Gaspar J. Machado
Gueorgui Smirnov
Irene Brito
Ricardo Costa

Álgebra Linear do ponto de vista geométrico

Coleção Educação | Ciências, Engenharia e Tecnologia

UMinho Editora

Educação
Ciências, Engenharia
e Tecnologia

UMinho Editora

AUTORES
Gaspar J. Machado
Guerorgui Smirnov
Irene Brito
Ricardo Costa

COORDENAÇÃO EDITORIAL
Manuela Martins

DESIGN
UMinho Editora

PAGINAÇÃO EM LATEX
Os Autores

EDIÇÃO UMinho Editora

LOCAL DE EDIÇÃO Braga 2025

ISBN 978-989-9074-77-4

ISBN DIGITAL 978-989-9074-78-1

DOI https://doi.org/10.21814/uminho.ed.174

Álgebra Linear do ponto de vista geométrico

Gaspar J. Machado
Gueorgui Smirnov
Irene Brito
Ricardo Costa

Coleção Educação I Ciências, Engenharia e Tecnologia

Escrevemos este livro para estudantes de cursos de Matemática Aplicada, Física e Engenharia, que estudam Álgebra Linear no primeiro ano da Universidade, o que explica a escolha da matéria e o nível de generalidade da apresentação. Incluímos no livro apenas os conceitos e resultados importantes para as aplicações da Matemática e as demonstrações foram feitas sem tentar atingir a máxima generalidade. Do nosso ponto de vista, o nome "Álgebra Linear" não é o mais adequado para chamar a esta disciplina. O nome mais adequado seria "Geometria de Espaços de Dimensão Finita". No entanto, decidimos dar ao livro o título "Álgebra Linear do Ponto de Vista Geométrico" para não confundir o leitor sobre o seu conteúdo. Salientamos, porém, que é daquele ponto de vista que expomos a matéria neste livro.

O livro contém doze capítulos. No primeiro capítulo introduzimos os conceitos principais do cálculo vetorial no plano. Estes conceitos são generalizados para o espaço de dimensão três no segundo capítulo. O espaço linear de dimensão n é o objeto principal de estudo no terceiro capítulo. No quarto capítulo estudamos espaços normados e Euclidianos. No capítulo cinco consideramos as aplicações lineares e estabelecemos a correspondência entre aplicações lineares e matrizes. Introduzimos a noção de determinante como um objeto geométrico no sexto capítulo. Os vetores e valores próprios são o tema do sétimo capítulo. No oitavo capítulo estudamos os espaços afins que generalizam os espaços de vetores considerados nos primeiros dois capítulos do livro. No nono capítulo consideramos formas quadráticas. Dedicamos o capítulo dez às cónicas e o capítulo onze às quádricas. O último capítulo é destinado ao leitor mais preparado. Nele demonstramos o Teorema Fundamental da Álgebra e o teorema sobre a forma canónica de Jordan.

As partes do livro dedicadas à ligação da Álgebra Linear com outros ramos da Matemática, tais como equações diferenciais ordinárias, cadeias de Márkov, análise funcional e equações em derivadas parciais, ou que têm matéria mais difícil, estão marcadas com um asterisco. O leitor interessado pode voltar a estas partes numa segunda leitura, pois este livro foi pensado como "livro de cabeceira" de um estudante durante todo o seu percurso académico.

O livro não tem exercícios. É aconselhável o uso do livro de exercícios *Seymour Lipschutz, Marc L. Lipson, Linear Algebra, Schaum's Outline Series, McGraw-Hill, 2009* para aprender melhor a matéria.

Escrevemos este texto como uma introdução à Álgebra Linear. Ao leitor interessado em conhecer os resultados mais profundos desta parte da Matemática podemos recomendar os seguintes livros: *Felix R. Gantmacher, The Theory of Matrices, AMS Chelsea Publishing, 1959; Georgi E. Shilov, Linear Algebra, Dover Publications, 1977; Paul Hal-*

mos, Finite-Dimensional vetor Spaces, Springer, 1987; Luís Magalhães, Álgebra Linear como Introdução a Matemática Aplicada, Texto Editora, 1996; Igor R. Shafarevich and Alexey O. Remizov, Linear Algebra and Geometry, Springer, 2013; Sheldon Axler, Linear Algebra Done Right, Springer, 2015.

Os autores agradecem os comentários e as sugestões que Maria Margarida Ferreira e Sofia Lopes fizeram à versão preliminar deste livro.

Universidade do Minho, julho de 2025

Gaspar J. Machado, Gueorgui Smirnov, Irene Brito e Ricardo Costa.

1. O plano

1.1. Introdução

Neste capítulo vamos introduzir um conjunto de definições e teoremas, algumas delas já conhecidas do curso de Matemática que o leitor teve na Escola, associados a objetos planos. Para facilitar a leitura, muitas vezes, nas demonstrações dos teoremas, vamos recorrer a figuras que esclarecem as ideias principais do raciocínio.

1.2. Vetores no plano

O conceito principal com que vamos trabalhar neste capítulo é o de vetor do plano, que designaremos apenas por vetor. Um vetor é caracterizado por um comprimento, uma direção e um sentido. Assim, um vetor pode ser representado por vários segmentos orientados com ponto inicial e ponto final diferentes desde que os seus comprimentos, direções e sentidos sejam iguais (este procedimento é semelhante à representação dos números racionais, em que, por exemplo, as representações $\frac{1}{2}$, $\frac{2}{4}$ e $\frac{100}{200}$ referem-se todas ao mesmo número racional). Assim, dado um comprimento, uma direção e um sentido, há uma infinidade de representações do mesmo vetor. Este conjunto de representações para um vetor é, o que em Matemática se diz, uma classe de equivalência. A notação que aqui vamos usar para representar vetores é a de identificar o vetor através de uma letra minúscula com uma seta por cima ou então identificando o ponto inicial e o ponto final de um representante do vetor e colocando também por cima uma seta. Temos, então, que todos os segmentos orientados representados na Fig. 1.1 referem-se a um mesmo vetor que pode ser designado por $\vec{a}$ ou $\overrightarrow{P_1P_1'}$ ou $\overrightarrow{P_2P_2'}$ ou $\overrightarrow{P_3P_3'}$.

Representaremos por $\mathbb{V}^2$ o conjunto de todos os vetores.

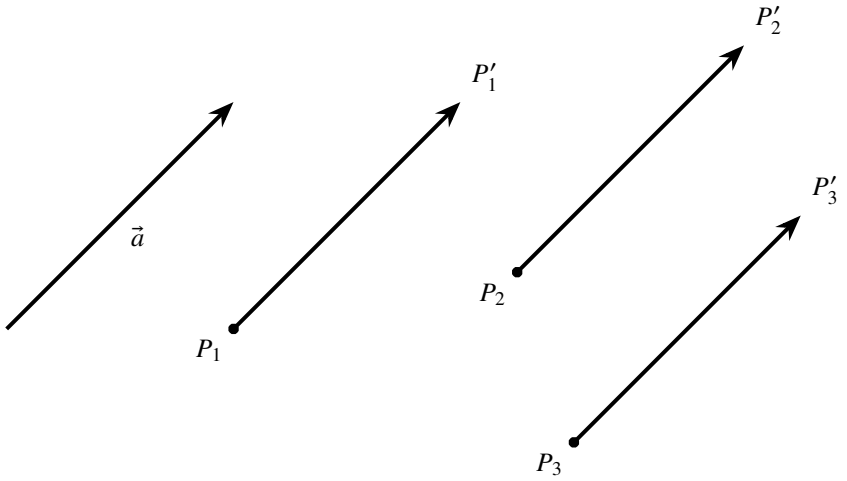


Figura 1.1
Definição de vetor através de segmentos orientados.

Se o ponto inicial e o ponto final de um segmento orientado do plano coincidirem, chamamos a esse caso particular vetor nulo, que representamos por $\vec{0}_2$. O vetor nulo

não tem nem a direção nem o sentido definidos e pode ser representado por qualquer ponto do plano.

Dizemos que dois vetores são paralelos (ou colineares) se têm a mesma direção e iremos considerar que o vetor nulo é paralelo a todos os vetores. Assim, os vetores representados na Fig. 1.2 são todos paralelos dois a dois. Para denotar que os vetores $\vec{a}$ e $\vec{b}$ são paralelos, escrevemos $\vec{a}||\vec{b}$.

