



Universidade Estadual de Campinas  
Instituto de Matemática, Estatística e  
Computação Científica - IMECC



Dissertação de Mestrado

*O uso da Programação Linear na  
separação de pontos*

por

**Eberson Paulo Trevisan**

Mestrado Profissional em Matemática

Orientadora: **Prof<sup>a</sup>. Dr<sup>a</sup>. Valéria Abrão de Podestá**

Junho de 2010  
Campinas - SP

# O uso da Programação Linear na separação de pontos

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por **Eberson Paulo Trevisan** e aprovada pela comissão julgadora.

Campinas, 23 de junho de 2010.



---

Prof<sup>a</sup>. Dr<sup>a</sup>. Valéria Abrão de Podestá  
Orientadora

Banca Examinadora:

**Prof<sup>a</sup>. Dr<sup>a</sup>. Valéria Abrão de Podestá** (IMECC - UNICAMP)

**Prof. Dr. Aurélio Ribeiro Leite de Oliveira** (IMECC - UNICAMP)

**Prof<sup>a</sup>. Dr<sup>a</sup>. Marta Ines Velazco Fontova** (FACCAMP)

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

**FICHA CATALOGRÁFICA ELABORADA PELA  
BIBLIOTECA DO IMECC DA UNICAMP**  
Bibliotecária: Miriam Cristina Alves – CRB8 / 5089

Trevisan, Eberson Paulo

T729u O uso da programação linear na separação de pontos/Eberson Paulo  
Trevisan -- Campinas, [S.P. : s.n.], 2010.

Orientadora : Valéria Abrão de Podestá

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

1. Programação Linear. 2. Simplex (Matemática). 3. Otimização  
matemática. I. Podestá, Valéria Abrão de. II. Universidade Estadual de  
Campinas. Instituto de Matemática, estatística e Computação Científica.  
III. Título.

Titulo em inglês: The use of linear programming in patterns separation.

Palavras-chave em inglês (Keywords): 1. Linear Programming 2. Simplex (Mathematics).  
3. Mathematical optimization.

Área de concentração: Programação Linear

Titulação: Mestre em Matemática

Banca examinadora: Profa. Dra. Valéria Abrão de Podestá (IMECC-UNICAMP)  
Prof. Dr. Aurélio Ribeiro Leite de Oliveira (IMECC - UNICAMP)  
Profa. Dra. Marta Inês Velasco Fontova (FACCAMP)

Data da defesa: 23/06/2010

Programa de Pós-Graduação: Mestrado Profissional em Matemática

**Dissertação de Mestrado Profissional defendida em 23 de junho de 2010 e  
aprovada pela Banca Examinadora composta pelos Profs. Drs.**



---

**Prof. (a). Dr (a). VALÉRIA ABRÃO DE PODESTÁ**



---

**Prof. (a). Dr (a). AURELIO RIBEIRO LEITE DE OLIVEIRA**



---

**Prof. (a). Dr (a). MARTA INES VELAZCO FONTOVA**

Dedico esse trabalho  
à minha esposa Andreia,  
grande amor da minha vida.

---

# Agradecimentos

---

Uma coisa é fato na vida, “*sonhos não se realizam sozinhos*”. Na realização deste sonho, a conclusão deste mestrado, algumas pessoas contribuíram de forma significativa. A todas essas esse singelo tributo, com certeza, é pouco para expressar o tamanho de minha gratidão.

Em primeiro lugar agradeço a Deus pela vida, companhia e pelas portas abertas pelas mãos Divinas.

A Prof.<sup>a</sup> Dr.<sup>a</sup> Valéria Abrão de Podestá, pela orientação desse trabalho e grandioso auxílio na implementação dos programas computacionais aqui utilizados, e principalmente pela paciência e extrema didática com que me conduzia frente aos erros cometidos durante essa *caminhada*.

Aos professores Dr. Aurélio Ribeiro Leite de Oliveira e Prof<sup>a</sup>. Dra. Marta Ines Velazco Fontova, pela participação na banca examinadora desse trabalho e pelas valiosas sugestões.

Ao professor Miguel Tadayuki Koga da UNEMAT pelo apoio e amizade, e por me hospedar em sua casa em Campinas nos primeiros módulos desse curso.

Minha esposa Andreia. Com certeza não teria conseguido terminar essa etapa sem seu apoio e entusiasmo, que sempre foram animadores, mesmo com você triste pelas minhas ausências em decorrência das viagens para realizar o curso. Como você mesmo costuma dizer, “*Você é meu porto seguro*”. Esse mestrado compartilho contigo.

A todos os meus colegas de mestrado, em especial a amiga Adriana, conterrânea de Mato Grosso, a quem agradeço muito pelo companheirismo durante essa caminhada.

À direção e colegas de trabalho da escola Rosmay Kara José, pelo apoio e compreensão durante meus afastamentos para realizar esse curso.

---

# Resumo

---

Neste trabalho são apresentados alguns assuntos iniciais da teoria de Programação Linear e o método Simplex. Mostramos também como a Programação Linear pode ser utilizada na separação de dois conjuntos de pontos (padrões), através de um modelo linear cuja solução é um hiperplano separador. Finalizamos o trabalho com a apresentação de alguns exemplos de aplicação da Programação Linear na separação de dois conjuntos linearmente separáveis e linearmente inseparáveis.

**Palavras-chave:** Programação Linear, Método Simplex, Separabilidade e Inseparabilidade Linear.

---

# Abstract

---

In this work we present some introductory issues from Linear Programming theory and the Simplex method. We also show how we can use Linear Programming in two patterns separation by constructing a linear model which solution is a separating hyperplane. Finally, we also present some examples of Linear Programming application in the linear separability and inseparability of two patterns sets.

**Keywords:** Linear Programming, Simplex Method, Linear Separability and Inseparability.

---

# Sumário

---

<b>Resumo</b>	<b>ix</b>
<b>Abstract</b>	<b>xi</b>
<b>Introdução</b>	<b>1</b>
<b>1 Introdução à Programação Linear</b>	<b>3</b>
1.1 Introdução e síntese histórica . . . . .	3
1.2 Problemas de Programação Linear: definições e propriedades . . . . .	5
1.2.1 Diferentes formas para representação de PPL . . . . .	8
1.2.2 Equivalência entre PPL's . . . . .	9
1.2.3 Pressupostos dos problemas de Programação Linear . . . . .	13
1.3 Resolução gráfica de problemas de Programação Linear . . . . .	14
1.3.1 Tipos de soluções de um PPL . . . . .	16
<b>2 O Método Simplex</b>	<b>19</b>
2.1 Fundamentação teórica do método Simplex . . . . .	19
2.1.1 Convexidade do conjunto de soluções de um PPL . . . . .	21
2.1.2 Solução básica de um PPL . . . . .	22
2.1.3 Teoremas fundamentais relacionados ao método Simplex . . . . .	25
2.2 O Método Simplex . . . . .	25
2.2.1 Teste de otimalidade e critério para entrada na base . . . . .	27
2.2.2 Critério para saída da base ou Teste da Razão . . . . .	29
2.2.3 Nova SBV - atualização . . . . .	31
2.2.4 O algoritmo Simplex . . . . .	32

<b>3</b>	<b>Separação de Padrões (Pontos) Via Programação Linear</b>	<b>43</b>
3.1	Introdução e definições preliminares . . . . .	43
3.2	Critérios de separabilidade linear . . . . .	46
3.3	Formulação de um Programa Linear para separação de padrões . . . . .	53
3.3.1	Algumas particularidades do problema . . . . .	61
<b>4</b>	<b>Exemplos de separação de pontos</b>	<b>65</b>
4.1	Exemplos de separação de pontos . . . . .	65
4.1.1	Exemplos com conjuntos linearmente separáveis . . . . .	68
4.1.2	Exemplos com conjuntos linearmente inseparáveis . . . . .	76
4.1.3	Separação por mais de um hiperplano . . . . .	82
4.2	Uma Aplicação da PL na separação de pontos . . . . .	91
	<b>Considerações Finais e Perspectivas Futuras</b>	<b>95</b>
<b>A</b>	<b>Apêndice 1 - Equivalente ao Teorema (3.1) usando Dualidade</b>	<b>99</b>
	<b>Referências Bibliográficas</b>	<b>101</b>

---

# Introdução

---

Dentro dos propósitos desse programa, o objetivo dessa dissertação é discutir a aplicação da Programação Linear na separação de pontos do  $\mathbb{R}^n$ , através de hiperplanos. A Programação Linear é um elemento da Pesquisa Operacional, trata-se de um ramo interdisciplinar da matemática aplicada, tendo aplicações em vários setores como: administração, economia, engenharias, física, biologia, medicina, e é claro na própria matemática, entre outras [11].

O problema de separar pontos via Programação Linear tem sido foco de estudo de alguns matemáticos há mais de 40 anos, como pode ser visto na cronologia dos artigos [14], [15]. Tem aplicação em vários ramos, como diagnóstico de câncer de mama [16], e *Authorship Determination* (Determinação de autoria) [9]. No entanto, não encontramos bibliografias em português que abordasse o assunto.

Para abordar o problema da separação de pontos por Programação Linear, faz-se necessário um pouco de conhecimento sobre a mesma. Contudo, durante minha formação acadêmica não cursei nenhuma disciplina propriamente dita sobre o assunto. Por isso foi necessário um bom tempo de estudo, reflexão e resolução de exercícios sobre o tema. Parte elementar disso é apresentada nos dois primeiros capítulos.

Esse trabalho está organizado em quatro capítulos assim distribuídos:

No Capítulo 1 apresentamos uma introdução à Programação Linear. Iniciamos com uma síntese histórica de seu desenvolvimento e apresentamos as principais definições, propriedades e alguns pressupostos. Finalizamos o capítulo com a discussão sobre a

solução gráfica de problemas lineares.

No Capítulo 2 apresentamos de modo sucinto o método Simplex. Esse capítulo está dividido em duas partes: na primeira, discutimos a fundamentação teórica do método Simplex, onde são apresentados fundamentos matemáticos sobre convexidade, solução básica e os principais teoremas relacionados ao método. Nessa parte também são apresentados elementos básicos de álgebra linear, sem maior detalhamento. A segunda parte aborda o método Simplex propriamente dito, a partir da fundamentação da primeira parte.

No Capítulo 3 abordamos a separação de pontos, ou padrões, via Programação Linear. Apresentamos algumas definições necessárias, estabelecemos alguns critérios para separabilidade linear de dois conjuntos e apresentamos as discussões, a respeito da formulação de um único problema de Programação Linear, capaz de gerar um hiperplano separador para dois conjuntos de pontos. Apresentamos também algumas propriedades e particularidades desse problema, como por exemplo, a obtenção de uma solução ótima sempre, o que conseqüentemente sempre irá gerar um hiperplano separador para os casos linearmente separáveis ou inseparáveis.

O Capítulo 4 foi destinado apenas para a apresentação de exemplos, onde aplicamos e resolvemos o problema de Programação Linear proposto para separação de pontos no Capítulo 3, a conjuntos linearmente separáveis e inseparáveis, discutindo algumas particularidades encontradas nos mesmos. Finalizamos o capítulo com uma breve apresentação da aplicação da separação de pontos no diagnóstico do câncer de mama discutido no artigo [16].

---

# INTRODUÇÃO À PROGRAMAÇÃO LINEAR

---

O objetivo desse capítulo é discutir previamente os chamados Problemas de Programação Linear (PPL's), algumas definições importantes a respeito dos mesmos, bem como suas principais propriedades, servindo assim de base para discussões futuras nos próximos capítulos.

## 1.1 Introdução e síntese histórica

A Programação Linear (PL) é uma técnica amplamente difundida para resolver problemas de otimização, além de ser um elemento importantíssimo na Pesquisa Operacional (PO). Tem suas aplicações, como ressaltam Maculan e Fampa [12], comuns em quase todos os setores do cotidiano, como por exemplo, nas indústrias, nos transportes, na saúde, na educação, na agricultura, nas finanças, na economia, nas administrações públicas e na medicina, como é visto em [15], [16], artigos que impulsionaram esse trabalho.

O desenvolvimento da Programação Linear, no início do século passado, foi estimulado basicamente por dois trabalhos científicos: o primeiro foi a publicação do teorema central da teoria dos jogos por John von Neumann em 1928 e o segundo foi a análise insumo-produto, proposta por Leontief, em seu artigo de 1936. Esse trabalho, intitulado *Quantitative input and output Relations in the Economic Systems of the United*

*States*, tratava-se na verdade de um modelo matricial linear que posteriormente veio a ser utilizado na forma de um PPL.

Um crescente interesse em obter um método geral para a solução de PPL's ocorreu nas décadas de 30 e 40 do século passado, após a formulação de alguns problemas sob a forma de restrições. Um dos primeiros a surgir foi o chamado PPL para organização e planejamento de produção, que foi formulado e resolvido pelo matemático e economista soviético Leonid V. Kantorovitch em 1939. Graças a esse trabalho, Kantorovitch recebeu o prêmio Nobel de economia em 1975. Em 1941, Hitchcock formulou o PPL de Transporte; já em 1945, Stigler formulou o PPL da Dieta.

Em 1947, George Dantzig e sua equipe da SCOOP<sup>1</sup>, forneceram os resultados teóricos não apenas de um modelo matemático inicial de um PPL, mas também um método de solução para tais problemas que foi denominado método Simplex. Com isso houve um crescente aumento no interesse por Programação Linear por parte dos matemáticos e economistas da época.

O interesse foi tanto que rapidamente novos avanços vieram, como por exemplo em 1951, Tucker obteve os primeiros resultados na teoria da dualidade que se tornou uma parte muito importante da PL. Um ano mais tarde, em 1952, Charnes e Lemkes desenvolveram o método Simplex Revisado, que foi muito importante para a implementação computacional do método Simplex [11].

Não podemos deixar de citar os trabalhos de Khachiyan em 1979 que, utilizando os métodos dos elipsóides, propôs um algoritmo no qual o número de iterações para resolver um PPL é limitado por uma função polinomial. Porém, conforme [12], seu desempenho prático deixa bastante a desejar. A publicação de um novo algoritmo por Karmarkar, em 1984, deu início a outra importante linha de pesquisa em PL - os *méto-*

---

<sup>1</sup> SCOOP é a abreviação do nome da equipe do projeto *Scientific Computation of Optimal Programs*, montado pela força aérea americana durante a 2<sup>a</sup> guerra, responsável por pesquisar modelos matemáticos para representar principalmente a indústria de serviços bélicos. Para maiores informações ver [11].

*dos de pontos interiores* [11]. Esses métodos vieram a partir da busca por um algoritmo que tivesse tempo de computação polinomial, já que Klee e Minty *apud* [12] ilustraram que o algoritmo Simplex pode apresentar um crescimento exponencial no tempo de computação em seu pior caso. No entanto, na prática, como é ressaltado em Lins e Calôba [11], o algoritmo Simplex exibe um tempo de computação polinomial de baixa ordem.

## 1.2 Problemas de Programação Linear: definições e propriedades

Um problema de Programação Linear (PPL) pode ser definido como um problema de otimização com função objetivo (FO) linear, onde a função está sujeita a um conjunto de restrições compostas por equações e/ou inequações também lineares. Além disso, as variáveis da função objetivo e das restrições são variáveis reais. Nesse sentido, de maneira geral, podemos definir um PPL como segue abaixo.

**Definição 1.1** (Definição geral de um PPL). *Um PPL é um problema de otimização que consiste em encontrar os valores das variáveis  $x_1, x_2, \dots, x_n$  que minimizam (ou maximizam) a função objetivo*

$$z = c_1x_1 + c_2x_2 + \dots + c_nx_n \quad (1.1)$$

*Sujeito a:*

$$\begin{cases} a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n = b_i, & i = 1, \dots, p \\ a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n \leq b_i, & i = p + 1, \dots, q \\ a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n \geq b_i, & i = q + 1, \dots, m \\ x_j \geq 0, & j = 1, 2, \dots, n \end{cases} \quad (1.2)$$

onde:

$x_j$ : nível da atividade  $j$  a ser determinada pela solução do PPL (variáveis);

$c_j$ : custo unitário associado à atividade  $j$ ;

$c_j x_j$ : custo total da atividade  $j$ ;

$b_i$ : valor limitante referente à restrição  $i$  (lado direito);

$a_{ij}$ : coeficiente que representa a contribuição unitária da atividade  $j$  à restrição  $i$ .

Naturalmente, nem todos os tipos de restrições precisam existir em um PPL. A função objetivo também é conhecida como função econômica ou função critério. As restrições  $x_j \geq 0$ ,  $j = 1, 2, \dots, n$ , são conhecidas como **restrições de não-negatividade**. Vamos definir ainda o que vem a ser uma solução viável de um PPL e uma solução ótima.

**Nota:** Usaremos a seguinte notação:  $x \in \mathbb{R}^n$  é um vetor coluna ( $n \times 1$ ) e  $x'$  é um vetor linha ( $1 \times n$ ).

**Definição 1.2** (Solução viável). *Um vetor  $x \in \mathbb{R}^n$  que satisfaz todas as restrições de um PPL é uma **solução viável** para esse PPL.*

**Definição 1.3.** *O conjunto de todos os vetores que satisfazem as restrições de um PPL é dito **conjunto das soluções viáveis**.*

Em um problema de Programação Linear que envolva apenas duas variáveis é sempre possível representar o conjunto das soluções viáveis através de uma região convexa<sup>2</sup> do plano. Ilustramos a seguir um exemplo de PPL que envolve duas variáveis e a representação de seu conjunto de soluções viáveis.

---

<sup>2</sup> À frente definiremos o que é uma região convexa e mostramos que o conjunto de soluções viáveis de um PPL é sempre uma região convexa.

**Exemplo 1.1.** Seja o problema de Programação Linear seguinte, do qual queremos representar graficamente o conjunto de soluções viáveis:

$$\begin{aligned} \max \quad & z = 2x_1 + 2x_2 \\ \text{s.a.} \quad & \begin{cases} x_1 + 2x_2 \leq 20 \\ 2x_1 + x_2 \leq 16 \\ x_1 \leq 6 \\ x_j \geq 0, \quad j = 1, 2 \end{cases} \end{aligned} \quad (1.3)$$

Para tal precisamos determinar o conjunto de todos os vetores  $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$  que satisfazem todas as restrições. Isso pode ser feito determinando os pontos do plano que satisfazem cada uma das restrições, e a intersecção desses pontos é o espaço procurado. Esse espaço é mostrado na figura abaixo, representado pela parte colorida.

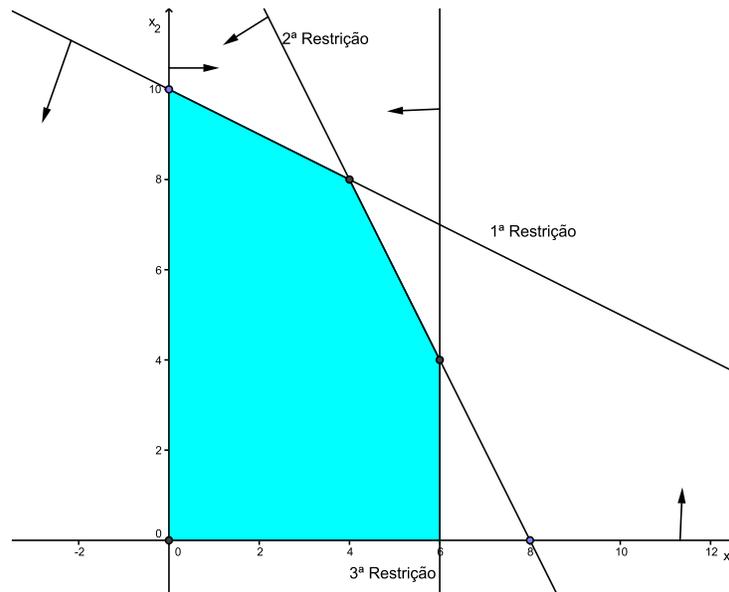


Figura 1.1: Representação do conjunto de pontos viáveis do Exemplo 1.1.

Na figura acima, cada seta na reta que representa o limite da restrição indica a parte do semiplano que satisfaz essa restrição. É sempre possível, para o caso em que o PPL envolve apenas duas variáveis, encontrar a solução ótima graficamente. Para tal estudo dedicamos uma seção própria um pouco mais à frente.

**Definição 1.4** (Solução ótima viável). *Considere um PPL qualquer cujo objetivo seja minimizar a FO. Seja  $X$  o conjunto de todas as soluções viáveis deste PPL. Dizemos que um vetor  $x^* \in X$  é uma **solução ótima viável** do PPL se, e somente se,  $c'x^* \leq c'x \forall x \in X$ . De modo análogo podemos definir uma solução ótima viável para um PPL de maximização.*

**Definição 1.5** (Valor ótimo). *O **valor ótimo** de um PPL é dado pela solução ótima viável aplicada na função objetivo.*

### 1.2.1 Diferentes formas para representação de PPL

Normalmente os problemas de Programação Linear são representados de formas mais compactas do que a apresentada na Definição 1.1, formas essas que facilitam a compreensão de certos aspectos especiais dos PPL's. Algumas literaturas como [22], [12] apresentam a notação abaixo<sup>3</sup>.

$$\begin{aligned} \min \quad & z = \sum_{j=1}^n c_j x_j \\ \text{s.a.} \quad & \begin{cases} \sum_{j=1}^n a_{ij} x_j \geq b_i, & i = 1, 2, \dots, m \\ x_j \geq 0, & j = 1, 2, \dots, n \end{cases} \end{aligned} \quad (1.4)$$

Um PPL qualquer também pode ser representado na forma matricial que, além de ser a mais compacta, é bastante útil na exploração da teoria da dualidade e por isso se faz presente em grande parte das obras que abordam Programação Linear. Assim o modelo (1.4) pode ser também representado como se segue:

$$\begin{aligned} \min \quad & z = c'x \\ \text{s.a.} \quad & \begin{cases} Ax \geq b \\ x \geq 0 \end{cases} \end{aligned} \quad (1.5)$$

<sup>3</sup> A estrutura do modelo (1.4) descreve o problema conhecido como *problema da dieta*, problema que pode referir-se a uma dieta humana ou pecuária. Para maiores informações consultar [11], [22].

Para essa notação temos os seguintes elementos:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}, \quad c = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}.$$

### 1.2.2 Equivalência entre PPL's

Quando o problema está da forma apresentada na Definição 1.1, dizemos que esse PPL está sob a forma mista. Apresentamos a seguir as definições de **forma padrão** e **forma canônica**.

**Definição 1.6** (Forma padrão de um PPL). *Dizemos que um problema de Programação Linear na forma*

$$\begin{aligned} & \min \text{ (max)} \quad z = c'x \\ & \text{s.a.} \quad \begin{cases} Ax = b \\ x \geq 0 \end{cases} \end{aligned} \quad (1.6)$$

com  $b \geq 0$ , está na **forma padrão**.

**Definição 1.7** (Forma canônica de um PPL). *Dizemos que um problema de Programação Linear em uma das formas abaixo*

$$\begin{aligned} & \max \quad z = c'x \\ & \text{s.a.} \quad \begin{cases} Ax \leq b \\ x \geq 0 \end{cases} \end{aligned} \quad (1.7)$$

ou

$$\begin{aligned} & \min \quad z = c'x \\ & \text{s.a.} \quad \begin{cases} Ax \geq b \\ x \geq 0 \end{cases} \end{aligned} \quad (1.8)$$

está na **forma canônica**.

Um problema de Programação Linear sempre pode ser transformado de uma forma na outra, esteja ele na forma padrão, canônica ou mista, como é apresentado no teorema abaixo adaptado de [11].

**Teorema 1.1.** *Qualquer PPL pode ser transformado em um PPL equivalente através das seguintes regras:*

a. *Desigualdades:*

$$\sum_{j=1}^n a_{ij}x_j \leq b_i \text{ é equivalente a } \sum_{j=1}^n a_{ij}x_j + x_{n+i} = b_i, \quad x_{n+i} \geq 0$$

$$\sum_{j=1}^n a_{ij}x_j \geq b_i \text{ é equivalente a } \sum_{j=1}^n a_{ij}x_j - x_{n+i} = b_i, \quad x_{n+i} \geq 0$$

b. *Variáveis sem restrição de sinal:*

$$x_j \text{ irrestrito de sinal é equivalente a } \bar{x}_j - \bar{\bar{x}}_j, \text{ onde } \bar{x}_j \geq 0 \text{ e } \bar{\bar{x}}_j \geq 0.$$

c. *Variáveis negativas:*

$$x_j \leq 0 \text{ é equivalente a } (-\bar{x}_j), \text{ com } \bar{x}_j \geq 0.$$

d. *Função Objetivo:*

$$\max z(x) \text{ equivale a } -\min -z(x).$$

**Demonstração:**

- (a) Na primeira desigualdade basta tomar  $x_{n+1} = b_i - \sum_{j=1}^n a_{ij}x_j$ , com  $x_{n+1} \geq 0$  e a igualdade é satisfeita. O mesmo procedimento pode ser tomado para a segunda desigualdade.
- (b) Basta notar que qualquer número sempre pode ser escrito como diferença de dois números positivos.
- (c) Basta notar que se  $x_j \leq 0$ , tomando  $x_j = -\bar{x}_j$ , temos que  $-\bar{x}_j \leq 0 \Leftrightarrow \bar{x}_j \geq 0$ .
- (d) Seja  $X$  o conjunto de todas as soluções viáveis do PPL e  $x^*$  a solução ótima do PPL, então temos que  $\max z(x) = z(x^*)$ . Mas, da Definição 1.4, temos que  $z(x^*) \geq z(x)$ ,  $\forall x \in X$ , portanto podemos afirmar que  $-z(x^*) \leq -z(x)$ ,  $\forall x \in X$ , de onde segue que  $-z(x^*) = \min(-z(x))$  implicando enfim que  $z(x^*) = -\min(-z(x))$ . ■

O item (d) desse teorema nos diz que através da minimização do inverso da função objetivo de um PPL de maximização, encontra-se o resultado ótimo com sinal contrário, sendo assim, ao adotar tal procedimento, é sempre necessário inverter novamente o sinal para então encontrar a resposta correta.

No item (a) do teorema, as variáveis  $x_{n+i}$  acrescidas nas desigualdades para a obtenção das igualdades são conhecidas como **variáveis de folga** (ou de **sobra**, no segundo caso). Com esse teorema um PPL qualquer pode ser facilmente transformado em um PPL equivalente na forma padrão. A grande importância disso se dá pelo fato que o método Simplex, largamente utilizado para se obter a solução de um PPL, utiliza o formato padrão, como será visto no próximo capítulo.

