas necessárias para a produção de 1.000 unidades de C_1 e C_2 pelos processos P_1 e P_2 e total de horas disponíveis por dia para cada tipo de processo.

Pelo enunciado e pela análise da tabela temos as seguintes restrições:

$$\begin{aligned} 3x + y &\leq 9 \\ 3x + 6y &\leq 24. \end{aligned}$$

Como x e y representam quantidades produzidas de chocolate, temos:

$$x \geq 0 \text{ e } y \geq 0.$$

Assim, o único gráfico que satisfaz as restrições:

$$\begin{cases} 3x + y \leq 9 \\ 3x + 6y \leq 24 \\ x \geq 0 \\ y \geq 0 \end{cases}$$

é o gráfico da alternativa e da figura 13.

■

Exemplo numérico 4 - Questão 89 (Vestibular UNESP - 2010)

Dado que o lucro na venda de uma unidade do chocolate produzido pelo processo P_1 é de R\$ 0,50, enquanto que o lucro na venda de uma unidade do chocolate produzido pelo processo P_2 é de R\$ 0,80, e se forem vendidas todas as unidades produzidas em um dia nos dois processos, no número máximo possíveis de horas, o lucro obtido, em reais, será:

- (A) 3.400,00.
- (B) 3.900,00.
- (C) 4.700,00.
- (D) 6.400,00.
- (E) 11.200,00.

Solução

Com a correção mencionada para o significado das variáveis o enunciado desta questão ficaria:

“Dado que o lucro na venda de uma unidade do chocolate C_1 é de R\$ 0,50, enquanto que o lucro na venda de uma unidade do chocolate C_2 é de R\$ 0,80, e se forem vendidas todas as unidades produzidas em um dia dos dois tipos de chocolates, no número máximo possíveis de horas, o lucro obtido, em reais, será:”

Para resolver esta questão, precisamos da função lucro por lote de 1.000 unidades, que representaremos por

$$f(x, y) = 500x + 800y.$$

Considerando o gráfico das restrições em que o problema está sujeito (Figura 14), temos que a solução ótima para o lucro desse conjunto compacto se encontra em um dos vértices.

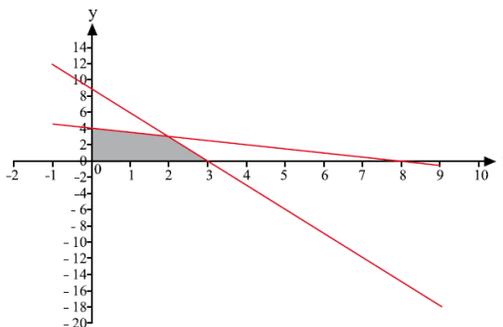


Figura 14 - Gráfico das restrições em que o problema está sujeito.

Apresentamos na Tabela 4 as coordenadas de cada um dos vértices do polígono formado pelas restrições do problema e os valores correspondentes para $f(x,y)$.

Vértice (x,y)	Valor de $f(x,y)$ em reais
(0,0)	0
(0,4)	3.200
(2,3)	3.400
(3,0)	1.500

Tabela 4 -Coordenadas de cada um dos vértices do polígono e os valores correspondentes para $f(x,y)$.

Notamos que o maior valor de $f(x,y)$ é 3.400, portanto, nas condições do problema, o lucro máximo é dado pela alternativa a.

Observação: Os alunos com conhecimentos de Ensino Médio deveriam perceber que as condições estabelecidas no problema exigiam o esgotamento das desigualdades:

$$\begin{aligned} 3x + y &\leq 9 \\ 3x + 6y &\leq 24. \end{aligned}$$

Isso ocorre quando as seguintes igualdades são satisfeitas:

$$\begin{aligned} 3x + y &= 9 \\ 3x + 6y &= 24. \end{aligned}$$

Com isso, o aluno não precisaria conhecer as particularidades da programação linear para resolver o problema. Bastaria resolver o sistema linear

$$\begin{cases} 3x + y = 9 \\ 3x + 6y = 24 \end{cases}$$

para achar o ponto que fornece a solução ótima pedida pela questão.

■

Observação: Note que o valor de cada variável deve ser inteiro. Isso dá origem a uma área chamada *programação inteira*.

Capítulo 2

Aspectos algébricos do PPL: ampliando a análise

Neste capítulo aprofundamos a análise das características algébricas intrínsecas aos PPL, iniciando com o caso em que se aplicaria o método dos multiplicadores de Lagrange (somente restrições de igualdade) para posteriormente trabalhar com problemas lineares gerais e a noção de espaço de exigência.

2.1 Problemas lineares sujeitos apenas a restrições de igualdade

Consideremos o seguinte problema:

$$\begin{array}{ll} \text{minimizar} & z = \mathbf{c}^T \mathbf{x} \\ \text{sujeito a} & \mathbf{Ax} = \mathbf{b} \end{array}$$

onde \mathbf{A} é uma matriz $m \times n$, \mathbf{b} um vetor coluna de dimensão m , \mathbf{c} e \mathbf{x} vetores coluna de dimensão n .

Observe que neste problema não existe a restrição de não negatividade $\mathbf{x} \geq \mathbf{0}$, portanto, para resolvê-lo utilizaremos o Método dos Multiplicadores de Lagrange ([Seção 1.4.5](#)).

Dessa forma, tomando $\boldsymbol{\lambda} \in \mathfrak{R}^m$, temos que em uma solução do problema o gradiente da função objetivo pode ser escrito como combinação linear dos gradientes das restrições, ou seja,

$$\mathbf{c} = \mathbf{A}^T \boldsymbol{\lambda}$$

em que $\lambda_1, \dots, \lambda_m$ serão os multiplicadores de Lagrange.

Assim, queremos resolver o sistema

$$\begin{cases} \mathbf{c} = \mathbf{A}^T \boldsymbol{\lambda} \\ \mathbf{A} \mathbf{x} = \mathbf{b} \end{cases} \Leftrightarrow \begin{cases} \mathbf{A}^T \boldsymbol{\lambda} = \mathbf{c} \\ \mathbf{A} \mathbf{x} = \mathbf{b} \end{cases},$$

que corresponde, na forma matricial, à seguinte equação:

$$\begin{pmatrix} 0 & \mathbf{A}^T \\ \mathbf{A} & 0 \end{pmatrix} \cdot \begin{pmatrix} \mathbf{x} \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} \mathbf{c} \\ \mathbf{b} \end{pmatrix}$$

Note que as variáveis \mathbf{x} e $\boldsymbol{\lambda}$ são desacopladas. Esse tipo de problema geralmente não apresenta solução (em geral $\mathbf{c} \notin \text{Im}(\mathbf{A}^T)$). E quando existe (quando $\mathbf{c} \in \text{Im}(\mathbf{A}^T)$), encontramos infinitas soluções pois o sistema $\mathbf{A} \mathbf{x} = \mathbf{b}$ é indeterminado, o que pouco ajuda em situações práticas, já que qualquer solução de $\mathbf{A} \mathbf{x} = \mathbf{b}$ minimiza (ou maximiza) z .

Isso será ilustrado no problema a seguir, com $\mathbf{x} \in \mathfrak{R}^2$ e $\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2)$:

$$\begin{aligned} \text{minimizar } z &= c_1 x_1 + c_2 x_2 \\ \text{sujeito a } a_1 x_1 + a_2 x_2 &= b. \end{aligned}$$

Pelo Método dos Multiplicadores de Lagrange, para obtermos uma solução é preciso que

$$\mathbf{c} = \mathbf{A}^T \boldsymbol{\lambda}$$

em que $\boldsymbol{\lambda} \in \mathfrak{R}$ é o multiplicador de Lagrange.

Assim, queremos resolver o sistema

$$\begin{cases} \mathbf{c} = \mathbf{A}^T \boldsymbol{\lambda} \\ \mathbf{A} \mathbf{x} = \mathbf{b} \end{cases},$$

que como vimos pode ser reescrito como

$$\begin{pmatrix} 0 & \mathbf{A}^T \\ \mathbf{A} & 0 \end{pmatrix} \cdot \begin{pmatrix} \mathbf{x} \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} \mathbf{c} \\ \mathbf{b} \end{pmatrix},$$

ou seja,

$$\begin{pmatrix} 0 & 0 & a_1 \\ 0 & 0 & a_2 \\ a_1 & a_2 & 0 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ \lambda \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ b \end{pmatrix}$$

O sistema

$$(2) \quad \begin{cases} a_1 x_1 + a_2 x_2 = b \\ \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \lambda = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \end{cases}$$

diz-nos que o vetor \mathbf{c} deve ser uma combinação linear do vetor \mathbf{a} , ou seja, o vetor \mathbf{c} deve pertencer à imagem de \mathbf{A}^T o que torna o problema solúvel apenas em um caso muito particular. Além disso, o sistema (2) é desacoplado, ou seja, a primeira equação depende apenas de \mathbf{x} , enquanto o segundo bloco depende apenas de λ . Isso faz com que a possibilidade de soluções seja pequena como pode ser verificado na Figura 15: apenas quando $\mathbf{c} \in \text{Im}(\mathbf{A}^T)$ é que haverá solução, neste caso, toda a reta $a_1 x_1 + a_2 x_2 = b$.

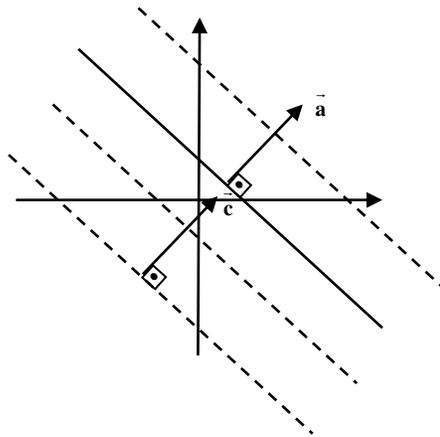


Figura 15 - O vetor \mathbf{c} é um múltiplo do vetor \mathbf{a} .

Percebemos que esse tipo de problema apenas com restrições lineares de igualdade é bastante peculiar e não reflete os modelos da programação linear, sendo de pouca utilidade para nós. Com a inclusão de desigualdades o modelo linear passa a acomodar melhor os problemas desejados. No entanto, a ferramenta dos multiplicadores de Lagrange sozinha não basta para resolvê-los. Isso ficará mais claro na próxima seção.

2.2 Problemas lineares sujeitos a restrições gerais

Como vimos, os problemas de programação linear requerem a condição de não negatividade para suas variáveis, o que impede o uso direto dos recursos utilizados para a resolução dos problemas da forma apresentada na seção anterior ([Seção 2.1](#)). Agora, faremos a representação do PPL na forma de somatórios e estudaremos alguns conceitos importantes para a compreensão das implicações trazidas por restrições lineares gerais.

2.3 Representação

Já vimos que um problema de programação linear pode ser escrito em notação matricial como:

$$\text{minimizar } z = \mathbf{c}^T \mathbf{x}$$

$$\text{sujeito a } \begin{cases} \mathbf{Ax} = \mathbf{b} \\ \mathbf{x} \geq \mathbf{0} \end{cases}$$

onde \mathbf{A} é uma matriz $m \times n$, em que a j -ésima coluna é denotada por $\mathbf{a}_{\cdot j}$, $\mathbf{b} \in \mathfrak{R}^m$ e $\mathbf{c}, \mathbf{x} \in \mathfrak{R}^n$.

Vamos representar essas equações na forma de somatório:

$$z = \mathbf{c}^T \mathbf{x} \Rightarrow (\mathbf{c}_1 \quad \mathbf{c}_2 \quad \cdots \quad \mathbf{c}_n) \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \Rightarrow z = \mathbf{c}_1 x_1 + \mathbf{c}_2 x_2 + \cdots + \mathbf{c}_n x_n \Rightarrow \sum_{j=1}^n \mathbf{c}_j x_j$$

e, como visto na Seção 1.2.1, $\mathbf{Ax} = \mathbf{b}$ e $\mathbf{x} \geq \mathbf{0}$ podem ser escritos como:

$$\sum_{j=1}^n \mathbf{a}_{\cdot j} x_j = \mathbf{b}$$

$$x_j \geq 0, j = 1, \dots, n.$$

Concluimos que $\mathbf{Ax} = \mathbf{b}$ ou $\sum_{j=1}^n \mathbf{a}_{\cdot j} x_j = \mathbf{b}$ tem solução se, e somente se, \mathbf{b} pode ser expresso como uma combinação linear dos vetores-coluna de \mathbf{A} , ou seja, \mathbf{b} está no espaço-coluna de \mathbf{A} .

Assim, nosso problema inicial pode ser reescrito da seguinte maneira:

$$(3) \quad \begin{aligned} & \text{minimizar} \quad z = \sum_{j=1}^n c_j x_j \\ & \text{sujeito a} \quad \begin{cases} \sum_{j=1}^n \mathbf{a}_{\cdot j} x_j = \mathbf{b} \\ x_j \geq 0, j = 1, \dots, n. \end{cases} \end{aligned}$$

2.4 O espaço de exigência

Um problema de programação linear pode ser analisado e resolvido geometricamente em um espaço normalmente chamado de *espaço de exigência* (cf.[2, p.19]).

O problema (3), pode ser interpretado da seguinte maneira: dados os vetores $\mathbf{a}_{\cdot 1}, \mathbf{a}_{\cdot 2}, \dots, \mathbf{a}_{\cdot n}$, nós queremos achar escalares não negativos x_1, x_2, \dots, x_n tais que $\sum_{j=1}^n \mathbf{a}_{\cdot j} x_j = \mathbf{b}$ e $z = \sum_{j=1}^n c_j x_j$ seja mínimo. Observe que o conjunto de vetores da forma $\sum_{j=1}^n \mathbf{a}_{\cdot j} x_j$, quando $x_1, x_2, \dots, x_n \geq 0$, é o cone gerado por $\mathbf{a}_{\cdot 1}, \mathbf{a}_{\cdot 2}, \dots, \mathbf{a}_{\cdot n}$. Assim, a viabilidade do problema está associada ao vetor \mathbf{b} pertencer a este cone (\mathbf{b} deve ser escrito como uma combinação linear não negativa dos vetores $\mathbf{a}_{\cdot 1}, \mathbf{a}_{\cdot 2}, \dots, \mathbf{a}_{\cdot n}$).

Na Figura 15 temos uma ilustração das duas situações possíveis. Em (a) o vetor \mathbf{b} pertence ao cone gerado pelos vetores $\mathbf{a}_{\cdot 1}, \mathbf{a}_{\cdot 2}$ e $\mathbf{a}_{\cdot 3}$, ou seja, o conjunto solução não é vazio. Em (b) isso não ocorre, portanto, temos um conjunto solução vazio.