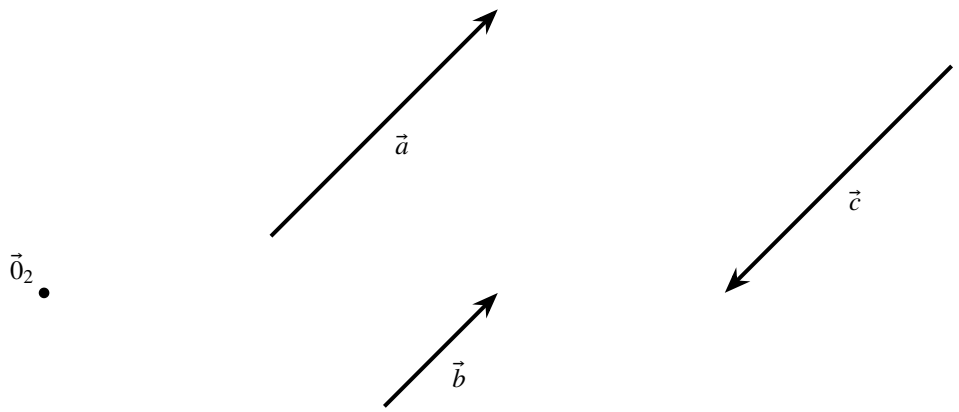


Figura 1.2
Vetores paralelos.

Dado um vetor não nulo $\vec{a}$, podemos construir um novo vetor com o mesmo comprimento e a mesma direção, mas com o sentido contrário, que chamamos vetor oposto (ou simétrico) de $\vec{a}$ e representamos por $-\vec{a}$ (veja a Fig. 1.3). O vetor oposto do vetor nulo é o próprio vetor nulo.

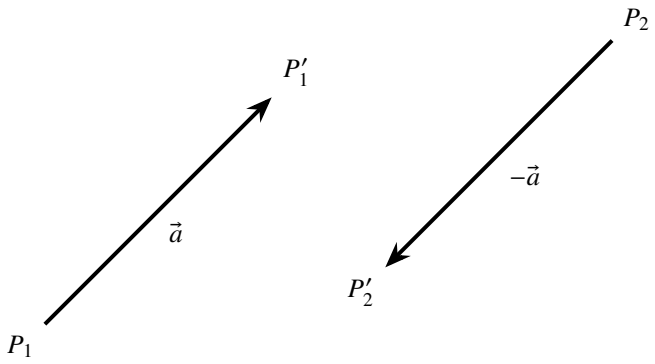
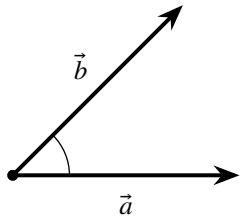


Figura 1.3
Vetor oposto.

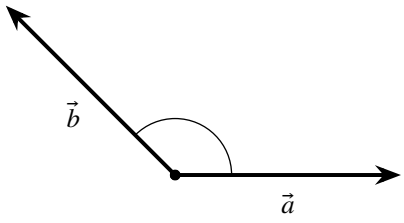
Representamos o comprimento do vetor $\vec{a}$ por $|\vec{a}|$, a que também chamamos *norma* de $\vec{a}$, e dizemos que é unitário se $|\vec{a}| = 1$. No caso do vetor nulo temos que $|\vec{0}_2| = 0$. A *distância* entre dois pontos P_1 e P_2 é a norma $|\overrightarrow{P_1P_2}|$. A distância entre o ponto P e o conjunto de pontos $\mathcal{P}$ é o ínfimo das normas $|\overrightarrow{PP'}|$, onde $P' \in \mathcal{P}$.

Representamos por $\angle(\vec{a}, \vec{b})$ o ângulo entre dois vetores não nulos $\vec{a}$ e $\vec{b}$, que é um valor do intervalo $[0, \pi]$ (ou $[0°, 180°]$), e temos que $\angle(\vec{a}, \vec{b}) = \angle(\vec{b}, \vec{a})$ (o ângulo entre dois vetores em que pelo menos um deles é o vetor nulo não está definido). No caso dos vetores terem a mesma direção e o mesmo sentido o ângulo é 0 e no caso de terem a mesma direção mas sentidos opostos o ângulo é π. Dizemos que dois vetores não nulos são ortogonais (ou perpendiculares) se o ângulo que formam é $\frac{\pi}{2}$. Se pelo

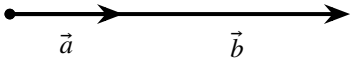
menos um deles é o vetor nulo dizemos também que são ortogonais. Usamos a notação $\vec{a} \perp \vec{b}$ para dizer que os vetores $\vec{a}$ e $\vec{b}$ são ortogonais. Na Fig. 1.4 estão representadas todas estas situações.



(a) $\angle(\vec{a},\vec{b}) < \frac{\pi}{2}$.

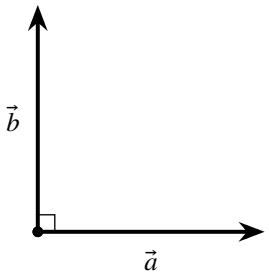


(b) $\angle(\vec{a},\vec{b}) > \frac{\pi}{2}$.



(c) $\angle(\vec{a},\vec{b}) = 0$.

(d) $\angle(\vec{a},\vec{b}) = \pi$.



(e) $\angle(\vec{a},\vec{b}) = \frac{\pi}{2}$.

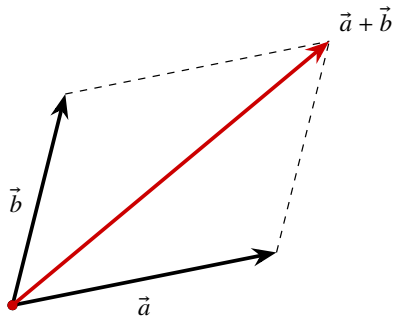
Figura 1.4
Ângulo entre dois vetores.

Vamos agora definir duas operações que envolvem vetores: a soma de vetores e a multiplicação de um vetor por um número real.

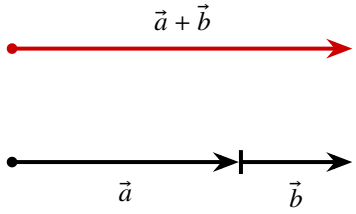
A soma dos vetores $\vec{a}$ e $\vec{b}$, que continua a ser um vetor do plano e que representamos por $\vec{a}+\vec{b}$, é igual ao outro vetor no caso de um deles ser o vetor nulo (ou seja, $\vec{a}+\vec{0}_2 = \vec{a}$ e $\vec{0}_2+\vec{0}_2 = \vec{0}_2$). No caso de ambos os vetores serem não nulos, temos:

- Se os vetores têm direções distintas, então a sua soma é obtida através da seguinte regra, conhecida por regra do paralelogramo: aplicamos os vetores $\vec{a}$ e $\vec{b}$ no mesmo ponto sendo a sua soma o vetor definido pela diagonal do paralelogramo que tem os vetores $\vec{a}$ e $\vec{b}$ como lados adjacentes a partir do ponto onde se aplicaram os vetores $\vec{a}$ e $\vec{b}$ (veja a Fig. 1.5(a)).

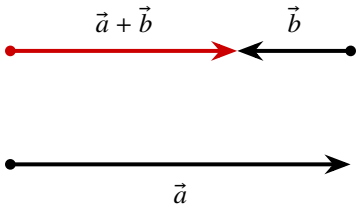
- Se os vetores têm a mesma direção e o mesmo sentido, então a sua soma é o vetor com a mesma direção e o mesmo sentido dos vetores $\vec{a}$ e $\vec{b}$ e com o comprimento igual à soma dos comprimentos dos vetores $\vec{a}$ e $\vec{b}$ (veja a Fig. 1.5(b)).
- Se os vetores têm a mesma direção, sentidos opostos e comprimentos diferentes, então a sua soma é o vetor com a direção dos vetores $\vec{a}$ e $\vec{b}$, com o sentido do vetor com maior comprimento e com o comprimento igual à diferença entre o comprimento do vetor com maior comprimento e o comprimento do vetor com menor comprimento (veja a Fig. 1.5(c)).
- Se os vetores têm a mesma direção, sentidos opostos e o mesmo comprimento, ou seja, a soma de vetores opostos, então a sua soma é o vetor nulo $\vec{0}_2$ (veja a Fig. 1.5(d)).