O inverso também pode ser feito, isto é, um problema na forma padrão pode ser colocado numa outra forma qualquer. Uma restrição de igualdade pode ser trocada por duas desigualdades:

$$a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n = b_i \iff \begin{cases} a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n \leq b_i \\ a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n \geq b_i \end{cases}$$

É interessante observar que quando transformamos um PPL com desigualdades para a forma padrão estamos aumentando o número de variáveis, ou seja, temos um aumento na dimensão do problema. Nesse sentido, como as duas formulações são equivalentes é sempre possível estabelecer a correspondência entre as duas formulações.

Vejamos melhor isso através das representações de um PPL na forma canônica e seu equivalente na forma padrão. A Figura 1.2 ilustra a correspondência entre os conjuntos de soluções viáveis desses PPL's.

Seja o PPL na forma canônica como segue abaixo:

$$\begin{aligned} \max \quad & z = x_1 + 2x_2 \\ & \begin{cases} x_1 + x_2 \leq 3 \\ x_1, x_2 \geq 0 \end{cases} \end{aligned} \quad (1.9)$$

Esse problema pode ser transformado em um PPL equivalente na forma padrão através do acréscimo da variável de folga  $x_3$ , ficando como segue:

$$\begin{aligned} \max \quad & z = x_1 + 2x_2 + 0x_3 \\ & \begin{cases} x_1 + x_2 + x_3 = 3 \\ x_1, x_2 \text{ e } x_3 \geq 0 \end{cases} \end{aligned} \quad (1.10)$$

Assim podemos definir a correspondência entre as duas formulações por:  $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$  do

problema (1.9) com  $\begin{bmatrix} x_1 \\ x_2 \\ 3 - x_1 - x_2 \end{bmatrix}$  do problema (1.10).

A Figura 1.2, mostra a relação estabelecida entre os conjuntos de soluções viáveis dos dois PPL's.

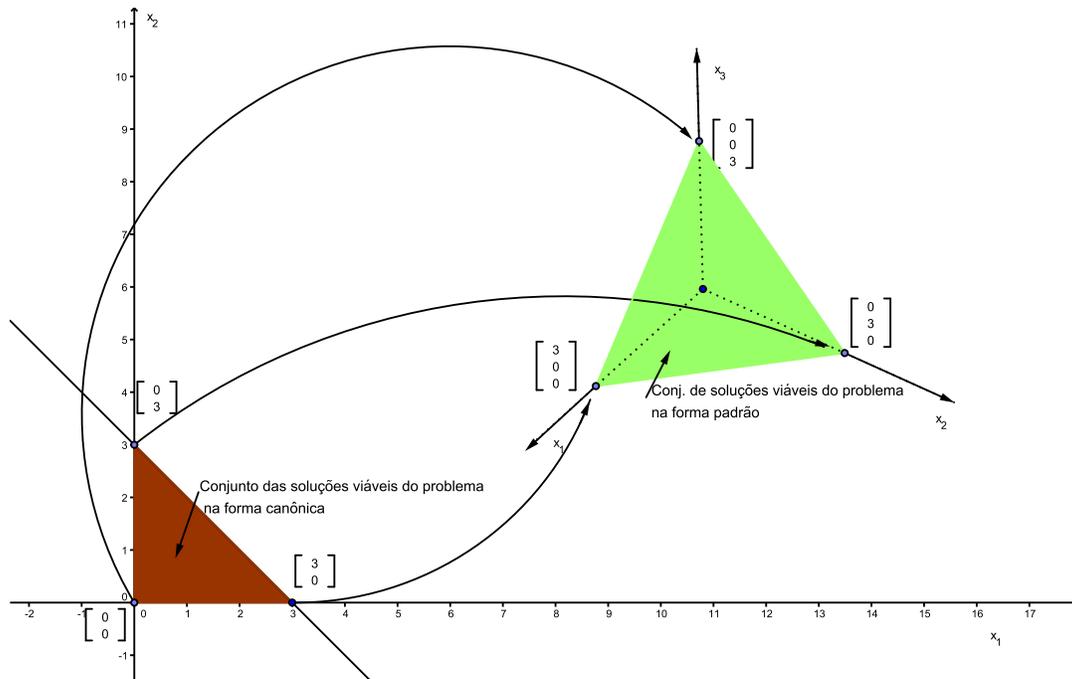


Figura 1.2: Representação da correspondência entre os conjuntos de soluções viáveis do PPL na forma canônica (1.9) e seu equivalente na forma padrão (1.10).

### 1.2.3 Pressupostos dos problemas de Programação Linear

Os problemas de Programação Linear de maneira geral apresentam certas limitações que precisam ser observadas; tais limitações segundo Lins e Calôba [11] e Puccini [22] devem-se aos seguintes pressupostos:

1. **Proporcionalidade:** nos modelos de Programação Linear assume-se que, na função objetivo, o custo de cada atividade é proporcional ao nível de produção  $x_j$ , sendo o custo unitário  $c_j$  o coeficiente de proporcionalidade. Essa hipótese nos diz que o custo unitário  $c_j$  independe do nível de produção  $x_j$ , sendo assim não se considera os custos iniciais da produção da atividade  $j$  e a chamada economia de escala. Como é reforçado em [22] isso não é válido na maioria dos problemas reais. Para os coeficientes das restrições  $a_{ij}$  também assume-se que eles são independentes do nível de produção  $x_j$ , para qualquer quantidade do recurso  $i$ .
2. **Aditividade:** assume-se que não existe interação entre as alternativas de atividades, tanto na função objetivo quanto nas restrições. Assim, por exemplo, o

lucro total de uma empresa será sempre igual à soma dos lucros obtidos em cada uma das atividades.

3. **Divisibilidade:** esse pressuposto assume que todas as variáveis de decisão podem ser fracionárias.
4. **Determinismo:** os coeficientes  $c_j$ ,  $a_{ij}$  e  $b_i$  são bem conhecidos, sem haver incerteza ou aleatoriedade entre eles.

## 1.3 Resolução gráfica de problemas de Programação Linear

Como já foi apresentado no Exemplo 1.1, num problema que envolve apenas duas variáveis é sempre possível representar graficamente seu conjunto de soluções viáveis. Na verdade, para esse tipo de problema, é sempre possível encontrar a solução ótima (caso tal solução exista). Problemas que envolvam 3 variáveis também podem ser resolvidos graficamente, porém como é ressaltado em [27], tal procedimento na maioria das vezes não é fácil. A partir de 4 variáveis a solução só é possível algebricamente.

Descreveremos tal procedimento, pois a partir dele fica mais fácil a compreensão de resultados associados a um PPL. Para iniciar esse estudo vamos tomar o Exemplo 1.1, que se trata de um problema de maximização.

A partir do conjunto de soluções viáveis do problema, que está representado pela Figura 1.1, precisamos determinar, caso exista, um ponto  $(x_1^*, x_2^*)$  que pertença a esse conjunto onde a função objetivo  $z = 2x_1 + 2x_2$  assumo o maior valor possível. Isso pode ser feito estabelecendo alguns valores para a função objetivo  $z$ , dados por suas curvas de nível, que nada mais são que uma família de retas paralelas. Para este problema podemos olhar, por exemplo, para as retas paralelas representadas pelas equações abaixo, que estão na Figura 1.3.

$$2x_1 + 2x_2 = 4$$

$$2x_1 + 2x_2 = 12$$

$$2x_1 + 2x_2 = 20$$

$$2x_1 + 2x_2 = 24$$

Utilizando o cálculo de várias variáveis, temos como fato indubitável, como é mostrado em [7], [10], [18] e [23], que as curvas de nível são perpendiculares ao vetor gradiente da função, que é dado por  $\nabla z = \left( \frac{\partial z}{\partial x_1}, \frac{\partial z}{\partial x_2} \right)$ . Em nosso exemplo temos  $\nabla z = (2, 2)$ . Sabendo ainda que o vetor gradiente indica o sentido de crescimento da função, para um problema de maximização teremos a solução ótima viável no último ponto do conjunto de soluções viáveis, no sentido do vetor gradiente, que é possível estar contido em uma curva de nível da função. A Figura 1.3 ilustra tal solução. Nela, a parte colorida representa o conjunto de soluções viáveis e as retas pontilhadas são as curvas de nível da função objetivo. Assim, a solução ótima viável será dada pelo ponto E de coordenadas (4, 8), tendo como valor ótimo para o problema  $z = 24$ .

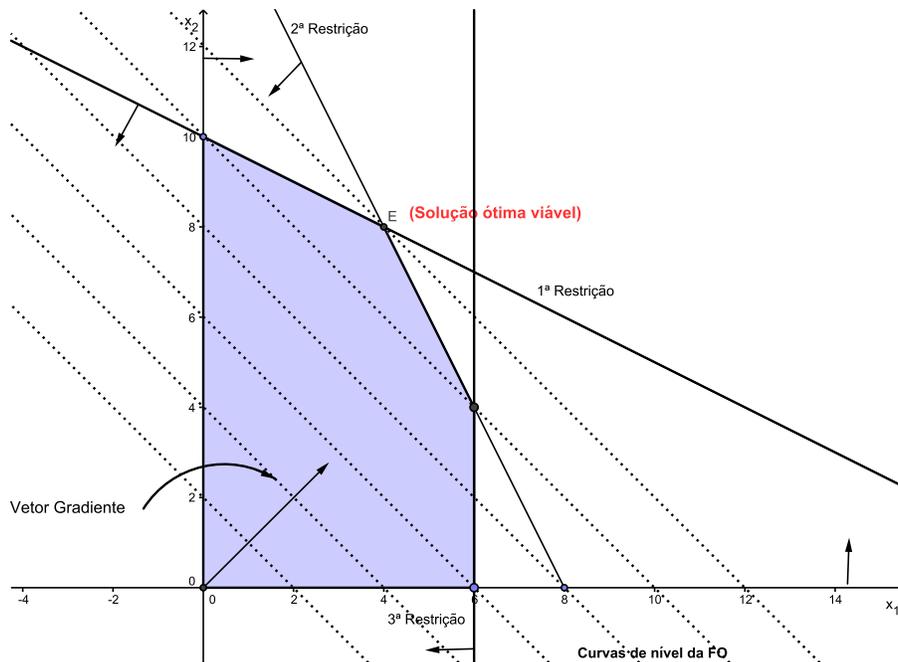


Figura 1.3: Representação da solução gráfica do PPL.

### 1.3.1 Tipos de soluções de um PPL

Aproveitemos a representação gráfica de um problema de Programação Linear para visualizar as possíveis situações que podem ocorrer com a solução de um problema. A saber, um problema de Programação Linear pode ser classificado nas seguintes situações:

1. PPL com **única solução ótima**, como no exemplo representado pela Figura 1.3.
2. PPL com **infinitas soluções ótimas**, que ocorre quando a função objetivo é paralela a alguma das restrições. Tal situação pode ser obtida alterando a função objetivo do problema (1.3) para  $z = 4x_1 + 2x_2$ , onde teríamos o mesmo conjunto de soluções viáveis, como é mostrado na Figura 1.4 a seguir, com a função objetivo paralela à segunda restrição.

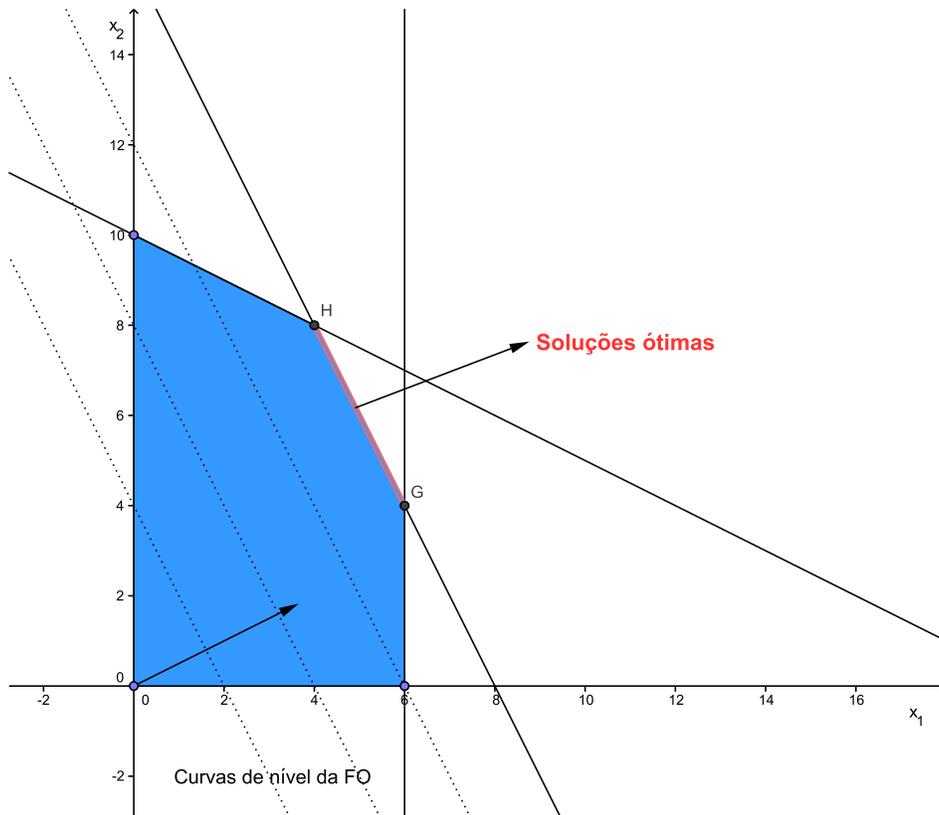


Figura 1.4: Representação gráfica de um PPL com infinitas soluções.

Nesse caso, todos os pontos do segmento de reta  $\overline{GH}$  fornecem uma solução ótima para o PPL, por isso o mesmo admite infinitas soluções ótimas.

3. PPL **sem solução viável**, também conhecido como PPL **inviável**, que ocorre quando não existe nenhuma solução que satisfaça simultaneamente todas as restrições, isso porque o lugar geométrico das soluções viáveis definidas pelo conjunto de restrições caracteriza um conjunto vazio. Tal situação pode ser obtida alterando o conjunto de restrições do PPL exemplo (1.3) para:

$$s.a. \begin{cases} 3x_1 + x_2 \leq 6 \\ x_1 + x_2 \geq 10 \\ x_j \geq 0 \quad j = 1, 2 \end{cases} \quad (1.11)$$

A Figura 1.5 ilustra esse caso.

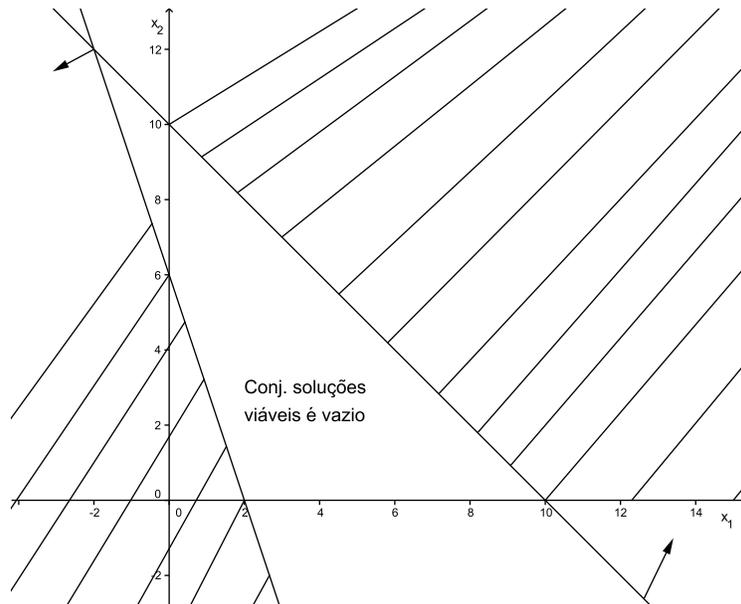


Figura 1.5: Representação gráfica de um PPL com infinitas soluções.

4. PPL **ilimitado**, quando o conjunto de soluções viáveis não é fechado e a função objetivo do PPL aumenta seu valor (problema de maximização) em uma direção sem jamais atingir uma fronteira da região viável, não sendo possível estabelecer

uma solução ótima. Nesse caso o PPL é dito ilimitado. O PPL a seguir exemplifica essa situação.

$$\begin{aligned} \max \quad & z = 2x_1 + 2x_2 \\ \text{s.a.} \quad & \begin{cases} -3x_1 + 4x_2 \leq 12 \\ x_1 - 4x_2 \leq 4 \\ x_j \geq 0, \quad j = 1, 2 \end{cases} \end{aligned} \quad (1.12)$$

A Figura 1.6 mostra a representação do conjunto de soluções viáveis e as curvas de nível da função objetivo para o problema (1.12).

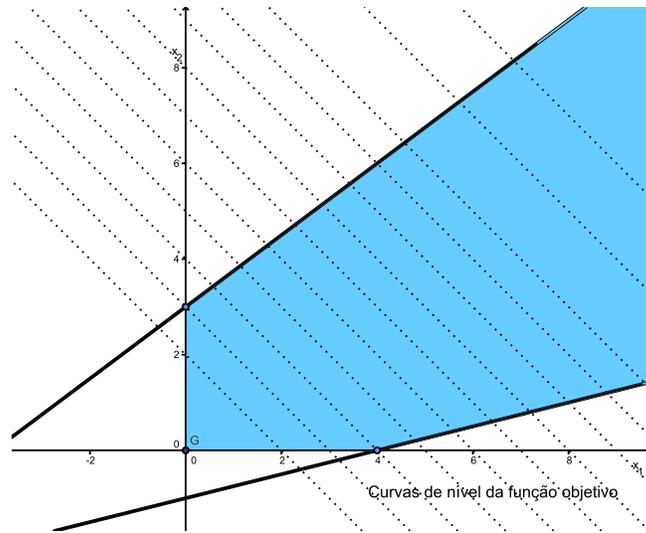


Figura 1.6: Representação gráfica de um PPL ilimitado.

É interessante observar que para esse problema (1.12), se a função objetivo fosse de minimização, o mesmo admitiria uma única solução viável ótima, a saber o ponto  $(0, 0)$  com valor ótimo zero. Ou seja, apenas pelo fato do conjunto de soluções viáveis não ser fechado, não se garante que um problema de Programação Linear tenha solução ótima ilimitada, na verdade ele pode continuar tendo uma ou até infinitas soluções ótimas com valor limitado.

# O MÉTODO SIMPLEX

O objetivo desse capítulo é discutir o chamado método Simplex, que trata-se de um procedimento para a resolução de Problemas de Programação Linear, desenvolvido por George B. Dantzig em 1947. Iniciaremos o capítulo com a apresentação de alguns fundamentos teóricos referentes ao método Simplex e posteriormente passaremos ao método propriamente dito.

## 2.1 Fundamentação teórica do método Simplex

Para compreender melhor o método Simplex, faz-se necessário apresentarmos alguns conceitos matemáticos sobre a estrutura algébrica da solução de um problema de Programação Linear. Parte desses conceitos é geralmente abordado em literatura que trata de assuntos relacionados à álgebra linear, bem como nos cursos que os utilizam como base.

Iniciemos com um conjunto de definições.

**Definição 2.1** ([25]). *Um vetor  $x \in \mathbb{R}^n$  é uma **combinação linear** (CL) dos vetores  $x^1, x^2, \dots, x^p \in \mathbb{R}^n$  se existirem escalares  $a_1, a_2, \dots, a_p \in \mathbb{R}$  tais que  $x = a_1x^1 + a_2x^2 + \dots + a_px^p = \sum_{i=1}^p a_ix^i$ .*

**Definição 2.2** ([27]). Um vetor  $x \in \mathbb{R}^n$  é uma **Combinação Linear Convexa (CLC)** dos vetores  $x^1, x^2, \dots, x^p \in \mathbb{R}^n$  se existirem escalares  $a_1, a_2, \dots, a_p \in \mathbb{R}$  tais que  $x = a_1x^1 + a_2x^2 + \dots + a_px^p = \sum_{i=1}^p a_ix^i$ , onde  $a_i \geq 0$ ,  $i = 1, 2, \dots, p$  e  $\sum_{i=1}^p a_i = 1$ .

Definiremos agora conjuntos convexos, pois logo mais a frente mostraremos que o conjunto de soluções viáveis de um PPL é um conjunto convexo. Esse fato é de suma importância dentro do desenvolvimento do método Simplex pois a partir da convexidade é possível garantir que uma solução ótima de um problema, quando existe, é atingida em um vértice do conjunto convexo.

**Definição 2.3** ([4]). Um conjunto  $X \subset \mathbb{R}^n$  é um **conjunto convexo** se para quaisquer  $x^1 \in X$  e  $x^2 \in X$ ,  $x = ax^1 + (1-a)x^2$  também  $\in X$ ,  $a \in [0, 1]$ . Ou seja, o segmento de reta que liga  $x^1$  a  $x^2$  está inteiramente contido em  $X$ .

A Figura (2.1) ilustra um conjunto convexo e um conjunto não convexo.

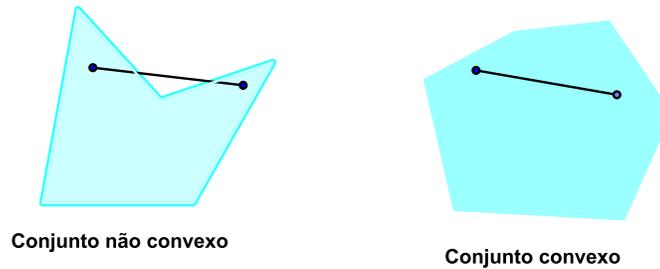


Figura 2.1: Representação em  $\mathbb{R}^2$  de um conjunto convexo e um conjunto não convexo.

**Definição 2.4** (Hiperplano). Um conjunto de pontos na forma  $\{x \in \mathbb{R}^n / a'x = b\}$ ,  $a \in \mathbb{R}^n$  (não nulo) e  $b \in \mathbb{R}$  é chamado de **hiperplano** em  $\mathbb{R}^n$ .

**Definição 2.5** (Semiespaço fechado). *Um conjunto de pontos na forma  $\{x \in \mathbb{R}^n / a'x \geq b\}$  ou  $\{x \in \mathbb{R}^n / a'x \leq b\}$ ,  $a \in \mathbb{R}^n$  (não nulo) e  $b \in \mathbb{R}$  é chamado de **semiespaço fechado** em  $\mathbb{R}^n$ .*

É interessante notar que um hiperplano separa o  $\mathbb{R}^n$  em dois semiespaços, representados pelas inequações:  $\sum_{i=1}^n a_i x_i \geq b$  e  $\sum_{i=1}^n a_i x_i \leq b$ .

### 2.1.1 Convexidade do conjunto de soluções de um PPL

Sabemos, da Definição 1.1, que as restrições de um PPL podem se apresentar nas formas de igualdades e/ou desigualdades (hiperplanos e semiespaços). Assim apresentaremos dois resultados importantes sobre hiperplanos e semiespaços que servem de base para a garantia da convexidade do conjunto de soluções de um PPL.

**Proposição 2.1.** *Um hiperplano  $X = \{x \in \mathbb{R}^n / c'x = z\}$ ,  $c \in \mathbb{R}^n$  e  $z \in \mathbb{R}$  é um conjunto convexo.*

**Demonstração:** Se  $x^1, x^2 \in X$ , segue que  $c'x^1 = z$  e  $c'x^2 = z$ ; então  $x = \alpha x^1 + (1 - \alpha)x^2 \in X$ ,  $\alpha \in [0, 1]$ , pois  $c'x = c'[\alpha x^1 + (1 - \alpha)x^2] = \alpha c'x^1 + (1 - \alpha)c'x^2 = \alpha z + (1 - \alpha)z = z$ . ■

**Proposição 2.2.** *Um semiespaço é um conjunto convexo.*

A demonstração dessa proposição é praticamente análoga à demonstração da proposição anterior, e pode ser encontrada em [5] e [19].

**Proposição 2.3.** *O conjunto de todas as combinações lineares convexas de um número finito de pontos convexas é convexo, isto é,*

$$X = \{x / x = \sum_{i=1}^r a_i x^i, a_i > 0, i = 1, 2, \dots, r \text{ e } \sum_{i=1}^r a_i = 1\}$$

*é um conjunto convexo.*

Outro fato importante sobre os conjuntos convexos é que a intersecção finita de conjuntos convexos ainda é um conjunto convexo, como é estabelecido pelo teorema à frente.

**Teorema 2.1.** *Se  $X_i$ ,  $i = 1, 2, \dots, r$ , for uma coleção finita de conjuntos convexos, então  $X = \bigcap_{i=1}^r X_i$  é um conjunto convexo.*

A demonstração da última proposição e do último teorema podem ser encontradas em [19], [22] e [27].

**Teorema 2.2.** *O conjunto de soluções viáveis de um problema de Programação Linear é um conjunto convexo.*

A demonstração segue dos resultados anteriores, uma vez que o conjunto de soluções viáveis de um PPL é formado pela intersecção de hiperplanos e semiespaços, ou seja, restrições do tipo  $a'x = b$ ,  $a'x \leq b$ ,  $a'x \geq b$ ,  $x \geq 0$ , que são conjuntos convexos.

Aos leitores interessados, demonstrações alternativas para esse último teorema podem ser encontradas em [5], [11], [20], [22] e [27].

Aproveitemos ainda, uma vez que estamos tratando sobre convexidade, para definir ponto extremo de um conjunto convexo, também conhecido como vértice do conjunto.

**Definição 2.6.** *Seja  $X$  um conjunto convexo. Um ponto  $x \in X$  é dito **vértice de  $X$**  (ou ponto extremo de  $X$ ) se  $\nexists x^1, x^2 \in X$ ,  $x^1 \neq x^2$ , tais que  $x = \alpha x^1 + (1 - \alpha)x^2$ ,  $\alpha \in (0, 1)$ , ou seja,  $x$  não pode ser escrito como combinação linear convexa de dois pontos distintos de  $X$ .*

### 2.1.2 Solução básica de um PPL

Dado um problema de Programação Linear qualquer, já vimos no Capítulo 1 que é sempre possível transformar suas restrições em igualdades. Assim, vamos sempre considerar o problema já em sua forma padrão:

$$\begin{aligned} \min \quad & z = c'x \\ \text{s.a.} \quad & \begin{cases} Ax = b \\ x \geq 0 \end{cases} \end{aligned} \quad (2.1)$$

onde  $A \in \mathbb{R}^{m \times n}$ ,  $x, c \in \mathbb{R}^n$ ,  $b \in \mathbb{R}^m$ ,  $m \leq n$ .