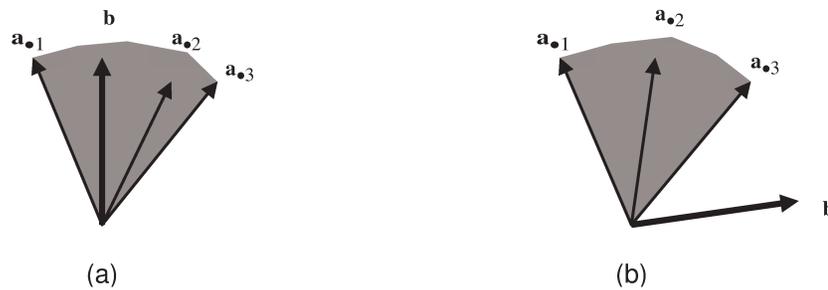


Figura 15 - Representação gráfica da região viável do *espaço de exigência*.
 a) Região viável não vazia; b) Região viável vazia.

Exemplo numérico 5 - Ilustrando o espaço de exigência

Considere os sistemas abaixo:

$$\text{Sistema A} \begin{cases} x_1 + 2x_2 + x_3 & = 1 \\ -x_1 + 3x_2 & + x_4 = 3 \\ x_1, & x_2, x_3, x_4 \geq 0 \end{cases}$$

e

$$\text{Sistema B} \begin{cases} x_1 + 2x_2 + x_3 & = -1 \\ -x_1 + 3x_2 & + x_4 = 3 \\ x_1, & x_2, x_3, x_4 \geq 0. \end{cases}$$

A Figura 16 mostra o espaço de exigência para ambos os sistemas. Para o *sistema A* o vetor **b** pertence ao cone gerado pelos vetores $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$, $\begin{pmatrix} 2 \\ 3 \end{pmatrix}$, $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ e $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ e, portanto admite solução. Para o *sistema B*, **b** não pertence ao cone, ou seja, o sistema é impossível (não possui solução).



Figura 16 - Representação do espaço de exigência dos sistemas A e B.
 a) O sistema A possui solução; b) O sistema B não possui solução.

2.5 Solução ótima

Agora vamos buscar a solução ótima do problema.

Estando satisfeitas as condições de viabilidade, ou seja,

$\sum_{j=1}^n \mathbf{a}_{\cdot j} x_j = \mathbf{b}$ e $x_j \geq 0$ com $j = 1, 2, \dots, n$, procuramos as variáveis x_1, x_2, \dots, x_n que minimizam $z = \sum_{j=1}^n c_j x_j$.

Podemos escrever o problema de programação linear da forma como segue, cf.[2, p.21].

Encontrar $x_1, x_2, \dots, x_n \geq 0$, tais que

$$\begin{pmatrix} c_1 \\ \mathbf{a}_{\cdot 1} \end{pmatrix} x_1 + \begin{pmatrix} c_2 \\ \mathbf{a}_{\cdot 2} \end{pmatrix} x_2 + \dots + \begin{pmatrix} c_n \\ \mathbf{a}_{\cdot n} \end{pmatrix} x_n = \begin{pmatrix} z \\ \mathbf{b} \end{pmatrix}$$

onde z deve ser minimizado. Portanto, buscamos representar o vetor $\begin{pmatrix} z \\ \mathbf{b} \end{pmatrix}$, para o

menor valor possível de z , no cone gerado pelos vetores $\begin{pmatrix} c_1 \\ \mathbf{a}_{\cdot 1} \end{pmatrix} + \begin{pmatrix} c_2 \\ \mathbf{a}_{\cdot 2} \end{pmatrix} + \dots + \begin{pmatrix} c_n \\ \mathbf{a}_{\cdot n} \end{pmatrix}$.

Dessa forma, apenas acrescentamos uma linha (função objetivo) ao sistema de condições de viabilidade. Isso leva ao aumento da dimensão do espaço de exigência de m para $m+1$.

Exemplo numérico 6 - Solução ótima finita

Consideremos o problema:

$$\text{minimizar } x_1 + 2x_2$$

$$\text{sujeito a } \begin{cases} 3x_1 + x_2 \leq 4 \\ x_1, x_2 \geq 0. \end{cases}$$

Adicionando a variável de folga $x_3 \geq 0$, o problema fica da seguinte forma:

Encontrar $x_1, x_2, x_3 \geq 0$, tais que

$$\begin{pmatrix} 1 \\ 3 \end{pmatrix} x_1 + \begin{pmatrix} 2 \\ 1 \end{pmatrix} x_2 + \begin{pmatrix} 0 \\ 1 \end{pmatrix} x_3 = \begin{pmatrix} z \\ 4 \end{pmatrix},$$

onde z deve ser minimizado. O cone gerado pelos vetores $\begin{pmatrix} 1 \\ 3 \end{pmatrix}$, $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$ e $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ é mostrado

na Figura 17. Nós queremos o vetor $\begin{pmatrix} z \\ 4 \end{pmatrix}$ desse cone com z mínimo.

Conseguimos a solução ótima para $z^* = 0$ com $x_1^* = x_2^* = 0$ e $x_3^* = 4$, ou seja, o vetor $\begin{pmatrix} z \\ 4 \end{pmatrix}$ é um múltiplo do vetor $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

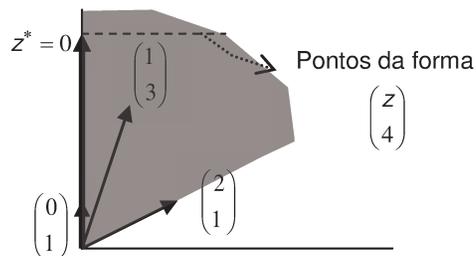


Figura 17 - Cone gerado pelos vetores $\begin{pmatrix} 1 \\ 3 \end{pmatrix}$, $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$ e $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

Exemplo numérico 7 - Problema ilimitado

Consideremos o problema:

$$\text{minimizar } -2x_1 - 2x_2$$

$$\text{sujeito a } \begin{cases} 4x_1 + 2x_2 \geq 3 \\ x_1, x_2 \geq 0. \end{cases}$$

Subtraindo a variável de folga $x_3 \geq 0$, o problema fica da seguinte forma:

Encontrar $x_1, x_2, x_3 \geq 0$, tais que

$$\begin{pmatrix} -2 \\ 4 \end{pmatrix} x_1 + \begin{pmatrix} -2 \\ 2 \end{pmatrix} x_2 + \begin{pmatrix} 0 \\ -1 \end{pmatrix} x_3 = \begin{pmatrix} z \\ 3 \end{pmatrix},$$

onde z deve ser minimizado. O cone gerado pelos vetores $\begin{pmatrix} -2 \\ 4 \end{pmatrix}$, $\begin{pmatrix} -2 \\ 2 \end{pmatrix}$ e $\begin{pmatrix} 0 \\ -1 \end{pmatrix}$ é

mostrado na Figura 18. Nós queremos o vetor $\begin{pmatrix} z \\ 3 \end{pmatrix}$ desse cone com z mínimo.

Observe que é possível achar pontos da forma $\begin{pmatrix} z \\ 3 \end{pmatrix}$ com z cada vez menor.

Portanto, o valor da solução ótima é $-\infty$, ou seja, o problema é ilimitado.

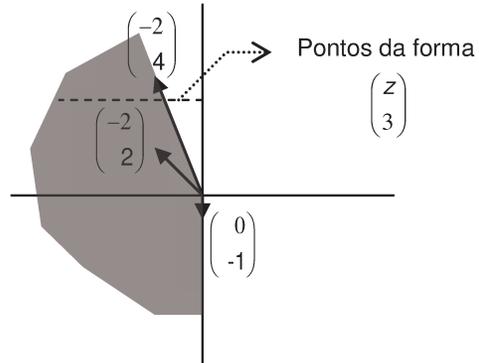


Figura 18 - Cone gerado pelos vetores $\begin{pmatrix} -2 \\ 4 \end{pmatrix}$, $\begin{pmatrix} -2 \\ 2 \end{pmatrix}$ e $\begin{pmatrix} 0 \\ -1 \end{pmatrix}$.

■

Capítulo 3

Dualidade e o Lema de Farkas

Iniciamos este capítulo apresentando elementos da dualidade dos PPL, em seguida mostramos as condições necessárias e suficientes de otimalidade do PPL e sua conexão com o Lema de Farkas. Uma prova para este conhecido resultado de alternativas é feita no final deste capítulo.

3.1 Dualidade

Para cada problema de programação linear existe outro problema linear associado a ele, que satisfaz algumas propriedades importantes e fornece informações úteis para a resolução do problema original. Este par de problemas é denominado par primal-dual. Como veremos na Seção 3.2, as terminologias primal e dual são relativas, pois [o dual do dual é o primal](#).

Vamos introduzir o conceito de dualidade por meio do nosso [exemplo numérico 2 \(Seção 1.5\)](#) cf.[15, p.196]. Nesse exemplo, procuramos a solução do seguinte PPL:

$$(4) \quad \begin{array}{l} \text{maximizar} \quad f(y_1, y_2) = 40y_1 + 60y_2 \\ \text{sujeito a} \quad \left\{ \begin{array}{l} 2y_1 + y_2 \leq 70 \\ y_1 + y_2 \leq 40 \\ y_1 + 3y_2 \leq 90 \\ y_1 \geq 0 \\ y_2 \geq 0 \end{array} \right. \end{array}$$

ou, matricialmente,

maximizar $\mathbf{b}^T \mathbf{y}$

sujeito a $\begin{cases} \mathbf{A}^T \mathbf{y} \leq \mathbf{c} \\ \mathbf{y} \geq \mathbf{0} \end{cases}$,

com $\mathbf{b}^T = (40, 60)$; $\mathbf{A}^T = \begin{pmatrix} 2 & 1 \\ 1 & 1 \\ 1 & 3 \end{pmatrix}$ e $\mathbf{c}^T = (70, 40, 90)$.

Temos na Figura 19 a representação gráfica da região viável de (4) com algumas curvas de nível.

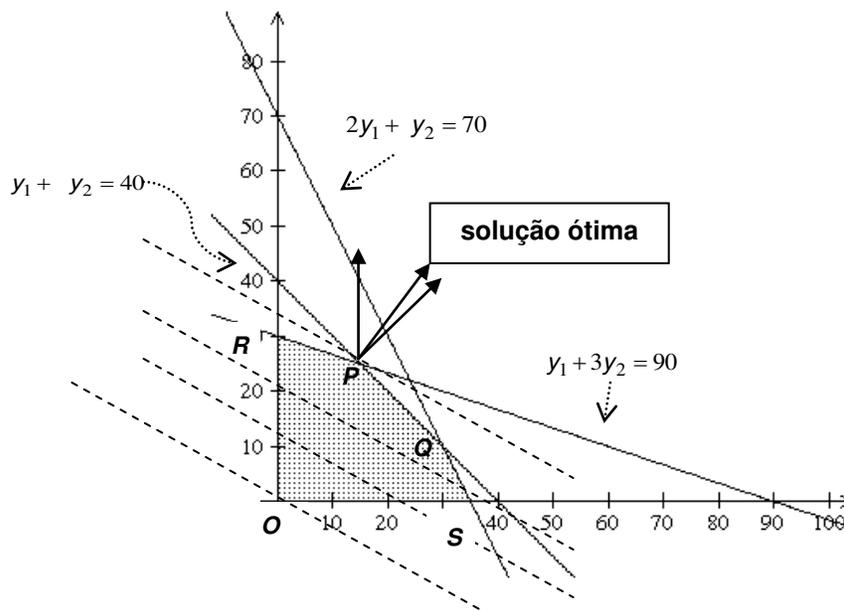


Figura 19 - Representação gráfica da região viável do problema (4) com algumas curvas de nível.

Estamos querendo maximizar $f(y_1, y_2)$ sobre a região poligonal $ORPQS$. Isso ocorrerá, como já verificado, no ponto P . Geometricamente, vemos que o vetor

gradiente ∇f que é perpendicular à reta $40y_1 + 60y_2 = 2100$ e que passa pelo ponto ótimo P está no cone gerado pelas normais exteriores às retas PQ e RP em P .

Portanto, temos que o ∇f pode ser escrito como uma combinação linear não-negativa dos dois vetores normais exteriores às retas PQ e RP , que são os gradientes das restrições ativas em P (satisfeitas na igualdade).

Assim, como o vetor gradiente da função objetivo e os gradientes das restrições mencionadas têm as direções

$$\begin{pmatrix} 40 \\ 60 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ e } \begin{pmatrix} 1 \\ 3 \end{pmatrix},$$

respectivamente, temos

$$\begin{pmatrix} 40 \\ 60 \end{pmatrix} = 30 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + 10 \begin{pmatrix} 1 \\ 3 \end{pmatrix}.$$

Observe que o vetor $\begin{pmatrix} 40 \\ 60 \end{pmatrix}$ é simplesmente o vetor \mathbf{b} da nossa função linear

$f(y_1, y_2) = 40y_1 + 60y_2 = \mathbf{b}^T \mathbf{y}$, que deve ser maximizada, enquanto os vetores $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ e $\begin{pmatrix} 1 \\ 3 \end{pmatrix}$

são os vetores-linha da matriz \mathbf{A}^T em $\mathbf{A}^T \mathbf{y} \leq \mathbf{c}$, que são satisfeitas como *igualdades* pelo vetor ótimo. Se \mathbf{y}^* maximiza $\mathbf{b}^T \mathbf{y}$ com as restrições $\mathbf{A}^T \mathbf{y} \leq \mathbf{c}$ e $\mathbf{y} \geq \mathbf{0}$, onde \mathbf{A}^T é uma matriz $n \times m$, então \mathbf{b}^T pode ser escrito como uma combinação linear não-negativa destas linhas de \mathbf{A}^T correspondentes a inequações realmente satisfeitas como igualdades na solução \mathbf{y}^* (que são chamadas de restrições *ativas* em \mathbf{y}^*).

Vamos analisar algebricamente o significado disso.