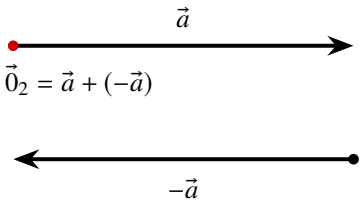
(a) Vetores com direções distintas.



(b) Vetores com a mesma direção e o mesmo sentido.



(c) Vetores com a mesma direção, sentidos opostos e comprimentos diferentes.



(d) Vetores com a mesma direção, sentidos opostos e comprimentos iguais.

Figura 1.5
Soma de dois vetores não nulos.

A multiplicação de um vetor $\vec{a}$ por um número α, que continua a ser um vetor do plano e que representamos por $\alpha\vec{a}$, é igual ao vetor nulo no caso de $\alpha = 0$ ou $\vec{a} = \vec{0}_2$ (ou seja, $0\vec{a} = \vec{0}_2$ e $\alpha\vec{0}_2 = \vec{0}_2$). No caso de $\vec{a} \neq \vec{0}_2$ e de $\alpha \neq 0$, a multiplicação de α por $\vec{a}$ dá um vetor com a mesma direção do vetor $\vec{a}$, com o comprimento $|\alpha||\vec{a}|$ e com o sentido do vetor $\vec{a}$ se $\alpha > 0$ ou com o sentido oposto se $\alpha < 0$ (veja a Fig. 1.6).

A cada vetor não nulo $\vec{a}$ é possível associar um vetor unitário com a mesma direção e o mesmo sentido de $\vec{a}$. Este novo vetor, que se chama versor de $\vec{a}$, é dado por $\frac{\vec{a}}{|\vec{a}|}$.

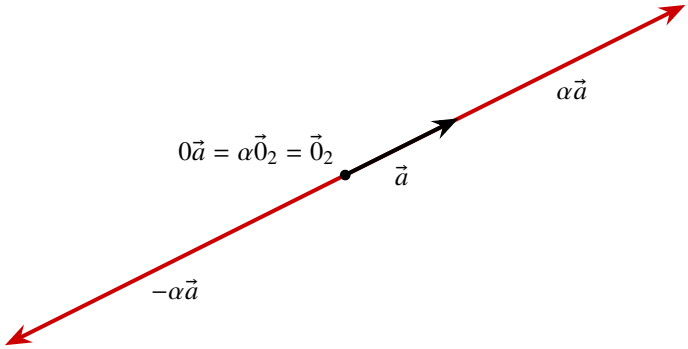


Figura 1.6
Multiplicação de um vetor $\vec{a}$ por um número α.

No teorema seguinte estabelecemos as propriedades da soma de vetores e da multiplicação de um vetor por um número.

Teorema 1.1

1. A soma de vetores é comutativa, isto é, quaisquer que sejam os vetores $\vec{a}$ e $\vec{b}$, temos que $\vec{a} + \vec{b} = \vec{b} + \vec{a}$.
2. A soma de vetores é associativa, isto é, quaisquer que sejam os vetores $\vec{a}$, $\vec{b}$ e $\vec{c}$, temos que $(\vec{a} + \vec{b}) + \vec{c} = \vec{a} + (\vec{b} + \vec{c})$, pelo que podemos escrever $\vec{a} + \vec{b} + \vec{c}$.
3. O vetor nulo é o único vetor tal que o resultado da sua soma a qualquer vetor dá o mesmo vetor a que se somou o vetor nulo, isto é, qualquer que seja o vetor $\vec{a}$, temos que $\vec{a} + \vec{0}_2 = \vec{a}$ (o vetor nulo é o elemento neutro da soma de vetores).
4. Dado um vetor, existe um único vetor tal que o resultado da sua soma é o vetor nulo. Temos, ainda, que esse único vetor é o vetor oposto do vetor dado, isto é, qualquer que seja o vetor $\vec{a}$, temos que $\vec{a} + (-\vec{a}) = \vec{0}_2$.
5. A soma de dois vetores é distributiva relativamente à multiplicação por um número, ou seja, a multiplicação da soma de dois vetores por um número é igual a primeiro multiplicar cada um dos vetores por aquele número e depois somar os dois vetores daí resultantes, isto é, qualquer que seja o número α e quaisquer que sejam os vetores $\vec{a}$ e $\vec{b}$, temos que $\alpha(\vec{a} + \vec{b}) = \alpha\vec{a} + \alpha\vec{b}$.
6. A soma de dois números é distributiva relativamente à multiplicação por um vetor, ou seja, a multiplicação da soma de dois números por um vetor é igual a primeiro multiplicar esse vetor por aqueles números e depois somar os dois vetores daí resultantes, isto é, quaisquer que sejam os números α e β e qualquer que seja o vetor $\vec{a}$, temos que $(\alpha + \beta)\vec{a} = \alpha\vec{a} + \beta\vec{a}$.
7. A multiplicação de números por um vetor é associativa, ou seja, a multiplicação do produto de dois números por um vetor é igual a primeiro

continua na página seguinte

continuação da página anterior

multiplicar esse vetor por um dos números e depois multiplicar o vetor daí resultante pelo outro número, isto é, quaisquer que sejam os números α e β e qualquer que seja o vetor $\vec{a}$, temos que $(\alpha\beta)\vec{a} = \alpha(\beta\vec{a})$.

8. A multiplicação de um vetor pelo número 1 é igual ao vetor, isto é, qualquer que seja o vetor $\vec{a}$, temos que $1\vec{a} = \vec{a}$ (o número 1 é o elemento neutro da multiplicação de um número por um vetor).

Demonstração. As demonstrações destas propriedades são tão simples que tudo se torna óbvio nas seguintes figuras.

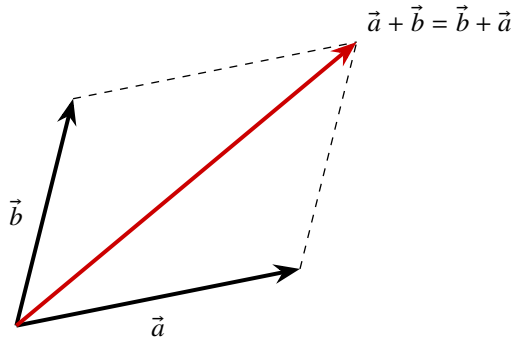


Figura 1.7
Comutatividade da soma de vetores.

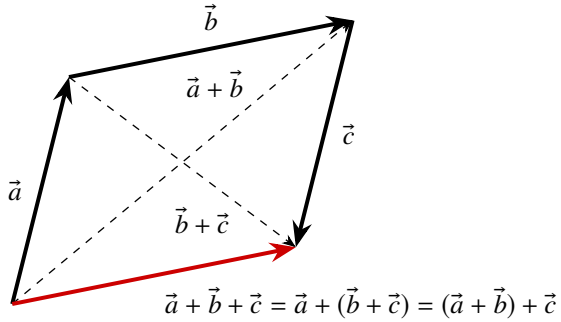


Figura 1.8
Associatividade da soma de vetores.

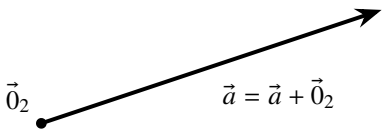


Figura 1.9
Unicidade do elemento neutro da soma de vetores — vetor nulo.

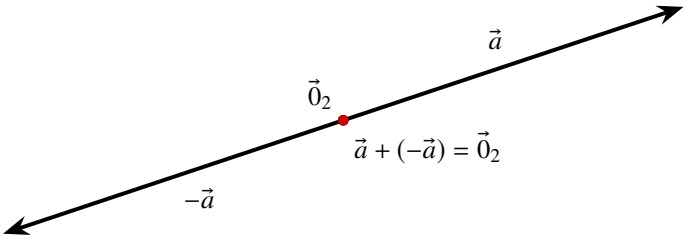


Figura 1.10
Unicidade de vetor oposto.

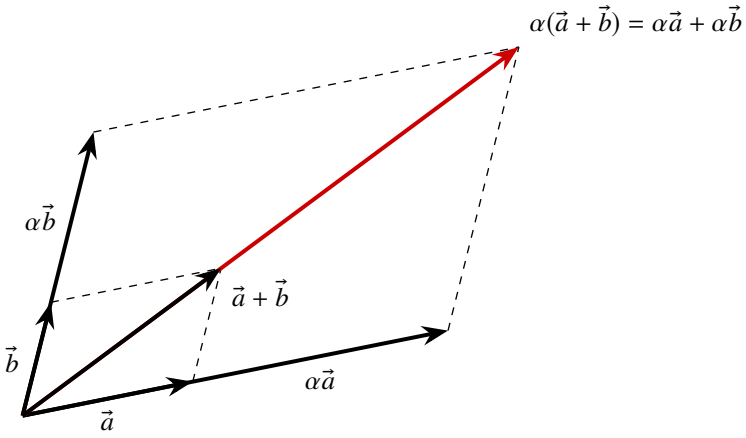


Figura 1.11
Distributividade da multiplicação de um número relativamente à soma de vetores.