Iniciemos analisando o sistema de equações lineares  $Ax = b$ . Vamos supor, sem perda de generalidade, que  $A$  tem posto  $m$ , ou seja,  $A$  possui  $m$  linhas linearmente independentes<sup>1</sup> (LI). Nesse caso, existe um conjunto de  $m$  índices, que chamaremos de  $I_B$ , das  $n$  colunas de  $A$ , de tal forma que a matriz  $B$  formada com as colunas relativas a esses índices tem posto  $m$ .

**Definição 2.7** (Matriz básica). *Seja  $A \in \mathbb{R}^{m \times n}$ . A submatriz  $B \in \mathbb{R}^{m \times m}$ , onde  $\det(B) \neq 0$  ou seja,  $B$  tem posto  $m$ , é dita **matriz base**<sup>2</sup> (ou matriz básica).*

Podemos ainda alocar os índices das colunas de  $A$  não pertencentes a  $I_B$  em um outro conjunto  $I_N$  e com as colunas de  $A$  referentes a esses índices formar uma matriz  $N \in \mathbb{R}^{m \times (n-m)}$ , essa chamada de **matriz não básica**.

Dada uma matriz  $A \in \mathbb{R}^{m \times n}$ , usando um pouco de análise combinatória é fácil ver que o número de submatrizes bases ( $m \times m$ ) relacionadas a  $A$  é limitado por  $\binom{n}{m} = \frac{n!}{m!(n-m)!}$ .

Sendo assim podemos particionar a matriz  $A$  da seguinte forma:  $A = (B | N)$ , onde  $B$  é a matriz base definida acima e  $N$  a matriz não básica, formada pelas  $(n - m)$  colunas restantes da matriz  $A$ .

De modo análogo podemos particionar os vetores  $x$  e  $c'$  em  $x = \begin{bmatrix} x_B \\ x_N \end{bmatrix}$  e  $c' = (c'_B | c'_N)$ , onde:

<sup>1</sup> Dizemos que um conjunto de vetores  $v^1, v^2, \dots, v^p \in \mathbb{R}^n$  é linearmente independente se  $\lambda_1 v^1 + \lambda_2 v^2 + \dots + \lambda_p v^p = 0$  implicar que  $\lambda_1 = \lambda_2 = \dots = \lambda_p = 0$  [25].

<sup>2</sup> A matriz  $B$  é chamada de matriz base pelo fato dos vetores que compõem suas colunas formarem uma base para  $\mathbb{R}^m$ .

- $x_B$  : vetor de  $m$  componentes, formado pelas variáveis do vetor  $x$  associadas à matriz base  $B$ , variáveis essas chamadas de **variáveis básicas** (VB).
- $x_N$  : vetor de  $(n-m)$  componentes, formado pelas variáveis do vetor  $x$  associadas à matriz não básica  $N$ , variáveis essas chamadas de **variáveis não básicas** (VNB).
- $c_B$  : vetor de  $m$  componentes, formado pelos coeficientes do vetor  $c$  referentes às variáveis básicas.
- $c_N$  : vetor de  $(n-m)$  componentes, formado pelos coeficientes do vetor  $c$  referentes às variáveis não básicas.

Dessa forma, podemos escrever o PPL (2.1) da seguinte maneira:

$$\begin{aligned} \min \quad & z = c'_B x_B + c'_N x_N \\ \text{s.a.} \quad & \begin{cases} Bx_B + Nx_N = b \\ x_B \geq 0, x_N \geq 0 \end{cases} \end{aligned} \quad (2.2)$$

Podemos explicitar  $x_B$  em função de  $x_N$ , multiplicando a restrição principal de (2.2) por  $B^{-1}$ , pois  $B$  é inversível, onde obtemos:

$$x_B = B^{-1}b - B^{-1}Nx_N \quad (2.3)$$

Tomando  $x_N = 0$  em (2.3) obtemos  $\bar{x}_B = B^{-1}b$ . Assim definimos:

**Definição 2.8** (Solução básica).  $\bar{x}$  é uma **solução básica** para  $Ax = b$  em (2.1) quando  $\bar{x}_N = 0$  e  $\bar{x}_B = B^{-1}b$ , ou seja,  $\bar{x} = \begin{bmatrix} \bar{x}_B \\ \bar{x}_N \end{bmatrix} = \begin{bmatrix} B^{-1}b \\ 0 \end{bmatrix}$ .

**Definição 2.9** (Solução básica viável). Uma solução básica em que  $x_B$  assume valores maiores ou iguais a zero é dita **Solução Básica Viável** (SBV).

### 2.1.3 Teoremas fundamentais relacionados ao método Simplex

Aproveitemos agora para enunciar alguns teoremas nos quais o método Simplex se baseia. Iniciaremos mostrando a relação entre uma solução básica viável de um PPL e um ponto extremo do conjunto convexo que representa suas restrições.

**Teorema 2.3.** *Dado um PPL com conjunto de soluções viáveis  $X$ ,  $\bar{x}$  é solução básica viável do problema, se, e somente se,  $\bar{x}$  é ponto extremo do conjunto  $X$ .*

**Teorema 2.4.** *Se um problema de Programação Linear admitir uma solução ótima, então pelo menos um ponto extremo do conjunto de pontos viáveis é solução ótima.*

**Teorema 2.5.** *Se a função objetivo de um PPL assume seu máximo (mínimo) em mais de um ponto extremo, então ela terá o mesmo valor para qualquer combinação linear convexa desses pontos.*

Isso significa que se um PPL possui mais do que uma solução ótima, então ele possui infinitas soluções ótimas. Ressaltamos ainda que as provas dos teoremas anteriores podem ser encontradas, por exemplo em [5], [11], [22] e [27].

Os resultados apresentados nessa seção dão suporte ao método Simplex que será discutido na próxima seção.

## 2.2 O Método Simplex

Os teoremas (2.3) e (2.4) são muito importantes na teoria de PL pois deles é possível concluir que a busca por uma solução ótima pode se restringir aos pontos extremos do conjunto de soluções viáveis do PPL, ou seja, não é necessário pesquisar soluções no interior deste conjunto. Além disso, uma vez que existe a correspondência entre ponto extremo e SBV, é suficiente gerar SBV's. No entanto, não é conveniente gerar todas as SBV's, cuja quantidade é limitada, como já vimos, por  $C_{n,m} = \frac{n!}{m!(n-m)!}$ , pois essa quantidade se torna muito grande quando o tamanho do problema cresce. E mais, podemos chegar a conclusões erradas quando o problema possui solução ótima ilimitada.

O método Simplex, que mostraremos resumidamente a seguir, apresenta uma maneira eficiente de gerar SBV's a partir de uma SBV dada. Em SBV's consecutivas, uma matriz base  $B$  se diferencia da seguinte por apenas uma coluna. É um processo iterativo, que repete os mesmos passos a partir de uma SBV inicial, até se chegar à uma SBV ótima. Este método também detecta se o PPL tem infinitas soluções, se o problema é ilimitado ou se o problema é inviável.

Para apresentarmos alguns resultados importantes do método Simplex, vamos utilizar a notação da partição da matriz  $A$  e do vetor  $x$  utilizada na seção anterior e construir uma nova forma de escrevermos o PPL. Vamos supor que temos disponível uma SBV. Iniciemos substituindo a expressão de  $x_B$  dada por (2.3) na função objetivo do PPL (2.2), de onde segue que esse PPL pode ser reescrito como:

$$\begin{aligned} \min \quad & z = c'_B B^{-1}b - (c'_B B^{-1}N - c'_N)x_N \\ \text{s.a.} \quad & \begin{cases} x_B = B^{-1}b - B^{-1}Nx_N \\ x_B \geq 0, x_N \geq 0 \end{cases} \end{aligned} \quad (2.4)$$

Vamos escrever a matriz  $A$  por colunas, ou seja,  $A = (a^1 \ a^2 \ \dots \ a^n)$  onde  $a^j = (a_{1j} \ a_{2j} \ \dots \ a_{mj})'$ ,  $a^j \in \mathbb{R}^m$ , representa uma coluna de  $A$  com  $j \in (I_B \cup I_N)$ . Definimos ainda  $z_j = c'_B B^{-1}a^j$ ,  $z_j \in \mathbb{R}$ , e  $\bar{z} = c'_B B^{-1}b = c'_B \bar{x}_B$ ,  $\bar{z} \in \mathbb{R}$ , como sendo o valor da função objetivo para a SBV corrente. Desta maneira a função objetivo do problema (2.4) pode ser reescrita como:

$$\min \quad z = \bar{z} - \sum_{j \in I_N} (z_j - c_j)x_j \quad (2.5)$$

Definindo  $y_j = B^{-1}a^j$ ,  $y_j \in \mathbb{R}^m$ , podemos escrever as restrições do problema (2.4) como:

$$x_B = \bar{x}_B - \sum_{j \in I_N} y_j x_j \quad (2.6)$$

### 2.2.1 Teste de otimalidade e critério para entrada na base

O teste de otimalidade é utilizado para verificar se uma SBV é ótima ou não. Caso a SBV disponível seja ótima, o método termina. Caso contrário, devemos indicar uma nova variável não básica (VNB) para entrar na base de modo que a função objetivo diminua (problema de mínimo). Nesse sentido, tal passo do método mostra-se interessado na melhoria da função objetivo. O próximo teorema estabelece esse critério para dizer se uma SBV é ótima a partir da análise do sinal de  $(z_j - c_j)$ .

**Teorema 2.6.** *Se  $\bar{x}_B \geq 0$  e  $(z_j - c_j) \leq 0, \forall j \in I_N$ , então o vetor  $x^* \in \mathbb{R}^n$ , onde  $x_{B_i}^* = \bar{x}_{B_i}, i = 1, 2, \dots, m$  e  $x_j^* = 0, j \in I_N$ , será uma solução ótima do PPL.*

**Demonstração:** Das hipóteses:  $(z_j - c_j) \leq 0$  e  $x_j \geq 0, \forall j \in I_N$  e  $\bar{x}_B \geq 0$ , segue que  $\sum_{j \in I_N} (z_j - c_j)x_j \leq 0$ . Assim, temos da função objetivo (2.5) do PPL, que  $z \geq \bar{z}$ , onde  $\bar{z} = c'_B \bar{x}_B = c'x^*$ . Isso quer dizer que  $z$  nunca será menor que  $\bar{z} = c'x^*$ , e sendo  $x^*$  claramente uma solução básica viável do PPL, segue da Definição 1.4 (solução ótima de um PPL), que  $x^*$  é solução ótima. ■

Esse teorema nos diz que, dada uma SBV associada a uma matriz base  $B$ , se tivermos  $(z_j - c_j) \leq 0, \forall j \in I_N$ , o problema de procurar uma solução ótima está resolvido. Assim sendo, ao se trabalhar com o método Simplex, de posse de uma SBV faz-se necessário verificar se  $(z_j - c_j) \leq 0, \forall j \in I_N$ . Quando esta condição é verificada, o método pode ser finalizado, já que estaremos de posse de uma solução ótima. É nesse sentido que as literaturas sobre o assunto chamam tal procedimento de **teste de otimalidade** ou **critério de otimalidade**.

Como comentamos anteriormente, no método Simplex uma matriz base  $B$  se diferencia da matriz base usada na iteração seguinte por apenas uma coluna. Como cada variável  $x_j$  pode ser associada à coluna  $a^j$  de  $A$ , podemos estabelecer um critério para a seleção de uma variável  $x_r$  para entrar na base. Para tal, os valores de  $(z_j - c_j), j \in I_N$  desempenham papéis fundamentais, pois além de garantirem a otimalidade de

uma SBV, podemos, a partir do estudo de seu sinal, definir tal critério, como veremos.

Caso o critério de otimalidade não seja satisfeito, existe pelo menos um índice  $r \in I_N$  tal que  $(z_r - c_r) > 0$ . Se tornarmos a variável não básica  $x_r$  positiva e mantivermos as outras variáveis não básicas nulas, ou seja,  $x_j = 0, \forall j \in I_N, j \neq r$ , então a função objetivo (2.5) do PPL pode ser expressada sem o somatório por:

$$z = \bar{z} - (z_r - c_r)x_r \quad (2.7)$$

Sendo  $x_r \geq 0$ , segue de (2.7) que, quando  $(z_r - c_r) > 0$ , o valor de  $z$  nessa iteração irá diminuir. Como queremos minimizar a função objetivo, é sensato admitir que  $x_r$  é candidata a entrar na base, ou seja, a coluna  $a^r$  é candidata a compor a matriz base  $B$ . Notamos ainda da equação (2.7), que quando  $(z_r - c_r) < 0$ , a função objetivo  $z$  aumenta nessa iteração, não sendo então conveniente colocar  $x_r$  na base. E por último, se escolhermos  $r \in I_N$ , tal que  $(z_r - c_r) = 0$ , o valor de  $z$  se manterá inalterado<sup>3</sup>.

Desta última observação e do Teorema (2.6), temos que se  $(z_j - c_j) \leq 0 \forall j \in I_N$  e existe algum  $k \in I_N$  tal que  $(z_k - c_k) = 0$  então o PPL admite mais de uma solução ótima, e nesse caso, pelo Teorema (2.5), admite infinitas soluções. Ou seja, temos aqui um critério para garantir quando um PPL tem infinitas soluções ótimas.

Faz-se necessário ressaltar que nem sempre existe um único  $r \in I_N$  tal que  $(z_r - c_r) > 0$ . Nesse caso, podemos usar um critério bem simples, como é apresentado em [5], para a escolha da variável  $x_r$  que entra na base:  $x_r$  é escolhida de forma que:

$$(z_r - c_r) = \max_j (z_j - c_j / z_j - c_j > 0) \quad (2.8)$$

Isso é, vamos tomar para entrar na base a variável que tiver o maior coeficiente positivo na função objetivo.

<sup>3</sup> Estamos supondo que o problema não apresenta solução degenerada. Uma solução básica é chamada de degenerada quando pelo menos uma componente de  $x_B$  é nula.

### 2.2.2 Critério para saída da base ou Teste da Razão

Como o critério para entrada na base estabelece uma variável  $x_r$  para entrar na base, e só podemos ter  $m$  variáveis básicas, é necessário agora estabelecer um critério para escolha de uma VB para deixar a base, tornando-se VNB.

A partir dos critérios de entrada na base e de saída da base, temos que o método Simplex, em cada iteração, troca uma coluna da matriz  $N$  por uma coluna da matriz  $B$ , “caminhando” assim por pontos extremos adjacentes.

Como vimos anteriormente, escolhemos a variável  $x_r$  para entrar na base com o objetivo de diminuir a função objetivo. Sendo assim, é conveniente que  $x_r$  assuma o maior valor possível, já que quanto maior for  $x_r$  mais a função objetivo  $z$  diminuirá. Porém precisamos manter a viabilidade da solução. Isso é feito pelo teste da razão, que nos diz o quanto a variável  $x_r$  pode crescer, assumindo que todas as VB devem permanecer não negativas.

A equação (2.6) nos diz que:

$$x_B = \bar{x}_B - \sum_{j \in I_N} y_j x_j$$

Essa equação pode ser escrita sem o somatório, pois somente a variável  $x_r$  que entrará na base assumirá um valor não negativo, as demais VNB permanecerão nulas. Assim temos que:

$$x_B = \bar{x}_B - y_r x_r \tag{2.9}$$

Alternativamente explicitando, as  $m$  componentes básicas, temos:

$$\begin{pmatrix} x_{B_1} \\ x_{B_2} \\ \vdots \\ x_{B_m} \end{pmatrix} = \begin{pmatrix} \bar{x}_{B_1} \\ \bar{x}_{B_2} \\ \vdots \\ \bar{x}_{B_m} \end{pmatrix} - \begin{pmatrix} y_{r_1} \\ y_{r_2} \\ \vdots \\ y_{r_m} \end{pmatrix} x_r \quad (2.10)$$

ou ainda, equivalentemente,

$$x_{B_i} = \bar{x}_{B_i} - y_{r_i} x_r, \quad i = 1, 2, \dots, m \quad (2.11)$$

Como comentamos, o método Simplex mantém a viabilidade da solução, ou seja, precisa manter  $x_{B_i} \geq 0$ ,  $i = 1, 2, \dots, m$ . Como  $x_r$  sai do valor zero e cresce para um valor positivo, é preciso analisar o que acontece com o valor de cada VB quando  $x_r$  cresce. É fácil ver, pela expressão (2.11), que isso depende do sinal de  $y_{r_i}$ ,  $i = 1, 2, \dots, m$ .

Analisando o sinal de  $y_{r_i}$ , temos:

- 1 - Se  $y_{r_i} < 0$ , então  $x_{B_i}$  aumenta, uma vez que já era positivo e continuará positivo com mais razão ainda. Isso quer dizer que  $x_{B_i}$  associado pode crescer indefinidamente com o valor de  $x_r$ .
- 2 - Se  $y_{r_i} = 0$ , então o valor de  $x_{B_i}$  permanecerá inalterado.
- 3 - Se  $y_{r_i} > 0$ , então o valor de  $x_{B_i}$  decresce à medida que  $x_r$  cresce. Mas  $x_{B_i}$  só pode decrescer até atingir o valor zero. Neste caso, para satisfazer a condição de não negatividade, vamos impor  $x_{B_i} \geq 0$ , isto é,

$$\begin{aligned} \bar{x}_{B_i} - y_{r_i} x_r &\geq 0 \Rightarrow \\ -y_{r_i} x_r &\geq -\bar{x}_{B_i} \Rightarrow \\ y_{r_i} x_r &\leq \bar{x}_{B_i} \Rightarrow \\ x_r &\leq \frac{\bar{x}_{B_i}}{y_{r_i}}, \quad i / y_{r_i} > 0 \end{aligned} \quad (2.12)$$

Os valores de  $\frac{\bar{x}_{B_i}}{y_{r_i}}$  limitam o crescimento da variável  $x_r$ . Como a nova VB só poderá crescer até que a primeira componente  $x_{B_i}$  atinja o valor zero, é necessário tomar o mínimo entre todos os  $\frac{\bar{x}_{B_i}}{y_{r_i}}$ ,  $y_{r_i} > 0$ . Se  $x_{B_s}$  for essa componente, então temos que:

$$x_{B_s} = \min_{1 \leq i \leq m} \left\{ \frac{\bar{x}_{B_i}}{y_{r_i}} \mid y_{r_i} > 0 \right\} \quad (2.13)$$

Nesse caso,  $x_{B_s}$  é a variável básica que deixará a base, tornando-se variável não básica, pois a mesma é a primeira a anular com o crescimento de  $x_r$ . A expressão (2.13) é conhecida na literatura como **teste da razão**.

**Observação 2.1.** Se  $y_{r_i} \leq 0$ ,  $i = 1, 2, \dots, m$ ,  $x_r$  pode crescer indefinidamente mantendo a factibilidade da solução. Logo, da função objetivo  $z = \bar{z} - (z_r - c_r)x_r$ , segue que se  $x_r$  crescer indefinidamente, o valor da função objetivo tenderá a menos infinito ( $z \rightarrow -\infty$ ), o que significa que o problema é ilimitado.

### 2.2.3 Nova SBV - atualização

Pelos critérios de entrada e de saída da base expostos anteriormente, sabemos como determinar a variável que entra na base,  $x_r$ , e a variável que sai da base,  $x_{B_s}$ . Devemos agora estabelecer como é feita a atualização dos elementos que são usados no método Simplex, ou seja, as matrizes  $B$  e  $N$ , e os vetores  $c_B, c_N, x_B$  e  $x_N$ . Essa atualização é feita como segue:

**B:** a nova matriz  $B$  é obtida a partir da matriz  $B$  anterior, trocando-se a coluna associada à variável de saída  $x_{B_s}$ , ou seja, a coluna  $a^{B_s}$ , pela coluna associada à variável de entrada  $x_r$ , ou seja, coluna  $a^r$ .

**N:** a nova matriz  $N$  é obtida a partir da matriz  $N$  anterior, trocando-se a coluna  $a^r$  pela coluna  $a^{B_s}$ .

$c_B$  e  $c_N$ : os vetores  $c_B$  e  $c_N$ , são atualizados segundo as trocas feitas em  $B$  e  $N$ , ou seja, o coeficiente  $c_{B_s}$  é substituído pelo coeficiente  $c_r$  em  $c_B$ , e o coeficiente  $c_r$  de  $c_N$  é substituído pelo coeficiente  $c_{B_s}$ .

$x_B$  e  $x_N$ : os vetores  $x_B$  e  $x_N$  também são atualizados segundo as trocas feitas em  $B$  e  $N$ , ou seja, a variável  $x_{B_s}$  é substituída pela variável  $x_r$  em  $x_B$ , e  $x_r$  é substituída por  $x_{B_s}$  em  $x_N$ .

Lembramos que o valor de  $\bar{x}_B$  continua sendo dado por  $\bar{x}_B = B^{-1}b$ , onde  $B^{-1}$  é a inversa da nova matriz  $B$ , ou, preferencialmente através da solução do sistema linear  $B\bar{x}_B = b$ . Do mesmo modo, o valor de  $\bar{z}$  continua sendo dado por  $\bar{z} = c'_B \bar{x}_B$ , onde  $c'_B$  e  $\bar{x}_B$  estão associados à nova matriz  $B$ .

### 2.2.4 O algoritmo Simplex

A idéia do algoritmo Simplex é, através dos critérios estabelecidos anteriormente, descrever uma sequência de passos para a obtenção de uma solução ótima de um PPL. O algoritmo Simplex pode ser escrito resumidamente pelos seguintes passos:

- 1° **Passo:** Determinar uma SBV inicial para o problema;
- 2° **Passo:** Aplicar o teste de otimalidade: caso a solução seja ótima, parar. Caso contrário, determinar uma VNB  $x_r$  para entrar na base;
- 3° **Passo:** Aplicar o teste da razão para determinar qual variável básica  $x_{B_s}$  deve sair da base;
- 4° **Passo:** Atualizar os elementos, obtendo uma nova SBV e retornar ao 2° passo.

Utilizando as equações que foram desenvolvidas anteriormente, podemos reescrever este algoritmo de maneira um pouco mais detalhada. Dessa forma, a solução ótima de um problema geral de PL do tipo (2.1) pode ser encontrada aplicando este algoritmo:

**Observação 2.2.** Note que, para a aplicação do método Simplex o problema deve estar na forma padrão.

**Algoritmo Simplex:**

**Inicialização:**

- Fornecer  $m$ ,  $n$ , a matriz  $A$ , os vetores  $b$  e  $c$  e uma SBV inicial para o problema, isto é,  $B$ ,  $N$ ,  $c_B$ ,  $c_N$ ,  $x_B$ ,  $x_N$ ,  $\bar{x}_B = B^{-1}b$ ,  $\bar{z} = c'_B \bar{x}_B$  e  $I_N$ .

**Repetir até parar:**

**Teste de otimalidade - entrada na base:**

- calcular  $z_j - c_j = c'_B B^{-1}a^j - c_j$ ,  $j \in I_N$ ;
- se  $z_j - c_j \leq 0$ ,  $\forall j \in I_N$ , então a SBV atual é ótima;
  - se  $z_j - c_j < 0$ ,  $\forall j \in I_N$ , o PPL possui uma única solução ótima  $\bar{x} = (\bar{x}_B \mid \underbrace{\bar{x}_N}_0)'$ . Parar.
  - se  $z_j - c_j = 0$ , para algum  $j \in I_N$ , o PPL possui soluções ótimas alternativas. Parar.

senão, escolher  $x_r$  para entrar na base tal que  $z_r - c_r > 0$ . Um critério que pode ser utilizado é escolher  $x_r$  tal que  $z_r - c_r = \max_{j \in I_N} \{z_j - c_j / z_j - c_j > 0\}$ ;

**Teste da razão - saída da base:**

- calcular  $y_r = B^{-1}a^r$  ou equivalentemente  $By_r = a^r$ ;
  - se  $y_r \leq 0$ , então o PPL é ilimitado. Parar.
  - senão  $x_{B_s} = \min_i \left\{ \frac{\bar{x}_{B_i}}{y_{r_i}} / y_{r_i} > 0 \right\}$ ,  $x_{B_s}$  é a VB que sai da base;

**Atualização - nova SBV:**

- $x_r$  entra na base;  $x_{B_s}$  sai da base;

- atualizar as matrizes  $B$  e  $N$  e os vetores  $c_B, c_N, x_B, x_N$ ;
- atualizar  $\bar{x}_B$  e  $\bar{z}$ .

Vamos aplicar o algoritmo apresentado acima para resolver um exemplo. Tomemos o PPL (1.3) proposto no Capítulo 1, Exemplo 1.1.

**Exemplo 2.1.** Seja o PPL:

$$\begin{array}{ll} \max & z = 2x_1 + 2x_2 \\ \text{s.a.} & \begin{cases} x_1 + 2x_2 \leq 20 \\ 2x_1 + x_2 \leq 16 \\ x_1 \leq 6 \\ x_j \geq 0, \quad j = 1, 2 \end{cases} \end{array} \quad (2.14)$$

Vamos encontrar a solução ótima através do método Simplex.

Conforme vimos no capítulo anterior, o problema tem solução ótima (4, 8), com valor ótimo  $z = 24$ . A região viável e a solução ótima estão destacadas na Figura 2.2.