Como \mathbf{b}^T é uma combinação linear não-negativa das linhas de \mathbf{A}^T , correspondentes às restrições ativas em \mathbf{y}^* , existe um vetor $\mathbf{x}^* \in \mathfrak{R}^n$, $\mathbf{x}^* \geq \mathbf{0}$ tal que

$$\mathbf{b}^T = \mathbf{x}^{*T} \mathbf{A}^T, \text{ ou seja, } \mathbf{A} \mathbf{x}^* = \mathbf{b},$$

e tal que as componentes de \mathbf{x}^* que multiplicam as linhas de \mathbf{A}^T correspondentes às restrições inativas em \mathbf{y}^* devem se anular. Suponha agora que \mathbf{x} é outro vetor n -dimensional qualquer satisfazendo às chamadas *restrições duais*

$$(5) \quad \mathbf{Ax} \geq \mathbf{b}, \quad \mathbf{x} \geq \mathbf{0}.$$

Como $\mathbf{A}^T \mathbf{y}^* \leq \mathbf{c}$ e $\mathbf{x} \geq \mathbf{0}$, segue-se que

$$\mathbf{x}^T \mathbf{c} \geq \mathbf{x}^T \mathbf{A}^T \mathbf{y}^*.$$

Como $\mathbf{Ax} \geq \mathbf{b} \Leftrightarrow (\mathbf{Ax})^T \geq \mathbf{b}^T \Leftrightarrow \mathbf{x}^T \mathbf{A}^T \geq \mathbf{b}^T$ e $\mathbf{y}^* \geq \mathbf{0}$, segue-se que

$$(6) \quad \mathbf{x}^T \mathbf{c} \geq \mathbf{x}^T \mathbf{A}^T \mathbf{y}^* \geq \mathbf{b}^T \mathbf{y}^*.$$

Além disso, como \mathbf{x}^* se anula nas componentes que correspondem às linhas das restrições inativas, enquanto que nas outras componentes as inequações $\mathbf{A}^T \mathbf{y}^* \leq \mathbf{c}$ são satisfeitas como igualdades, segue-se que

$$(7) \quad \mathbf{x}^{*T} \mathbf{c} = \mathbf{x}^{*T} \mathbf{A}^T \mathbf{y}^* = \mathbf{b}^T \mathbf{y}^*.$$

Mas então (6) e (7) juntamente afirmam que

$$\mathbf{x}^T \mathbf{c} \geq \mathbf{x}^{*T} \mathbf{c} = \mathbf{b}^T \mathbf{y}^*$$

sempre que \mathbf{x} satisfaz (5). Ou seja, \mathbf{x}^* é solução do PPL

$$(8) \quad \begin{array}{l} \text{minimizar} \quad \mathbf{c}^T \mathbf{x} \\ \text{sujeito a} \quad \begin{cases} \mathbf{Ax} \geq \mathbf{b} \\ \mathbf{x} \geq \mathbf{0} \end{cases} \end{array}$$

e o valor mínimo deste programa (8) é igual ao valor máximo do PPL original

$$(9) \quad \begin{array}{l} \text{maximizar} \quad \mathbf{b}^T \mathbf{y} \\ \text{sujeito a} \quad \begin{cases} \mathbf{A}^T \mathbf{y} \leq \mathbf{c} \\ \mathbf{y} \geq \mathbf{0}. \end{cases} \end{array}$$

O problema original (9) é chamado problema *primal*, e o problema (8) que acabamos de definir é chamado problema *dual*. As relações entre estes dois problemas são muito importantes tanto na teoria como computacionalmente.

Embora tenhamos motivado a formulação do problema dual geométrica e algebricamente, uma vez definidos estes dois problemas, podemos passar a explorar suas relações a partir dos princípios básicos da otimização linear.

Para completar a análise do Exemplo numérico 2 iniciada na Seção 1.5, e do problema dual associado, vamos explicitar o significado econômico dos coeficientes que definem o problema, e das respectivas variáveis.

Nos problemas (8) e (9), as componentes dos vetores \mathbf{b} , \mathbf{c} e da matriz \mathbf{A} possuem os seguintes significados:

b_i : lucro por unidade fabricada do produto P_i ($i=1,2$);

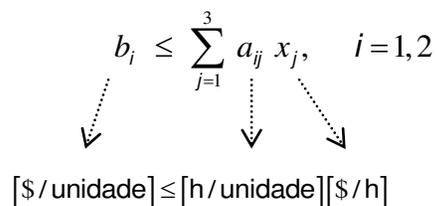
a_{ij} : número de horas empregadas pela máquina j para produzir uma unidade do produto i ($j=1,2,3$ e $i=1,2$);

c_j : total de horas semanais disponíveis para a máquina j ($j=1,2,3$);

As variáveis de (9) – y_i – descrevem o número de unidades fabricadas do produto P_i ($i=1,2$), enquanto as variáveis do problema (8) – x_j – correspondem ao preço de utilização da máquina j por hora adicional ($j=1,2,3$).

Assim, as restrições de (9) $\mathbf{A}^T \mathbf{y} \leq \mathbf{c}$ modelam o tempo disponível e as restrições de (8) $\mathbf{Ax} \geq \mathbf{b}$ modelam a produção justa - o lucro obtido supera o valor de mercado competitivamente.

Fazendo uma análise dimensional,

$$b_i \leq \sum_{j=1}^3 a_{ij} x_j, \quad i=1,2$$


$$[\$/\text{unidade}] \leq [h/\text{unidade}][\$/h]$$

completamos a interpretação econômica deste par de problemas primal-dual.

3.2 Representações do problema dual

Como vimos, a cada problema primal existe outro problema associado chamado de problema dual.

A representação desses problemas pode ser feita de duas formas equivalentes cf.[2, p.243]. São elas:

- Forma padrão da dualidade

Suponha que o programa linear primal (P) é dado na forma padrão:

$$\begin{aligned} \text{P:} \quad & \text{minimizar} \quad \mathbf{c}^T \mathbf{x} \\ & \text{sujeito a} \quad \begin{cases} \mathbf{Ax} = \mathbf{b} \\ \mathbf{x} \geq \mathbf{0}. \end{cases} \end{aligned}$$

Então o programa linear dual (D) é definido como:

$$\begin{aligned} \text{D:} \quad & \text{maximizar} \quad \mathbf{b}^T \mathbf{y} \\ & \text{sujeito a} \quad \begin{cases} \mathbf{A}^T \mathbf{y} \leq \mathbf{c} \\ \mathbf{y} \text{ irrestrito.} \end{cases} \end{aligned}$$

Exemplo numérico 8 - Dual de PPL na forma padrão

Considere o seguinte problema e explicita seu dual.

$$\begin{aligned} \text{P:} \quad & \text{minimizar} \quad 6x_1 + 8x_2 \\ & \text{sujeito a} \quad \begin{cases} 3x_1 + x_2 - x_3 = 4 \\ 5x_1 + 2x_2 - x_4 = 7 \\ x_1, x_2, x_3, x_4 \geq 0. \end{cases} \end{aligned}$$

Então o programa linear dual (D) é definido como:

$$D: \quad \text{maximizar} \quad 4y_1 + 7y_2$$

$$\text{sujeito a} \quad \begin{cases} 3y_1 + 5y_2 \leq 6 \\ y_1 + 2y_2 \leq 8 \\ -y_1 \leq 0 \\ -y_2 \leq 0 \\ y_1, y_2 \text{ irrestritos} \end{cases},$$

ou seja,

$$\text{maximizar} \quad 4y_1 + 7y_2$$

$$\text{sujeito a} \quad \begin{cases} 3y_1 + 5y_2 \leq 6 \\ y_1 + 2y_2 \leq 8 \\ y_1, y_2 \geq 0. \end{cases}$$

■

- Forma canônica da dualidade

Suponha que o programa linear primal (P) é dado na forma canônica:

$$P: \quad \text{minimizar} \quad \mathbf{c}^T \mathbf{x}$$

$$\text{sujeito a} \quad \begin{cases} \mathbf{Ax} \geq \mathbf{b} \\ \mathbf{x} \geq \mathbf{0}. \end{cases}$$

Então o programa linear dual (D) é definido como:

$$D: \quad \text{maximizar} \quad \mathbf{b}^T \mathbf{y}$$

$$\text{sujeito a} \quad \begin{cases} \mathbf{A}^T \mathbf{y} \leq \mathbf{c} \\ \mathbf{y} \geq \mathbf{0}. \end{cases}$$

Exemplo numérico 9 - Dual de PPL na forma canônica

Considere o seguinte problema e explicita seu dual.

$$P: \quad \text{minimizar} \quad 6x_1 + 8x_2$$

$$\text{sujeito a} \quad \begin{cases} 3x_1 + x_2 \geq 4 \\ 5x_1 + 2x_2 \geq 7 \\ x_1, x_2 \geq 0. \end{cases}$$

Então o programa linear dual (D) é definido como:

$$D: \quad \text{maximizar} \quad 4y_1 + 7y_2$$

$$\text{sujeito a} \quad \begin{cases} 3y_1 + 5y_2 \leq 6 \\ y_1 + 2y_2 \leq 8 \\ y_1, y_2 \geq 0. \end{cases}$$

■

Mesmo sendo equivalentes essas representações, cada uma delas pode ser mais útil do que a outra em diferentes contextos. A forma padrão, por exemplo, é bastante usada na resolução do PPL pelo Método Simplex.

Observe que a principal variação entre as duas maneiras de representação consiste que na forma padrão as restrições lineares gerais a que o problema está sujeito são equações, enquanto que na forma canônica essas condições aparecem como desigualdades. Muitas vezes será necessário mudarmos de uma forma para outra, o que pode ser feito com o acréscimo das variáveis de folga para levar da forma

canônica para a padrão ([Seção 1.1](#)), ou pela reescrita das restrições gerais de igualdade como duas desigualdades $\left(\mathbf{Ax} = \mathbf{b} \Rightarrow \begin{cases} \mathbf{Ax} \leq \mathbf{b} \\ \mathbf{Ax} \geq \mathbf{b} \end{cases} \Rightarrow \begin{pmatrix} \mathbf{A} \\ -\mathbf{A} \end{pmatrix} \mathbf{x} \leq \begin{pmatrix} \mathbf{b} \\ -\mathbf{b} \end{pmatrix} \right)$ para levar da forma padrão para a canônica.

Vejamos mais dois exemplo numéricos:

Exemplo numérico 10 - Dual do problema do exemplo 2

Ache o dual do programa linear (4), repetido a seguir:

$$\text{maximizar } 40y_1 + 60y_2$$

$$\text{sujeito a } \begin{cases} 2y_1 + y_2 \leq 70 \\ y_1 + y_2 \leq 40 \\ y_1 + 3y_2 \leq 90 \\ y_1 \geq 0 \\ y_2 \geq 0 \end{cases}$$

Solução

Podemos reescrever o problema dado na forma canônica:

$$\text{minimizar } -40y_1 - 60y_2$$

$$\text{sujeito a } \begin{cases} -2y_1 - y_2 \geq -70 \\ -y_1 - y_2 \geq -40 \\ -y_1 - 3y_2 \geq -90 \\ y_1 \geq 0 \\ y_2 \geq 0 \end{cases}$$

cujo dual é dado por (considerando x como o vetor das variáveis duais):

$$\text{maximizar } -70x_1 - 40x_2 - 90x_3$$

$$\text{sujeito a } \begin{cases} -2x_1 - x_2 - x_3 \leq -40 \\ -x_1 - x_2 - 3x_3 \leq -60 \\ x_1 \geq 0 \\ x_2 \geq 0 \\ x_3 \geq 0. \end{cases}$$

que é equivalente a

$$\text{minimizar } 70x_1 + 40x_2 + 90x_3$$

(10)

$$\text{sujeito a } \begin{cases} 2x_1 + x_2 + x_3 \geq 40 \\ x_1 + x_2 + 3x_3 \geq 60 \\ x_1 \geq 0 \\ x_2 \geq 0 \\ x_3 \geq 0. \end{cases}$$

■

Exemplo numérico 11 - O dual do dual

Ache o dual do programa linear (10).

Solução

Como o problema (10) está na forma canônica seu dual pode ser imediatamente escrito como

$$(11) \quad \begin{array}{l} \text{maximizar} \quad 40y_1 + 60y_2 \\ \\ \text{sujeito a} \quad \left\{ \begin{array}{l} 2y_1 + y_2 \leq 70 \\ y_1 + y_2 \leq 40 \\ y_1 + 3y_2 \leq 90 \\ y_1 \geq 0 \\ y_2 \geq 0, \end{array} \right. \end{array}$$

note que (11) é o PPL inicial (4) do exemplo 10, ou seja, o *dual do dual é igual ao primal!*



Verifiquemos esse resultado de forma geral.

Considere o programa linear primal (P) dado na forma canônica:

$$\begin{array}{l} \text{P:} \quad \text{minimizar} \quad \mathbf{c}^T \mathbf{x} \\ \\ \text{sujeito a} \quad \left\{ \begin{array}{l} \mathbf{Ax} \geq \mathbf{b} \\ \mathbf{x} \geq \mathbf{0}. \end{array} \right. \end{array}$$

Vimos que o seu dual (D) é definido como:

$$\begin{array}{l} \text{D:} \quad \text{maximizar} \quad \mathbf{b}^T \mathbf{y} \\ \\ \text{sujeito a} \quad \left\{ \begin{array}{l} \mathbf{A}^T \mathbf{y} \leq \mathbf{c} \\ \mathbf{y} \geq \mathbf{0}. \end{array} \right. \end{array}$$

Podemos reescrevê-lo na forma:

$$\text{minimizar } -\mathbf{b}^T \mathbf{y}$$

$$\text{sujeito a } \begin{cases} -\mathbf{A}^T \mathbf{y} \geq -\mathbf{c} \\ \mathbf{y} \geq \mathbf{0} \end{cases},$$

cujo dual é dado por (considerando \mathbf{x} como o vetor das variáveis dual):

$$\text{maximizar } -\mathbf{c}^T \mathbf{x}$$

$$\text{sujeito a } \begin{cases} -\mathbf{A}\mathbf{x} \leq -\mathbf{b} \\ \mathbf{x} \geq \mathbf{0} \end{cases},$$

mas isto é o mesmo que

$$\text{minimizar } \mathbf{c}^T \mathbf{x}$$

$$\text{sujeito a } \begin{cases} \mathbf{A}\mathbf{x} \geq \mathbf{b} \\ \mathbf{x} \geq \mathbf{0} \end{cases}$$

que é exatamente o problema primal original. Concluimos, portanto, que o *dual do dual é igual ao primal!*

3.3 O Lema de Farkas

O Lema de Farkas é um teorema de alternativa, que envolve dois sistemas de desigualdades lineares relacionados e afirma que, um ou outro tem solução, mas nunca ambos.

O enunciado do *Lema de Farkas*, estabelecido em 1901 cf.[6], é:

Lema de Farkas

Seja \mathbf{A} uma matriz $m \times n$ e $\mathbf{b} \in \mathfrak{R}^m$, ou

I. $\mathbf{Ax} = \mathbf{b}$ e $\mathbf{x} \geq \mathbf{0}$

tem solução com $\mathbf{x} \in \mathfrak{R}^n$, ou

II. $\mathbf{b}^T \mathbf{y} < \mathbf{0}$ e $\mathbf{A}^T \mathbf{y} \geq \mathbf{0}$

tem solução com $\mathbf{y} \in \mathfrak{R}^m$, mas nunca ocorrem ambos os casos.

Esse lema pode ser interpretado geometricamente da seguinte forma: dado um cone e um vetor, ou o vetor pertence ao cone ou existe um hiperplano separando o vetor do cone, mas nunca ocorrem as duas possibilidades simultaneamente.

Para viabilizar a interpretação geométrica desse lema reescrevemos I e II da seguinte maneira cf.[12, p.17]:

- De I temos

$$I'. \sum_{j=1}^n \mathbf{a}_{\bullet j} x_j = \mathbf{b} \text{ e } x_j \geq 0, j = 1, \dots, n;$$

- De II temos

$$\mathbf{b}^T \mathbf{y} < \mathbf{0} \text{ e } \mathbf{a}_{\bullet i}^T \mathbf{y} \geq \mathbf{0}, i = 1, \dots, n, \text{ ou equivalentemente, sem perda de generalidade}$$

$$II'. \mathbf{b}^T \mathbf{y} > \mathbf{0} \text{ e } \mathbf{a}_{\bullet i}^T \mathbf{y} \leq \mathbf{0}, i = 1, \dots, n,$$

onde $\mathbf{a}_{\bullet j}$ é a j -ésima coluna de \mathbf{A} e $\mathbf{a}_{\bullet i}^T$ é a i -ésima linha de \mathbf{A}^T .