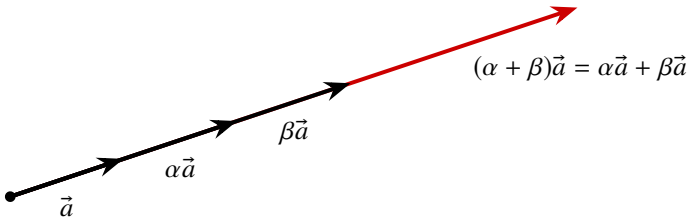


Figura 1.12
Distributividade da multiplicação de um vetor relativamente à soma de números.

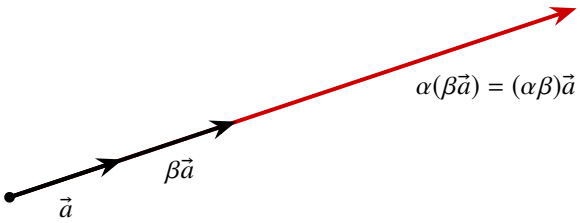


Figura 1.13
Associatividade da multiplicação de números por um vetor.

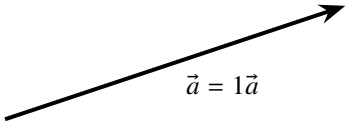


Figura 1.14
Elemento neutro da multiplicação de um número por um vetor.

■

1.3. Combinação linear, dependência linear, base, coordenadas e dimensão

Vamos agora caracterizar numericamente os vetores por forma a desenvolver um formalismo de cálculo das operações com vetores de uma maneira alternativa às definições gráficas dadas na secção anterior. Esta nova abordagem, conhecida por método das coordenadas, é fundamental para a resolução dos problemas que pretendemos tratar.

Sendo $\mathbb{R}^2$ o conjunto dos pares ordenados cujas componentes são números reais, ou seja, $\mathbb{R}^2 = \{(\alpha, \beta) : \alpha, \beta \in \mathbb{R}\}$, podemos estabelecer uma correspondência biunívoca entre o "espaço abstrato" $\mathbb{R}^2$ e o "espaço físico" $\mathbb{V}^2$, em que os vetores podem ter um significado físico representando, por exemplo, uma força ou uma velocidade. Assim, dados dois vetores $\vec{a}$ e $\vec{b}$ não paralelos (o que implica que nem $\vec{a}$ nem $\vec{b}$ são o vetor nulo), existe um único par ordenado $(\alpha, \beta) \in \mathbb{R}^2$ tal que qualquer vetor $\vec{v}$ se pode escrever como

$$\vec{v} = \alpha\vec{a} + \beta\vec{b}. \tag{1.1}$$

Dizemos neste caso que o vetor $\vec{v}$ é uma *combinação linear* dos vetores $\vec{a}$ e $\vec{b}$ e vamos representar por $\text{Lin}\{\vec{a}, \vec{b}\}$ o conjunto de todas as combinações lineares dos vetores $\vec{a}$ e $\vec{b}$.

Fixemos um ponto do plano, que denotamos por O e que chamamos *origem*. A partir de agora, vamos representar todos os vetores fazendo com que o seu ponto inicial coincida com o ponto O. Em particular, aplicando o vetor nulo ao ponto O obtemos o próprio ponto O. Através deste processo podemos estabelecer uma correspondência biunívoca entre os pontos e os vetores do plano da seguinte maneira: ao ponto A fazemos corresponder o vetor $\overrightarrow{OA}$ que vamos considerar como o representante do vetor $\vec{a}$ que é o conjunto de todos os segmentos orientados com a direção, o sentido e o comprimento do vetor $\overrightarrow{OA}$. Assim, o plano pode ser encarado como um conjunto de pontos ou um conjunto de vetores.

Consideremos pontos A e B diferentes de O. Os vetores $\vec{a}$ e $\vec{b}$ são representados pelos vetores $\overrightarrow{OA}$ e $\overrightarrow{OB}$, respetivamente. Em vez dos representantes $\overrightarrow{OA}$ e $\overrightarrow{OB}$ vamos usar os símbolos $\vec{a}$ e $\vec{b}$ para simplificar a escrita.

Sejam $\vec{a}$ e $\vec{b}$ vetores paralelos. Então, pelo menos um deles é múltiplo do outro, ou seja, existem números α e β não ambos nulos tais que $\alpha\vec{a} + \beta\vec{b} = \vec{0}_2$, ou seja, o vetor nulo é uma combinação linear dos vetores $\vec{a}$ e $\vec{b}$. Dizemos, neste caso, que os vetores $\vec{a}$ e $\vec{b}$ são *linearmente dependentes*. No caso dos vetores $\vec{a}$ e $\vec{b}$ não serem paralelos, então o vetor nulo $\vec{0}_2$ só é uma combinação linear dos vetores $\vec{a}$ e $\vec{b}$ com $\alpha = 0$ e $\beta = 0$. Dizemos neste caso que os vetores $\vec{a}$ e $\vec{b}$ são *linearmente independentes*.

Se os vetores $\vec{a}$ e $\vec{b}$ são linearmente independentes, então qualquer vetor $\vec{v}$ é uma combinação linear desses dois vetores dada pela fórmula (1.1). Neste caso dizemos que $\mathcal{B} = \{\vec{a}, \vec{b}\}$ é uma *base* de $\mathbb{V}^2$ (veja a Fig. 1.15) e que α e β são as *coordenadas* do vetor $\vec{v}$ nesta base e usamos a notação $\vec{v}_{\mathcal{B}} = (\alpha, \beta)$ ou simplesmente $\vec{v} = (\alpha, \beta)$ no

caso de não haver ambiguidade da definição da base. Como qualquer base em $\mathbb{V}^2$ tem dois elementos, dizemos que a dimensão de $\mathbb{V}^2$ é dois e escrevemos $\dim(\mathbb{V}^2) = 2$.

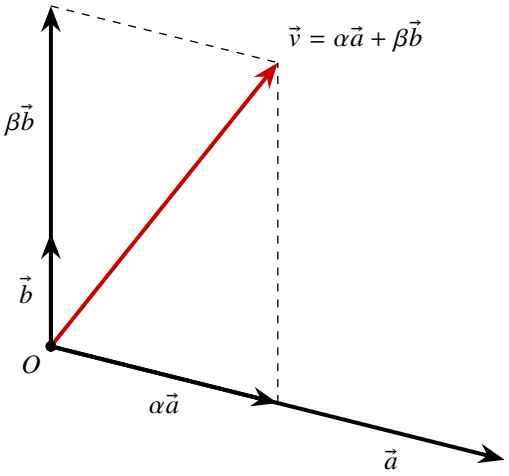


Figura 1.15
Coordenadas de um vetor.

A representação de vetores numa base é única. De facto, se

$$\alpha\vec{a} + \beta\vec{b} = \alpha'\vec{a} + \beta'\vec{b},$$

então

$$(\alpha - \alpha')\vec{a} + (\beta - \beta')\vec{b} = \vec{0}_2.$$

Como os vetores da base são linearmente independentes, temos $\alpha - \alpha' = 0$ e $\beta - \beta' = 0$.

Vamos voltar agora às operações soma de vetores e multiplicação de um vetor por um número. Sejam $\mathcal{B} = \{\vec{a}, \vec{b}\}$ uma base de $\mathbb{V}^2$ e $\vec{x}$ e $\vec{y}$ vetores dados nesta base por $\vec{x}_{\mathcal{B}} = (x^1, x^2)$ e $\vec{y}_{\mathcal{B}} = (y^1, y^2)$. (Neste livro vamos representar as coordenadas de um vetor, por exemplo, $\vec{z}$, por índices superiores: z^1, z^2, ...) Então,

$$(\vec{x} + \vec{y})_{\mathcal{B}} = (x^1 + y^1, x^2 + y^2)$$

e

$$(\alpha\vec{x})_{\mathcal{B}} = (\alpha x^1, \alpha x^2).$$

Representamos nas Figs. 1.16 e 1.17 a soma de vetores e multiplicação de um vetor por um número, respetivamente, em termos de coordenadas na base $\mathcal{B}$.