Para iniciar a aplicação do método Simplex, o problema precisa ser colocado na forma padrão. Para isso, acrescentamos as variáveis de folga  $x_3, x_4$  e  $x_5$ . Além disso, como apresentamos o método Simplex para problema de mínimo, vamos transformar a função objetivo, obtendo:

$$\begin{array}{ll} \max & z = 2x_1 + 2x_2 = -\min(-z) = -2x_1 - 2x_2 \\ \text{s.a.} & \begin{cases} x_1 + 2x_2 + x_3 = 20 \\ 2x_1 + x_2 + x_4 = 16 \\ x_1 + x_5 = 6 \\ x_j \geq 0, \quad j = 1, \dots, 5 \end{cases} \end{array} \quad (2.15)$$

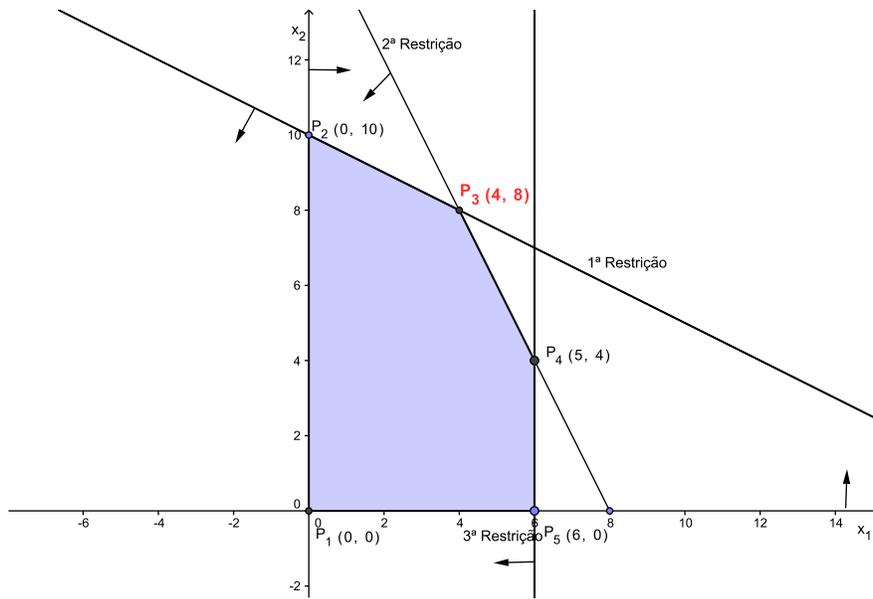


Figura 2.2: Representação do conjunto de pontos viáveis e solução ótima do Exemplo 2.1.

Desta maneira, temos que:

$$A = \begin{bmatrix} 1 & 2 & 1 & 0 & 0 \\ 2 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 20 \\ 16 \\ 6 \end{bmatrix}, \quad c' = \begin{bmatrix} -2 & -2 & 0 & 0 & 0 \end{bmatrix}$$

Observando a matriz  $A$ , vemos uma SBV inicial óbvia, onde as VB são  $x_3$ ,  $x_4$  e  $x_5$ , pois as colunas correspondentes a essas variáveis formam a matriz identidade.

Assim,

$$B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} a^3 & a^4 & a^5 \end{bmatrix}; \quad N = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} a^1 & a^2 \end{bmatrix}$$

$$x_B = \begin{bmatrix} x_3 \\ x_4 \\ x_5 \end{bmatrix}; \quad \bar{x}_B = B^{-1}b = Ib = \begin{bmatrix} 20 \\ 16 \\ 6 \end{bmatrix}; \quad x_N = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$c'_B = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}; \quad c'_N = \begin{bmatrix} -2 & -2 \end{bmatrix}; \quad I_N = \{1, 2\}$$

A SBV inicial é  $\bar{x}' = \begin{bmatrix} 0 & 0 & 20 & 16 & 6 \end{bmatrix}$ , que corresponde ao ponto  $P_1$  na Figura 2.2. O valor da função objetivo é dado por:

$$\bar{z} = c'_B \bar{x}_B = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 20 \\ 16 \\ 6 \end{bmatrix} = 0.$$

### 1ª iteração

#### Teste de Otimalidade — entrada na base

Calcular  $z_j - c_j = c'_B \underbrace{B^{-1}a^j}_{y_j} - c_j$ ,  $j = 1, 2$ .

$$\text{para } j = 1: \quad y_1 = B^{-1}a^1 = a^1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$$

$$z_1 - c_1 = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} - (-2) = 2 > 0 \quad (x_1 \text{ é candidata a entrar na base}).$$

$$\text{para } j = 2: \quad y_2 = B^{-1}a^2 = a^2 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$$

$$z_2 - c_2 = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} - (-2) = 2 > 0 \text{ (} x_2 \text{ é candidata a entrar na base).}$$

Podemos escolher tanto  $x_1$  quanto  $x_2$  para entrar na base. Vamos escolher  $x_2$ , isto é,  $x_r = x_2$  ( $r = 2$ ).

### Teste da razão — saída da base

A coluna atualizada de  $x_2$  é  $y_2 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$ . Vamos definir qual VB sai da base quando  $x_2$  entra na base.

Das equações (2.9) e (2.10) temos que  $x_B = \bar{x}_B - y_2 x_2$ , que queremos manter não negativo, ou seja:

$$\begin{pmatrix} x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 20 \\ 16 \\ 6 \end{pmatrix} - \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} x_2 = \begin{pmatrix} 20 - 2x_2 \\ 16 - x_2 \\ 6 \end{pmatrix} \geq 0 \quad (2.16)$$

$$\begin{aligned} \text{De onde segue que: } 20 - 2x_2 \geq 0 &\Rightarrow x_2 \leq 10 \\ 16 - x_2 \geq 0 &\Rightarrow x_2 \leq 16 \end{aligned}$$

Isso significa que  $x_2$  só poderá crescer até  $x_2 = 10$ , limite que é dado pela primeira VB,  $x_3$ , que deverá então sair da base. Neste caso,  $x_{B_s} = x_3$ .

**Observação 2.3.** O teste da razão também pode ser feito diretamente da expressão (2.13):

$$x_{B_s} = \min_{1 \leq i \leq 3} \left\{ \frac{\bar{x}_{B_i}}{y_{2_i}} / y_{2_i} > 0 \right\} = \min_{1 \leq i \leq 3} \left\{ \frac{20}{2}, \frac{16}{1} \right\} = 10$$

dado pela primeira VB, isto é,  $x_{B_s} = x_3$  é a VB que sai da base.

### Nova SBV — atualização

Temos que:  $x_2$  entra na base,  $x_2 = 10$ .

$x_3$  sai da base,  $x_3 = 0$ .

Colocando  $x_2 = 10$  em (2.16), obtemos:  $x_3 = 0$ ,  $x_4 = 6$ ,  $x_5 = 6$ . Portanto:

$$B = \begin{bmatrix} a^2 & a^4 & a^5 \end{bmatrix}; \quad N = \begin{bmatrix} a^1 & a^3 \end{bmatrix}; \quad I_N = \{1, 3\}$$

$$x_B = \begin{bmatrix} x_2 \\ x_4 \\ x_5 \end{bmatrix}; \quad \bar{x}_B = \begin{bmatrix} 10 \\ 6 \\ 6 \end{bmatrix}; \quad x_N = \begin{bmatrix} x_1 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$c'_B = \begin{bmatrix} -2 & 0 & 0 \end{bmatrix}; \quad c'_N = \begin{bmatrix} -2 & 0 \end{bmatrix}$$

$$\bar{z} = c'_B \bar{x}_B = \begin{bmatrix} -2 & 0 & 0 \end{bmatrix} \begin{bmatrix} 10 \\ 6 \\ 6 \end{bmatrix} = -20$$

A nova SBV é  $\bar{x}' = \begin{bmatrix} 0 & 10 & 0 & 6 & 6 \end{bmatrix}$ , que corresponde ao ponto  $P_2$  na Figura 2.2.

Podemos notar que o valor da função objetivo diminuiu de zero para -20. De fato,  $(z_2 - c_2)x_2 = -2 \times 10 = -20$ .

### 2ª iteração

#### Teste de Otimalidade — entrada na base

Calcular  $z_j - c_j = c'_B y_j - c_j$ ,  $j = 1, 3$ .

$$\text{para } j = 1: y_1 = B^{-1}a^1 \Leftrightarrow By_1 = a^1 \Leftrightarrow \begin{bmatrix} 2 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_{1_1} \\ y_{1_2} \\ y_{1_3} \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \Leftrightarrow$$

$$y_1 = \begin{bmatrix} 1/2 \\ 3/2 \\ 1 \end{bmatrix}$$

$$z_1 - c_1 = \begin{bmatrix} -2 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1/2 \\ 3/2 \\ 1 \end{bmatrix} - (-2) = -1 + 2 = 1 > 0 \text{ (} x_1 \text{ é} \\ \text{candidata a entrar na base).}$$

$$\text{para } j = 3: y_3 = B^{-1}a^3 \Leftrightarrow By_3 = a^3 \Leftrightarrow \begin{bmatrix} 2 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_{3_1} \\ y_{3_2} \\ y_{3_3} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \Leftrightarrow$$

$$y_3 = \begin{bmatrix} 1/2 \\ -1/2 \\ 0 \end{bmatrix}$$

$$z_3 - c_3 = \begin{bmatrix} -2 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1/2 \\ -1/2 \\ 0 \end{bmatrix} - 0 = -1 \text{ (} x_3 \text{ não é candidata a} \\ \text{entrar na base).}$$

Neste caso, a única candidata a entrar na base é  $x_1$ , isto é,  $x_r = x_1$  ( $r = 1$ ).

### Teste da razão — saída da base

A coluna atualizada de  $x_1$  é  $y_1 = \begin{bmatrix} 1/2 \\ 3/2 \\ 1 \end{bmatrix}$ . Vamos definir qual variável básica sai da base quando  $x_1$  entra na base, utilizando diretamente a expressão (2.13), onde temos:

$$x_{B_s} = \min_{1 \leq i \leq 3} \left\{ \frac{\bar{x}_{B_i}}{y_{1_i}} / y_{1_i} > 0 \right\} = \min_{1 \leq i \leq 3} \left\{ \frac{10}{1/2}, \frac{6}{3/2}, \frac{6}{1} \right\} = 4$$

Como 4 está associado à segunda VB, ou seja, está associada à  $x_4$ , temos que  $x_4$  sai da base, logo  $x_{B_s} = x_4$ .

### Nova SBV — atualização

Temos que:  $x_1$  entra na base,  $x_1 = 4$ .

$x_4$  sai da base,  $x_4 = 0$ .

Portanto:

$$B = \begin{bmatrix} a^2 & a^1 & a^5 \end{bmatrix}; \quad N = \begin{bmatrix} a^4 & a^3 \end{bmatrix}; \quad I_N = \{3, 4\}$$

$$x_B = \begin{bmatrix} x_2 \\ x_1 \\ x_5 \end{bmatrix}; \quad x_N = \begin{bmatrix} x_3 \\ x_4 \end{bmatrix}$$

$$\bar{x}_B = B^{-1}b \Leftrightarrow B\bar{x}_B = b \Leftrightarrow \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_2 \\ x_1 \\ x_5 \end{bmatrix} = \begin{bmatrix} 20 \\ 16 \\ 6 \end{bmatrix} \Leftrightarrow \bar{x}_B = \begin{bmatrix} 8 \\ 4 \\ 2 \end{bmatrix}$$

$$c'_B = \begin{bmatrix} -2 & -2 & 0 \end{bmatrix}; \quad c'_N = \begin{bmatrix} 0 & 0 \end{bmatrix}$$

$$\bar{z} = c'_B \bar{x}_B = \begin{bmatrix} -2 & -2 & 0 \end{bmatrix} \begin{bmatrix} 8 \\ 4 \\ 2 \end{bmatrix} = -16 - 8 = -24$$

A nova SBV é  $\bar{x}' = \begin{bmatrix} 4 & 8 & 0 & 0 & 2 \end{bmatrix}$ , que corresponde ao ponto  $P_3$  na Figura 2.2. Note que o valor da função objetivo diminui de -20 para -24.

### 3ª iteração

#### Teste de Otimalidade — entrada na base

Calcular  $z_j - c_j = c'_B y_j - c_j$ ,  $j = 3, 4$ .

$$\text{para } j = 3: y_3 = B^{-1}a^3 \Rightarrow By_3 = a^3 \Leftrightarrow \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} y_{3_1} \\ y_{3_2} \\ y_{3_3} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \Leftrightarrow$$

$$y_3 = \begin{bmatrix} 2/3 \\ -1/3 \\ 1/3 \end{bmatrix}$$

$$z_3 - c_3 = \begin{bmatrix} -2 & -2 & 0 \end{bmatrix} \begin{bmatrix} 2/3 \\ -1/3 \\ 1/3 \end{bmatrix} - 0 = \frac{-4}{3} + \frac{2}{3} = -\frac{2}{3} < 0$$

( $x_3$  não é candidata a entrar na base)

$$\text{para } j = 4: y_4 = B^{-1}a^4 \Rightarrow By_4 = a^4 \Leftrightarrow \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} y_{4_1} \\ y_{4_2} \\ y_{4_3} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \Leftrightarrow$$

$$y_4 = \begin{bmatrix} -1/3 \\ 2/3 \\ -2/3 \end{bmatrix}$$

$$z_4 - c_4 = \begin{bmatrix} -2 & -2 & 0 \end{bmatrix} \begin{bmatrix} -1/3 \\ 2/3 \\ -2/3 \end{bmatrix} - 0 = \frac{2}{3} - \frac{4}{3} = -\frac{2}{3} < 0$$

( $x_4$  não é candidata a entrar na base)

Como  $z_j - c_j < 0$  para  $j = 3, 4$ , temos que  $\bar{x}' = [4, 8, 0, 0, 2]$  é solução ótima para o PPL, sendo o valor ótimo dado por  $-(-24)=24$ . Esse valor corresponde ao mesmo valor encontrado graficamente no capítulo anterior.

Como podemos ver, o algoritmo Simplex necessita no primeiro passo que seja dada uma SBV inicial. Em alguns problemas, tal SBV inicial é facilmente determinada. Por exemplo quando existem  $m$  colunas da matriz  $A$  formando uma matriz identidade ( $m \times m$ ), que tomamos como matriz base  $B$ . Em alguns problemas, tal situação apresenta-se trivialmente com o acréscimo das variáveis de folga. No entanto, em outros isso não ocorre, ou ainda pior, em alguns problemas pode nem haver solução viável.

No entanto, existe um método eficiente para se determinar uma SBV inicial, conhecido como método **Simplex de duas fases**, ou simplesmente **método das duas fases**. Esse método consiste em aplicar o próprio método Simplex a um problema de PL chamado de “problema auxiliar”, que trata-se de um problema de minimização montado a partir do problema original, com o acréscimo de variáveis chamadas de **variáveis artificiais** (positivas).

A solução ótima do problema auxiliar nos fornece uma SBV inicial para o problema original, ou então, nos diz se o problema original é um problema inviável. De posse da SBV inicial, podemos resolver o problema original através da aplicação do método Simplex discutido anteriormente.

Como nosso objetivo é apresentar o método Simplex para sua posterior aplicação ao problema de separação de pontos que discutiremos no próximo capítulo, e para tal discussão não necessitamos do método das duas fases, optamos, por não apresentá-lo. Porém o leitor interessado pode encontrar sobre este método em várias referências, tais como [11], [12], [20], [22] e [27].

# SEPARAÇÃO DE PADRÕES (PONTOS) VIA PROGRAMAÇÃO LINEAR

---

O objetivo desse capítulo é discutir critérios de separação de pontos, ou separação de padrões, como é comumente conhecido na literatura, através de Programação Linear, e definir um problema capaz de separar linearmente dois conjuntos de padrões caso esses sejam disjuntos e propor uma separação linear que minimize a média dos erros de classificação caso os conjuntos não sejam disjuntos.

## 3.1 Introdução e definições preliminares

Definimos padrões como sendo uma amostra confiável de traços, atos, tendências, ou quaisquer outros tipos de características observáveis de um determinado objeto ou ser vivo. Uma separação de padrões nada mais é que a obtenção de um critério para a distinção entre os elementos de dois conjuntos de padrões. Usualmente, considerando uma separação matemática, os padrões são representados por pontos no espaço Euclidiano  $\mathbb{R}^n$ . Dessa forma, uma maneira matemática de se obter uma separação é construir um hiperplano ou uma superfície não linear de  $\mathbb{R}^n$  de modo que cada conjunto de padrões esteja localizado de um lado do hiperplano ou da superfície.

Assim, vamos supor que são dados dois conjuntos de padrões  $\bar{E}$  e  $\bar{F}$ , onde o conjunto  $\bar{E}$  é formado por  $m$  padrões e o conjunto  $\bar{F}$  é formado por  $k$  padrões. Cada padrão é constituído de  $n$  observações (traços, atos, tendências, etc, citados na definição de padrões) que são representadas por  $n$  escalares reais. Separar esses conjuntos por um hiperplano de  $\mathbb{R}^n$  é encontrar tal superfície de forma que os  $m$  pontos que representam os padrões do conjunto  $\bar{E}$  estejam de um lado desse hiperplano, enquanto os  $k$  pontos que representam os padrões do conjunto  $\bar{F}$  estejam do outro lado desse hiperplano. Naturalmente, nem todos os conjuntos de pontos são linearmente separáveis, como pode ser visto na Figura 3.1.

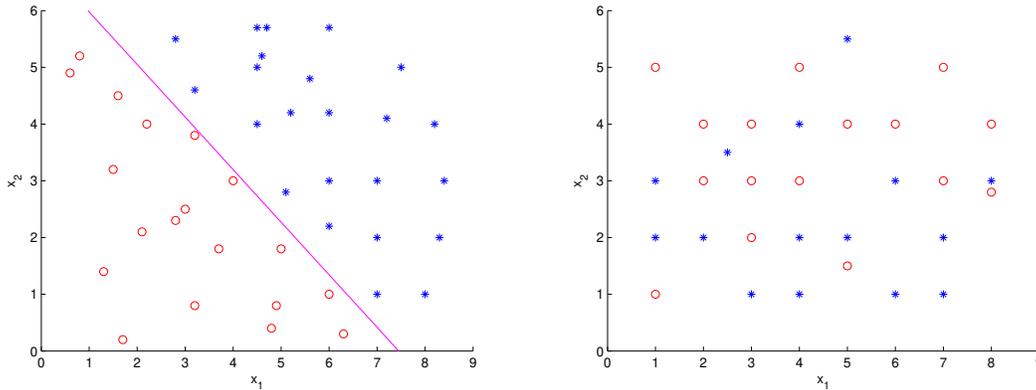


Figura 3.1: Representação em  $\mathbb{R}^2$  de dois conjuntos de padrões linearmente separáveis e linearmente inseparáveis, respectivamente.

Como usualmente os padrões são representados por pontos de  $\mathbb{R}^n$ , podemos formar com esses conjuntos de padrões uma matriz não vazia de números reais, chamada de matriz padrão, conforme Mangasarian [14]. Cada linha dessa matriz define um único padrão, composto de  $n$  números reais que são as representações das observações. Assim sendo, os  $m$  pontos que representam o conjunto de padrões  $\bar{E}$  são representados por uma matriz  $E \in \mathbb{R}^{m \times n}$ , onde cada uma das  $m$  linhas são os  $m$  pontos do conjunto  $\bar{E}$ . De modo análogo podemos representar os  $k$  padrões de  $\bar{F}$  por uma matriz  $F \in \mathbb{R}^{k \times n}$  onde as  $k$  linhas da matriz  $F$  são os  $k$  pontos de  $\bar{F}$ . Chamaremos de  $a_{ij}$  os elementos de  $E$  e de  $b_{ij}$  os elementos de  $F$ . Vejamos um exemplo de dois conjuntos de padrões.

**Exemplo 3.1.** Segundo a OMC (Organização Mundial da Saúde), o Índice de Massa Corpórea (IMC) que relaciona o peso de uma pessoa com sua altura ao quadrado, indica a quantidade aproximada de gordura geral no corpo. Atualmente, os valores considerados normais vão de 18,5 a 24,9  $kg/m^2$ , para homens. A partir de 25  $kg/m^2$ , configura-se sobrepeso e, acima de 30, obesidade. Sendo assim, segundo o IMC podemos classificar um grupo de homens em dois grupos, a saber, normais e sobrepeso, conforme a tabela abaixo:

Altura	Peso	IMC	Classificação
170	73	25,25	Sobrepeso
171	75	25,64	Sobrepeso
160	71	27,71	Sobrepeso
165	70	25,71	Sobrepeso
169	75	26,25	Sobrepeso
162	67	25,52	Sobrepeso
168	71	25,15	Sobrepeso
164	62	23,05	Normal
162	65	24,76	Normal
168	65	23,03	Normal
170	69	23,87	Normal
166	67	24,31	Normal
172	71	23,99	Normal

Tabela 3.1: Classificação de um grupo de homens quanto ao IMC.

Nesse caso, sendo o IMC o fator referente à classificação de um homem estar com sobrepeso ou normal, podemos considerar sobrepeso como um padrão e normal como outro padrão, e assim, a partir das observações de altura e peso definir as matrizes de padrões  $E$  e  $F$  como abaixo:

$$E = \begin{bmatrix} 170 & 73 \\ 171 & 75 \\ 160 & 71 \\ 165 & 70 \\ 169 & 75 \\ 162 & 67 \\ 168 & 71 \end{bmatrix}, \quad F = \begin{bmatrix} 164 & 62 \\ 162 & 65 \\ 168 & 65 \\ 170 & 69 \\ 166 & 67 \\ 172 & 71 \end{bmatrix}$$

Nesse exemplo temos  $n = 2$  (apenas duas características a saber, altura e peso),  $m = 7$  (quantidade de pontos em  $E$ ) e  $k = 6$  (quantidade de pontos em  $F$ ).

## 3.2 Critérios de separabilidade linear

O problema de separação de dois conjuntos de padrões consiste em determinar um único hiperplano

$$x'w - \gamma = 0 \quad (3.1)$$

onde  $w$  é um vetor de  $\mathbb{R}^n$  normal ao hiperplano e  $\gamma$  é um escalar real, tal que:

$$Ew - e\gamma > 0 \quad (3.2)$$

e

$$Fw - e\gamma < 0 \quad (3.3)$$

onde  $e$  é um vetor de uns de dimensão apropriada, ou seja, dimensão  $m$  em (3.2) e dimensão  $k$  em (3.3). Assim, podemos definir a separação de dois conjuntos de padrões como segue abaixo.

**Definição 3.1** (Separabilidade Linear). *Sejam os conjuntos  $\bar{E}$  e  $\bar{F}$  representados pelas matrizes  $E \in \mathbb{R}^{m \times n}$  e  $F \in \mathbb{R}^{k \times n}$  respectivamente. Dizemos que esses conjuntos são linearmente separáveis se, e somente se, existir  $w \in \mathbb{R}^n$  e  $\gamma \in \mathbb{R}$  tal que (3.2) e (3.3) sejam satisfeitos.*

Caso não existam  $w$  e  $\gamma$  satisfazendo a definição acima,  $\bar{E}$  e  $\bar{F}$  são ditos linearmente inseparáveis. Para facilitar a linguagem não vamos mais fazer distinção entre conjuntos de padrões e matriz padrão.

Em nosso Exemplo 3.1, é fácil ver graficamente que os padrões sobrepeso e normal são linearmente separáveis, pois é possível passar uma infinidade de retas que separam completamente os conjuntos de padrões representados pelas observações de peso e altura.

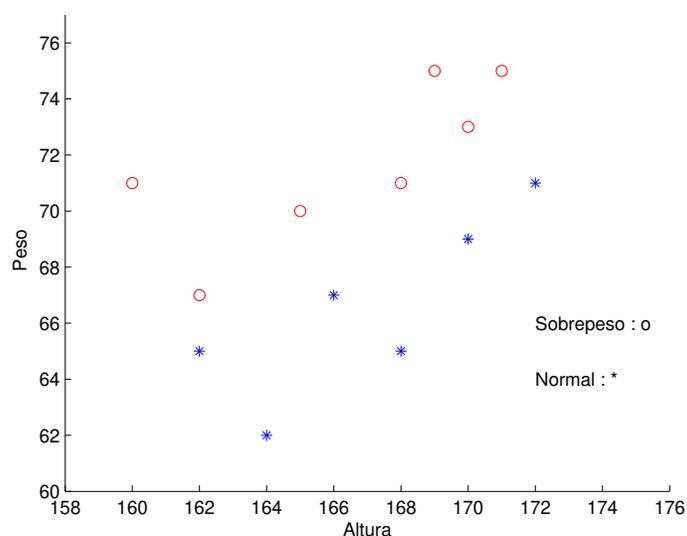


Figura 3.2: Representação em  $\mathbb{R}^2$  dos conjuntos  $\bar{E}$  e  $\bar{F}$  do Exemplo 3.1, que deixa visualmente clara a separação desses conjuntos.

Porém, na prática, os conjuntos de padrões geralmente tem dimensões maiores do que 3, o que exclui a possibilidade de verificação gráfica quanto à separabilidade ou não. Por isso é interessante determinar um método para avaliar se dois conjuntos de padrões são ou não linearmente separáveis. Para estabelecer tal critério iniciemos estabelecendo o lema seguinte.

**Lema 3.1** ([14]). *Os conjuntos de padrões  $\bar{E}$  e  $\bar{F}$  representados pelas matrizes  $E \in \mathbb{R}^{m \times n}$  e  $F \in \mathbb{R}^{k \times n}$  respectivamente, são linearmente separáveis se, e somente se, existir  $c \in \mathbb{R}^n$  e constantes  $\alpha$  e  $\beta$  tal que*

$$Ec - e\alpha \geq 0 \quad (3.4)$$

$$-Fc + e\beta \geq 0 \quad (3.5)$$

$$\alpha - \beta > 0 \quad (3.6)$$

$$e \geq c \geq -e \quad (3.7)$$

onde  $e$  é um vetor coluna de uns de dimensão  $m$  em (3.4),  $k$  em (3.5) e  $n$  em (3.7).

**Demonstração:** ( $\Rightarrow$ ) Se existem  $c$ ,  $\alpha$  e  $\beta$  satisfazendo as condições de (3.4) à (3.7), então, tomando  $w = c$  e  $\gamma = \frac{1}{2}(\alpha + \beta)$  teremos (3.2) e (3.3) satisfeitos e com isso  $\bar{E}$  e  $\bar{F}$  são linearmente separáveis.