O sistema I' requer que o vetor \mathbf{b} seja uma combinação linear não negativa dos vetores $\mathbf{a}_{\bullet 1}, \mathbf{a}_{\bullet 2}, \dots, \mathbf{a}_{\bullet n}$, ou seja, \mathbf{b} deve pertencer ao cone gerado pelos vetores-coluna de \mathbf{A} . O sistema II' requer que achemos um vetor $\mathbf{y} \in \mathfrak{R}^m$ que faça um ângulo estritamente agudo com o vetor \mathbf{b} e obtuso com os vetores $\mathbf{a}_{\bullet 1}, \mathbf{a}_{\bullet 2}, \dots, \mathbf{a}_{\bullet m}$.*

* Isso decorre do fato que $\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \cdot \|\mathbf{v}\| \cos \theta$, sendo $\theta \in [0, \pi]$ o ângulo entre os vetores \mathbf{u} e \mathbf{v} .

Portanto, para $\mathbf{u} \cdot \mathbf{v} > 0$, temos que $\cos \theta > 0 \Rightarrow \theta < \pi/2$, que é um ângulo estritamente agudo. E para $\mathbf{u} \cdot \mathbf{v} \leq 0$, temos que $\cos \theta \leq 0 \Rightarrow \pi/2 \leq \theta \leq \pi$, que é um ângulo obtuso.

Na Figura 19 temos um caso simples com $m=2$, $n=3$, em que I' tem solução e, portanto pelo lema de Farkas, II' não tem solução.

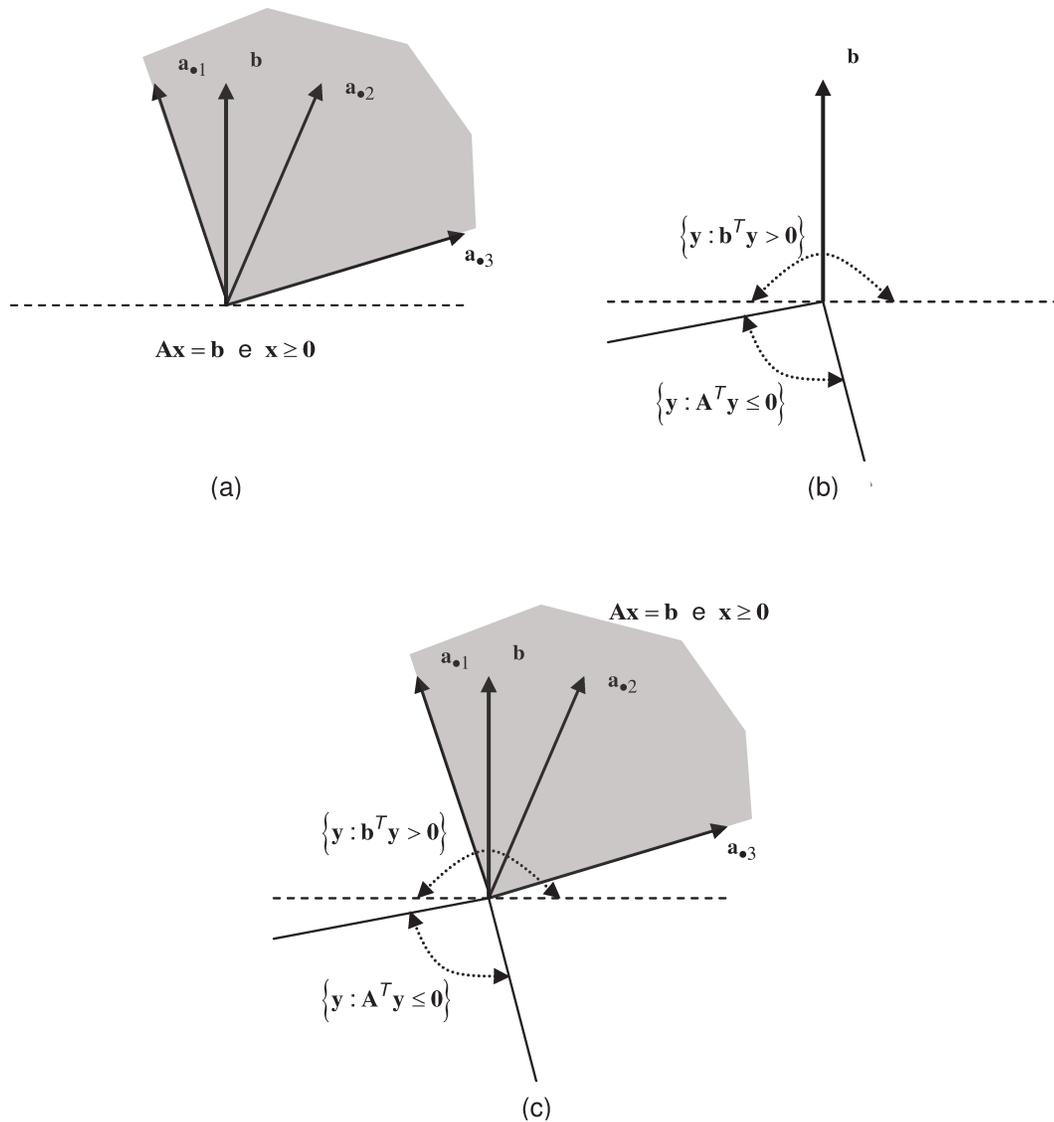


Figura 19 - Interpretação geométrica do Lema de Farkas:
a) I' tem solução; b) II' não tem solução; c) Ilustração com (a) e (b).

Já na Figura 20 temos um caso (também com $m=2, n=3$) em que II' tem solução e, portanto pelo lema de Farkas, I' não tem solução.

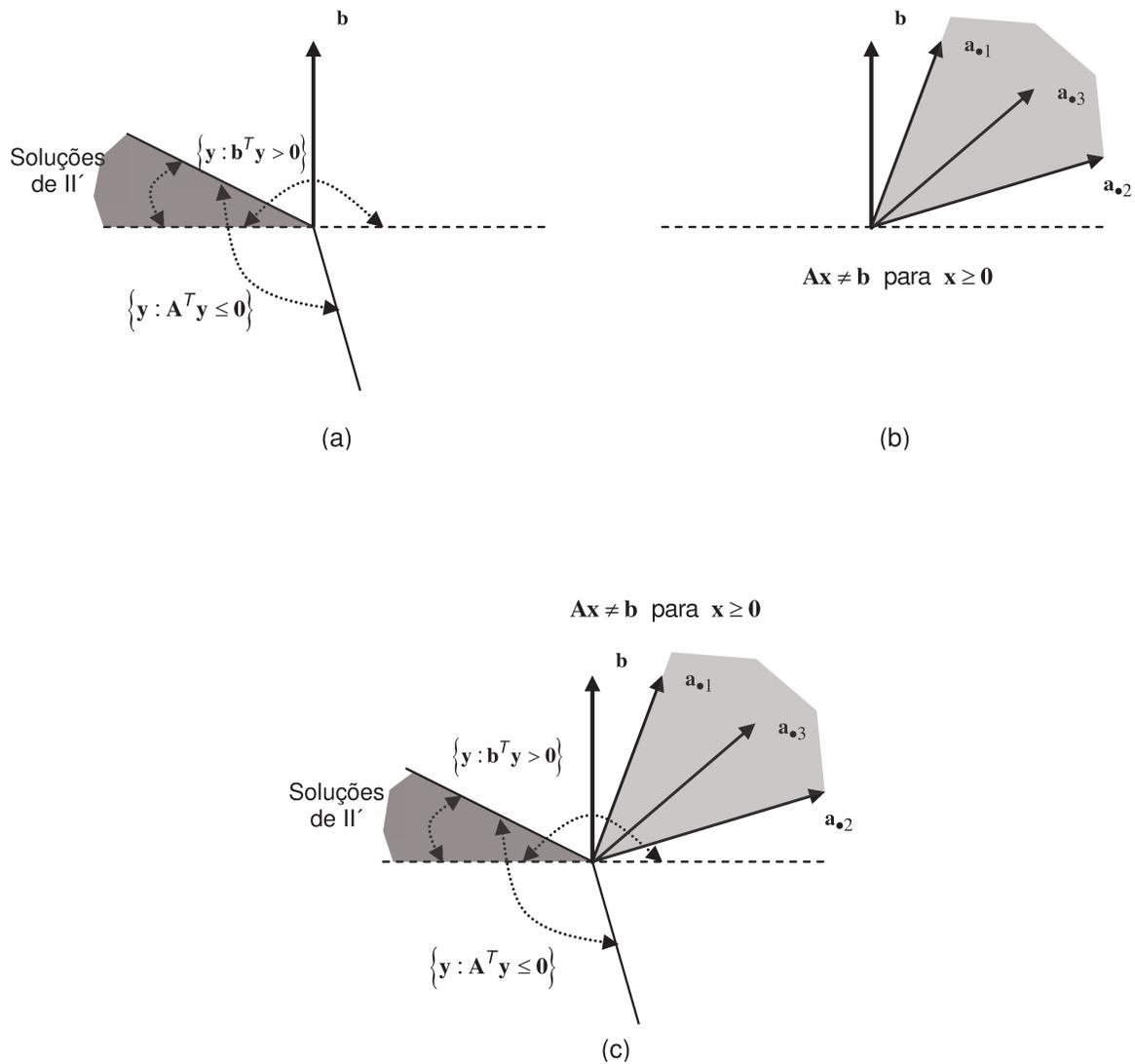


Figura 20 - Interpretação geométrica do Lema de Farkas:
a) II' tem solução; b) I' não tem solução; c) ilustração com (a) e (b).

3.4 O critério de otimalidade da programação linear: uma aplicação do Lema de Farkas

Os teoremas de alternativas são muito usados na matemática. Como exemplo da sua importância e em conexão com o lema deste trabalho vamos apresentar uma aplicação do Lema de Farkas (cf.[12, p.18]) para obter as condições de otimalidade do problema de programação linear no seguinte formato:

$$(12) \quad \begin{array}{ll} \text{maximizar} & \mathbf{b}^T \mathbf{y} \\ \text{sujeito a} & \mathbf{A}^T \mathbf{y} \leq \mathbf{c} \end{array}$$

onde \mathbf{b} e \mathbf{c} são vetores do \mathfrak{R}^m e \mathfrak{R}^n , respectivamente, e \mathbf{A}^T , de ordem $n \times m$, é a transposta da matriz \mathbf{A} . Vamos denotar o conjunto viável de (12) por $Y = \{\mathbf{y} : \mathbf{y} \in \mathfrak{R}^m, \mathbf{A}^T \mathbf{y} \leq \mathbf{c}\}$ e sua solução ótima por $\bar{\mathbf{y}}$. Vale notar que o formato (12) é justamente o dual de um PPL na forma padrão. Observe que todo PPL pode ser escrito dessa forma, após transformações convenientes, se necessário.

Critério de otimalidade da programação linear

As soluções ótimas de um PPL são caracterizadas pelas seguintes condições:

Condição necessária

Seja $\bar{\mathbf{y}}$ uma solução do problema de programação linear (12).

Então existe $\bar{\mathbf{x}} \in \mathfrak{R}^n$ tal que o par $(\bar{\mathbf{y}}, \bar{\mathbf{x}})$ satisfaz

- (13) $\mathbf{A}^T \bar{\mathbf{y}} \leq \mathbf{c}$ (viabilidade dual)
 (14) $\mathbf{A} \bar{\mathbf{x}} = \mathbf{b}$
 (15) $\bar{\mathbf{x}} \geq \mathbf{0}$ } (viabilidade primal)
 (16) $\mathbf{b}^T \bar{\mathbf{y}} = \mathbf{c}^T \bar{\mathbf{x}}$ (complementaridade)

Condição suficiente

Se algum $\bar{\mathbf{y}} \in \mathfrak{R}^m$ e algum $\bar{\mathbf{x}} \in \mathfrak{R}^n$ satisfazem (13) a (16) então $\bar{\mathbf{y}}$ é solução de (12).

Agora vamos apresentar a demonstração dessas condições, sendo que para provarmos a condição necessária vamos utilizar o Lema de Farkas.

Demonstração:

Condição necessária

Seja $\bar{\mathbf{y}}$ solução de (12). Defina os conjuntos de índices P , Q e N como segue:

$$P \cup Q = N = \{1, \dots, n\}; \quad P = \{i : \mathbf{a}_{\cdot i}^T \bar{\mathbf{y}} = c_i\}; \quad Q = \{i : \mathbf{a}_{\cdot i}^T \bar{\mathbf{y}} < c_i\}$$

e assumamos que P e Q contém p e q elementos, respectivamente.

Temos que P corresponde ao conjunto de índices das equações ativas e Q ao conjunto de índices das equações inativas do sistema formado por $\mathbf{A}^T \bar{\mathbf{y}} \leq \mathbf{c}$.

Então, usando a notação $\mathbf{A}_I = \{\mathbf{a}_{\cdot i}; i \in I\}$ e $\mathbf{c}_I = \{c_i; i \in I\}$, esse sistema pode ser reescrito como

(17) $\mathbf{A}_P^T \bar{\mathbf{y}} = \mathbf{c}_P,$

(18) $\mathbf{A}_Q^T \bar{\mathbf{y}} < \mathbf{c}_Q.$

Se $P = \emptyset$, segue que $\mathbf{A}^T \bar{\mathbf{y}} < \mathbf{c} \Leftrightarrow \mathbf{A}^T \bar{\mathbf{y}} - \mathbf{c} < \mathbf{0} \Leftrightarrow \mathbf{A}^T \bar{\mathbf{y}} - \mathbf{c} < -\delta \mathbf{e} < \mathbf{0}$ para algum número real $\delta > 0$, onde $\mathbf{e} \in \mathfrak{R}^n$ é o vetor de uns n -dimensional. Então para cada $\mathbf{y} \in \mathfrak{R}^m$, podemos encontrar um número real $\alpha > 0$ tal que $\mathbf{A}^T (\bar{\mathbf{y}} + \alpha \mathbf{y}) - \mathbf{c} < -\delta \mathbf{e} + \alpha \mathbf{A}^T \mathbf{y} \leq \mathbf{0}$ e, portanto, $\bar{\mathbf{y}} + \alpha \mathbf{y} \in Y$. Podemos interpretar geometricamente a última desigualdade da maneira como segue.

Seja a região viável Y representada pela Figura 21.