1.4. Sistemas de coordenadas

Escolhemos uma origem O e uma base $\mathcal{B}$ de $\mathbb{V}^2$. Ao par $(O, \mathcal{B})$ chamamos um *sistema de coordenadas* de $\mathbb{V}^2$.

Consideremos, em $\mathbb{V}^2$, dois sistemas de coordenadas $(O, \mathcal{B})$ e $(O', \mathcal{B}')$, onde $\mathcal{B} = \{\vec{v}_1, \vec{v}_2\}$ e $\mathcal{B}' = \{\vec{v}'_1, \vec{v}'_2\}$, com $\vec{v}_0 = \overrightarrow{O'O}$ (veja a Fig. 1.18).

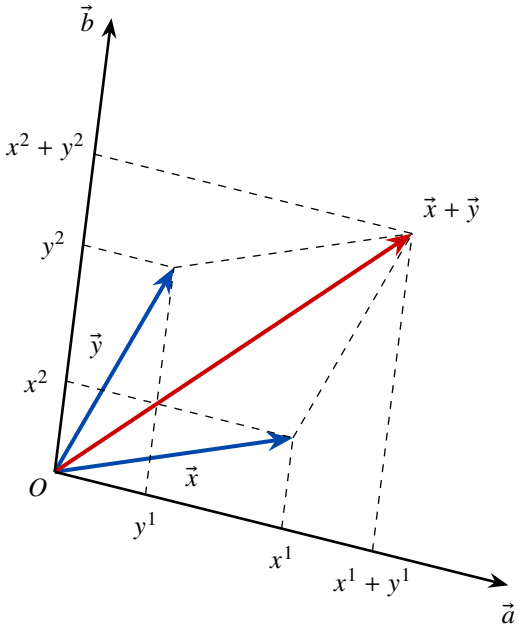


Figura 1.16
Soma de dois vetores na base $\mathcal{B} = \{\vec{a}, \vec{b}\}$.

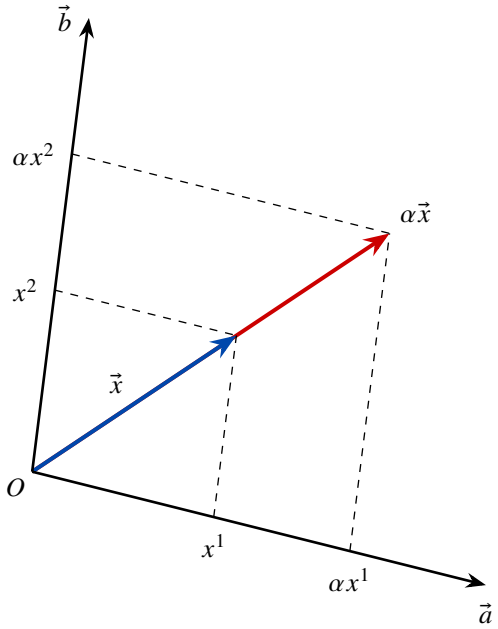


Figura 1.17
Multiplicação de um vetor por um número na base $\mathcal{B} = \{\vec{a}, \vec{b}\}$.

Consideremos um vetor $\vec{v}$ que no sistema de coordenadas $(O, \mathcal{B})$ tem a forma

$$\vec{v} = x^1\vec{v}_1 + x^2\vec{v}_2. \tag{1.2}$$

Índices de vetores são representados por índices inferiores. Vamos ver como escrever o mesmo vetor no sistema de coordenadas $(O', \mathcal{B}')$. (Esta passagem de um sistema de coordenadas para outro é muito importante para a resolução de muitos problemas.)

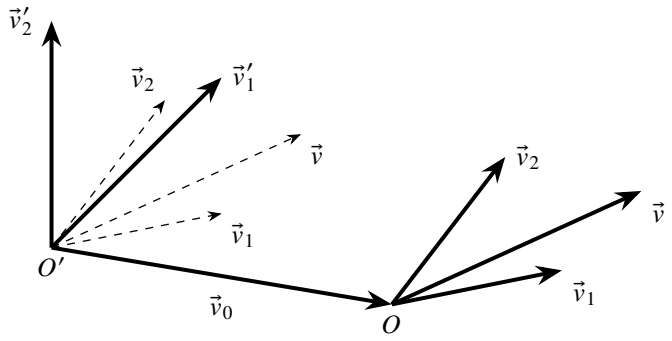


Figura 1.18
Mudança do sistema de coordenadas.

Suponhamos que na base $\mathcal{B}'$ temos as representações

$$\vec{v}_0 = a_1^0\vec{v}_1' + a_2^0\vec{v}_2',$$
$$\vec{v}_1 = a_1^1\vec{v}_1' + a_2^1\vec{v}_2',$$
$$\vec{v}_2 = a_1^2\vec{v}_1' + a_2^2\vec{v}_2'.$$

(Na Fig. 1.18, os vetores $\vec{v}_1$, $\vec{v}_2$ e $\vec{v}$ estão representados por linhas tracejadas quando o seu ponto inicial é O'.) No sistema de coordenadas $(O', \mathcal{B}')$, em vez de (1.2), temos

$$\begin{aligned}\vec{v} &= x^1\vec{v}_1 + x^2\vec{v}_2 + \vec{v}_0 = x^1(a_1^1\vec{v}_1' + a_2^1\vec{v}_2') + x^2(a_1^2\vec{v}_1' + a_2^2\vec{v}_2') + a_1^0\vec{v}_1' + a_2^0\vec{v}_2' \\ &= (a_1^1x^1 + a_1^2x^2 + a_1^0)\vec{v}_1' + (a_2^1x^1 + a_2^2x^2 + a_2^0)\vec{v}_2'.\end{aligned}$$

Portanto, as coordenadas do vetor $\vec{v}$ no sistema de coordenadas $(O', \mathcal{B}')$ são

$$(x^1)' = a_1^1x^1 + a_1^2x^2 + a_1^0,$$
$$(x^2)' = a_2^1x^1 + a_2^2x^2 + a_2^0,$$

ou seja,

$$\vec{v} = (a_1^1x^1 + a_1^2x^2 + a_1^0, a_2^1x^1 + a_2^2x^2 + a_2^0).$$

Consideremos um exemplo simples de passagem de um sistema de coordenadas para um outro.

Exemplo 1.1. Temos duas bases $\mathcal{B} = \{\vec{e}_1, \vec{e}_2\}$ e $\mathcal{B}' = \{\vec{e}_1', \vec{e}_2'\}$ em $\mathbb{V}^2$ com a mesma origem (ou seja, com $\vec{v}_0 = (0,0)$), e três vetores $(\vec{e}_1')_{\mathcal{B}} = (1,2)$, $(\vec{e}_2')_{\mathcal{B}} = (2,-3)$ e $\vec{a}_{\mathcal{B}} = (1,1)$. Vamos determinar $(\vec{e}_1)_{\mathcal{B}'}$, $(\vec{e}_2)_{\mathcal{B}'}$ e $\vec{a}_{\mathcal{B}'}$. Como $\vec{e}_1' = \vec{e}_1 + 2\vec{e}_2$ e $\vec{e}_2' = 2\vec{e}_1 - 3\vec{e}_2$, da primeira igualdade obtemos $\vec{e}_1 = \vec{e}_1' - 2\vec{e}_2$. Substituindo na segunda igualdade temos $\vec{e}_2' = 2\vec{e}_1' - 7\vec{e}_2$. Logo, $\vec{e}_2 = \frac{2}{7}\vec{e}_1' - \frac{1}{7}\vec{e}_2'$ e $\vec{e}_1 = \frac{3}{7}\vec{e}_1' + \frac{2}{7}\vec{e}_2'$, ou seja, $(\vec{e}_1)_{\mathcal{B}'} = (\frac{3}{7}, \frac{2}{7})$ e $(\vec{e}_2)_{\mathcal{B}'} = (\frac{2}{7}, -\frac{1}{7})$. Da igualdade $\vec{e}_1 + \vec{e}_2 = \frac{3}{7}\vec{e}_1' + \frac{2}{7}\vec{e}_2' + \frac{2}{7}\vec{e}_1' - \frac{1}{7}\vec{e}_2'$, obtemos $\vec{e}_1 + \vec{e}_2 = \frac{5}{7}\vec{e}_1' + \frac{1}{7}\vec{e}_2'$, isto é, $\vec{a}_{\mathcal{B}'} = (\frac{5}{7}, \frac{1}{7})$. □