( $\Leftarrow$ ) Inversamente, se  $\bar{E}$  e  $\bar{F}$  são linearmente separáveis, então existem  $w$  e  $\gamma$  satisfazendo (3.2) e (3.3). Notemos que ao menos uma componente de  $w$  deve ser diferente de zero, pois caso  $w = 0$ , (3.2) e (3.3) nos levariam a  $-\gamma > 0$  e  $-\gamma < 0$ , o que é obviamente um absurdo. Seja  $\delta = \max_{i=1, \dots, n} |w_i|$ , assim fazendo a divisão de (3.2) e (3.3) por  $\delta$  obtemos:

$$(Ew/\delta) - e\gamma/\delta > 0 \quad (3.8)$$

$$(Fw/\delta) - e\gamma/\delta < 0 \quad (3.9)$$

Note que nesse caso vale a relação

$$e \geq w/\delta \geq -e \quad (3.10)$$

Vamos definir agora

$$\alpha = \min_{i \in 1,2,\dots,m} \sum_{j=1}^n (a_{ij}w_j/\delta) \quad (3.11)$$

$$\beta = \max_{i \in 1,2,\dots,k} \sum_{j=1}^n (b_{ij}w_j/\delta) \quad (3.12)$$

$$c = w/\delta \quad (3.13)$$

Assim, de (3.8), (3.9), (3.11) e (3.12) segue que

$$\alpha > \beta \Rightarrow \alpha - \beta > 0 \quad (3.14)$$

Do mesmo modo, segue de (3.8) à (3.13) que

$$e \geq c \geq -e \quad (3.15)$$

$$Ec - e\alpha = (Ew/\delta) - e \min_{i \in 1,2,\dots,m} \sum_{j=1}^n (a_{ij}w_j/\delta) \geq 0 \quad (3.16)$$

$$Fc - e\beta = (Fw/\delta) - e \max_{i \in 1,2,\dots,k} \sum_{j=1}^n (b_{ij}w_j/\delta) \leq 0 \quad (3.17)$$

Portanto as condições (3.16), (3.17), (3.14) e (3.15) são exatamente as condições (3.4) à (3.7). ■

A partir do lema anterior, Mangasarian [14] estabelece um critério para dizer se dois conjuntos de padrões são linearmente separáveis utilizando um problema de Programação Linear que tem como função objetivo  $(\alpha - \beta)$  e como restrições as desigualdades (3.4), (3.5) e (3.7). Isso fica bem claro ao observarmos o teorema à frente.

**Teorema 3.1** ([14] Condição de Separabilidade Linear). *Uma condição necessária e suficiente para separabilidade linear dos padrões dos conjuntos  $E$  e  $F$  é:*

$$\theta(E, F) > 0 \quad (3.18)$$

onde  $\theta(E, F)$  é a solução do problema de Programação Linear:

$$\begin{aligned} & \max_{\alpha, \beta, c} (\alpha - \beta) \\ \text{s.a.} & \begin{cases} Ec - e\alpha \geq 0 \\ -Fc + e\beta \geq 0 \\ e \geq c \geq -e \end{cases} \end{aligned} \quad (3.19)$$

**Demonstração:** Se  $\theta(E, F) > 0$ , então (3.4) à (3.7) são obviamente satisfeitos pela solução do problema de Programação Linear (3.19) e como consequência disso  $E$  e  $F$  são linearmente separáveis. Inversamente, se  $E$  e  $F$  são linearmente separáveis, então existe algum  $c$ ,  $\alpha$  e  $\beta$  satisfazendo (3.4) à (3.7). Logo  $c$ ,  $\alpha$  e  $\beta$  formam uma solução viável para o problema de Programação Linear (3.19) e torna, portanto,  $\alpha - \beta > 0$  de onde segue que  $\theta(E, F) > 0$ . ■

Do teorema acima segue imediatamente o corolário à frente.

**Corolário 3.1** ([14] Condição de Inseparabilidade Linear). *Uma condição necessária e suficiente para inseparabilidade linear dos conjuntos de padrões de  $E$  e  $F$  é que  $\theta(E, F) = 0$ .*

**Demonstração:** A prova segue imediatamente do teorema 3.1 pela observação que  $c = 0$ ,  $\alpha = 0$  e  $\beta = 0$  formam uma solução viável do problema de Programação Linear (3.19) e, neste caso  $\theta(E, F) = 0$ . ■

Podemos notar que o problema de Programação Linear (3.19) tem dois aspectos importantes que merecem ser ressaltados, conforme o próprio Mangasarian [14] apresenta, a saber:

1. Este problema sempre possui a solução viável  $c = 0$ ,  $\alpha = 0$  e  $\beta = 0$ .
2. As suas soluções são limitadas superiormente por

$$\max_{i \in 1, 2, \dots, m} \sum_{j=1}^n |a_{ij}| + \max_{i \in 1, 2, \dots, k} \sum_{j=1}^n |b_{ij}|$$

O primeiro desses itens nos diz que o conjunto de soluções viáveis do PPL não é vazio, enquanto que o segundo item garante que o PPL nunca será ilimitado, sendo assim, entre as quatro possíveis situações que podem ocorrer com a solução de um PPL, conforme apresentado na subseção 1.3.1 do Capítulo 1, esse PPL só admite duas situações, a saber, possui uma única solução ou possui infinitas soluções.

A título de exemplo, vejamos uma simples aplicação do Teorema 3.1.

**Exemplo 3.2.** Aplicando o Teorema 3.1 ao Exemplo 3.1 para dizer se os conjuntos de padrões são linearmente separáveis ou não, basta resolver o problema de Programação Linear (3.19) onde:

$$E = \begin{bmatrix} 170 & 73 \\ 171 & 75 \\ 160 & 71 \\ 165 & 70 \\ 169 & 75 \\ 162 & 67 \\ 168 & 71 \end{bmatrix} \quad F = \begin{bmatrix} 164 & 62 \\ 162 & 65 \\ 168 & 65 \\ 170 & 69 \\ 166 & 67 \\ 172 & 71 \end{bmatrix}$$

Usando o software Matlab [6], com a opção de utilização do método Simplex, a solução ótima apresentada é  $w' = (-0,67; 1)$ ,  $\alpha = -41$  e  $\beta = -43$ . Logo, o valor ótimo da função objetivo é 2, assim sendo,  $E$  e  $F$  são linearmente separáveis, como já havíamos verificado visualmente. E nesse caso, a reta  $-0,67x_1 + x_2 = -42$  (onde  $-42 = \frac{\alpha + \beta}{2}$ ) separa linearmente esses dois conjuntos, conforme é apresentado na Figura 3.3. Podemos notar também que a solução ótima não é única, ou seja, existem outras retas que separam estes conjuntos.

É possível, recorrendo ao princípios de dualidade em Programação Linear, estabelecer um teorema dual ao Teorema 3.1, assim como um novo critério para inseparabilidade. Como em nosso texto não discutimos sobre dualidade, optamos por apresentar

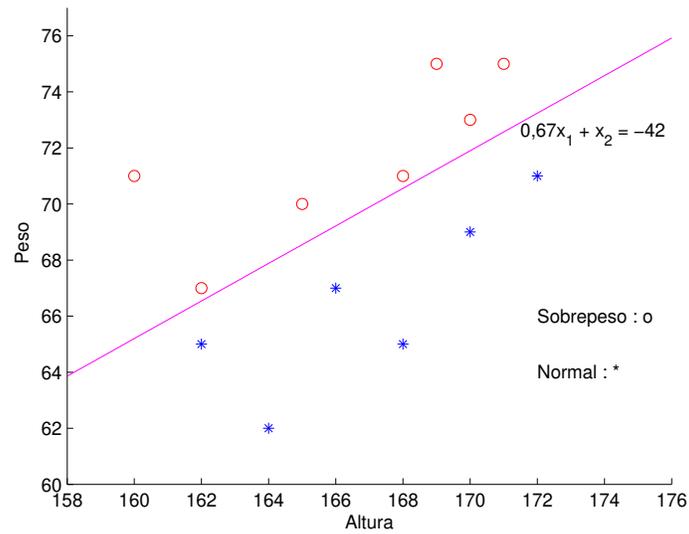


Figura 3.3: Representação em  $\mathbb{R}^2$  dos conjuntos  $\bar{E}$  e  $\bar{F}$  do Exemplo 3.1 juntamente com a reta de separação obtida a partir da aplicação do Teorema 3.1.

esses resultados em nosso apêndice.

Com o exposto até o momento é possível identificar se dois conjuntos de padrões são linearmente separáveis ou não, analisando a solução de um problema de Programação Linear. Além disso, se os dois conjuntos forem linearmente separáveis, a partir da solução do problema de Programação Linear (3.19) é possível estabelecer um hiperplano que separa linearmente esses conjuntos.

No entanto, na maioria dos problemas envolvendo situações reais, os conjuntos de padrões não são linearmente separáveis. Nesse caso, é necessário estabelecer um critério não apenas para identificar quando dois conjuntos são ou não linearmente separáveis mas também, quando não são, o ideal seria conseguir subsídios para gerar um hiperplano que minimize a média das violações dos pontos pertencentes a cada conjunto. Neste caso, teremos uma separação parcial dos conjuntos.

### 3.3 Formulação de um Programa Linear para separação de padrões

Para resolver problemas de separação de padrões, Bennett e Mangasarian [3] propõem a solução de um único problema de Programação Linear, cuja formulação é superior às formulações anteriores, e com as seguintes propriedades:

1. Se os conjuntos de pontos  $\bar{E}$  e  $\bar{F}$  forem disjuntos, a partir da solução do problema de Programação Linear é gerado um hiperplano que separa estritamente esses conjuntos.
2. Se os conjuntos de pontos  $\bar{E}$  e  $\bar{F}$  não forem disjuntos, a partir da solução do problema de Programação Linear é gerado um hiperplano que minimiza a média dos erros de classificação dos pontos.
3. Não há necessidade de impor restrições ao problema para incluir qualquer caso particular a ser considerado.

Novamente, se os pontos dos conjuntos  $\bar{E}$  e  $\bar{F}$  de padrões forem representados pelas matrizes  $E \in \mathbb{R}^{m \times n}$  e  $F \in \mathbb{R}^{k \times n}$  respectivamente, o problema de Programação Linear proposto por Mangasarian e outros em [3] e [16] é dado por:

$$\begin{aligned} \min_{w, \gamma, y, z} \quad & \frac{1}{m} e' y + \frac{1}{k} e' z \\ \text{s.a.} \quad & \begin{cases} Ew + y - e\gamma \geq e \\ -Fw + z + e\gamma \geq e \\ y \geq 0, z \geq 0 \end{cases} \end{aligned} \quad (3.20)$$

onde  $w \in \mathbb{R}^n$ ,  $y \in \mathbb{R}^m$ ,  $z \in \mathbb{R}^k$  e  $\gamma \in \mathbb{R}$  e o vetor  $e$  é o vetor de uns com dimensão apropriada.

Nesse caso  $w'x = \gamma$  é o hiperplano separador citado nas propriedades anteriores, onde  $w$  é o vetor normal ao hiperplano de separação e o número real  $\gamma$  é responsável pela localização do hiperplano.

O problema de Programação Linear (3.20) é baseado em um problema de otimização com norma. Antes de apresentarmos esse problema, vamos definir dois itens que o compõem.

**Definição 3.2.** *Se  $x$  é um vetor de  $\mathbb{R}^n$  então vamos definir:*

1.  $(x_i)_+ = \max_{i=1,2,\dots,n} \{x_i, 0\}$
2.  $\|x\|_1 = \sum_{i=1}^n |x_i|$ .

Agora sim, podemos enunciar o problema de otimização com norma comentado como sendo o problema de minimização que segue:

$$\min_{w,\gamma} \frac{1}{m} \|(-Ew + e\gamma + e)_+\|_1 + \frac{1}{k} \|(Fw - e\gamma + e)_+\|_1 \tag{3.21}$$

A justificativa para a equivalência do problema de otimização (3.21) com o problema de Programação Linear (3.20) segue imediatamente do lema abaixo, que relaciona o problema de minimização com norma a um problema de minimização restrita desprovido de norma e acrescido de funções.

**Lema 3.2** ([3]). *Seja  $g : \mathbb{R}^n \mapsto \mathbb{R}^m$ ,  $h : \mathbb{R}^n \mapsto \mathbb{R}^k$  e seja  $S$  um subconjunto de  $\mathbb{R}^n$ . Os problemas*

$$\min_{x \in S} \|g(x)_+\|_1 + \|h(x)_+\|_1 \tag{3.22}$$

e

$$\begin{aligned} & \min_{x \in S} e'y + e'z \\ \text{s.a.} & \begin{cases} y \geq g(x) \\ z \geq h(x) \\ y \geq 0, z \geq 0 \end{cases} \end{aligned} \tag{3.23}$$

*possuem soluções idênticas.*

**Demonstração:** Como  $g(x)_+ \geq g(x)$  e  $h(x)_+ \geq h(x)$ , podemos estabelecer a equivalência observando que para o problema (3.23) os valores ótimos de  $y$ ,  $z$  e  $x$  são dados

pelas igualdades  $y = g(x)_+$  e  $z = h(x)_+$ . ■

Utilizando o Lema 3.2 Bennett e Mangasarian [3] puderam afirmar a equivalência entre o problema de otimização com norma (3.21) e o problema de Programação Linear (3.20) tomando  $\lambda_1 = \frac{1}{m}$  e  $\lambda_2 = \frac{1}{k}$  na proposição que segue abaixo.

**Proposição 3.1** ([3]). *Para  $\lambda_1 > 0$  e  $\lambda_2 > 0$ , o problema de minimização*

$$\min_{w,\gamma} \lambda_1 \| (-Ew + e\gamma + e)_+ \|_1 + \lambda_2 \| (Fw - e\gamma + e)_+ \|_1 \tag{3.24}$$

*é equivalente ao problema de Programação Linear*

$$\begin{aligned} \min_{w,\gamma,y,z} & \lambda_1 e'y + \lambda_2 e'z \\ \text{s.a.} & \begin{cases} Ew - e\gamma + y \geq e \\ -Fw + e\gamma + z \geq e \\ y \geq 0, z \geq 0 \end{cases} \end{aligned} \tag{3.25}$$

**Demonstração:** Basta tomar  $g(x) = -Ew + e\gamma + e$  e  $h(x) = Fw - e\gamma + e$  no Lema 3.2 para obtermos a equivalência procurada. ■

A próxima proposição apresenta-nos uma nova definição para separabilidade linear considerando um caso especial.

**Proposição 3.2** ([9]). *Se  $E \in \mathbb{R}^{m \times n}$  e  $F \in \mathbb{R}^{k \times n}$  representam pontos dos conjuntos  $\bar{E}$  e  $\bar{F}$  respectivamente, então  $\bar{E}$  e  $\bar{F}$  são linearmente separáveis se, e somente se, existe  $w \in \mathbb{R}^n$  e  $\gamma \in \mathbb{R}$  tal que*

$$Ew \geq e\gamma + e \tag{3.26}$$

$$Fw \leq e\gamma - e$$

**Demonstração:** ( $\Rightarrow$ ) Suponha que  $E$  e  $F$  são linearmente separáveis. Seja  $E_i$  a linha  $i$  de  $E$  e  $F_i$  a linha  $i$  de  $F$ . Então de (3.2) e (3.3) existem  $v \in \mathbb{R}^n$  e  $\gamma \in \mathbb{R}$  tal que  $E_iv > \gamma$  e  $F_iv < \gamma$ .

Se colocarmos:

$$\eta = \min_{1 \leq i \leq m} E_i v - \max_{1 \leq i \leq k} F_i v$$

$$w = \frac{2v}{\eta}$$

$$\text{e } \gamma = \frac{1}{\eta} \left( \min_{1 \leq i \leq m} E_i v + \max_{1 \leq i \leq k} F_i v \right)$$

Então

$$\begin{aligned} \min_{1 \leq i \leq m} E_i w &= \left( \frac{2}{\eta} \right) \left( \min_{1 \leq i \leq m} E_i v \right) = \left( \frac{1}{\eta} \right) \left( \min_{1 \leq i \leq m} E_i v + \min_{1 \leq i \leq m} E_i v \right) \\ &= \underbrace{\left( \frac{1}{\eta} \right) \left( \min_{1 \leq i \leq m} E_i v + \max_{1 \leq i \leq k} F_i v \right)}_{\gamma} + \underbrace{\left( \frac{1}{\eta} \right) \left( \min_{1 \leq i \leq m} E_i v - \max_{1 \leq i \leq k} F_i v \right)}_{\eta} = \gamma + 1 \end{aligned}$$

De modo análogo temos que

$$\begin{aligned} \max_{1 \leq i \leq k} F_i w &= \left( \frac{2}{\eta} \right) \left( \max_{1 \leq i \leq k} F_i v \right) = \left( \frac{1}{\eta} \right) \left( \max_{1 \leq i \leq k} F_i v + \max_{1 \leq i \leq k} F_i v \right) \\ &= \underbrace{\left( \frac{1}{\eta} \right) \left( \max_{1 \leq i \leq k} F_i v + \min_{1 \leq i \leq m} E_i v \right)}_{\gamma} - \underbrace{\left( \frac{1}{\eta} \right) \left( \max_{1 \leq i \leq k} F_i v - \min_{1 \leq i \leq m} E_i v \right)}_{\eta} = \gamma - 1 \end{aligned}$$

Desse modo

$$E_i w \geq \gamma + 1$$

e

$$F_i w \leq \gamma - 1$$

de onde segue que:

$$E w \geq e \gamma + e$$

e

$$F w \leq e \gamma - e.$$

( $\Leftrightarrow$ ) Supondo que  $E w \geq e \gamma + e$  e  $F w \leq e \gamma - e$ ,

como

$$E w - e \gamma > E w - e \gamma - e \geq 0$$

então

$$Ew - e\gamma > 0$$

e

$$Fw - e\gamma < Fw - e\gamma + e \leq 0$$

então

$$Fw - e\gamma < 0.$$

Sendo assim o hiperplano  $w'x = \gamma$  separa linearmente os conjuntos  $E$  e  $F$ . ■

Nas condições dessa proposição os conjuntos  $\bar{E}$  e  $\bar{F}$  são linearmente separáveis se existe um hiperplano que separa esses conjuntos e está pelo menos a uma unidade do ponto mais próximo de  $\bar{E}$  e pelo menos uma unidade do ponto mais próximo de  $\bar{F}$  [9]. Essa distância de uma unidade se dá pela soma de uns (vetor  $e$ ) aos termos  $-Ew + e\gamma$  e  $Fw - e\gamma$ .

Aproveitando a Proposição (3.2), é possível estabelecer que os conjuntos de padrões  $E$  e  $F$  são linearmente separáveis somente se o valor mínimo de (3.21) é zero, e ainda mais, se isso acontecer, a solução  $w = 0$ ,  $\gamma \in \mathbb{R}$  não pode ser ótima, como é mostrado no teorema abaixo.

**Teorema 3.2** ([3]). *Os conjuntos  $\bar{E}$  e  $\bar{F}$  representados pela matrizes  $E \in \mathbb{R}^{m \times n}$  e  $F \in \mathbb{R}^{k \times n}$  respectivamente, são linearmente separáveis se, e somente se, o valor mínimo de (3.21) é zero, caso em que a solução  $w = 0$ ,  $\gamma \in \mathbb{R}$  não pode ser ótima.*

**Demonstração:** Basta notar que pela definição (3.2), o mínimo de (3.21) é zero se, e somente se,  $-Ew + e\gamma + e \leq 0$  e  $Fw - e\gamma + e \leq 0$ , o que equivale à separabilidade linear de  $\bar{E}$  e  $\bar{F}$ . Para ver que  $w = 0$ ,  $\gamma \in \mathbb{R}$  não pode ser ótima para (3.21) basta notar que se  $w = 0$  em (3.21) obtemos  $\min_{\gamma}(1 + \gamma)_+ + (1 - \gamma)_+ = 2 > 0$ , o que obviamente contradiz o fato requerido de que o mínimo de (3.21) é zero para a separação linear desses conjuntos. ■

Desse resultado segue imediatamente o importante corolário.

**Corolário 3.2** ([9]). *Os conjuntos  $\bar{E}$  e  $\bar{F}$  representados pelas matrizes  $E \in \mathbb{R}^{m \times n}$  e  $F \in \mathbb{R}^{k \times n}$  respectivamente, são linearmente separáveis se, e somente se, a solução do problema de Programação Linear (3.20) é zero.*

**Demonstração:** Seja  $(\bar{w}, \bar{\gamma}, \bar{y}, \bar{z})$  a solução ótima de (3.20). Nesse caso, a função objetivo de (3.20) é zero se, e somente se,  $\bar{y} = 0$  e  $\bar{z} = 0$ , de onde segue que  $E\bar{w} - e\bar{\gamma} - e \geq 0$  e  $-F\bar{w} + e\bar{\gamma} - e \geq 0$ , implicando assim na separabilidade linear de  $\bar{E}$  e  $\bar{F}$ . ■

O teorema e o corolário anteriores estabelecem um critério para dizer se dois conjuntos de pontos são linearmente separáveis de acordo com a solução do PPL (3.20), já que, quando obtivermos como solução o valor ótimo da função objetivo igual a zero, isto é, os vetores  $y = z = 0$ , os conjuntos serão linearmente separáveis, com  $w \neq 0$ .

Os mesmos também nos dizem que o problema de PL (3.20) sempre irá gerar um hiperplano de separação  $w'x = \gamma$  para os conjuntos linearmente separáveis  $\bar{E}$  e  $\bar{F}$ , satisfazendo assim a primeira propriedade enunciada no início dessa seção. Fato importante de se mencionar é que, segundo Bennett [2], a constante 1 somada em cada  $-E_i w + \gamma$  e em cada  $F_i w - \gamma$  assegura que nenhum ponto em nenhuma das classes de padrões estará diretamente sobre o plano de separação, isso para o caso linearmente separável.

Já para os conjuntos  $\bar{E}$  e  $\bar{F}$  linearmente inseparáveis o problema de otimização (3.21) irá gerar um hiperplano de separação  $w'x = \gamma$  que minimiza as médias das violações dos pontos de  $\bar{E}$  que estão do "lado errado" do hiperplano  $w'x = \gamma + 1$ , e dos pontos de  $\bar{F}$  que estão do "lado errado" do hiperplano  $w'x = \gamma - 1$ . Segundo Bennett [2], a constante 1 que dá a localização dos planos  $w'x = \gamma + 1$  e  $w'x = \gamma - 1$  pode ser simplesmente considerada como um fator positivo qualquer, e portanto, pode ser substituída por qualquer  $\xi > 0$ . Assim o problema de PL (3.20) irá gerar ainda um hiperplano  $w'x = \gamma$  que minimiza os erros de classificação para qualquer  $\xi > 0$ .

É mais fácil entender o porquê da geração de um hiperplano que minimiza as médias das violações dos pontos, para o caso linearmente inseparável, a partir da solução do problema de Programação Linear (3.20), quando olhamos para (3.21), que também poderia ser escrito como

$$\min_{w,\gamma} \left\{ \frac{1}{m} \sum_{i=1}^m (-E_i w + \gamma + 1)_+ + \frac{1}{k} \sum_{j=1}^k (F_j w - \gamma + 1)_+ \right\} \quad (3.27)$$

Dessa última expressão vemos que, quando  $-E_i w + \gamma + 1 \leq 0$  para algum  $i = 1, 2, \dots, m$ , ou seja, quando o ponto representado por  $E_i$  está do lado certo do hiperplano  $w'x = \gamma + 1$ , o termo dado por esse  $i$  irá contribuir com zero na primeira parte do somatório acima. Quando  $-E_i w + \gamma + 1 \geq 0$  para algum  $i = 1, 2, \dots, m$ , caso em que o ponto representado por  $E_i$  está do lado errado do hiperplano  $w'x = \gamma + 1$ , o termo correspondente a esse  $i$  irá contribuir com um valor positivo na primeira parte do somatório acima. O mesmo ocorre para os pontos do conjunto  $\bar{F}$ . As contribuições positivas formarão a média das violações dos pontos que estão do lado errado do hiperplano, conforme propõe a segunda propriedade de (3.20) do início da seção. Neste caso, o que ocorre é que os valores de  $y, z \neq 0$  representam as violações e mesmo assim um plano é gerado. A Figura 3.4 ajuda a ilustrar melhor tal situação para o caso linearmente inseparável em  $\mathbb{R}^2$ .

Em formulações anteriores do PPL (3.20) havia a necessidade de se colocar restrições extras para garantir a obtenção de  $w \neq 0$  na solução. O PPL proposto por Smith [24], citado em [3], propõe os pesos  $\lambda_1 = \lambda_2 = \frac{1}{m+k}$ , e obtém algumas vezes  $w = 0$  como solução. No entanto, os problemas baseados em situações reais geralmente possuem conjuntos linearmente inseparáveis, e a obtenção de uma solução com  $w = 0$  não nos permite traçar hiperplanos.

O Teorema 3.3 à frente garante que, quando  $w = 0$  é solução ótima do PPL (3.20), este PPL apresenta outra solução ótima com  $w \neq 0$  (caso de soluções ótimas alternati-

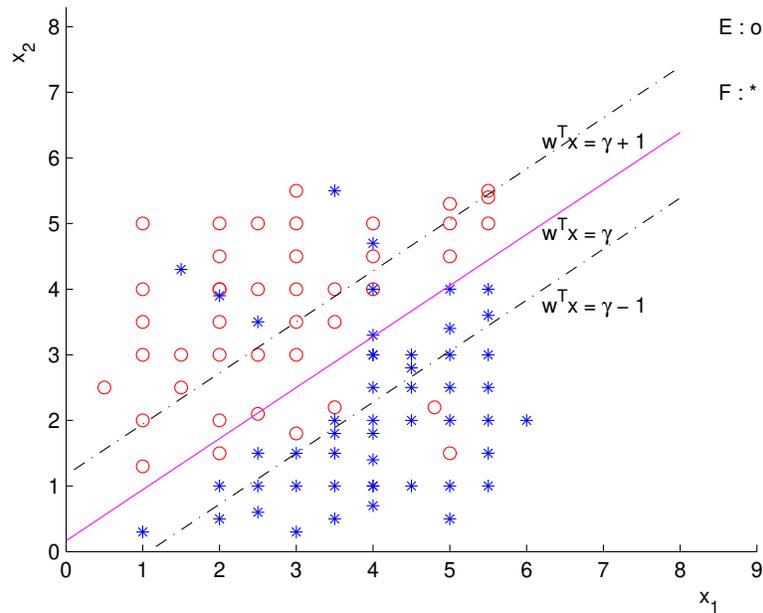


Figura 3.4: Representação de dois conjuntos linearmente inseparáveis, adaptado de [2] (Obs: nesta figura usamos  $w^T$  em vez de  $w'$ ).

vas), de modo que um plano sempre pode ser traçado também para o caso linearmente inseparável. Isto justifica a propriedade 3 apresentada no início dessa seção.

**Teorema 3.3** (Não unicidade da solução para  $w$  nulo). *A solução ( $w = 0, \gamma, y, z$ ) para (3.20) não é única.*

A prova desse teorema pode ser encontrada em [3].