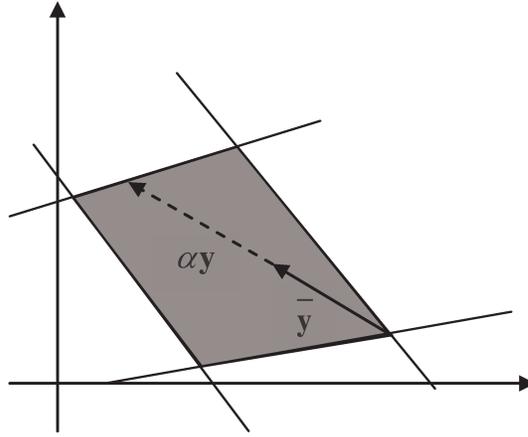


Figura 21 - Região viável $\mathbf{A}^T \mathbf{y} \leq \mathbf{c}$.

Como $\bar{\mathbf{y}}$ corresponde ao valor ótimo procurado, que por (18) pertence ao interior do conjunto viável Y , temos que qualquer que seja $\mathbf{y} \in \mathfrak{R}^m$ existe um múltiplo dele que somado a $\bar{\mathbf{y}}$ também pertence à região viável (Figura 21).

Já que $\bar{\mathbf{y}}$ é a solução máxima de (12), temos que para cada $\mathbf{y} \in \mathfrak{R}^m$ existe um $\alpha > 0$ tal que, pela otimalidade de $\bar{\mathbf{y}}$, $\mathbf{b}^T \bar{\mathbf{y}} \geq \mathbf{b}^T (\bar{\mathbf{y}} + \alpha \mathbf{y})$, pois $(\bar{\mathbf{y}} + \alpha \mathbf{y}) \in Y$, assim $-\mathbf{b}^T \bar{\mathbf{y}} \leq -\mathbf{b}^T (\bar{\mathbf{y}} + \alpha \mathbf{y})$. Dessa forma, $\mathbf{b}^T \mathbf{y} \leq 0$ para cada $\mathbf{y} \in \mathfrak{R}^m$ o que implica que $\mathbf{b} = \mathbf{0}$. Para chegarmos a essa conclusão basta tomarmos para um i fixo pertencente a $N = \{1, 2, \dots, m\}$, $y_i = b_i$, $y_{j \neq i} = 0$. Então $\mathbf{b}^T \mathbf{y} \leq 0$ implica que $(b_i)^2 \leq 0$. Portanto, $b_i = 0$. Repetindo esse processo para cada $i \in N$, chegaremos em $\mathbf{b} = \mathbf{0}$.

Assim, não existiria uma função objetivo e conseqüentemente o PPL se restringiria a um problema de viabilidade. Realmente, como já vimos, a solução de um PPL, com função objetivo não nula, nunca se encontra no interior da região viável.

Se $P \neq \emptyset$, então asseguramos que o sistema

$$(19) \quad \mathbf{A}_P^T \mathbf{y} \leq \mathbf{0} \text{ e } \mathbf{b}^T \mathbf{y} > 0$$

não tem solução $\mathbf{y} \in \mathfrak{R}^m$. Se (19) tivesse uma solução \mathbf{y} , então $\alpha \mathbf{y}$ também seria uma solução para todo $\alpha > 0$. De fato, considere o ponto $\bar{\mathbf{y}} + \alpha \mathbf{y}$, quando \mathbf{y} é uma solução de (19) e $\alpha > 0$. Então

$$(20) \quad -\mathbf{b}^T (\bar{\mathbf{y}} + \alpha \mathbf{y}) = -\mathbf{b}^T \bar{\mathbf{y}} - \alpha \mathbf{b}^T \mathbf{y} < -\mathbf{b}^T \bar{\mathbf{y}} \text{ para } \alpha > 0 \text{ e } \mathbf{b}^T \mathbf{y} > 0 \quad (\text{por 19})$$

$$(21) \quad \mathbf{A}_P^T (\bar{\mathbf{y}} + \alpha \mathbf{y}) \leq \mathbf{c}_P \Leftrightarrow \mathbf{A}_P^T (\bar{\mathbf{y}} + \alpha \mathbf{y}) - \mathbf{c}_P \leq \mathbf{0} \Leftrightarrow \mathbf{A}_P^T (\bar{\mathbf{y}} + \alpha \mathbf{y}) - \mathbf{c}_P = \mathbf{A}_P^T \bar{\mathbf{y}} + \alpha \mathbf{A}_P^T \mathbf{y} - \mathbf{c}_P = \\ = \mathbf{c}_P + \alpha \mathbf{A}_P^T \mathbf{y} - \mathbf{c}_P = \alpha \mathbf{A}_P^T \mathbf{y} \leq \mathbf{0} \text{ para } \alpha > 0 \text{ e } \mathbf{A}_P^T \mathbf{y} \leq \mathbf{0} \quad (\text{por 17 e 19})$$

$$(22) \quad \mathbf{A}_Q^T (\bar{\mathbf{y}} + \alpha \mathbf{y}) - \mathbf{c}_Q \leq -\delta \mathbf{e} + \alpha \mathbf{A}_Q^T \mathbf{y} \leq \mathbf{0} \text{ para algum } \alpha > 0 \quad (\text{por 18})$$

onde a primeira inequação de (22) segue da definição de \mathbf{e} como um vetor de uns de dimensão q e $-\delta = \max_{i \in Q} (\mathbf{A}_i^T \bar{\mathbf{y}} - c_i) < 0$ e a segunda desigualdade de (22) é válida para algum $\alpha > 0$ porque $-\delta < 0$.

As relações de (20) a (22) implicam que $\bar{\mathbf{y}} + \alpha \mathbf{y} \in Y$ e $-\mathbf{b}^T (\bar{\mathbf{y}} + \alpha \mathbf{y}) < -\mathbf{b}^T \bar{\mathbf{y}}$, o que contradiz a suposição que $\bar{\mathbf{y}}$ é uma solução de (12). Portanto, (19) não tem solução $\mathbf{y} \in \mathfrak{R}^m$ e pelo [Lema de Farkas](#) o sistema

$$(23) \quad \mathbf{A}_P \mathbf{x} = \mathbf{b} \text{ e } \mathbf{x} \geq \mathbf{0}$$

deve ter uma solução $\mathbf{x} \in \mathfrak{R}^p$. Se $\mathbf{0} \in \mathfrak{R}^q$, temos então

$$(24) \quad \mathbf{A}_P \mathbf{x} + \mathbf{A}_Q \mathbf{0} = \mathbf{b} \text{ e } \mathbf{x} \geq \mathbf{0} \text{ e}$$

$$(25) \quad \mathbf{c}_P^T \mathbf{x} + \mathbf{c}_Q^T \mathbf{0} = \mathbf{x}^T \mathbf{c}_P = \mathbf{x}^T \mathbf{A}_P^T \bar{\mathbf{y}} = (\mathbf{A}_P \mathbf{x})^T \bar{\mathbf{y}} = \mathbf{b}^T \bar{\mathbf{y}}. \quad (\text{por 17 e 23})$$

Seja $\bar{\mathbf{x}} \in \mathfrak{R}^n$ tal que $\bar{\mathbf{x}} = (\bar{\mathbf{x}}_P, \bar{\mathbf{x}}_Q)$, onde $\bar{\mathbf{x}}_P = \mathbf{x} \in \mathfrak{R}^p$, $\bar{\mathbf{x}}_Q = \mathbf{0} \in \mathfrak{R}^q$, a condição (24) converte-se nas condições (14) e (15) e a condição (25) torna-se a condição $\mathbf{b}^T \bar{\mathbf{y}} = \mathbf{c}^T \bar{\mathbf{x}}$. A condição (13) implica $\bar{\mathbf{y}} \in Y$. Isso completa a prova da necessidade.

Condição suficiente

Se $\bar{\mathbf{y}} \in \mathfrak{R}^m$ e $\bar{\mathbf{x}} \in \mathfrak{R}^n$ satisfazem (13) a (16) e se $\mathbf{y} \in Y$, por (13) temos que $\bar{\mathbf{y}} \in Y$.

Agora

$$\mathbf{b}^T \bar{\mathbf{y}} \geq \mathbf{b}^T \mathbf{y} \Rightarrow \mathbf{b}^T \bar{\mathbf{y}} - \mathbf{b}^T \mathbf{y} \geq 0 \Rightarrow \mathbf{b}^T \mathbf{y} - \mathbf{b}^T \bar{\mathbf{y}} \leq 0 \Rightarrow -\mathbf{b}^T \mathbf{y} - (-\mathbf{b}^T \bar{\mathbf{y}}) \geq 0, \text{ pois } -\mathbf{b}^T \bar{\mathbf{y}} = \min_{\mathbf{y} \in Y} (-\mathbf{b}^T \mathbf{y}).$$

Assim,

$$\begin{aligned} -\mathbf{b}^T \mathbf{y} - (-\mathbf{b}^T \bar{\mathbf{y}}) &= && \text{(por 14)} \\ &= -\bar{\mathbf{x}}^T \mathbf{A}^T \mathbf{y} + \mathbf{b}^T \bar{\mathbf{y}} && \text{(por 16)} \\ &= -\bar{\mathbf{x}}^T \mathbf{A}^T \mathbf{y} + \mathbf{c}^T \bar{\mathbf{x}} = \\ &= -\bar{\mathbf{x}}^T \mathbf{A}^T \mathbf{y} + \bar{\mathbf{x}}^T \mathbf{c} = \\ &= -\bar{\mathbf{x}}^T (\mathbf{A}^T \mathbf{y} - \mathbf{c}) \geq 0 && \text{(por 15)} \end{aligned}$$

temos que $\mathbf{A}^T \mathbf{y} - \mathbf{c} \leq \mathbf{0} \Leftrightarrow \mathbf{A}^T \mathbf{y} \leq \mathbf{c}$, portanto, $\mathbf{y} \in Y$, o que completa a prova da suficiência.

□

A condição (16) pode ser interpretada como segue:

$$\begin{aligned} \mathbf{c}^T \bar{\mathbf{x}} &= \mathbf{b}^T \mathbf{y} \\ \mathbf{c}^T \bar{\mathbf{x}} - \mathbf{b}^T \mathbf{y} &= 0 \\ \mathbf{c}^T \bar{\mathbf{x}} - \mathbf{y}^T \mathbf{b} &= 0 \\ \mathbf{c}^T \bar{\mathbf{x}} - \mathbf{y}^T \mathbf{A} \bar{\mathbf{x}} &= 0 && \text{(por 14)} \\ (\mathbf{c}^T - \mathbf{y}^T \mathbf{A}) \bar{\mathbf{x}} &= 0 \\ (\mathbf{c} - \mathbf{A}^T \mathbf{y})^T \bar{\mathbf{x}} &= 0. \end{aligned}$$

(26)

Como

$$(27) \quad \begin{aligned} \mathbf{A}^T \mathbf{y} &\leq \mathbf{c} \text{ então} \\ \mathbf{A}^T \mathbf{y} + \mathbf{z} &= \mathbf{c} \text{ para } \mathbf{z} \geq \mathbf{0}, \text{ ou seja,} \\ \mathbf{z} &= \mathbf{c} - \mathbf{A}^T \mathbf{y} \end{aligned}$$

onde \mathbf{z} é a variável de folga. Substituindo (27) em (26) chegamos em

$$\mathbf{z}^T \bar{\mathbf{x}} = \mathbf{0} \text{ para } \mathbf{z} \geq \mathbf{0} \text{ e } \bar{\mathbf{x}} \geq \mathbf{0},$$

sendo essa a expressão que gera a terminologia da condição de complementaridade, também usual em programação não-linear.

3.5 Uma prova do Lema de Farkas

Vamos apresentar uma prova do Lema de Farkas utilizando como ferramenta a álgebra linear - cálculo vetorial e matricial (cf.[17]). A abordagem geral utilizada será feita conforme [4], mas em vez de definir um método de restrições ativas para provar a existência de soluções ótimas para problemas de quadrados mínimos com restrição de não negatividade (a parte mais difícil da prova), usaremos uma partição natural da região viável em sub-regiões disjuntas, seguida por uma simples investigação de cada sub-região. No final desta secção incluímos um mapa conceitual que ilustra o encadeamento dos resultados e definições utilizados na prova.

Na dissertação de Eustáquio [5], o autor desenvolve a prova apresentando duas versões para o Lema: versão geométrica (Lema 2.8, p.40) e versão algébrica (Lema 2.11, p.44). Mostra que as suas versões são equivalentes, com base em cones gerados por um número finito de elementos e seu polar. O Lema de Farkas é utilizado então para obter as condições de otimalidade de problemas gerais de programação não linear.

Retomando o enunciado do Lema de Farkas temos:

Lema de Farkas

Seja \mathbf{A} uma matriz $m \times n$ e $\mathbf{b} \in \mathfrak{R}^m$, ou

$$(28) \quad \mathbf{Ax} = \mathbf{b} \text{ e } \mathbf{x} \geq \mathbf{0}$$

tem solução com $\mathbf{x} \in \mathfrak{R}^n$, ou

$$(29) \quad \mathbf{b}^T \mathbf{y} < \mathbf{0} \text{ e } \mathbf{A}^T \mathbf{y} \geq \mathbf{0}$$

tem solução com $\mathbf{y} \in \mathfrak{R}^m$, mas nunca ocorrem ambos os casos.

O primeiro sistema (28) está intimamente relacionado com o seguinte problema linear de quadrados mínimos com restrições de não-negatividade das variáveis:

$$(30) \quad \text{minimizar } \frac{1}{2} \|\mathbf{Ax} - \mathbf{b}\|^2 \text{ sujeito a } \mathbf{x} \geq \mathbf{0}.$$

Conforme mostrado em [4], a prova do Lema de Farkas pode ser baseada nos seguintes lemas:

Lema 3.1

Existe sempre uma solução ótima $\bar{\mathbf{x}}$ do problema (30).

Lema 3.2

O ponto $\bar{\mathbf{x}}$ é uma solução ótima do problema (30) se e somente se

$$(i) \quad \bar{\mathbf{x}} \geq \mathbf{0}$$

$$(ii) \quad \mathbf{A}^T (\mathbf{A}\bar{\mathbf{x}} - \mathbf{b}) \geq \mathbf{0}$$

$$(iii) \quad \bar{\mathbf{x}}^T \mathbf{A}^T (\mathbf{A}\bar{\mathbf{x}} - \mathbf{b}) = 0$$

A prova do Lema de Farkas

Primeiramente vamos assumir que existe uma solução $\bar{\mathbf{x}}$ do sistema (28).

Então $\mathbf{b}^T \mathbf{y} = (\mathbf{A}\bar{\mathbf{x}})^T \mathbf{y} = \bar{\mathbf{x}}^T (\mathbf{A}^T \mathbf{y})$ e se $\mathbf{A}^T \mathbf{y} \geq \mathbf{0}$ então $\mathbf{b}^T \mathbf{y} \geq \mathbf{0}$.

Assim, não existe uma solução do sistema (29).

Em seguida, vamos supor que o sistema (28) não possui solução.

Seja $\bar{\mathbf{x}}$ uma solução ótima do problema (30) (que existe pelo Lema 3.1) e seja $\bar{\mathbf{y}} = \mathbf{A}\bar{\mathbf{x}} - \mathbf{b}$. Então $\bar{\mathbf{y}} \neq \mathbf{0}$, já que não há solução de (28).