1.5. Projeção de um vetor sobre uma reta orientada, produto escalar e base canónica

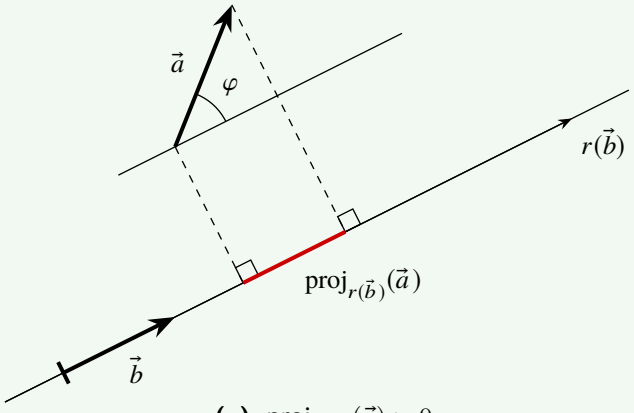
Nesta secção vamos começar por definir a projeção de um vetor sobre uma reta orientada para depois introduzir a definição de produto escalar.

Definição 1.1. Reta orientada

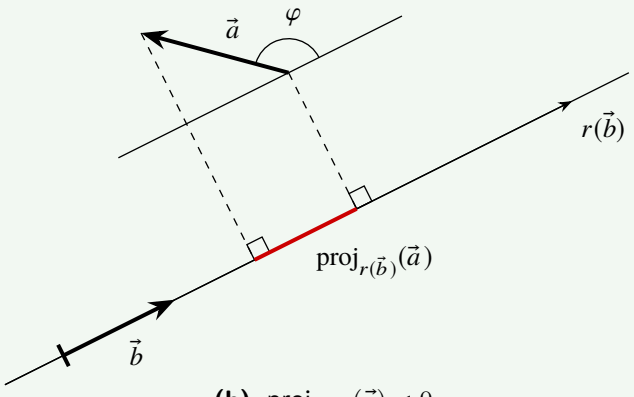
Seja $\vec{a}$ um vetor não nulo. A uma reta paralela ao vetor $\vec{a}$ podemos associar o sentido desse vetor. Nesse caso dizemos que a reta é *orientada* pelo vetor $\vec{a}$ e usamos a notação $r(\vec{a})$.

Definição 1.2. Projeção de um vetor sobre uma reta orientada

Sejam $\vec{a}$ um vetor e $r(\vec{b})$ a reta orientada pelo vetor não nulo $\vec{b}$. A *projeção do vetor* $\vec{a}$ sobre a reta orientada $r(\vec{b})$, que representamos por $\text{proj}_{r(\vec{b})}(\vec{a})$, é o número resultante da multiplicação da norma do vetor $\vec{a}$ pelo cosseno do ângulo formado entre os vetores $\vec{a}$ e $\vec{b}$, ou seja, $\text{proj}_{r(\vec{b})}(\vec{a}) = |\vec{a}| \cos\varphi$, $\varphi = \angle(\vec{a}, \vec{b})$ (veja a Fig. 1.19).



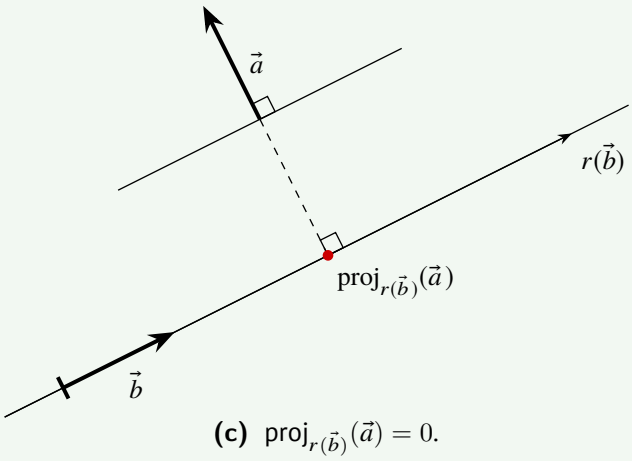
(a) $\text{proj}_{r(\vec{b})}(\vec{a}) > 0$.



(b) $\text{proj}_{r(\vec{b})}(\vec{a}) < 0$.

continua na página seguinte

continuação da página anterior



(c) $\text{proj}_{r(\vec{b})}(\vec{a}) = 0.$

Figura 1.19
Projeção de um vetor sobre uma reta orientada.

Temos, então, as seguintes propriedades.

Teorema 1.2

1. A projeção da soma de vetores é igual à soma das projeções destes vetores, isto é, sendo $\vec{a}_1, \ldots, \vec{a}_m$ vetores e $\vec{b}$ um vetor não nulo, temos que

$$\text{proj}_{r(\vec{b})}(\vec{a}_1 + \cdots + \vec{a}_m) = \text{proj}_{r(\vec{b})}(\vec{a}_1) + \cdots + \text{proj}_{r(\vec{b})}(\vec{a}_m).$$

2. A projeção do resultado da multiplicação de um vetor por um número é igual à multiplicação da projeção do vetor pelo número, isto é, sendo α um número, $\vec{a}$ um vetor e $\vec{b}$ um vetor não nulo, temos que

$$\text{proj}_{r(\vec{b})}(\alpha\vec{a}) = \alpha\,\text{proj}_{r(\vec{b})}(\vec{a}).$$

Demonstração. As demonstrações destas propriedades são uma consequência imediata da definição de projeção de um vetor sobre uma reta orientada e são ilustradas nas Figs. 1.20 e 1.21.

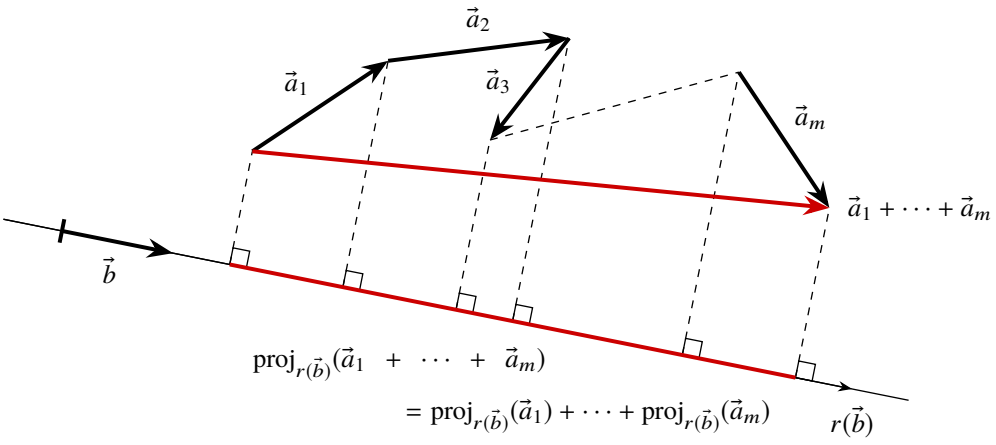


Figura 1.20
A projeção da soma de vetores é igual à soma das projeções.

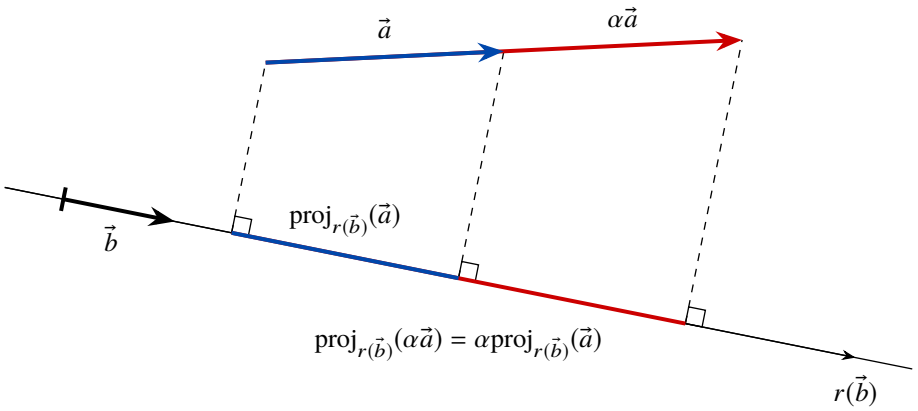


Figura 1.21
A projeção do resultado da multiplicação de um vetor por um número é igual ao resultado da multiplicação do número pela projeção do vetor.