Assim sendo, o PPL (3.20) aqui proposto é mais completo e eficiente que o PPL proposto no Teorema 3.1 da seção anterior. Além disso, ele é mais completo também que o PPL proposto por Smith [24], já que não é interessante obter a solução  $w = 0$  para conjuntos linearmente inseparáveis, pois, neste caso, não existe o hiperplano separador. Nesse sentido é essencial que a solução do PPL seja  $w \neq 0$ , justificando a importância do teorema anterior.

### 3.3.1 Algumas particularidades do problema

O problema de PL (3.20) é apresentado com as restrições a partir das matrizes  $E$  e  $F$ , o que facilita a identificação dos dados, uma vez que queremos separar os conjuntos de pontos associados a estas matrizes. Porém, para visualizar melhor certas características do problema, podemos mudar a forma de apresentar a função objetivo e as restrições do mesmo, utilizando matrizes particionadas. Assim, o PPL (3.20) pode ser escrito como:

$$\begin{aligned}
 \min \quad & \left[ 0 \mid (1/m)e \mid (1/k)e \mid 0 \right] \begin{bmatrix} w \\ y \\ z \\ \gamma \end{bmatrix} \\
 s.a. \quad & \left\{ \begin{array}{l} \left[ \begin{array}{c|c|c|c} E & I_m & 0 & -e \\ \hline -F & 0 & I_k & e \end{array} \right] \begin{bmatrix} w \\ y \\ z \\ \gamma \end{bmatrix} \geq \begin{bmatrix} e \\ e \end{bmatrix} \\ w, \gamma \text{ irrestritos de sinal; } y, z \geq 0 \end{array} \right. \end{aligned} \tag{3.28}$$

onde  $I_m$  é a matriz identidade de ordem  $m$ ,  $I_k$  é a matriz identidade de ordem  $k$  e  $e$  é um vetor de uns de dimensão apropriada.

Ou ainda, opcionalmente utilizando uma única matriz  $D$  e um único vetor de variáveis  $x$ , podemos escrever:

$$\begin{aligned}
 \min \quad & c'x \\
 s.a. \quad & \left\{ Dx \geq b \right. \end{aligned} \tag{3.29}$$

onde:

$x = [w \ y \ z \ \gamma]'$ ,  $x \in \mathbb{R}^p$ , com  $p = n + m + k + 1$ ,  $y, z \geq 0$ ,  $w, \gamma$  irrestritos de sinal.

$$c = [0 \mid \frac{1}{m}e \mid \frac{1}{k}e \mid 0]'; \quad c \in \mathbb{R}^p$$

$$b = (1 \ \dots \ 1); \quad b \in \mathbb{R}^{m+k}$$

$$D = \left[ \begin{array}{c|c|c|c} \overbrace{\phantom{E}}^w & \overbrace{\phantom{I_m}}^y & \overbrace{\phantom{0}}^z & \overbrace{\phantom{-e}}^\gamma \\ \hline E & I_m & 0 & -e \\ \hline -F & 0 & I_k & e \end{array} \right]; \quad D \in \mathbb{R}^{(m+k) \times p}$$

Observando a matriz  $D$ , que representa as restrições do problema, podemos ver que é conveniente resolver o PPL (3.20), proposto para separação de padrões, pelo método Simplex, pois ele apresenta uma matriz base  $B = I_{(m+k)}$  quando tomamos como variáveis básicas as variáveis associadas aos vetores  $y$  e  $z$ .

Faz-se necessário mencionar que, como  $w$  e  $\gamma$  são livres de sinal, poderíamos escrever, conforme foi discutido no primeiro capítulo,  $w = \bar{w} - \bar{\bar{w}}$  e  $\gamma = \bar{\gamma} - \bar{\bar{\gamma}}$ , com  $\bar{w}, \bar{\bar{w}}, \bar{\gamma}, \bar{\bar{\gamma}} \geq 0$ , acrescentando assim, para cada componente de  $w$ , mais uma coluna na matriz  $D$ , além de mais uma coluna pela componente  $\gamma$ , passando então a ter  $(2n + m + k + 2)$  colunas, continuando com as  $(m + k)$  linhas. No entanto, para problemas com um número muito grande de colunas, dobrar o número de colunas relacionadas a  $w$ , para trabalhar apenas com variáveis positivas, pode aumentar bastante a dimensão do problema, podendo requerer maior tempo computacional na resolução. Nesse sentido, pode ser melhor assumir que as componentes de  $w$ , e até mesmo  $\gamma$ , estejam limitadas a um intervalo que seja conveniente. Nos problemas que resolvemos, tal estratégia se mostrou satisfatória, como será visto no próximo capítulo em nossos exemplos, embora o *software* utilizado tivesse a opção de se usar variáveis livres de sinal.

Outro ponto que merece ser ressaltado quando trabalhamos com conjuntos linearmente inseparáveis é que, a partir do hiperplano  $w'x = \gamma$ , gerado pela solução do PPL

(3.20), novos hiperplanos podem ser gerados na tentativa de separar os conjuntos com um número finito de hiperplanos. Em outras palavras, podemos utilizar recursivamente o mesmo procedimento — resolução de um PPL — para um ou mais semiespaços do tipo  $w'x > \gamma$  ou  $w'x < \gamma$ , quando ocorre a mistura de pontos de  $E$  e  $F$  no mesmo semiespaço. Para isso, podemos construir novos hiperplanos levando em conta as regiões já determinadas pelos hiperplanos anteriores. Assim, obtida a solução do PPL (3.20), dependendo se  $y > 0$  ou  $z > 0$ , escolhemos o semiespaço  $w'x < \gamma$  ou  $w'x > \gamma$  e descartamos os pontos que ficaram no outro semiespaço. Isso significa que as matrizes  $E$  e  $F$  se alteram (diminuem) a cada novo hiperplano. Esse procedimento será melhor explicado no próximo capítulo.



---

# EXEMPLOS DE SEPARAÇÃO DE PONTOS

---

O objetivo desse capítulo é apresentar alguns conjuntos de pontos e a partir deles aplicar a teoria discutida no Capítulo 3, verificando se os mesmos são linearmente separáveis ou não e encontrar o hiperplano separador para cada caso. Aproveitamos ainda para comentar um pouco sobre a utilização da separação de padrões no diagnóstico do câncer de mama, aplicação que é apresentada no artigo de Mangasarian, Street e Wolberg [16], e que motivou nosso trabalho.

## 4.1 Exemplos de separação de pontos

Apresentaremos aqui alguns conjuntos relativamente simples, onde resolveremos o problema de PL (3.20) apresentado no Capítulo 3, para verificarmos se os conjuntos são linearmente separáveis ou não e encontrar o hiperplano separador. Para isso é necessário a utilização de um *software* que resolva problemas de PL, e que, nos casos possíveis possibilite a visualização do plano (ou reta) separador e dos dois conjuntos de pontos.

No início de nosso trabalho, utilizamos o *software* ProLin [17], que é distribuído gratuitamente para fins educacionais. Resolvemos o PPL (3.20) primeiramente para conjuntos pequenos, utilizando esse *software*. Porém, a medida em que fomos aumentando a quantidade de pontos em cada conjunto, a utilização desse *software* se tornou inviável, pois o mesmo requer que todas as entradas das restrições e da função objetivo

sejam dadas manualmente. Além disso, o ProLin apenas resolve o PPL, e não possibilita a representação de gráficos.

Inicialmente foi mais complicado a representação gráfica dos pontos e da reta separadora, pois, de posse dos dados do problema e da solução fornecida pelo ProLin, recorriamos ao *software* Winplot [26] para realizar a representação. O Winplot também é um *software* livre<sup>1</sup> de fácil manipulação e muito utilizado para esses fins. Porém, o mesmo também necessita que a entrada dos pontos seja feita manualmente, assim como a reta separadora, o que também dificultou muito sua utilização quando aumentamos o número de pontos de nossos exemplos (testes).

Frente a essa dificuldade de entrar com os dados do problema para conjuntos com um número maior de pontos, optamos por utilizar o *software* MATLAB, que se encontra disponível nos laboratórios do IMECC.

O MATLAB [6], é um *software* de alta performance voltado para o cálculo numérico. Uma das grandes vantagens deste *software* é o fato dele integrar assuntos como análise numérica, cálculo de matrizes, processamento de sinais, construção de gráficos, entre outros recursos, como por exemplo o procedimento LINPROG utilizado para resolver problemas de Programação Linear em um ambiente relativamente fácil de se usar.

Para resolver nossos exemplos utilizando o MATLAB, montamos uma rotina que, a partir da entrada das matrizes  $E \in \mathbb{R}^{m \times n}$  e  $F \in \mathbb{R}^{k \times n}$ , nos fornece a matriz  $D$  de restrições do problema (3.20). Fornecemos ainda como entrada, os limites superiores e inferiores para as variáveis livres de sinal, e utilizando o LINPROG com a opção pelo **método Simplex** (essa opção foi utilizada em todos os exemplos), o *software* apresenta como resposta o valor ótimo da função objetivo e das variáveis  $w$ ,  $y$ ,  $z$ , (vetores) e  $\gamma$  (escalar). Para o caso de  $n = 2$  o programa representa o gráfico dos pontos e a reta

---

<sup>1</sup> *Software* livre é a denominação que se dá a qualquer programa de computador que pode ser usado, copiado, estudado e redistribuído sem restrição, oposto de *software* restritivo ou *software* proprietário.

separadora.

As variáveis livres de sinal comentadas acima são as que compõem o vetor  $w$  e o escalar  $\gamma$ . Poderíamos, conforme foi comentado no Capítulo 1, reescrever cada componente do vetor  $w$  e o escalar  $\gamma$  como a diferença de duas variáveis positivas, ou seja,  $w_1 = \bar{w}_1 - \underline{w}_1$ ,  $w_2 = \bar{w}_2 - \underline{w}_2$ ,  $\dots$ ,  $w_n = \bar{w}_n - \underline{w}_n$ ,  $\bar{w}_i, \underline{w}_i \geq 0 \forall i = 1, 2, \dots, n$ , e  $\gamma = \bar{\gamma} - \underline{\gamma}$  com  $\bar{\gamma}, \underline{\gamma} \geq 0$ . O que ocorre, é que com isso, seria necessário dobrar as  $n$  colunas da matriz  $D$  associadas à variável  $w$ , além da coluna associada a variável  $\gamma$ , o que não seria adequado para problemas grandes.

Para se resolver o PPL (3.20) utilizando o MATLAB e o LINPROG, é necessário definir os valores dos limitantes superior e inferior de cada variável do problema. Esses limitantes representam os extremos do intervalo de valores que cada componente pode assumir. Por exemplo, se colocássemos como limite superior o valor 10 e limite inferior -8 para a componente de  $w_1$  do vetor  $w$ , isso que dizer que  $w_1$  pode assumir qualquer valor do intervalo  $[-8, 10]$ . É necessário ressaltar que o MATLAB permite definir variáveis livres de sinal, isto é, intervalos de variação do tipo  $(-\infty, \infty)$ , através do comando *inf*. Na verdade, sabemos que, computacionalmente, um intervalo real do tipo  $(-\infty, \infty)$  é substituído por uma representação numérica destes limitantes, modelando assim um intervalo de variação que atenda às necessidades do problema sem grande prejuízos.

Testamos inúmeros exemplos de conjuntos de pontos com a rotina construída no MATLAB, apesar de no presente texto apresentarmos apenas alguns desses que exemplificam certas características encontradas. Ocorre que, com os exemplos, percebemos que nem sempre usar um intervalo do tipo  $(-\infty, \infty)$  é satisfatório, e encontramos bons resultados usando um valor fixo bem menor para os limites desse intervalo. Variando os comprimentos desse intervalo, pudemos encontrar soluções diferentes inclusive em sua qualidade, como pode ser observado nas soluções gráficas apresentadas.

Ressaltamos ainda que é possível atribuir valores diferentes para os limitantes de cada componente do vetor  $w$  e para  $\gamma$ . Contudo, através das soluções de nossos exem-

plos, percebemos que era viável utilizar um intervalo de variação igual para cada componente do vetor, o que facilita também a entrada dos dados para a aplicação da rotina construída. Para  $\gamma$ , percebemos em certos exemplos, que o hiperplano proposto não era satisfatório quando este assumia o mesmo intervalo de variação de  $w$ . Por isso, estabelecemos para  $\gamma$  um intervalo de variação 3 vezes maior que o de  $w$ . Por exemplo, se escolhermos que as componentes de  $w$  assumem valores no intervalo  $[-2, 2]$ ,  $\gamma$  poderá assumir qualquer valor do intervalo  $[-6, 6]$ . Chegamos a este valor experimentalmente, depois de vários testes computacionais. Esse procedimento foi aplicado a todos os nossos exemplos e pelo menos nesses, um hiperplano separador sempre foi gerado.

#### 4.1.1 Exemplos com conjuntos linearmente separáveis

Vamos iniciar os exemplos mostrando conjuntos linearmente separáveis. Optamos por conjuntos de pontos de  $\mathbb{R}^2$ , pelo fato dos mesmos possibilitarem um apelo visual para compreender melhor como o plano (reta) separador se comporta. Lembramos que para conjuntos linearmente separáveis, a solução do problema (3.20), fornece um hiperplano (reta em nossos exemplos) que separa completamente os dois conjuntos analisados.

**Exemplo 4.1.** Iniciamos com dois conjuntos de pontos estritamente separáveis e com um número pequeno de pontos: 9 pontos no conjunto  $E$  e 7 no conjunto  $F$ . Sem fazer distinção entre os conjuntos e as matrizes que os representam, temos:

$$E = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 2 & 1 \\ 2 & 2 \\ 2 & 3 \\ 2 & 4 \\ 3 & 3 \\ 3 & 4 \end{bmatrix}, \quad F = \begin{bmatrix} 4 & 1 \\ 4 & 2 \\ 4 & 3 \\ 5 & 2 \\ 5 & 3 \\ 5 & 4 \\ 6 & 2 \end{bmatrix}$$

Como cada linha das matrizes representa um ponto  $(x_1, x_2)$ , é fácil perceber que os conjuntos são linearmente separáveis, já que qualquer reta paralela ao eixo  $x_2$ , com  $x_1$  entre os valores 3 e 4, separa linearmente esses conjuntos.

Apesar do pequeno número de pontos escolhidos para esse exemplo, ao montarmos o PPL (3.20) correspondente a esses pontos, vemos a necessidade da utilização dos recursos computacionais para a busca da solução do problema, pois mesmo com  $n = 2$ , como é o caso desse problema, o número de variáveis do PPL aumenta à medida em que o número de pontos aumenta. Para as matrizes  $E$  e  $F$  escolhidas acima ( $m = 9$ ,  $k = 7$ ,  $n = 2$ ), o PPL a ser resolvido é:



A estrutura do problema também estabelece a necessidade da utilização de um *software* que permita a fácil manipulação de matrizes de tamanho relativamente grande, de modo que, para problemas maiores, não exista a necessidade de se entrar manualmente com o valor de cada elemento. No *software* MATLAB, a matriz  $D$  dos coeficientes do PPL pode ser facilmente construída.

**Observação 4.2.** Para todos os exemplos deste texto, apresentaremos os resultados obtidos pelo MATLAB com um arredondamento de duas casas decimais.

Resolvendo esse problema com o MATLAB e utilizando como limites inferior e superior para as variáveis livres de sinal, os valores  $-2$  e  $2$ , respectivamente para  $w$ , o que nos fornece para  $\gamma$ , os limitantes  $-6$  e  $6$ , conforme comentado anteriormente, obtemos como resposta:

- Valor ótimo da FO: zero, o que indica, segundo os resultados apresentados no capítulo anterior, que os conjuntos são linearmente separáveis;
- $w = (w_1, w_2) = (-2; -0,33)$ ;
- $\gamma = -6$ ;
- $y \in \mathbb{R}^9$  e  $z \in \mathbb{R}^7$  vetores nulos dos respectivos espaços.

Nesse caso a reta separadora será dada por  $x_2 = \left( \frac{2x_1 - 6}{0,33} \right)$ . A Figura 4.1 ilustra os conjuntos de pontos e a reta separadora para esses limites.

Como podemos ver, essa não é a única reta separadora para esses conjuntos, pois existem infinitas retas que os separam linearmente. Mudando apenas os limites das variáveis livres de sinal obtemos outras retas separadoras; por exemplo, alterando um pouco os valores dos limites para  $-3 \leq w_1 \leq 3$ ,  $3 \leq w_2 \leq 3$  e  $-9 \leq \gamma \leq 9$ , a reta separadora é completamente alterada, inclusive passa de crescente, como era na Figura 4.1, para decrescente. A Figura 4.2 representa os conjuntos de pontos e a reta separadora

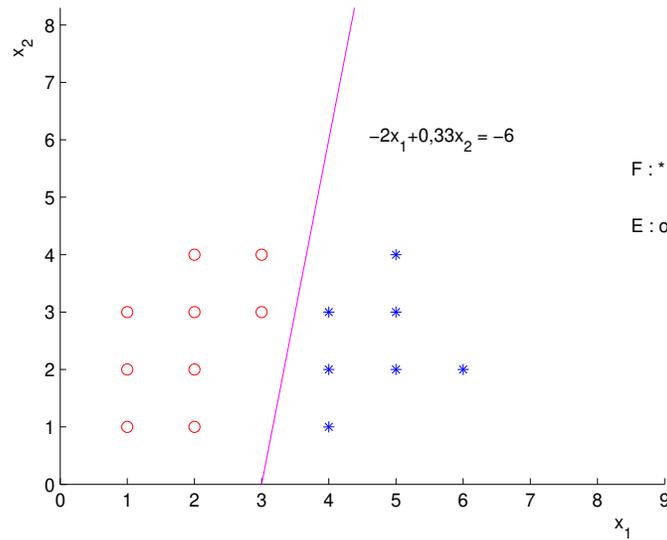


Figura 4.1: Representação dos conjuntos linearmente separáveis do Exemplo 4.1 e da reta separadora com limites inferior e superior pequenos.

para esses novos limites. Com esses limites o MATLAB apresenta os seguintes valores:

- Valor ótimo da FO: zero (linearmente separáveis);
- $w = (w_1, w_2) = (-2, 46; -0, 15)$ ;
- $\gamma = -9$ ;
- $y \in \mathbb{R}^9$  e  $z \in \mathbb{R}^7$  vetores nulos dos respectivos espaços.
- reta separadora  $x_2 = - \left( \frac{2,46x_1 - 9}{0,15} \right)$

Mudando novamente os limites das variáveis livres de sinal, por exemplo, colocando como limites para  $w_1$  e  $w_2$  -20 e 20 e -60 e 60 para  $\gamma$  e resolvendo novamente o problema obtemos como resposta:

- Valor ótimo da FO: zero (linearmente separável);
- $w = (w_1, w_2) = (-14, 23; -4, 07)$ ;
- $\gamma = -60$ ;

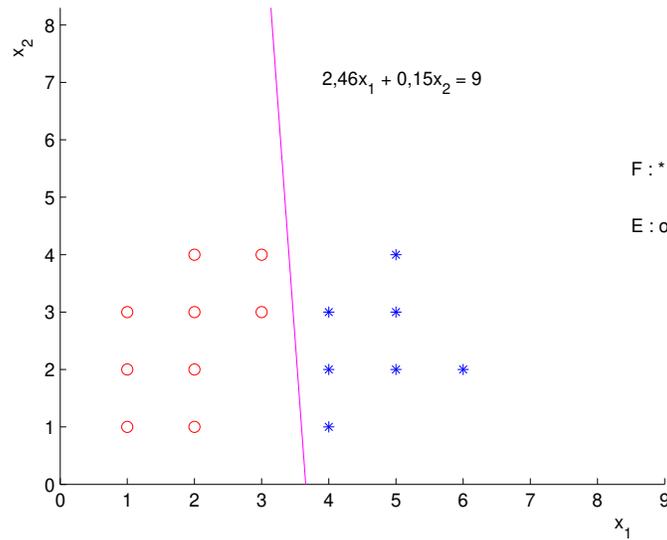


Figura 4.2: Representação dos conjuntos linearmente separáveis do Exemplo 4.1 e da reta separadora com  $w$  e  $\gamma$  entre  $[-3, 3]$  e  $[-9, 9]$  respectivamente.

- $y \in \mathbb{R}^9$  e  $z \in \mathbb{R}^7$  vetores nulos dos respectivos espaços.

Ou seja, a reta separadora será dada por  $x_2 = -\left(\frac{14,23x_1 - 60}{4,07}\right)$ . A Figura 4.3 ilustra os conjuntos de pontos e a reta separadora para esses limites.

Como podemos observar, quanto maior for o comprimento do intervalo  $[l, u]$ , onde  $l$  e  $u$  indicam os limites inferior e superior para as variáveis livres de sinal, mais a reta separadora tende a se aproximar de alguns pontos dos conjuntos. Porém, para os casos linearmente separáveis, como é comentado por [2] e apresentado no capítulo anterior, a reta nunca irá conter nenhum ponto desses conjuntos.

Ocorre também que, quando acrescentamos um novo ponto do conjunto  $E$ , em um dos semiespaços que só continha pontos de  $F$ , se os conjuntos ainda continuarem linearmente separáveis, como era esperado, a reta separadora se ajusta mantendo a separação dos conjuntos. Isso pode ser observado, acrescentando, por exemplo, o ponto  $(3, 3; 4)$  ao conjunto  $E$  no presente exemplo, mantendo o comprimento do intervalo dos limitantes em  $[-20, 20]$  para  $w$  e  $[-60, 60]$  para  $\gamma$ , intervalo esse, em que a reta separadora está próxima de alguns pontos de ambos os conjuntos, como pode ser observado na

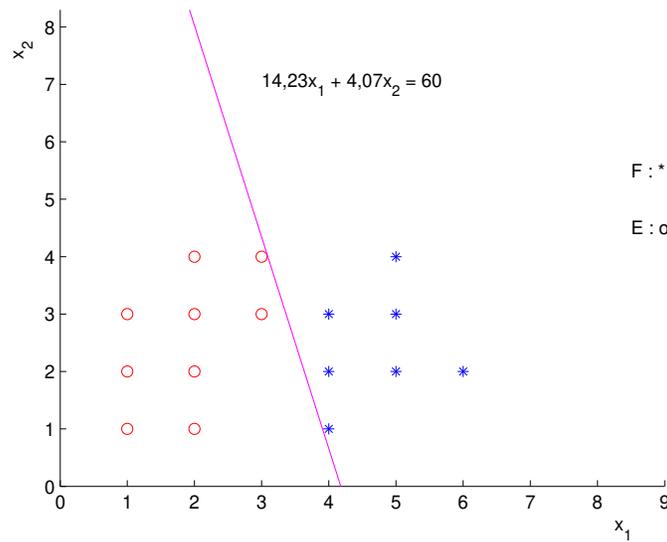


Figura 4.3: Representação dos conjuntos linearmente separáveis do Exemplo 4.1 e da reta separadora com limite inferior e superior maiores.

Figura 4.3. Resolvendo o PPL novamente, encontramos uma reta que mantém a separabilidade, como é ilustrado na Figura 4.4.

Percebemos, ao resolver vários desses problemas com o MATLAB, que o valor ótimo de  $\gamma$  muitas vezes é o próprio limitante inferior desta variável. Para problemas com poucos pontos, quando alteramos este limitante, o valor ótimo de  $\gamma$  é proporcional ao anterior (como é o caso do presente exemplo). Isso não acontece com o valor de  $w$ , que não se altera na mesma proporção — caso contrário não seria garantida a separabilidade — no entanto, não é uma regra o valor ótimo para  $\gamma$  ser igual ao seu limitante inferior, como poderemos ver em outros exemplos.

Quanto à melhor escolha dos limitantes para as variáveis livres de sinal fica, a princípio, a cargo da necessidade de cada exemplo. Se quisermos uma separação onde o plano separador passe mais perto de alguns pontos dos conjuntos, devemos atribuir um intervalo  $[l, u]$  com comprimento maior, se quisermos o plano separador passando mais afastado dos pontos dos conjuntos, situação que aparentemente parece ser melhor, devemos trabalhar um intervalo de comprimento menor. Uma melhor discussão sobre

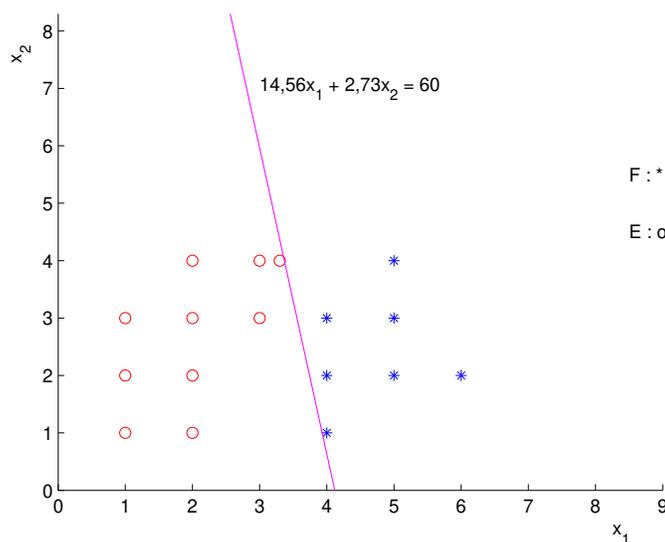


Figura 4.4: Representação dos conjuntos linearmente separáveis do Exemplo 4.1 e da reta separadora após acréscimo do ponto  $(3, 3; 4)$ .

a influência desses limitantes no plano separador fica como sugestão para trabalhos futuros.