Pelo Lema 3.2, $\bar{\mathbf{x}} \geq \mathbf{0}$, $\mathbf{A}^T \bar{\mathbf{y}} \geq \mathbf{0}$ e $\bar{\mathbf{x}}^T \mathbf{A}^T \bar{\mathbf{y}} = 0$, o que implica que $\mathbf{b}^T \bar{\mathbf{y}} = (\mathbf{A}\bar{\mathbf{x}} - \bar{\mathbf{y}})^T \bar{\mathbf{y}} = \bar{\mathbf{x}}^T \mathbf{A}^T \bar{\mathbf{y}} - \bar{\mathbf{y}}^T \bar{\mathbf{y}} = 0 - \|\bar{\mathbf{y}}\|^2 < 0$.

Assim, $\bar{\mathbf{y}}$ é uma solução do sistema (29). □

Resta provarmos os Lemas 3.1 e 3.2. Isso será feito usando álgebra linear elementar e raciocínio lógico.

Problemas de quadrados mínimos e a prova do Lema 3.2

Antes de provar o Lema 3.2, fazemos uma prova elementar de um resultado bem conhecido de problemas de quadrados mínimos irrestritos. Este resultado, também revisitado, será usado para provar o Lema 3.1.

Seja o seguinte problema quadrático na variável $\mathbf{x} = (x_1, \dots, x_n)^T$:

$$(31) \quad \min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}) = \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 = \frac{1}{2} (\mathbf{A}\mathbf{x} - \mathbf{b})^T (\mathbf{A}\mathbf{x} - \mathbf{b}).$$

Então para todos os vetores $\bar{\mathbf{x}}$ e \mathbf{d} ,

$$\begin{aligned} f(\bar{\mathbf{x}} + \mathbf{d}) &= \frac{1}{2} \|\mathbf{A}\bar{\mathbf{x}} - \mathbf{b} + \mathbf{A}\mathbf{d}\|^2 = \frac{1}{2} \|\mathbf{A}\bar{\mathbf{x}} - \mathbf{b}\|^2 + \frac{1}{2} (\mathbf{A}\bar{\mathbf{x}} - \mathbf{b})^T \mathbf{A}\mathbf{d} + \frac{1}{2} (\mathbf{A}\mathbf{d})^T (\mathbf{A}\bar{\mathbf{x}} - \mathbf{b}) + \frac{1}{2} \|\mathbf{A}\mathbf{d}\|^2 = \\ &= f(\bar{\mathbf{x}}) + \mathbf{d}^T \mathbf{A}^T (\mathbf{A}\bar{\mathbf{x}} - \mathbf{b}) + \frac{1}{2} \|\mathbf{A}\mathbf{d}\|^2. \end{aligned}$$

Com $\mathbf{d} = \mathbf{x} - \bar{\mathbf{x}}$, essa identidade pode ser escrita como

$$(32) \quad f(\mathbf{x}) = f(\bar{\mathbf{x}}) + (\mathbf{x} - \bar{\mathbf{x}})^T \mathbf{A}^T (\mathbf{A}\bar{\mathbf{x}} - \mathbf{b}) + \frac{1}{2} \|\mathbf{A}(\mathbf{x} - \bar{\mathbf{x}})\|^2,$$

que, portanto, vale para todos os vetores \mathbf{x} e $\bar{\mathbf{x}}$ de \mathfrak{R}^n .

O resultado a seguir será usado na prova do Lema 3.2.

Lema 3.3

O ponto $\bar{\mathbf{x}}$ minimiza $f(\mathbf{x})$, definida por (31), se e somente se $\mathbf{A}^T \mathbf{A}\bar{\mathbf{x}} = \mathbf{A}^T \mathbf{b}$.

Existe um único ponto $\bar{\mathbf{x}}$ que minimiza $f(\mathbf{x})$ se e somente se as colunas de \mathbf{A} são linearmente independentes.

Demonstração: Seja $\bar{\mathbf{g}} = \mathbf{A}^T (\mathbf{A}\bar{\mathbf{x}} - \mathbf{b})$. Então, por (32),

$$(33) \quad f(\mathbf{x}) - f(\bar{\mathbf{x}}) = (\mathbf{x} - \bar{\mathbf{x}})^T \bar{\mathbf{g}} + \frac{1}{2} \|\mathbf{A}(\mathbf{x} - \bar{\mathbf{x}})\|^2.$$

Vamos assumir primeiro que $\mathbf{A}^T \mathbf{A}\bar{\mathbf{x}} = \mathbf{A}^T \mathbf{b}$. Então $\bar{\mathbf{g}} = \mathbf{0}$, que por (33) implica que $f(\mathbf{x}) - f(\bar{\mathbf{x}}) = \frac{1}{2} \|\mathbf{A}(\mathbf{x} - \bar{\mathbf{x}})\|^2 \geq 0$ para todo \mathbf{x} , e então $\bar{\mathbf{x}}$ minimiza $f(\mathbf{x})$.

Agora, vamos assumir que $\mathbf{A}^T \mathbf{A}\bar{\mathbf{x}} \neq \mathbf{A}^T \mathbf{b}$. Então $\bar{\mathbf{g}} \neq \mathbf{0}$, como $\bar{\mathbf{g}} = \nabla f(\bar{\mathbf{x}})$ temos

$-\nabla f(\bar{\mathbf{x}})^T \bar{\mathbf{g}} < 0 \Leftrightarrow -\bar{\mathbf{g}}^T \bar{\mathbf{g}} < 0$, ou seja, $-\bar{\mathbf{g}}$ é direção de descida. Tomando $\mathbf{x} = \bar{\mathbf{x}} - t\bar{\mathbf{g}}$ em

(33) temos

$$\begin{aligned} f(\bar{\mathbf{x}} - t\bar{\mathbf{g}}) - f(\bar{\mathbf{x}}) &= (\bar{\mathbf{x}} - t\bar{\mathbf{g}} - \bar{\mathbf{x}})^T \bar{\mathbf{g}} + \frac{1}{2} \|\mathbf{A}(\bar{\mathbf{x}} - t\bar{\mathbf{g}} - \bar{\mathbf{x}})\|^2 \\ &= (-t\bar{\mathbf{g}})^T \bar{\mathbf{g}} + \frac{1}{2} \|\mathbf{A}t\bar{\mathbf{g}}\|^2 \\ &= -t\bar{\mathbf{g}}^T \bar{\mathbf{g}} + \frac{1}{2} t^2 \|\mathbf{A}\bar{\mathbf{g}}\|^2 \\ &= t \left(-\|\bar{\mathbf{g}}\|^2 + \frac{1}{2} t \|\mathbf{A}\bar{\mathbf{g}}\|^2 \right) < 0; \end{aligned}$$

pois $t > 0$ e $-\|\bar{\mathbf{g}}\|^2 + \frac{1}{2}t\|\mathbf{A}\bar{\mathbf{g}}\|^2 < 0$, para t suficientemente pequeno. Portanto, $\bar{\mathbf{x}}$ não minimiza $f(\mathbf{x})$.

Provando agora a unicidade do ponto $\bar{\mathbf{x}}$. Vamos mostrar que se as colunas de \mathbf{A} são linearmente independentes, então existe um único ponto $\bar{\mathbf{x}}$ que minimiza $f(\mathbf{x})$.

Note que se as colunas de \mathbf{A} são linearmente independentes então o posto desta matriz é igual a n , o que implica que $\mathbf{A}^T\mathbf{A}$ é não singular, ou seja, $\mathbf{A}^T\mathbf{A}$ admite inversa.

Como $\bar{\mathbf{x}}$ minimiza $f(\mathbf{x})$, já provamos que $\mathbf{A}^T\mathbf{A}\bar{\mathbf{x}} = \mathbf{A}^T\mathbf{b}$.

Assim,

$$\begin{aligned} (\mathbf{A}^T\mathbf{A})^{-1}(\mathbf{A}^T\mathbf{A})\bar{\mathbf{x}} &= (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b} \\ \Rightarrow \bar{\mathbf{x}} &= (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}. \end{aligned}$$

Portanto, $\bar{\mathbf{x}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$ de maneira única.

Vamos provar agora que se existe um único ponto $\bar{\mathbf{x}}$ que minimiza $f(\mathbf{x})$ então as colunas de \mathbf{A} são linearmente independentes.

Se existe um único ponto $\bar{\mathbf{x}}$ que minimiza $f(\mathbf{x})$, vimos que $\mathbf{A}^T\mathbf{A}\bar{\mathbf{x}} = \mathbf{A}^T\mathbf{b}$.

Suponha, por contradição, que \mathbf{A} não tenha posto completo. Então $\exists \mathbf{y} \neq \mathbf{0}$ tal que $\mathbf{A}\mathbf{y} = \mathbf{0}$ e portanto $\mathbf{A}^T(\mathbf{A}\mathbf{y}) = \mathbf{0}$. Vemos então que $\mathbf{A}^T\mathbf{A}\bar{\mathbf{x}} + \mathbf{A}^T\mathbf{A}\mathbf{y} = \mathbf{A}^T\mathbf{b} \Rightarrow \mathbf{A}^T\mathbf{A}(\bar{\mathbf{x}} + \mathbf{y}) = \mathbf{A}^T\mathbf{b}$, ou seja, $\nabla f(\bar{\mathbf{x}}) = \nabla f(\bar{\mathbf{x}} + \mathbf{y}) = 0$ e $f(\bar{\mathbf{x}}) = f(\bar{\mathbf{x}} + \mathbf{y})$. Logo, o ponto $(\bar{\mathbf{x}} + \mathbf{y})$ também seria um minimizador para $f(\mathbf{x})$, o que contraria a hipótese de unicidade de $\bar{\mathbf{x}}$.

Portanto, \mathbf{A} tem posto completo, ou seja, suas colunas são linearmente independentes.

□

Prova do Lema 3.2

Com $f(\mathbf{x}) = \frac{1}{2} \|\mathbf{Ax} - \mathbf{b}\|^2$ e $\bar{\mathbf{g}} = \mathbf{A}^T (\mathbf{A}\bar{\mathbf{x}} - \mathbf{b})$ como acima, o problema (30) torna-se:

minimizar $f(\mathbf{x})$ sujeito a $\mathbf{x} \geq \mathbf{0}$,

enquanto as três condições do Lema 3.2 tornam-se:

$$(34) \quad \bar{\mathbf{x}} \geq \mathbf{0}, \quad \bar{\mathbf{g}} \geq \mathbf{0} \text{ e } \bar{\mathbf{x}}^T \bar{\mathbf{g}} = 0.$$

Primeiro, vamos assumir que todas as condições (34) são satisfeitas. Então (33) implica que

$$f(\mathbf{x}) - f(\bar{\mathbf{x}}) = \mathbf{x}^T \bar{\mathbf{g}} - \bar{\mathbf{x}}^T \bar{\mathbf{g}} + \frac{1}{2} \|\mathbf{A}(\mathbf{x} - \bar{\mathbf{x}})\|^2 = \mathbf{x}^T \bar{\mathbf{g}} + \frac{1}{2} \|\mathbf{A}(\mathbf{x} - \bar{\mathbf{x}})\|^2 \geq \mathbf{x}^T \bar{\mathbf{g}} \geq 0 \text{ para todo } \mathbf{x} \geq \mathbf{0},$$

o que mostra que $\bar{\mathbf{x}}$ é uma solução ótima do problema (30).

Em seguida, suponhamos que $\bar{\mathbf{x}} < \mathbf{0}$. Então $\bar{\mathbf{x}}$ é inviável e, portanto, não é ótimo para (30).

Em seguida, suponhamos que $\bar{\mathbf{x}} \geq \mathbf{0}$, mas $\bar{g}_k < 0$ para algum k .

Então existe um escalar t tal que $t > 0$ e $\frac{1}{2} t \|\mathbf{Ae}_k\|^2 < -\bar{g}_k$, onde

$\mathbf{e}_k^T = (0, \dots, 1, \dots, 0) \in \mathbb{R}^n$ com 1 na k -ésima posição.

Agora, $\bar{\mathbf{x}} + t\mathbf{e}_k \geq \mathbf{0}$ e, por (33),

$$\begin{aligned} f(\bar{\mathbf{x}} + t\mathbf{e}_k) - f(\bar{\mathbf{x}}) &= (\bar{\mathbf{x}} + t\mathbf{e}_k - \bar{\mathbf{x}})^T \bar{\mathbf{g}} + \frac{1}{2} \|\mathbf{A}(\bar{\mathbf{x}} + t\mathbf{e}_k - \bar{\mathbf{x}})\|^2 \\ &= t\mathbf{e}_k^T \bar{\mathbf{g}} + \frac{1}{2} t^2 \|\mathbf{Ae}_k\|^2 \\ &= t \left(\bar{g}_k + \frac{1}{2} t \|\mathbf{Ae}_k\|^2 \right) < 0, \end{aligned}$$

pois $t > 0$ e $\bar{g}_k + \frac{1}{2} t \|\mathbf{Ae}_k\|^2 < 0$, mostrando que $\bar{\mathbf{x}}$ não é uma solução ótima do problema (30).

Finalmente, vamos assumir que $\bar{\mathbf{x}} \geq \mathbf{0}$ e $\bar{\mathbf{g}} \geq \mathbf{0}$, mas $\bar{x}_k > 0$ e $\bar{g}_k > 0$ para algum k . Então existe um escalar t tal que $0 < t < \bar{x}_k$ e $\frac{1}{2}t\|\mathbf{A}\mathbf{e}_k\|^2 < \bar{g}_k$.

Agora, $\bar{\mathbf{x}} - t\mathbf{e}_k \geq \mathbf{0}$ e, por (33),

$$\begin{aligned} f(\bar{\mathbf{x}} - t\mathbf{e}_k) - f(\bar{\mathbf{x}}) &= (\bar{\mathbf{x}} - t\mathbf{e}_k - \bar{\mathbf{x}})^T \bar{\mathbf{g}} + \frac{1}{2} \|\mathbf{A}(\bar{\mathbf{x}} - t\mathbf{e}_k - \bar{\mathbf{x}})\|^2 \\ &= -t\mathbf{e}_k^T \bar{\mathbf{g}} + \frac{1}{2} t^2 \|\mathbf{A}\mathbf{e}_k\|^2 \\ &= t \left(-\bar{g}_k + \frac{1}{2} t \|\mathbf{A}\mathbf{e}_k\|^2 \right) < 0, \end{aligned}$$

pois $t > 0$ e $-\bar{g}_k + \frac{1}{2}t\|\mathbf{A}\mathbf{e}_k\|^2 < 0$, mostrando que $\bar{\mathbf{x}}$ não é uma solução ótima do problema (30).

□

A prova do Lema 3.1

Agora, vamos apresentar uma prova elementar do Lema 3.1. Primeiramente algumas notações:

Suponha que σ é um subconjunto do conjunto de índices $\{1, \dots, n\}$.