■

Teorema 1.3

Sejam $\vec{a}$ um vetor e $\vec{b}$ um vetor não nulo. Então, $\text{proj}_{r(\vec{b})}(\vec{a}) = -\text{proj}_{r(-\vec{b})}(\vec{a})$.

Demonstração. Para variar, vamos fazer esta demonstração de uma maneira analítica, apesar de uma demonstração gráfica também ser possível e mais simples. Atendendo à definição da projeção de um vetor sobre uma reta orientada, temos:

$$\begin{aligned} -\text{proj}_{r(-\vec{b})}(\vec{a}) &= -(|\vec{a}|\cos(\angle(\vec{a},-\vec{b}))) = -(|\vec{a}|(-\cos(\angle(\vec{a},\vec{b})))) \\ &= |\vec{a}|\cos(\angle(\vec{a},\vec{b})) = \text{proj}_{r(\vec{b})}(\vec{a}). \end{aligned}$$

■

Vamos agora definir uma operação muito importante.

Definição 1.3. Produto escalar

Seja $\vec{a}$ um vetor do plano. Dizemos que o *produto escalar* entre $\vec{a}$ e $\vec{0}_2$, que representamos por $\langle\vec{a},\vec{0}_2\rangle$ ou $\vec{a}\cdot\vec{0}_2$, é 0. Sejam, agora, $\vec{a}$ e $\vec{b}$ dois vetores não nulos. Dizemos que o produto escalar entre $\vec{a}$ e $\vec{b}$, que representamos por $\langle\vec{a},\vec{b}\rangle$ ou $\vec{a}\cdot\vec{b}$, é igual ao resultado da multiplicação do comprimento do vetor $\vec{b}$ e da projeção do vetor $\vec{a}$ sobre a reta gerada pelo vetor $\vec{b}$. Temos, então:

$$\langle\vec{a},\vec{b}\rangle = \begin{cases} 0, & \text{se } \vec{a}=\vec{0}_2 \text{ ou } \vec{b}=\vec{0}_2, \\ |\vec{b}|\,\text{proj}_{r(\vec{b})}(\vec{a}), & \text{caso contrário.} \end{cases}$$

Ao produto escalar também chamamos *produto interno*.

Apresentamos, agora, algumas propriedades do produto escalar.

Teorema 1.4

Sejam $\vec{a}$ e $\vec{b}$ vetores não nulos do plano e φ o ângulo que formam. Então, $\langle\vec{a},\vec{b}\rangle = |\vec{a}||\vec{b}|\cos\varphi$.

Demonstração. Atendendo à definição da projeção de um vetor sobre uma reta orientada, temos:

$$\langle \vec{a}, \vec{b} \rangle = |\vec{b}| \operatorname{proj}_{r(\vec{b})}(\vec{a}) = |\vec{b}||\vec{a}| \cos \varphi = |\vec{a}||\vec{b}| \cos \varphi.$$

■

Notemos que, se o ângulo entre os vetores $\vec{a}$ e $\vec{b}$ for $\varphi = \frac{\pi}{2}$, então pelo teorema anterior temos $\langle \vec{a}, \vec{b} \rangle = 0$. Portanto, dois vetores são ortogonais se e só se o seu produto escalar é zero.

Teorema 1.5

Sejam $\vec{a}$, $\vec{b}$ e $\vec{c}$ vetores e α um número. Então:

1. $\langle \vec{a}, \vec{b} \rangle = \langle \vec{b}, \vec{a} \rangle$;
2. $\langle \alpha\vec{a}, \vec{b} \rangle = \alpha \langle \vec{a}, \vec{b} \rangle$;
3. $\langle \vec{a}, \vec{b} + \vec{c} \rangle = \langle \vec{a}, \vec{b} \rangle + \langle \vec{a}, \vec{c} \rangle$.

Demonstração.

1. Vamos fazer a demonstração considerando dois casos.
   - **(a)** Se $\vec{a} = \vec{0}_2$ ou $\vec{b} = \vec{0}_2$, então a propriedade é imediata.
   - **(b)** Se $\vec{a} \neq \vec{0}_2$ e $\vec{b} \neq \vec{0}_2$, e considerando $\varphi = \angle(\vec{a}, \vec{b})$, temos:

     $$\langle \vec{a}, \vec{b} \rangle = |\vec{a}||\vec{b}| \cos \varphi = |\vec{b}||\vec{a}| \cos \varphi = \langle \vec{b}, \vec{a} \rangle.$$

2. Vamos fazer a demonstração considerando três casos.
   - **(a)** Se $\alpha = 0$ ou $\vec{a} = \vec{0}_2$ ou $\vec{b} = \vec{0}_2$, então a propriedade é imediata.
   - **(b)** Se $\alpha > 0$ e $\vec{a} \neq \vec{0}_2$ e $\vec{b} \neq \vec{0}_2$, temos:

     $$\begin{aligned} \langle \alpha\vec{a}, \vec{b} \rangle &= \langle \vec{b}, \alpha\vec{a} \rangle = |\alpha\vec{a}| \operatorname{proj}_{r(\alpha\vec{a})}(\vec{b}) \\ &= |\alpha||\vec{a}| \operatorname{proj}_{r(\alpha\vec{a})}(\vec{b}) = \alpha|\vec{a}| \operatorname{proj}_{r(\vec{a})}(\vec{b}) = \alpha \langle \vec{a}, \vec{b} \rangle. \end{aligned}$$

   - **(c)** Se $\alpha < 0$ e $\vec{a} \neq \vec{0}_2$ e $\vec{b} \neq \vec{0}_2$, temos:

     $$\begin{aligned} \langle \alpha\vec{a}, \vec{b} \rangle &= \langle \vec{b}, \alpha\vec{a} \rangle = |\alpha\vec{a}| \operatorname{proj}_{r(\alpha\vec{a})}(\vec{b}) = |\alpha||\vec{a}| \operatorname{proj}_{r(\alpha\vec{a})}(\vec{b}) \\ &= \alpha|\vec{a}|(-\operatorname{proj}_{r(\vec{a})}(\vec{b})) = \alpha|\vec{a}| \operatorname{proj}_{r(\vec{a})}(\vec{b}) = \alpha \langle \vec{a}, \vec{b} \rangle. \end{aligned}$$

3. Vamos fazer a demonstração considerando quatro casos.
   - **(a)** Se $\vec{a} = \vec{0}_2$, então a propriedade é imediata.
   - **(b)** Se $\vec{b} + \vec{c} = \vec{0}_2$ com $\vec{b} = \vec{0}_2$ e $\vec{c} = \vec{0}_2$, então a propriedade é imediata.
   - **(c)** Se $\vec{b} + \vec{c} = \vec{0}_2$ com $\vec{b} \neq \vec{0}_2$ e $\vec{c} \neq \vec{0}_2$, então temos

     $$\langle \vec{a}, \vec{b} \rangle + \langle \vec{a}, \vec{c} \rangle = \langle \vec{a}, \vec{b} \rangle + \langle \vec{a}, -\vec{b} \rangle = \langle \vec{a}, \vec{b} \rangle - \langle \vec{a}, \vec{b} \rangle = 0 = \langle \vec{a}, \vec{b} + \vec{c} \rangle.$$

(d) Se $\vec{a} \neq \vec{0}_2$ e $\vec{b} + \vec{c} \neq \vec{0}_2$, então temos

$$\begin{aligned}\langle \vec{a}, \vec{b} + \vec{c} \rangle &= \langle \vec{b} + \vec{c}, \vec{a} \rangle = |\vec{a}| \operatorname{proj}_{r(\vec{a})}(\vec{b} + \vec{c}) \\ &= |\vec{a}|(\operatorname{proj}_{r(\vec{a})}(\vec{b}) + \operatorname{proj}_{r(\vec{a})}(\vec{c})) \\ &= |\vec{a}| \operatorname{proj}_{r(\vec{a})}(\vec{b}) + |\vec{a}| \operatorname{proj}_{r(\vec{a})}(\vec{c}) = \langle \vec{b}, \vec{a} \rangle + \langle \vec{c}, \vec{a} \rangle \\ &= \langle \vec{a}, \vec{b} \rangle + \langle \vec{a}, \vec{c} \rangle.\end{aligned}$$

■

As bases mais utilizadas são as *bases ortonormais*, ou seja, bases formadas por dois vetores ortogonais e unitários. Fixemos uma base ortonormal $\mathcal{B} = \{\vec{e}_1, \vec{e}_2\}$ em $\mathbb{V}^2$. A esta base vamos chamar *base canónica* (veja a Fig. 1.22). Obviamente as coordenadas destes vetores na base $\mathcal{B}$ são $\vec{e}_1 = (1, 0)$ e $\vec{e}_2 = (0, 1)$. Quando exprimimos as coordenadas de um vetor $\vec{a}$ numa base canónica escrevemos $\vec{a} = (a^1, a^2)$ (veja a Fig. 1.23).