**Exemplo 4.2.** Vamos considerar agora dois conjuntos de pontos de  $\mathbb{R}^2$  um pouco maiores que os conjuntos do exemplo anterior: um conjunto  $E$  com 39 pontos e um conjunto  $F$  com 35 pontos, representados a seguir através das matrizes transpostas de  $E$  e  $F$  (optamos por representar esses conjuntos através das matrizes transpostas simplesmente para otimizar espaço nesse texto).

$$E' = \begin{bmatrix} 4 & 4,3 & 2,5 & 4,5 & 5 & 7 & 6 & 6,5 & 7,5 & 5 & 5,5 & 5,5 & 4,5 & 3,5 & 3,4 & 4 & 6,5 & 4 & 6 \\ 1 & 2,5 & 1,4 & 1 & 1,4 & 2 & 3 & 2,7 & 3,5 & 3 & 4 & 2,8 & 4 & 2 & 3 & 3 & 4,5 & 4 & 5,5 \\ 5 & 5 & 6 & 6 & 5,5 & 5,7 & 3 & 3 & 3,5 & 4,5 & 5 & 5 & 5,5 & 6,5 & 5,5 & 6,5 & 5 & 4,5 & 5,5 & 4,5 \\ 5 & 5,5 & 1 & 2 & 1,4 & 2,3 & 1 & 1,3 & 2,5 & 4,5 & 4 & 4,5 & 4,5 & 6 & 6 & 3,5 & 2 & 1,5 & 3,5 & 3,5 \end{bmatrix},$$

$$F' = \begin{bmatrix} 0,5 & 0,5 & 0,5 & 0,7 & 0,3 & 1,5 & 1 & 1 & 1,3 & 1,5 & 1,5 & 1,5 & 2 & 2,5 & 1 & 4 & 3 & 2 \\ 2,4 & 4,5 & 6 & 3 & 1 & 1,5 & 6 & 2,5 & 3,3 & 3,5 & 4,3 & 5,2 & 4,4 & 2,5 & 4 & 5,5 & 6 & 6 \\ \\ 1 & 1,5 & 2 & 2,5 & 2,6 & 2 & 2,5 & 2,5 & 3,2 & 3 & 1 & 1,5 & 4 & 2 & 3 & 2 & 3,5 \\ 5 & 3 & 3 & 3,5 & 6 & 5,7 & 5,5 & 4,7 & 4,5 & 4 & 1 & 2 & 6 & 5 & 5,3 & 4 & 4,5 \end{bmatrix}$$

Resolvendo o problema de Programação Linear para separação de pontos proposto neste trabalho, utilizando o MATLAB, obtemos os valores:

- Valor ótimo da FO: zero (linearmente separável);
- $w = (w_1, w_2) = (3, 23; -2, 82)$ ;
- $\gamma = 4, 53$ ;
- $y \in \mathbb{R}^{39}$  e  $z \in \mathbb{R}^{35}$  vetores nulos dos respectivos espaços.

Logo, a reta separadora será dada pela equação  $x_2 = -\left(\frac{-3, 23x_1 + 4, 53}{2, 82}\right)$ . A Figura 4.5 mostra a representação dos conjuntos de pontos e da reta separadora correspondente.

Para esse exemplo, como o espaço que separa os conjuntos é menor, a variação nos limitantes das variáveis livres de sinal não influencia significativamente na posição da reta separadora, apesar de ainda assim haver infinitas retas que separam linearmente esses dois conjuntos.

### 4.1.2 Exemplos com conjuntos linearmente inseparáveis

Para os casos linearmente inseparáveis, lembramos que o problema de PL (3.20) gera um hiperplano que minimiza a média das violações dos pontos. Para esse caso, alguns pontos podem estar contidos sobre o hiperplano, diferentemente de quando os

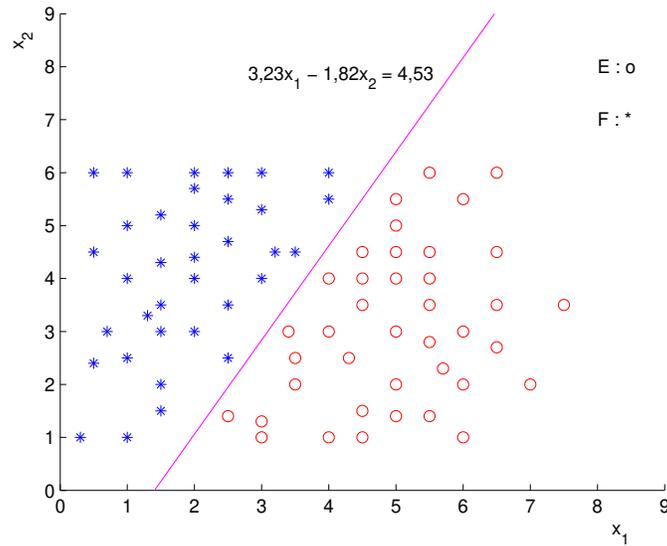


Figura 4.5: Representação dos conjuntos linearmente separáveis do Exemplo 4.2 e da reta separadora.

conjuntos são linearmente separáveis.

Temos também que o valor ótimo da função objetivo não será zero, ou seja, será algum valor positivo, uma vez que os vetores  $y$  e  $z$  não serão vetores nulos simultaneamente. Vejamos um exemplo.

**Exemplo 4.3.** Vamos considerar os conjuntos de pontos abaixo:  $E$  com 16 pontos e  $F$  com 14 pontos.

$$E' = \begin{bmatrix} 1,7 & 4,8 & 4,9 & 6 & 3,2 & 1,3 & 0,4 & 2,1 & 3 & 1,5 & 2,2 & 3,2 & 1 & 0,6 & 0,8 & 4 \\ 0,2 & 0,3 & 1,5 & 0,2 & 0,8 & 1,4 & 2,2 & 2,1 & 1,8 & 3,2 & 4 & 4,8 & 4,5 & 4,9 & 5,2 & 6 \end{bmatrix}$$

$$F' = \begin{bmatrix} 0,8 & 2,8 & 3,2 & 4,7 & 4,5 & 4,1 & 5,2 & 5,6 & 6 & 6 & 5 & 6 & 2,5 & 2,5 \\ 6,7 & 6,5 & 4,6 & 5,7 & 6,7 & 2 & 6,2 & 6,8 & 4,2 & 5,7 & 3 & 3 & 3 & 6 \end{bmatrix}$$

Resolvendo o problema de PL proposto novamente com o *software* MATLAB, obtemos como resposta os valores:

- Valor ótimo da FO: 0,76 (valor positivo indicando que os pontos não são linearmente separáveis);
- $w = (w_1, w_2) = (-0,63; -0,73)$ ;
- $\gamma = -4,93$ ;
- $y' = (0, 0, 0, 47, 0, 0, 0, 0, 0, 0, 0, 0, 0, 36, 1, 93, 0, 0, 0, 34, 2, 95)$ ,  $y \in \mathbb{R}^{16}$ ;
- $z' = (0, 56, 0, 0, 57, 0, 0, 1, 89, 0, 0, 0, 0, 0, 60, 0, 2, 18, 0)$ ,  $z \in \mathbb{R}^{14}$ .

Desses valores, segue que a reta separadora será dada pela equação  $x_2 = -\left(\frac{0,63x_1 - 4,93}{0,74}\right)$ , conforme pode ser visto na Figura 4.6.

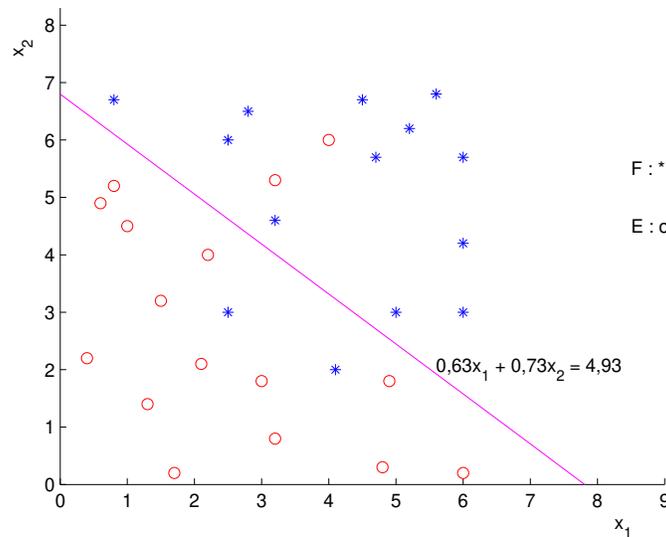


Figura 4.6: Representação dos conjuntos linearmente inseparáveis do Exemplo 4.3 e de sua melhor reta “separadora”.

Para obter essa separação fixamos os limites inferiores e superiores das variáveis livres de sinal em  $-50$  e  $50$ , respectivamente. Chegamos a alterar esses valores, mas a

reta separadora não se alterou significativamente.

Temos apenas dois pontos de cada um dos conjuntos do lado “errado” da reta; no entanto, temos 5 componentes positivas em cada um dos vetores  $y$  e  $z$ . O que ocorre é que os pontos que se encontram dentro da faixa de  $w'x = \gamma + 1$  e  $w'x = \gamma - 1$  de ambos os lados da reta, também contribuem com uma componente positiva para os vetores. Como pode ser visto na Figura 4.7, temos 3 pontos de cada conjunto dentro dessa faixa. Quando os pontos não se localizam na faixa mas estão do lado “certo” do hiperplano, a componente do vetor  $y$  ou  $z$  associada a esse ponto é zero na solução. Quando o ponto estiver dentro da faixa, ou do lado errado do hiperplano, será atribuído um valor positivo para a componente do vetor  $y$  ou  $z$  associada a esse ponto. No próximo exemplo, com apenas um ponto do lado errado, podemos observar esse fato.

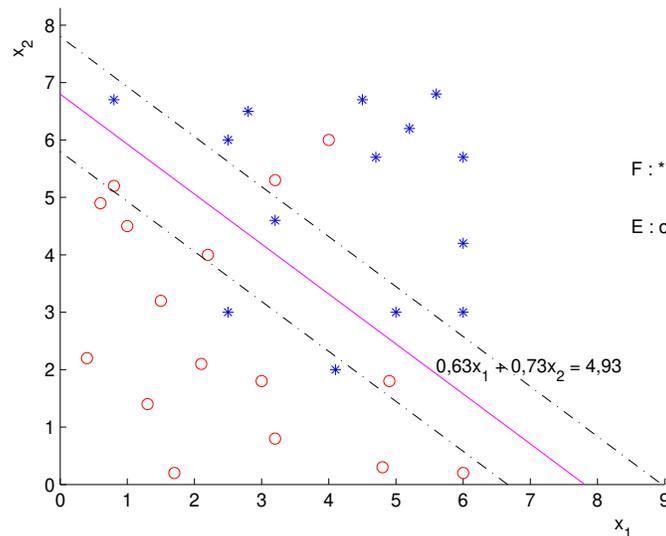


Figura 4.7: Representação dos conjuntos linearmente inseparáveis do Exemplo 4.3 e de sua melhor reta “separadora” com as faixas  $w'x = \gamma + 1$  e  $w'x = \gamma - 1$ .

**Exemplo 4.4.** Vamos considerar os conjuntos  $E$  com 4 pontos e  $F$  com 3 pontos como se segue:

$$E = \begin{bmatrix} 3 & 4 \\ 4,30 & 4,50 \\ 4,50 & 2 \\ 3 & 5,50 \end{bmatrix}, \quad F = \begin{bmatrix} 4 & 2 \\ 4,50 & 3,50 \\ 5 & 3 \end{bmatrix}$$

Resolvendo o problema de PL (3.20) associado encontramos os valores:

- Valor ótimo da FO: 0,92;
- $w = (w_1, w_2) = (-0,91; -0,91)$ ;
- $\gamma = -0,82$ ;
- $y' = (0, 0, 2,45, 0)$ ;
- $z' = (0, 0, 19, 0)$ ;
- Reta separadora  $x_2 = -\left(\frac{0,91x_1 - 0,82}{0,91}\right)$ .

A Figura 4.8 mostra os pontos e a reta “separadora” para esses pequenos conjuntos.

Veja que o ponto do conjunto  $E$  do lado errado da reta tem componente de  $y$  com o valor 2.45, e o valor 0.19 de  $z$  representa o ponto de  $F$  que está dentro da faixa.

Alterando apenas um ponto do conjunto  $E$ , a saber o ponto  $(4,30, 4,50)$  para  $(3,50, 4,50)$ , a reta “separadora” proposta mantém apenas um ponto do lado errado e deixa todos os pontos fora da faixa, ilustrando o que comentamos acima. Os resultados obtidos para esta alteração foram:

- Valor ótimo da FO: 0,69;
- $w = (w_1, w_2) = (-1,50; -0,50)$ ;

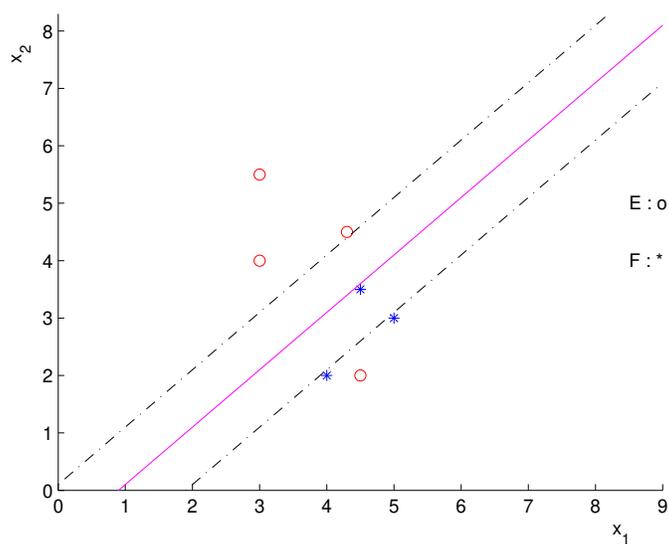


Figura 4.8: Representação dos conjuntos linearmente inseparáveis do Exemplo 4.4 e de sua melhor reta “separadora” com as faixas  $w'x = \gamma + 1$  e  $w'x = \gamma - 1$ .

- $\gamma = -4$ ;
- $y' = (0, 0, 2,75, 0)$ ;
- $z' = (0, 0, 0)$ .
- Reta separadora  $x_2 = \left( \frac{1,50x_1 - 4}{0,50} \right)$ .

A Figura 4.9 representa os conjuntos após alteração e a reta “separadora” proposta.

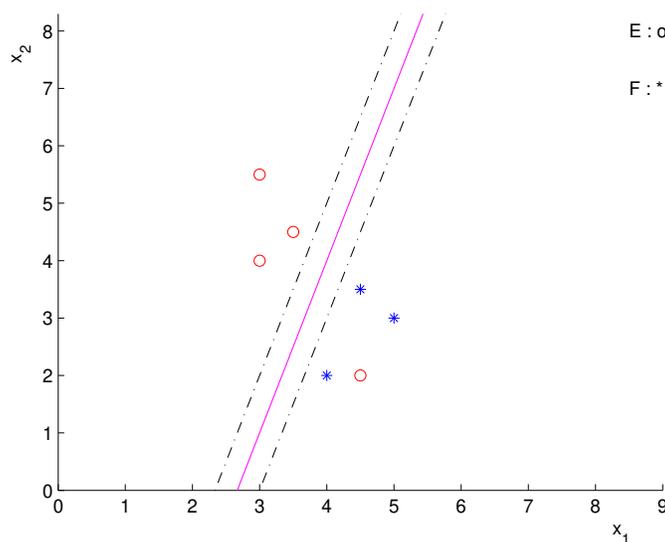


Figura 4.9: Representação dos conjuntos linearmente inseparáveis do Exemplo 4.4, alterando um ponto, e sua melhor reta “separadora” com as faixas  $w'x = \gamma + 1$  e  $w'x = \gamma - 1$ .

### 4.1.3 Separação por mais de um hiperplano

Como comentado no final do capítulo anterior, quando dois conjuntos de pontos são linearmente inseparáveis, é sempre possível considerar os pontos que estão de um lado do hiperplano separador como dois novos conjuntos e aplicar o mesmo procedimento a esses conjuntos, obtendo assim uma segunda separação. O mesmo pode ser feito com os pontos que estão do outro lado do hiperplano. Esse procedimento pode ser repetido quantas vezes for necessário.

Nesse caso, dados dois conjuntos de pontos  $E \in \mathbb{R}^{m \times n}$  e  $F \in \mathbb{R}^{k \times n}$ , obtido o hiperplano  $w'x = \gamma$  através da solução do PPL (3.20), desde que esses sejam linearmente inseparáveis, podemos aplicar o mesmo procedimento para um dos semiespaços:  $w'x > \gamma$  ou  $w'x < \gamma$ .

Tal procedimento pode ser realizado a partir do que segue:

Sabemos que cada valor das componentes do vetor  $y$  está associado a um único ponto representado na matriz  $E$  (linhas de  $E$ ), e cada componente do vetor  $z$  está associada a um único ponto de  $F$  (linha de  $F$ ). Ou seja, se  $y > 0$  ou  $z > 0$ , os pontos não são linearmente separáveis. Assim temos:

- se existe  $z_i > 1$ ,  $i \in \{1, 2, \dots, k\}$ , então os pontos de  $F$  não estão todos no mesmo semiespaço  $w'x < \gamma$ , logo o semiespaço  $w'x > \gamma$  contém pontos dos dois conjuntos.
- se existe  $y_j > 1$ ,  $j \in \{1, 2, \dots, m\}$ , então os pontos de  $E$  não estão todos no semiespaço  $w'x > \gamma$ , ou seja, o semiespaço  $w'x < \gamma$  contém pontos dos dois conjuntos.

Assim, se queremos aplicar novamente o procedimento para separação de pontos, nos semiespaços estabelecidos pelo primeiro hiperplano separador podemos proceder como segue:

- se existe  $z_i > 1$ ,  $i \in \{1, 2, \dots, k\}$ , tomar os pontos que estão no semiespaço  $w'x > \gamma$  e ignorar os pontos que estão no outro semiespaço. Construir uma nova matriz  $F^1$  onde as linhas serão compostas pelos pontos de  $F$  tais que  $F_i x > \gamma$ ,  $i \in \{1, 2, \dots, k\}$  (lembrando que  $F_i$  nada mais é que uma linha de  $F$ , a qual representa um ponto desse conjunto). Construir também, uma nova matriz  $E^1$  onde as linhas serão compostas pelos pontos de  $E$  tais que  $E_j x > \gamma$ ,  $j \in \{1, 2, \dots, m\}$ . Aplicar novamente o PPL (3.20) a essas duas novas matrizes, obtendo assim um novo hiperplano.
- se existe  $y_j > 1$ ,  $j \in \{1, 2, \dots, m\}$ , tomar os pontos que estão no semiespaço  $w'x < \gamma$  e ignorar os pontos que estão no outro semiespaço. Construir uma nova matriz  $E^2$  com os pontos de  $E$  tais que  $E_j x < \gamma$ ,  $j \in \{1, 2, \dots, m\}$ . Construir uma nova matriz  $F^2$  com os pontos de  $F$  tais que  $F_i x < \gamma$ ,  $i \in \{1, 2, \dots, k\}$ . Aplicar novamente o PPL (3.20) a essas duas novas matrizes, obtendo assim um novo hiperplano.

Vamos aplicar o exposto acima em um exemplo.

**Exemplo 4.5.** Vejam os conjuntos  $E$  com 8 pontos e  $F$  também com 8 pontos conforme segue:

$$E = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1,50 & 2,50 \\ 2 & 2 \\ 2 & 3 \\ 3 & 4 \\ 3,50 & 2,50 \end{bmatrix}, \quad F = \begin{bmatrix} 2,20 & 0,50 \\ 3 & 1 \\ 3 & 2,10 \\ 3,50 & 1,50 \\ 4 & 1 \\ 4 & 2 \\ 5 & 2,50 \\ 2,50 & 3 \end{bmatrix}$$

Resolvendo PPL (3.20), com o MATLAB encontramos como solução os valores:

- Valor ótimo da FO: 0,56 (linearmente inseparável);
- $w = (w_1, w_2) = (-1,30; 0,87)$ ;
- $\gamma = -1,43$ ;
- $y' = (0, 0, 0, 0, 0, 43, 0, 0, 1, 95)$ ;
- $z' = (0, 0, 0, 35, 0, 0, 0, 0, 1, 78)$ .
- Reta separadora  $x_2 = \left( \frac{-1,30x_1 - 1,43}{0,87} \right)$  (1ª reta).

A Figura 4.10 representa os conjuntos de pontos  $E$  e  $F$  e a reta “separadora” acima.

**Observação 4.3.** Na Figura 4.10, o lado esquerdo da reta representa o semiespaço  $-1,30x_1 + 0,87x_2 > -1,43$ , e o lado direito da reta representa o semiespaço  $-1,30x_1 + 0,87x_2 < -1,43$ .

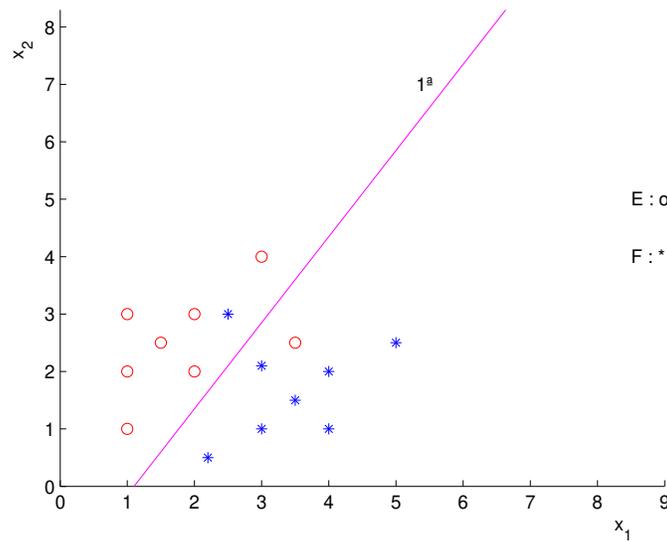


Figura 4.10: Representação dos conjuntos linearmente inseparáveis do Exemplo 4.5 com a 1ª reta “separadora”.

Como  $z_8 > 1$ , no semiespaço  $-1,30x_1 + 0,87x_2 > -1,43$  existem pontos dos dois conjuntos de pontos (note que existe mais uma componente de  $z$  positiva (0.35); porém como esse valor é menor que 1, o ponto a ele associado está dentro da faixa entre as retas  $w'x = \gamma$  e  $w'x = \gamma - 1$ ). Descartando os pontos que estão no outro semiespaço podemos montar as novas matrizes:

$$E^1 = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1,50 & 2,50 \\ 2 & 2 \\ 2 & 3 \\ 3 & 4 \end{bmatrix}, \quad F^1 = \begin{bmatrix} 2,50 & 3 \end{bmatrix}$$

Resolvendo novamente o PPL (3.20) agora com essas duas novas matrizes obtemos os valores:

- Valor ótimo da FO: 0,57 (novamente linearmente inseparável);
- $w = (w_1, w_2) = (-5; 0,50)$ ;
- $\gamma = -10$ ;
- $y' = (0, 0, 0, 0, 0, 0, 4)$ ;
- $z' = (0)$ .
- Reta separadora  $x_2 = \left(\frac{-5x_1 - 10}{0,50}\right)$  (2ª reta).

A Figura 4.11 representa os conjuntos de pontos  $E$  e  $F$ , além da 1ª e da 2ª retas.

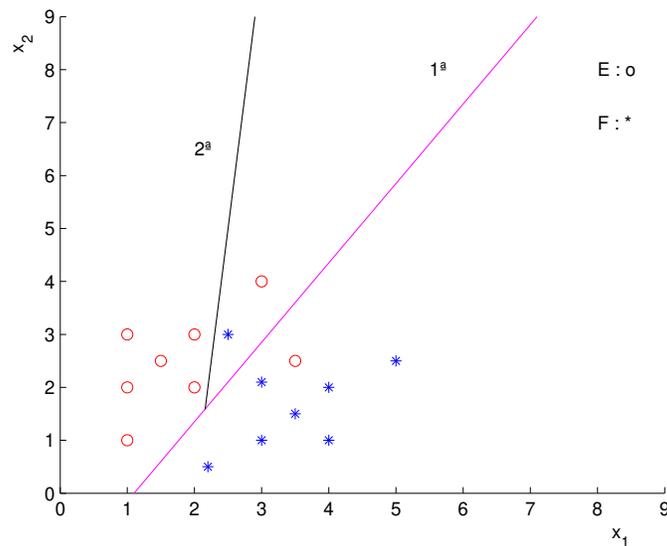


Figura 4.11: Representação dos conjuntos linearmente inseparáveis do Exemplo 4.5 com as 1ª e 2ª retas "separadoras".

Observando novamente a solução encontrada na 1ª reta, vemos que  $y_8 > 1$ . Isso indica que no semiespaço  $-1,30x_1 + 0,87x_2 < -1,43$  existem pontos dos dois conjuntos de pontos. Descartando os pontos que estão no outro semiespaço podemos montar as novas matrizes:

$$E^2 = \begin{bmatrix} 3,50 & 2,50 \end{bmatrix}, \quad F^2 = \begin{bmatrix} 2,20 & 0,50 \\ 3 & 1 \\ 3 & 2,10 \\ 3,50 & 1,50 \\ 4 & 1 \\ 4 & 2 \\ 5 & 2,50 \end{bmatrix}$$

Resolvendo novamente o PPL (3.20) com essas duas novas matrizes ( $E^2$  e  $F^2$ ) obtemos os valores:

- Valor ótimo da FO: 0 (linearmente separável);
- $w = (w_1, w_2) = (-1, 33; 6, 66)$ ;
- $\gamma = 11$ ;
- $y' = (0)$ ;
- $z' = (0, 0, 0, 0, 0, 0, 0, 0)$ .
- Reta separadora  $x_2 = \left( \frac{1, 33x_1 + 11}{6, 66} \right)$  (3ª reta).

A Figura 4.12 representa os conjuntos  $E$  e  $F$ , assim como as 3 retas separadoras propostas.

Note que nessa última aplicação, já foi possível separar linearmente os pontos que se encontravam no semiespaço  $-1, 30x_1 + 0, 87x_2 < -1, 43$ , enquanto que a segunda reta ainda não separou linearmente os pontos do semiespaço  $-1, 30x_1 + 0, 87x_2 > -1, 43$ .

Note também que encontramos as retas com aplicação recursiva do PPL (3.20), mas optamos por representar apenas as semirretas a partir da primeira reta separadora, para melhorar a visualização, e também porque a parte da reta que está do outro lado não tem sentido algum para o novo problema, pois foi obtida a partir dos pontos que estavam do lado oposto.