Vamos indicar o número de índices de σ por $|\sigma|$. Seja \mathbf{A}_σ a matriz $m \times |\sigma|$ com colunas $\{\mathbf{a}_{\cdot j}\}_{j \in \sigma}$, onde $\mathbf{a}_{\cdot j}$ é a j -ésima coluna de \mathbf{A} , e seja $\mathbf{x}_\sigma \in \mathbb{R}^{|\sigma|}$ um vetor com componentes $\{x_j\}_{j \in \sigma}$ (na mesma ordem das colunas de \mathbf{A}_σ).

Além disso, sejam as regiões X_σ e X_σ^+ definidas por

$$X_\sigma = \{\mathbf{x}^T = (x_1, \dots, x_n); x_j = 0 \text{ para } j \notin \sigma\} \text{ e}$$

$$X_\sigma^+ = \{\mathbf{x}^T = (x_1, \dots, x_n); x_j > 0 \text{ para } j \in \sigma, x_j = 0 \text{ para } j \notin \sigma\}.$$

Note que existem 2^n diferentes subconjuntos σ de $\{1, \dots, n\}$. As correspondentes 2^n diferentes regiões X_σ^+ são disjuntas duas a duas, e sua união é igual a toda a região viável do problema (30), já que para cada ponto $\mathbf{x} \geq \mathbf{0}$ existe um

único subconjunto σ tal que $\mathbf{x} \in X_\sigma^+$ (o subconjunto σ definido por $x_j > 0$ para $j \in \sigma$ e $x_j = 0$ para $j \notin \sigma$).

Seja P_σ o seguinte problema de quadrados mínimos com algumas variáveis fixas em zero:

$$P_\sigma : \text{minimizar } f(\mathbf{x}) = \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2, \text{ sujeito a } \mathbf{x} \in X_\sigma.$$

Se $\sigma = \emptyset$ então $X_\sigma^+ = X_\sigma = \{\mathbf{0}\}$ e $\bar{\mathbf{x}} = \mathbf{0}$ é a única solução ótima de P_σ .

Se $\sigma \neq \emptyset$ então existe uma única solução ótima de P_σ se e somente se as colunas de \mathbf{A}_σ são linearmente independentes. Isso decorre do Lema 3.3, uma vez que para resolver P_σ basta deixar $x_j = 0$ para $j \notin \sigma$ e minimizar $\frac{1}{2} \|\mathbf{A}_\sigma \mathbf{x}_\sigma - \mathbf{b}\|^2$ com relação a \mathbf{x}_σ , isto é, resolver o sistema normal reduzido $\mathbf{A}_\sigma^T \mathbf{A}_\sigma \mathbf{x}_\sigma = \mathbf{A}_\sigma^T \mathbf{b}$.

Definição 3.1

O subconjunto σ é chamado *subconjunto interessante* se P_σ tem uma única solução ótima $\bar{\mathbf{x}}$, e esta solução satisfaz $\bar{\mathbf{x}} \in X_\sigma^+$. Neste caso, $\bar{\mathbf{x}}$ é chamado de *ponto interessante*.

Caso contrário, se P_σ não tem uma única solução ótima, ou se P_σ tem uma única solução ótima $\bar{\mathbf{x}}$, mas $\bar{\mathbf{x}} \notin X_\sigma^+$, o subconjunto σ é chamado *subconjunto amplo*^{*}. Neste caso, não existe um ponto interessante correspondente a σ .

* No texto original, este subconjunto foi denominado “boring”, e poderia ser traduzido de maneira mais literal para “desinteressante” ou “chato”. Preferimos a nomenclatura de *subconjunto amplo* pois a fronteira da região X_σ^+ correspondente pode conter um ponto interessante, conforme veremos no Lema 3.4.

Já que $\mathbf{0}$ é sempre um ponto interessante (correspondente a $\sigma = \emptyset$), e existe no máximo um ponto interessante para cada subconjunto σ , sempre teremos pelo menos um e no máximo 2^n pontos interessantes. Nós vamos mostrar que pelo menos um desses pontos interessantes é uma solução ótima do problema (30).

Lema 3.4

Suponha que σ é um subconjunto interessante e que $\bar{\mathbf{x}}$ é o correspondente ponto interessante. Então $f(\bar{\mathbf{x}}) \leq f(\mathbf{x})$ para todo $\mathbf{x} \in X_\sigma^+$.

Suponha que σ é um subconjunto amplo. Então, para cada ponto dado $\mathbf{x} \in X_\sigma^+$, existem um subconjunto estrito $\bar{\sigma} \subset \sigma$ e um ponto $\bar{\mathbf{x}} \in X_{\bar{\sigma}}^+$ tais que $f(\bar{\mathbf{x}}) \leq f(\mathbf{x})$.

Demonstração: Se σ é um subconjunto interessante e $\bar{\mathbf{x}}$ é o ponto interessante correspondente, então $f(\bar{\mathbf{x}}) \leq f(\mathbf{x})$ para todo $\mathbf{x} \in X_\sigma$. Portanto, $f(\bar{\mathbf{x}}) \leq f(\mathbf{x})$ para todo $\mathbf{x} \in X_\sigma^+$, pois X_σ^+ é um subconjunto de X_σ .

Suponha agora que σ é um subconjunto amplo tal que as colunas de \mathbf{A}_σ são linearmente dependentes. Então existe um $\mathbf{d} \in X_\sigma$, com pelo menos uma componente estritamente negativa, tal que $\mathbf{A}\mathbf{d} = \mathbf{0}$.

Suponha que $\mathbf{x} \in X_\sigma^+$. Seja $\bar{t} = \min_j \{x_j / (-d_j); d_j < 0\}$. Note que $\bar{t} > 0$.

Então $\mathbf{x} + t\mathbf{d} \in X_\sigma^+$ para todo $t \in [0, \bar{t}]$, enquanto $\mathbf{x} + \bar{t}\mathbf{d} \in X_{\bar{\sigma}}^+$ para algum subconjunto estrito $\bar{\sigma} \subset \sigma$, já que $x_j + \bar{t}d_j = 0$ para pelo menos um $j \in \sigma$.

Além disso, $f(\mathbf{x} + t\mathbf{d}) = \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{b} + t\mathbf{A}\mathbf{d}\|^2 = \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 = f(\mathbf{x})$ para todo t .

Assim, se $\bar{\mathbf{x}} = \mathbf{x} + \bar{t}\mathbf{d}$ então $f(\bar{\mathbf{x}}) = f(\mathbf{x})$ e $\bar{\mathbf{x}} \in X_{\bar{\sigma}}^+$ com $\bar{\sigma} \subset \sigma$.

Suponha agora que σ é um subconjunto amplo tal que as colunas de A_σ são linearmente independentes, mas $\bar{x}_j \leq 0$ para pelo menos um $j \in \sigma$, onde \bar{x} é a única solução ótima de P_σ .

Suponha que $\mathbf{x} \in X_\sigma^+$. Seja $\bar{t} = \min_j \{x_j / (x_j - \bar{x}_j); x_j > \bar{x}_j\}$. Note que $\bar{t} \in]0,1[$.

Então $\mathbf{x} + t(\bar{\mathbf{x}} - \mathbf{x}) \in X_\sigma^+$ para todo $t \in [0, \bar{t}[$, enquanto $\mathbf{x} + \bar{t}(\bar{\mathbf{x}} - \mathbf{x}) \in X_\sigma^+$ para algum subconjunto estrito $\bar{\sigma} \subset \sigma$, já que $x_j + \bar{t}(\bar{x}_j - x_j) = 0$ para pelo menos um $j \in \sigma$. Além disso, a função quadrática $f(\mathbf{x} + t(\bar{\mathbf{x}} - \mathbf{x}))$, de variável t , é minimizada unicamente por $t=1$ e, portanto, $f(\mathbf{x} + t(\bar{\mathbf{x}} - \mathbf{x})) < f(\mathbf{x})$ para todo $t \in]0,1[$. Em particular, $f(\mathbf{x} + \bar{t}(\bar{\mathbf{x}} - \mathbf{x})) < f(\mathbf{x})$.

Assim, se $\bar{\mathbf{x}} = \mathbf{x} + \bar{t}(\bar{\mathbf{x}} - \mathbf{x})$ então $f(\bar{\mathbf{x}}) \leq f(\mathbf{x})$ e $\bar{\mathbf{x}} \in X_\sigma^+$ com $\bar{\sigma} \subset \sigma$.

□

Definição 3.2

Um ponto interessante $\bar{\mathbf{x}}$ é chamado *melhor ponto interessante* se $f(\bar{\mathbf{x}}) \leq f(\mathbf{x})$ para todo ponto interessante \mathbf{x} .

Note que sempre existe pelo menos um melhor ponto interessante, pois sempre há pelo menos um e no máximo 2^n pontos interessantes.

Lema 3.5

Se $\bar{\mathbf{x}}$ é um melhor ponto interessante e σ é um subconjunto interessante qualquer, então $f(\bar{\mathbf{x}}) \leq f(\mathbf{x})$ para cada $\mathbf{x} \in X_\sigma^+$.

Demonstração: Suponhamos que $\mathbf{x} \in X_\sigma^+$, quando σ é um subconjunto interessante. Já que σ é um subconjunto interessante, há um ponto interessante correspondente $\mathbf{x}^* \in X_\sigma^+$. Então, pelo Lema 3.4 $f(\mathbf{x}) \geq f(\mathbf{x}^*) \geq f(\bar{\mathbf{x}})$.

□

Lema 3.6

Se $\bar{\mathbf{x}}$ é o melhor ponto interessante e σ é qualquer subconjunto amplo então $f(\bar{\mathbf{x}}) \leq f(\mathbf{x})$ para cada $\mathbf{x} \in X_\sigma^+$.

Demonstração: Suponhamos que $\mathbf{x} \in X_\sigma^+$, quando σ é um subconjunto amplo. Então, pelo Lema 3.4, existem um subconjunto $\bar{\sigma} \subset \sigma$ e um ponto $\bar{\mathbf{x}} \in X_{\bar{\sigma}}^+$ tais que $f(\bar{\mathbf{x}}) \leq f(\mathbf{x})$.

Nota-se que $|\bar{\sigma}| \leq |\sigma| - 1$, de modo que existem menos componentes estritamente positivos em $\bar{\mathbf{x}}$ do que em \mathbf{x} . Se $\bar{\sigma}$ é um subconjunto interessante então, pelo Lema 3.5, $f(\bar{\mathbf{x}}) \leq f(\bar{\mathbf{x}}) \leq f(\mathbf{x})$, e este caso está provado.

Caso contrário, se $\bar{\sigma}$ é um subconjunto amplo, vamos repetir os argumentos acima, mas agora a partir do ponto $\bar{\mathbf{x}} \in X_{\bar{\sigma}}^+$ em vez do ponto $\mathbf{x} \in X_\sigma^+$. Como o número de componentes estritamente positivas do vetor das variáveis é estritamente decrescente para cada repetição, no máximo n repetições serão necessárias antes que um subconjunto interessante seja atingido (pois $\sigma = \emptyset$ é um subconjunto interessante), e então o Lema 3.5 pode ser aplicado.

□

Provando o Lema 3.1

Seja $\bar{\mathbf{x}}$ o melhor ponto interessante, e suponhamos que $\mathbf{x} \geq \mathbf{0}$. Então existe um único subconjunto σ tal que $\mathbf{x} \in X_\sigma^+$, ou seja, o subconjunto σ definido por $j \in \sigma$ para $x_j > 0$ e $j \notin \sigma$ para $x_j = 0$.

Se σ é um subconjunto interessante, decorre do Lema 3.5 que $f(\bar{\mathbf{x}}) \leq f(\mathbf{x})$, enquanto que, se σ é um subconjunto amplo, decorre do Lema 3.6, que $f(\bar{\mathbf{x}}) \leq f(\mathbf{x})$.

A conclusão é que $\bar{\mathbf{x}}$ é uma solução ótima do problema (30).

□

Apresentamos a seguir um Mapa Conceitual (Figura 22), o qual ilustra a sequência de lemas e definições adotada para a demonstração do Lema de Farkas, e como se interrelacionam:

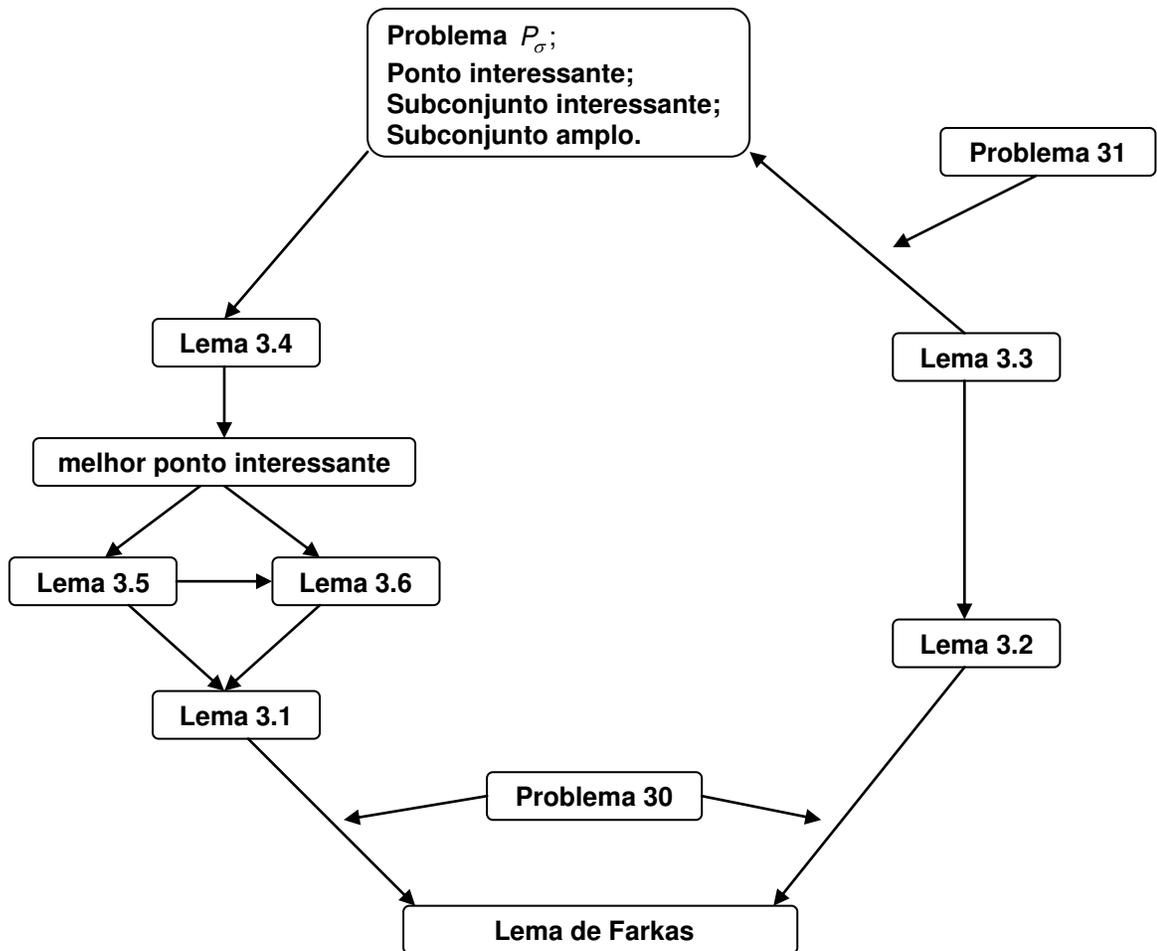


Figura 22 - Mapa Conceitual ilustrando a sequência de lemas e definições, adotada para a demonstração do Lema de Farkas.