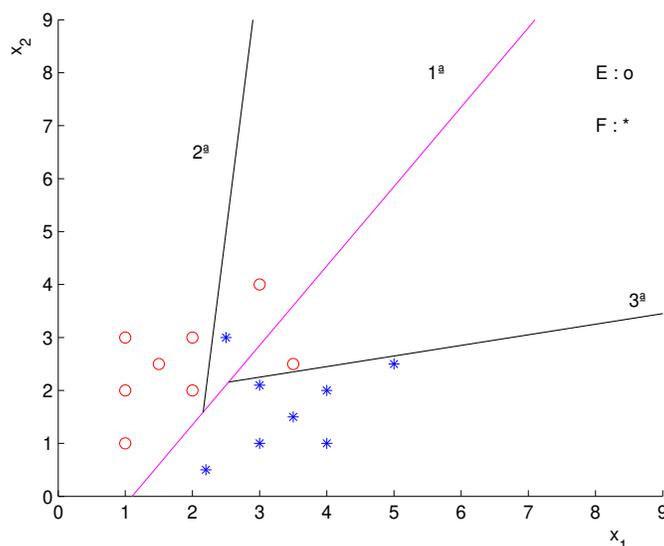


Figura 4.12: Representação dos conjuntos linearmente inseparáveis do Exemplo 4.5 com as 1ª, 2ª e 3ª retas “separadoras”.

Nesse exemplo, após obtidas as três retas separadoras, o espaço ( $\mathbb{R}^2$ ) foi dividido em quatro regiões, onde em duas delas só temos pontos do conjunto  $E$ , em uma delas só temos pontos do conjunto  $F$  e apenas uma região contínua com pontos dos dois conjuntos. Opcionalmente, poderíamos aplicar novamente o procedimento descrito acima a essa região que contém pontos dos dois conjuntos.

Vejamos mais um exemplo com dois conjuntos linearmente inseparáveis onde traçaremos a reta “separadora” e mais duas semirretas separadoras, exemplo esse com um número maior de pontos.

**Exemplo 4.6.** Vamos considerar os dois conjuntos de pontos como segue abaixo ( $m = 28$  e  $k = 28$ ):

$$E' = \begin{bmatrix} 1 & 1 & 2 & 1 & 5 & 2 & 1 & 0,5 & 1,5 & 1 & 0,5 & 0,8 & 1 & 0,5 & 0,5 & 0,8 & 6 & 1,5 & 1,5 \\ 1 & 2 & 2 & 3 & 3 & 4 & 4 & 3 & 2 & 3,5 & 2 & 4 & 4,5 & 0,5 & 1 & 2,4 & 3,5 & 3 & 4 \end{bmatrix}$$

$$F' = \begin{bmatrix} 3 & 3,5 & 2,5 & 1,5 & 2 & 2,5 & 2,5 & 2,5 & 2,3 & 5,5 \\ 3 & 4,5 & 1 & 2,5 & 1 & 3 & 3,5 & 2 & 1,8 & 3,8 \end{bmatrix}$$

$$F' = \begin{bmatrix} 4 & 5 & 6 & 2 & 5 & 1,8 & 4 & 5 & 6 & 7 & 7 & 7,5 & 7,5 & 7 & 7 & 6,5 & 6,5 & 5,5 & 5,5 \\ 1 & 2 & 2 & 2,8 & 4 & 3,2 & 3 & 1 & 1 & 3 & 3,5 & 4 & 3,5 & 2 & 1 & 1,5 & 2,5 & 2,5 & 1 \\ 4,5 & 4 & 4,5 & 7,5 & 5,5 & 6 & 7 & 6,5 & 2,5 \\ 4 & 2,5 & 0,5 & 2,5 & 1,8 & 4,6 & 4,5 & 5 & 3,8 \end{bmatrix}$$

Resolvendo o problema de PL associado obtivemos, no MATLAB, os seguintes valores:

- Valor ótimo da FO: 0,64;
- $w = (w_1, w_2) = (-0,76; 0,19)$ ;
- $\gamma = -2,33$ ;
- Reta separadora  $x_2 = \left( \frac{0,76x_1 - 2,33}{0,19} \right)$  (1ª reta).

A solução deste problema está representada na Figura 4.13, onde o lado esquerdo da reta “separadora” representa o semiespaço  $0,76x_1 + 0,19x_2 > -2,33$  e o lado esquerdo da reta “separadora” representa o semiespaço  $0,76x_1 + 0,19x_2 < -2,33$ .

Tomando os pontos que estão em um dos lados da reta e montando um novo problema de separação destes conjuntos, iremos obter uma nova reta que separará a média dos pontos que estão desse lado da reta obtida anteriormente. Esse mesmo procedimento pode ser feito com os pontos que estão do outro lado da reta.

Para este exemplo, os pontos de ambos os lados da reta ainda são inseparáveis linearmente, como pode ser observado na Figura 4.13. Aplicando o procedimento descrito anteriormente, encontraremos duas novas retas, obtendo os valores:

**Lado esquerdo (2ª reta):**

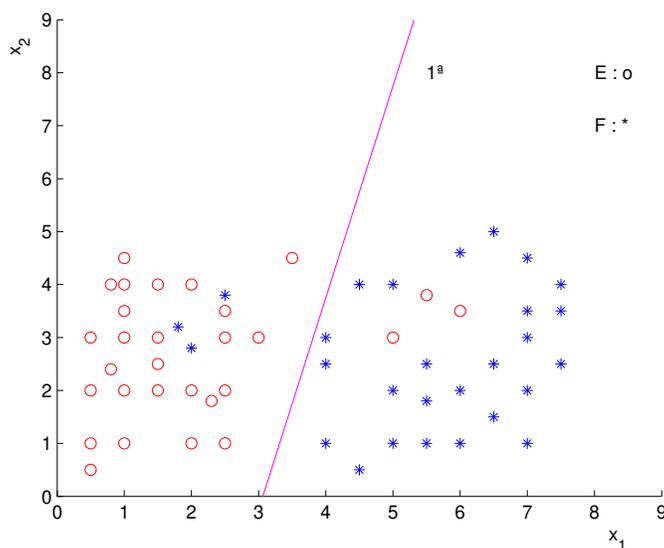


Figura 4.13: Representação dos conjuntos linearmente inseparáveis do Exemplo 4.6 e da 1ª reta “separadora”.

- Valor ótimo da FO: 0,53;
- $w = (w_1, w_2) = (-3; -1,50)$ ;
- $\gamma = -10$ ;
- Reta separadora  $x_2 = -\left(\frac{3x_1 - 10}{1,50}\right)$ .

**Lado direito (3ª reta):**

- Valor ótimo da FO: 1,08;
- $w = (w_1, w_2) = (-1,05; 1,23)$ ;
- $\gamma = -2,33$ ;
- Reta separadora  $x_2 = \left(\frac{1,05x_1 - 2,54}{1,23}\right)$ .

A Figura 4.14 representa os conjuntos de pontos, a primeira reta “separadora” e as semirretas “separadoras” encontradas após o procedimento descrito acima.

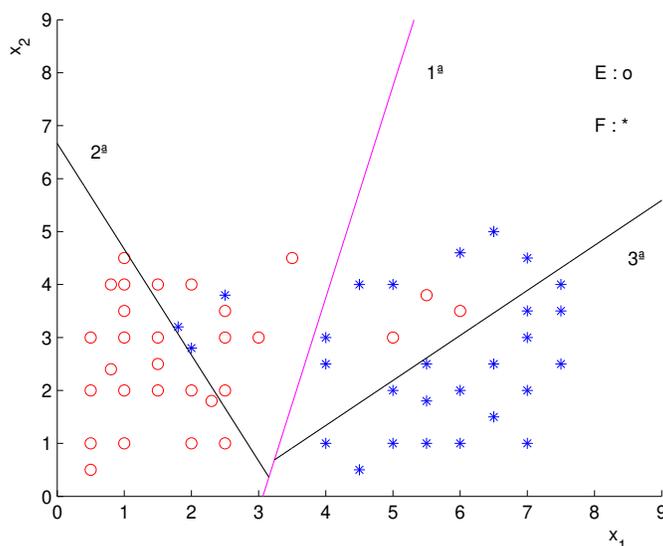


Figura 4.14: Representação dos conjuntos linearmente inseparáveis do Exemplo 4.6 com as 1ª, 2ª e 3ª retas “separadoras”.

Como podemos ver na segunda separação, de ambos os lados da reta os conjuntos continuam linearmente inseparáveis, sendo assim, poderíamos repetir novamente o processo, montando e resolvendo um PPL com os pontos que estão agora de ambos os lados das semirretas que ainda têm pontos inseparáveis. Contudo, com essas três retas (uma reta e duas semirretas) encontradas, o espaço onde os pontos se localizam já foi dividido em 4 regiões: duas delas contêm pontos exclusivos de cada conjunto, e duas regiões continuam com pontos de ambos os conjuntos.

## 4.2 Uma Aplicação da PL na separação de pontos

Como comentamos anteriormente, a idéia de discutir um pouco o uso da PL na separação de pontos se deu após conhecermos o artigo de Mangasarian, Street e Wolberg intitulado "*Breast Cancer Diagnosis and Prognosis via Linear Programming* [16]. Nesse artigo, os autores propõem duas aplicações de PL: uma para o diagnóstico do câncer de mama, utilizando o problema de PL apresentado no capítulo anterior para separação de pontos, e a outra, no prognóstico do câncer de mama, utilizando um outro

problema de PL não discutido em nosso trabalho. Vamos, nessa seção, comentar sobre a aplicação da PL no diagnóstico do câncer de mama, proposta no artigo citado acima.

A idéia central dessa aplicação é utilizar a PL em um conjunto de testes compostos de resultados conhecidos (maligno ou benigno), para aumentar o rigor e a objetividade do diagnóstico do câncer de mama. Esse tipo de câncer é muito comum, sendo que em 1994, quando o artigo foi publicado, ele era a segunda maior causa de morte por câncer entre as mulheres nos Estados Unidos. Como uma vítima de câncer tem suas chances de sobreviver aumentada quando a doença é detectada em seu início, aumentar o rigor e a objetividade no diagnóstico do câncer de mama é de extrema importância.

Segundo o artigo, a maioria dos casos de câncer de mama são detectados pelas pacientes a partir de um nódulo no seio, e a maioria dos nódulos mamários são benignos. Para confirmar a existência de um câncer de mama ou não, a *University of Wisconsin Hospitals*, onde a pesquisa foi realizada, dispunha de três métodos para diagnóstico: mamografia, FNA (*fine needle aspirate*) com interpretação visual e a biópsia cirúrgica.

Ocorre que a capacidade de diagnosticar corretamente quando o câncer está presente varia muito de método para método. Para a mamografia, essa capacidade varia de 68% à 79%, ou seja, está muito longe do ideal. Para o FNA, a capacidade varia de 65% à 98%, apresentando então uma variação muito grande, além de que é inviável apresentar um diagnóstico com 65% de chances de ser maligno. Já a biópsia cirúrgica traz 100% de certeza, no entanto é um procedimento muito caro.

A idéia proposta é melhorar a variação do diagnóstico por FNA. Esse método usa um aspirador de agulha fina, que é utilizado para coletar células de dentro do nódulo. Esse material é colocado sobre uma lâmina de vidro, onde é adicionado um corante para destacar os núcleos das células. Uma câmera de vídeo montada em um microscópio transfere a imagem das células para um computador, onde as mesmas são analisadas. A Figura 4.15 mostra uma dessas imagens coletadas.

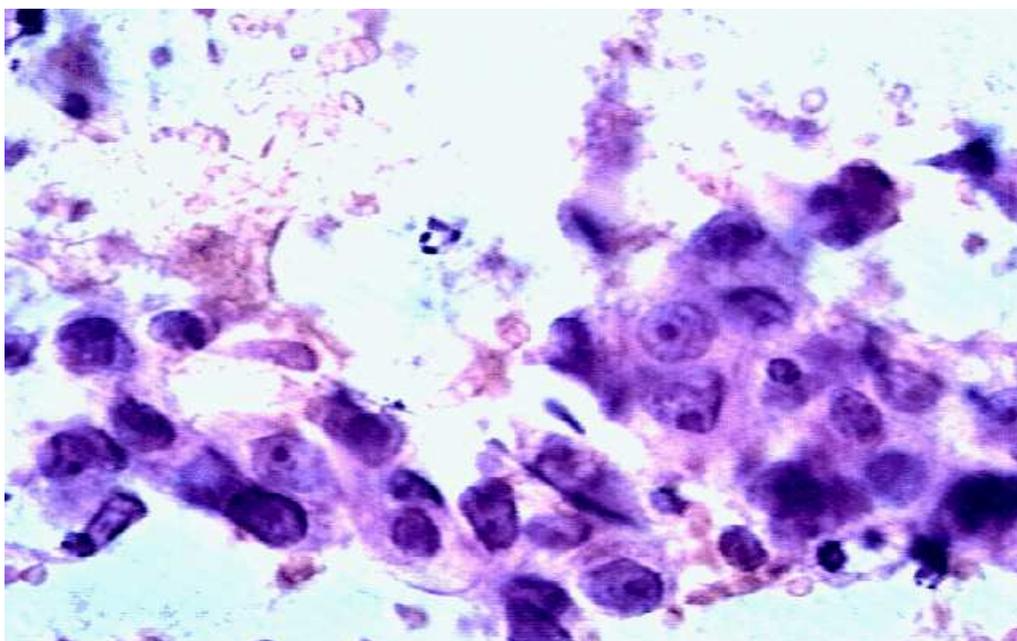


Figura 4.15: Representação da células analisadas no Xcyt; imagem extraída de [http://ftp.cs.wisc.edu/math-prog/cpo-dataset/machine-learn/cancer/cancer\\_images/](http://ftp.cs.wisc.edu/math-prog/cpo-dataset/machine-learn/cancer/cancer_images/)

A análise é feita em um programa de computador chamado Xcyt, que utiliza uma curva conveniente para determinar os limites exatos dos núcleos das células, e a partir daí, consegue extrair várias informações sobre seus núcleos, tais como: área, raio, perímetro, simetria, tamanho das cavidades, textura, variação dos níveis de cinza, maior tamanho, entre outros. Ao todo, até 30 características podem ser obtidas.

Para melhorar a objetividade do FNA, os pesquisadores utilizaram os dados coletados no Xcyt descritos acima para gerar um vetor de 30 componentes ( $n = 30$ ) para cada paciente. Essa coleta foi realizada em 569 pacientes, onde o resultado do diagnóstico, quanto a ser maligno ou benigno, já era conhecido. Os casos malignos haviam sido previamente confirmados por biópsia, enquanto que os casos benignos haviam sido confirmados por outros exames e acompanhamento médico, e alguns pela própria biópsia.

Esses 569 vetores eram compostos por 357 vetores que representavam os casos benignos ( $m = 357$ ), e 212 vetores que representavam os casos malignos ( $k = 212$ ), que

foram utilizados como um conjunto de testes, para classificar e diagnosticar casos futuros. Neste processo, os vetores eram separados utilizando Programação Linear através da solução do problema (3.20).

Feita a primeira separação, conforme foi visto na seção anterior, a separação era aplicada recursivamente para ambos os lados da superfície recém criada, mapeando assim regiões de vetores malignos e benignos. Um novo caso a ser diagnosticado também era transformado em um vetor e comparado com essas regiões. Caso o novo vetor se localizasse em uma região de vetores malignos, seria mais provável que essa paciente tivesse um câncer maligno, e vice versa.

Além dessa comparação entre as regiões, o diagnóstico se baseia também no cálculo da probabilidade de um nódulo ser maligno ou não, probabilidade essa proporcionada pelo próprio *software* Xcyt, mas que foge ao objetivo desse trabalho. Assim, com esses procedimentos, era possível melhorar a precisão do diagnóstico do câncer.

Após a implementação desse estudo no diagnóstico do câncer de mama na *University of Wisconsin Hospitals*, o artigo [16] afirma que os próximos 131 exames realizados tiveram um êxito de 100% de acerto nos diagnósticos, onde foram confirmados 94 casos benignos e 37 malignos.

---

# Considerações finais e perspectivas futuras

---

Neste trabalho desenvolvemos um estudo buscando compreender como a Programação Linear pode ser utilizada na separação de pontos do  $\mathbb{R}^n$  através de hiperplanos. A idéia de desenvolver esse estudo surgiu após conhecermos o artigo [16], onde um modelo de Programação Linear é utilizado para separar conjuntos de células (padrões) de pacientes com nódulos malignos e benignos, com o intuito de aumentar a objetividade do diagnóstico do câncer de mama.

No entanto, para se discutir o assunto, havia a necessidade de se conhecer um pouco de Programação Linear, e como a separação de pontos proposta se dá a partir da solução de um problema de Programação Linear, era essencial conhecer também um método eficiente de solução para tal problema. Nesse sentido, surgiu a necessidade de se estudar o método Simplex.

Como durante minha formação acadêmica não tive a oportunidade de estudar Programação Linear, nem o método Simplex propriamente dito, a elaboração desse trabalho proporcionou-me um aprendizado significativo neste tópico da matemática, que tem uma vasta área de aplicação multidisciplinar. Para estudarmos sobre separação de pontos via Programação Linear não encontramos referenciais em português, o que proporcionou mais um aprendizado, o de ler e traduzir os textos utilizados.

Utilizamos neste trabalho o *software* MATLAB para a implementação computacional de uma rotina para resolver problemas de Programação Linear e representar graficamente os conjuntos de pontos e a reta separadora. Faz-se necessário ressaltar

que, durante minha formação acadêmica não tive contato com este *software*. Assim sendo, para poder utilizá-lo, foi necessário um bom tempo de estudos e testes. Apesar de haver outras opções de *softwares*, o MATLAB se mostrou mais eficiente frente às nossas necessidades e frente ao tempo disponível para realização desse trabalho.

Apesar de não estarmos familiarizados com o MATLAB, como o mesmo encontra-se disponível nos laboratórios de informática do IMECC, foi possível aproveitar esses espaços para realizar parte do trabalho. Também recebemos valiosas contribuições de colegas que utilizam esses ambientes e estão mais familiarizados com a linguagem computacional utilizada pelo *software*, o que contribuiu significativamente para a realização desse trabalho.

Antes de utilizarmos o MATLAB, iniciamos trabalhando com outros *softwares*, tais como o ProLin e o Winplot, que são *softwares* livres. Embora eles não se mostrassem adequados aos nossos objetivos, tivemos um aprendizado que possivelmente poderemos utilizar em aulas futuras ou em outros trabalhos.

Utilizamos também o *software* GeoGebra para construir alguns gráficos ilustrativos utilizados nos Capítulos 1 e 2. Apesar de já possuímos um certo grau de conhecimento sobre o *software*, que foi utilizado em uma das disciplinas do próprio mestrado, tivemos a oportunidade de incorporar novos conhecimentos.

Para a apresentação do trabalho no formato atual, foi necessário aprendermos a trabalhar com *software* Texmaker, que é um editor para Latex livre, o que sem dúvida requereu bastante tempo e dedicação, pois tratava-se de algo totalmente novo.

Pudemos também perceber que, a separação de pontos via Programação Linear, pode ser utilizada em várias aplicações, apesar de que no presente trabalho só comentamos sobre uma. Poderia até mesmo ser utilizada como mais um instrumento na motivação do estudo da Programação Linear.

Vale ainda ressaltar que o modelo linear (3.20) é um modelo, além de completo, simples. A existência atual de *softwares* eficientes que resolvem PPL's num tempo curto nos facilitou bastante, no sentido de reduzir nosso trabalho de programação.

Como perspectivas futuras, pretendemos socializar esse trabalho em forma de seminários ou um mini-curso sobre Programação Linear, principalmente na nossa região do interior de Mato Grosso, onde o tema parece ainda não ter despertado muito interesse acadêmico. E, possivelmente, utilizá-lo futuramente nas aulas, como princípio motivador para o estudo da álgebra, já que a Programação Linear utiliza-se de muitos fundamentos dessa área.

Gostaríamos também de futuramente aplicar a rotina que construímos no MATLAB a dados reais de células de câncer de mama, como proposto no artigo [16]. Tal aplicação não foi realizada por falta de dados e o curto prazo para realização desse trabalho.

Fica ainda como perspectivas futuras analisar matematicamente a influência dos limites superiores e inferiores para as variáveis livres de sinal, e se possível estabelecer um padrão para a melhor escolha desses dados, contribuindo assim um pouco mais para a discussão sobre o uso da Programação Linear na separação de pontos por hiperplanos.

# APÊNDICE 1 - EQUIVALENTE AO TEOREMA (3.1) USANDO DUALIDADE

---



---

Recorrendo ao princípios de dualidade em Programação Linear, Mangasarian [14] estabelecem um teorema dual ao Teorema (3.1), conforme segue à frente.

**Teorema A.1** ([14] Separabilidade Linear Dual). *Uma condição necessária e suficiente para separabilidade linear dos conjuntos de padrões  $E$  e  $F$  é que:*

$$\varphi(E, F) > 0 \tag{A.1}$$

onde  $\varphi(E, F)$  é a solução do problema de Programação Linear:

$$\begin{array}{l} \min_{u,v,p} (e'p) \\ \text{s.a.} \left\{ \begin{array}{l} -E'u + F'v + p \geq 0 \\ E'u - F'v + p \geq 0 \\ e'u = 1 \\ e'v = 1 \\ u \geq, v \geq 0 \end{array} \right. \end{array} \tag{A.2}$$

com  $u \in \mathbb{R}^m$ ,  $v \in \mathbb{R}^k$  e  $p \in \mathbb{R}^n$ .

De modo análogo feito com o teorema (3.1) podemos enunciar a partir do teorema (A.1) um corolário para estabelecer um critério de inseparabilidade linear, como segue abaixo.

**Corolário A.1** ([14] Inseparabilidade Linear Dual). *Uma condição necessária e suficiente para inseparabilidade linear dos conjuntos de padrões  $E$  e  $F$  é que  $\varphi(E, F) = 0$ .*

**Demonstração:** A demonstração desse corolário segue do teorema (A.1) observando que, das duas primeiras restrições de (A.2), temos  $p \geq 0$ , fato esse que implica que  $e'p \geq 0$  e com isso  $\varphi(E, F) \geq 0$ . ■

---

## Referências Bibliográficas

---

- [1] Andrade, L. N., *Breve Introdução ao Latex 2<sub>ε</sub>*- disponível em <ftp://mat.ufpb.br/pub/textos/tex/brece21.zip>, 2000.
- [2] Bennett, K. P., *Decision Tree Construction Via Linear Programming*, In Proceedings of the 4th Midwest Artificial Intelligence and Cognitive Science Society Conference, 97-101, 1992.
- [3] Bennett, K. P. and Mangasarian, O. L., *Robust Linear Programming Discrimination of Two Linear inseparable sets*, Optimization Methods and Software, 23-34, 1992.
- [4] Boldrini, J. L., Costa, S. R., Figueiredo, V. L. e Wetzler, H. G., *Álgebra Linear*, Editora Harbra Ltda, 3. ed., 1980.
- [5] Bregalda, P. F., Oliveria, A. A. F. e Bornstein, C. T., *Introdução à Programação Linear*, Editora Campus, 2. ed., 1983.
- [6] Coral, A. M., Santos, M. R. e Bastos, T. D. A., *Curso de Matlab*, Universidade Federal de Santa Catarina - Departamento de Engenharia de Produção de Sistemas, disponível em <http://www.profwillian.com/numerico/ApostiladeMATLAB-31pags.pdf>, 1991.
- [7] Finney, R. L. e Weir, M. D. e Giordano, F. R., *Cálculo de George B. Thomas Jr. volume 2*, Editora Perarson Education do Brasil, 2. ed., 2002.
- [8] Howard, A. e Rorres, C., *Álgebra Linear com aplicações*, Bookman, 2001.
- [9] Kaluzny, B. e Toussaint, G. T., *Pattern Recognition via Linear Programming*. Disponível em <http://cgm.cs.mcgill.ca/beezer/cs644/>, acesso em novembro/2009.

- [10] Lima, E. L., *Curso de análise*, Editora IMPA, volume 2, 10 ed., 2008.
- [11] Lins, M. P. E. and Calôba, G. M., *Programação Linear: com aplicações em teoria dos jogos e avaliação de desempenho (Data Envelopment Analysis)*, Editora Interciência, 2006.
- [12] Maculan, N. F. e Fampa, M. H. C., *Otimização Linear*, Editora Universidade de Brasília, 2006.
- [13] Maculan, N. F. e Pereira, M. V. F., *Programação Linear*, Editora Atlas, 1986.
- [14] Mangasarian, O. L., *Linear and Nonlinear Separation of Patterns by Linear Programming*, Operations Research, 13, 444-452, 1965.
- [15] Mangasarian, O. L., Setiono, R. and Wolberg, W. H., *Pattern Recognition via Linear Programming: theory and Application to Medical Diagnosis*, In proceedings of the Workshop Large-Scale Numerical Optimization, 22-31, 1990. SIAM.
- [16] Mangasarian, O. L., Street, W. N. and Wolberg, W. H., *Breast Cancer Diagnosis and Prognosis via Linear Programming*, Mathematical Programming Technical Report, 94-10, 1994.
- [17] Matos, J. e Matos, R., *ProLin - Programação Linear Mista*. Disponível em <http://prolin.no.sapo.pt/>, acesso em junho/2009.
- [18] McCallum, W. G., Hughes-Hallet, D. e Gleason, A. M., *Cálculo de Várias Variáveis*, Editora Edgard Blucher Ltda, 1<sup>a</sup> ed., 1997.
- [19] Passos, A. N., *Estudos em programação linear*, Dissertação de Mestrado Profissional em Matemática - Universidade Estadual de Campinas, 2009.
- [20] Passos, E. J. P. F., *Programação Linear Como Instrumento da Pesquisa Operacional*, Editora Atlas, 2008.
- [21] Prado, D. S., *Programação Linear*, INDG Tecnologia e serviços Ltda, Série Pesquisa Operacional, volume 1, 5. ed., 2007.

- [22] Puccini, A. L., *Introdução à programação linear*, Livros técnicos e científicos editora S. A., 1978.
- [23] Shenk, A., *Cálculo e geometria analítica vol. 2*, Editora Campus, 2. ed., 1987.
- [24] Smith, F. W., *Pattern Classifier Design by Linear Programming*, IEEE Transaction on Computers C-17, 4, 367-372, 1968.
- [25] Ulhoa, C. F. e Lourenço, M. L., *Um Curso de Álgebra Linear*, Editora da Universidade de São Paulo, 2. ed., 2005.
- [26] *Winplot*. Disponível em <http://www.gregosetrianos.mat.br/sofwimplot>, acesso em junho/2009.
- [27] Yoshida, L. K., *Programação Linear: métodos quantitativos*, Editora Atual, 1987.