Considerações finais e perspectivas futuras

Nessa pesquisa procuramos desenvolver um texto que contribuísse para uma melhor compreensão dos Problemas de Programação Linear (PPL) através da apresentação dos principais conceitos que os envolvem de forma objetiva e contextualizada. A aproximação do Cálculo à Álgebra Linear foi feita de forma natural, favorecendo a construção do conhecimento pelo leitor.

Iniciamos o texto apresentando os aspectos essenciais dos PPL e desenvolvemos a análise do conjunto viável através de linhas e colunas, o que foi muito útil para os estudos feitos nos capítulos seguintes.

A visão geométrica feita na Seção 1.3 ilustrou as diversas possibilidades de conjuntos viáveis, o que facilitou a introdução dos conceitos do Cálculo (vetor gradiente, derivada direcional, máximos e mínimos sobre conjunto compacto, etc.). Vimos que a utilização dos Multiplicadores de Lagrange não foi suficiente para conseguirmos resolver os PPL gerais, o que nos levou à busca de mais elementos para caracterizar as soluções.

Com o estudo de problemas lineares sujeitos a restrições gerais e a introdução de elementos da dualidade em PL chegamos à caracterização das soluções por meio de condições necessárias e suficientes de otimalidade, cuja demonstração emprega o Lema de Farkas. Para completar o texto incluímos uma prova para este conhecido resultado de alternativas, baseada em um trabalho recente de Svanberg que emprega Álgebra Linear e Cálculo.

Nesse estudo pudemos ampliar nossos conhecimentos matemáticos e também didáticos, pois tivemos como um dos objetivos, produzir um texto acessível a alunos que possuem conhecimentos básicos do Cálculo, Geometria e Álgebra Linear.

Como perspectivas futuras vemos a continuação dos estudos em um programa de Doutorado em Matemática Aplicada (na área de Otimização), com o objetivo de tornar os conteúdos matemáticos cada vez mais significativos e acessíveis aos estudantes.

Referências Bibliográficas

- [1] H. Anton, C. Rorres, *Álgebra Linear com Aplicações*. 8.ed. - Porto Alegre: Bookman, 2001.
- [2] M. S. Bazaraa, J. J. Jarvis & H. D. Sherali, *Linear Programming and Network Flows*. New York:John Wiley, 1990.
- [3] J. L. Boldrini, S. I. R Costa, V. L. Figueiredo, H. G. Wetzler, *Álgebra Linear*. 3.ed. - São Paulo: Harper & Row do Brasil, 1980.
- [4] A. Dax, *An elementary proof of Farkas' lemma*. SIAM Review, 39(3):503-507, 1997.
- [5] R. G. Eustáquio, *Condições de Otimalidade e de Qualificação para Problemas de Programação Não Linear*, Dissertação (Mestrado Acadêmico em Engenharia), UFPR, Curitiba, 2007.
Disponível em <http://www.ppgmne.ufpr.br/arquivos/diss/167.pdf> (último acesso em 02/07/2010).
- [6] J. Farkas. Über die Theorie der einfachen Ungleichungen. *Journal für die Reine und Angewandte Mathematik*, 124:1-27, 1901.
- [7] H. L. Guidorizzi, *Um Curso de Cálculo*, volume 1 e 2, 5.ed., [reimpr.]. - Rio de Janeiro: LTC, 2008.
- [8] E. L. Lima, *Análise Real*, volume 2, 3.ed., Rio de Janeiro: IMPA, 2007.
- [9] D. G. Luenberger, *Linear and Nonlinear Programming*, 3.ed., Reading, Massachussets: Addison-Wesley, 1984.

- [10] W. G. McCallum, D. Hughes-Hallett, A. M. Gleason et al, *Cálculo de Várias Variáveis*, tradução Elza F. Gomide. - São Paulo: Edgard Blücher Ltda, 1997.
- [11] N. Maculan Filho & M. V. F. Pereira, *Programação Linear*, São Paulo: Editora Atlas, 1980.
- [12] O.L. Mangasarian, *Nonlinear Programming*, Philadelphia: SIAM, 1994.
- [13] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*, Philadelphia:SIAM, 2000.
- [14] A. R. Monticeli, *Um estudo sobre sistemas de inequações lineares*, Dissertação (Mestrado Profissional em Matemática), IMECC, Unicamp, Campinas, 2010.
- [15] B. Noble, J. W. Daniel, *Álgebra Linear Aplicada*, 2.ed., tradução João Pitombeira de Carvalho. - Rio de Janeiro: Prentice/ Hall do Brasil, 1986.
- [16] J. Stewart, *Cálculo*, volume 2, 5.ed., tradução Antônio Carlos Moretti, Antônio Carlos Gilli Martins. - São Paulo: Thomson Learning, 2007.
- [17] K. Svanberg, *Farkas'lemma derived by elementary linear algebra*, Technical Report TRITA-MAT-2008-OS7, Department of Mathematics, Royal Institute of Technology (KTH), Stockholm, Sweden (6 p.).
Disponível em <http://www.math.kth.se/~krille/farkas.pdf> (último acesso em 17/06/2010).

Apêndice A

Vértices (pontos extremos)

A determinação dos vértices do conjunto viável dos Problemas de Programação Linear (PPL) é de grande importância, pois é neles que a função objetivo atinge seu valor máximo ou mínimo, quando existem.

Chamamos de *vértices* (pontos extremos) àqueles pertencentes à fronteira de uma região viável que são interseções de dois hiperplanos da fronteira.

A1. Teorema Fundamental da Programação Linear

Para mostrarmos a relevância que os vértices possuem na resolução de (PPL) vamos provar cf.[3, p368] o *Teorema Fundamental da Programação Linear*. Iniciaremos com a apresentação de dois lemas.

Lema 1A

Seja $f(x_1, \dots, x_n) = c_1x_1 + \dots + c_nx_n$ e seja P um ponto interior a um segmento \overline{AB} do \mathbb{R}^n , isto é, $P = \lambda A + (1 - \lambda)B$, $0 < \lambda < 1$.

Então teremos $f(A) \leq f(P) \leq f(B)$ ou $f(B) \leq f(P) \leq f(A)$.

Demonstração: Como $f(x_1, \dots, x_n) = c_1x_1 + \dots + c_nx_n$ é linear e $P = \lambda A + (1 - \lambda)B$, temos que $f(P) = f(\lambda A + (1 - \lambda)B) = \lambda f(A) + (1 - \lambda)f(B)$.

Suponhamos $f(A) \leq f(B)$.

Como $f(P) = \lambda f(A) + (1 - \lambda)f(B)$ temos

$$\lambda f(A) + (1 - \lambda)f(A) \leq f(P) \leq \lambda f(B) + (1 - \lambda)f(B).$$

Portanto, $f(A) \leq f(P) \leq f(B)$.

Da mesma forma, se tivermos $f(B) \leq f(A)$ mostramos que:

$$f(B) \leq f(P) \leq f(A).$$

□

O lema que acabamos de demonstrar nos diz que os valores extremos de uma função afim são assumidos nos pontos extremos dos segmentos.

Uma consequência deste resultado é o lema a seguir.

Lema 2A

Seja $f(x_1, \dots, x_n) = c_1x_1 + \dots + c_nx_n$. Se dentre os valores que f assumir num segmento \overline{AB} do \mathfrak{R}^n , o valor máximo (mínimo) for assumido num ponto P interior deste segmento, então f será constante em \overline{AB} .

Demonstração: Da demonstração anterior para o caso $f(A) \leq f(B)$ se tivermos $f(P) = f(B)$, obtemos:

$$\lambda f(A) + (1 - \lambda)f(B) = \lambda f(B) + (1 - \lambda)f(B),$$

$$f(A) = f(B).$$

Assim, $f(A) = f(B) = f(P)$, ou seja, f é constante em \overline{AB} .

Da mesma forma, se tivermos $f(B) = f(A)$ mostramos que $f(B) = f(A) = f(P)$, ou seja, f é constante em \overline{AB} .

Portanto, se dentre os valores que f assumir num segmento \overline{AB} do \mathfrak{R}^n , o valor máximo (mínimo) for assumido num ponto P interior deste segmento, então f será constante em \overline{AB} .

□

Agora, vamos provar o *Teorema Fundamental da Programação Linear*, reescrito a seguir para facilitar a leitura:

Teorema 1.1

Seja $f(x_1, \dots, x_n) = c_1x_1 + \dots + c_nx_n$ definida numa região poliedral convexa G do \mathbb{R}^n . Suponha que f assuma um valor máximo (mínimo) nesta região. Então, se G possui vértice(s), este valor máximo (mínimo) será assumido num vértice.

Demonstração (Para regiões G do \mathbb{R}^2 .): Supondo que o valor máximo (mínimo) de f seja assumido num ponto P de G .

Considerando todas as regiões poliedrais convexas possíveis do \mathbb{R}^2 , pode-se ter:

- i) P é um vértice. (Neste caso o teorema já estará provado.)
- ii) P está numa aresta. Do Lema 2A, f assumirá este valor máximo (mínimo) em toda aresta. Como por hipótese a região G possui vértice(s) esta aresta conterá um vértice v obrigatoriamente. Portanto, $f(P) = f(v)$.
- iii) P está no interior de G . Neste caso, f será constante em toda região G .

De fato:

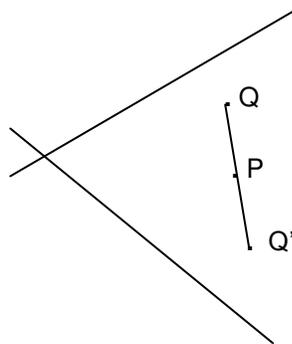


Figura 23 - Ilustração da região poliedral convexa G , com os pontos P , Q e Q' .

Seja Q outro ponto interior de G . Como G é poliedral convexa, o segmento QP está contido em G . Além disso, como P é interior, podemos considerar um prolongamento QQ' deste ainda contido em G . Do Lema 2A segue que f é constante em QQ' e, portanto, $f(P) = f(Q)$.

Observação 1A: A prova deste teorema para n -variáveis vai envolver um número maior de possibilidades: o ponto inicial de máximo (mínimo) poderá estar:

- i) num vértice;
- ii) numa aresta (solução de $(n-1)$ -equações). Neste caso, o máximo (mínimo) será assumido em toda a aresta que é um subconjunto de dimensão 1;
- iii) numa face (solução de $(n-2)$ -equações). Neste caso, o máximo (mínimo) será assumido em toda a face que é um subconjunto de dimensão 2;

.

.

.

ni) num ponto interior da região G (que tem dimensão n). Neste caso, a função será constante em toda região.

Pode se mostrar que, em todos estes casos, quando a região tem vértices, conseguimos um vértice v onde a função assume seu máximo (mínimo), $f(v) = f(P)$.

Observação 2A: É este teorema que permite, nos casos em que, pela natureza da função, já sabemos que ela assume máximo (mínimo) e encontrá-lo apenas determinando seus valores nos vértices da região poliedral convexa.

□

Assim, cf.[1, p.373] se a região viável de um PPL é não-vazia e limitada, então a função objetivo atinge tanto um valor máximo quanto um valor mínimo e estes ocorrem nos vértices da região viável. Se a região viável é ilimitada, então a função objetivo pode ou não atingir valores máximo ou mínimo; contudo, se atingir um máximo ou um mínimo, este ocorrerá em um vértice.

Apêndice B

Conjunto Compacto

De acordo com [8, p.5] vamos apresentar as definições que envolvem um conjunto compacto.

B1. Bola fechada e conjunto limitado

Dados o ponto $a \in \mathfrak{R}^n$ e o número real $r > 0$, a *bola fechada* de centro a e raio r é o conjunto $B[a; r]$ assim definido:

$$B[a; r] = \{x \in \mathfrak{R}^n; \|x - a\| \leq r\}.$$

Diz-se que o conjunto $X \subset \mathfrak{R}^n$ é *limitado* quando está contido em alguma bola $B[a; r]$.

Podemos dizer que um conjunto limitado é aquele que é finito em extensão.

B2. Conjunto fechado

Diz-se que o ponto a é *aderente* ao conjunto $X \subset \mathfrak{R}^n$ quando existe uma sequência de pontos $x_k \in X$ tais que $\lim x_k = a$.

Chama-se *fecho* do conjunto $X \subset \mathfrak{R}^n$ ao conjunto \bar{X} formado por todos os pontos aderentes a X . Portanto $a \in \bar{X} \Leftrightarrow a = \lim x_k, x_k \in X$.

Dizer que $a \in \bar{X}$ é o mesmo que afirmar que a é aderente a X .

Um conjunto $F \in \mathfrak{R}^n$ chama-se *fechado* quando $\bar{F} = F$, isto é, quando o limite de toda sequência convergente de pontos de F é ainda um ponto de F .

Todo ponto $x \in X$ é aderente a X , pois é limite da sequência constante (x, x, \dots) .

Como para os intervalos fechados os extremos do intervalo estão contidos no intervalo, um *conjunto fechado* de \mathfrak{R}^n contém todos os seus pontos da fronteira. Um ponto da fronteira de F é um ponto $x \in \mathfrak{R}^n$ tal que qualquer bola aberta com centro em x contém pontos de F e pontos não pertencentes a F . Por exemplo, a bola

$$F = \{(x_1, \dots, x_n); x_1^2 + \dots + x_n^2 \leq 1\}$$

constituída de todos os pontos sobre e dentro da bola $x_1^2 + \dots + x_n^2 = 1$, é um conjunto fechado porque contém todos os seus pontos da fronteira (que são os pontos sobre a bola $x_1^2 + \dots + x_n^2 = 1$). Se um único ponto da fronteira não pertencer ao conjunto, esse não será fechado.

B3. Conjunto compacto

Chamamos de *conjunto compacto* ao conjunto $C \subset \mathfrak{R}^n$ que é limitado e fechado.

Índice

bola fechada e conjunto limitado, 84
cone convexo, 7
conjunto compacto, 10, 84, 85
conjunto fechado, 84, 85
conjunto viável, 4
derivada direcional e gradiente, 15, 16
dualidade, 44
espaço de exigência, 38
forma canônica da dualidade, 50, 51
forma padrão da dualidade, 49
função objetivo, 3
Lagrange, 18, 19, 20, 21
Lema de Farkas, 55, 64
poliedro, 7
politopo, 7
posto de uma matriz, 8
região poliedral convexa fechada, 10
Teorema
 Fundamental da Programação Linear, 10, 82
 de Weierstrass, 17
variáveis de folga, 3
vértices, 10, 80
vetor gradiente e as curvas de nível, 16, 17
vetor gradiente, 14