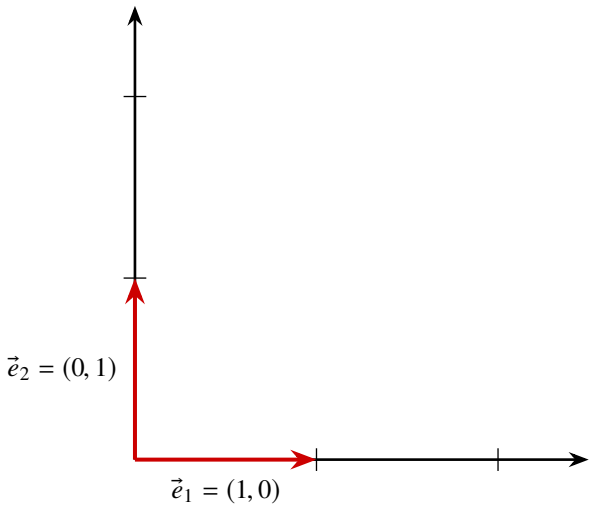


Figura 1.22
Base canónica.

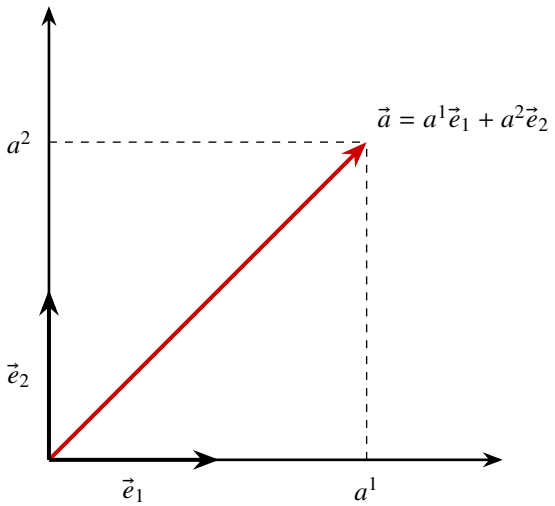


Figura 1.23
Coordenadas de um vetor numa base canónica.

Considerando uma base canónica, a soma dos dois vetores $\vec{a} = (a^1, a^2)$ e $\vec{b} = (b^1, b^2)$ é dada por (veja a Fig. 1.24)

$$\vec{a} + \vec{b} = (a^1 + b^1, a^2 + b^2)$$

e a multiplicação do vetor $\vec{a} = (a^1, a^2)$ pelo número α é dada por (veja a Fig. 1.25)

$$\alpha\vec{a} = (\alpha a^1, \alpha a^2).$$

De facto, temos

$$\begin{aligned}\vec{a} + \vec{b} &= a^1\vec{e}_1 + a^2\vec{e}_2 + b^1\vec{e}_1 + b^2\vec{e}_2 = (a^1 + b^1)\vec{e}_1 + (a^2 + b^2)\vec{e}_2 \\ &= (a^1 + b^1, a^2 + b^2)\end{aligned}$$

e

$$\alpha\vec{a} = \alpha(a^1\vec{e}_1 + a^2\vec{e}_2) = (\alpha a^1)\vec{e}_1 + (\alpha a^2)\vec{e}_2 = (\alpha a^1, \alpha a^2).$$

O comprimento do vetor $\vec{a} = (a^1, a^2)$, atendendo ao Teorema de Pitágoras, é dado por (veja a Fig. 1.26)

$$|\vec{a}| = \sqrt{(a^1)^2 + (a^2)^2}.$$

Se se considerar uma base canónica, o produto escalar pode ser calculado atendendo ao seguinte teorema.

Teorema 1.6

Sejam, numa base canónica, os vetores $\vec{a} = (a^1, a^2)$ e $\vec{b} = (b^1, b^2)$. Então,

$$\langle \vec{a}, \vec{b} \rangle = a^1b^1 + a^2b^2.$$

Demonstração. Pelo Teorema 1.5, temos:

$$\begin{aligned}\langle \vec{a}, \vec{b} \rangle &= \langle a^1\vec{e}_1 + a^2\vec{e}_2, b^1\vec{e}_1 + b^2\vec{e}_2 \rangle \\ &= a^1b^1\langle \vec{e}_1, \vec{e}_1 \rangle + a^1b^2\langle \vec{e}_1, \vec{e}_2 \rangle + a^2b^1\langle \vec{e}_2, \vec{e}_1 \rangle + a^2b^2\langle \vec{e}_2, \vec{e}_2 \rangle \\ &= a^1b^1 \times 1 + a^1b^2 \times 0 + a^2b^1 \times 0 + a^2b^2 \times 1 = a^1b^1 + a^2b^2.\end{aligned}$$

■

Se $\vec{a} \neq \vec{0}_2$ e $\vec{b} \neq \vec{0}_2$, podemos determinar a projeção do vetor $\vec{a}$ sobre a reta orientada $r(\vec{b})$ usando a expressão

$$\text{proj}_{r(\vec{b})}(\vec{a}) = \frac{\langle \vec{a}, \vec{b} \rangle}{|\vec{b}|} = \frac{a^1b^1 + a^2b^2}{\sqrt{(b^1)^2 + (b^2)^2}}.$$

Vamos ver como o cálculo vetorial pode ser utilizado na resolução de problemas geométricos.

Exemplo 1.2. Consideremos um triângulo ABC e a sua altura AH (veja a Fig. 1.27). Suponhamos que são dados os comprimentos dos lados do triângulo $|\overrightarrow{AB}|$, $|\overrightarrow{AC}|$ e $|\overrightarrow{BC}|$. Vamos encontrar o comprimento da altura AH, isto é, a norma do vetor $\vec{h} = \overrightarrow{AH}$.

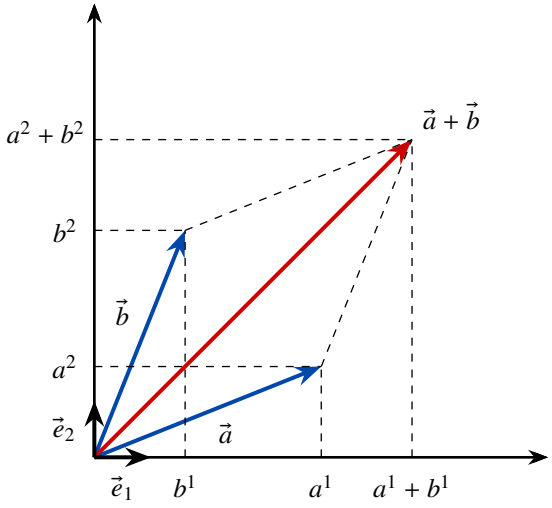


Figura 1.24
Soma de dois vetores considerando uma base canónica.

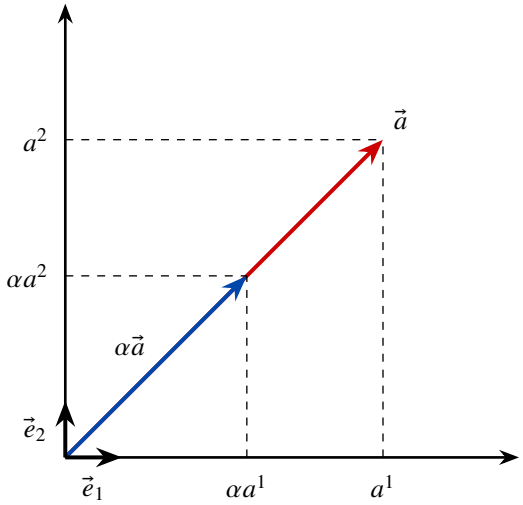


Figura 1.25
Multiplicação de um vetor por um número considerando uma base canónica.

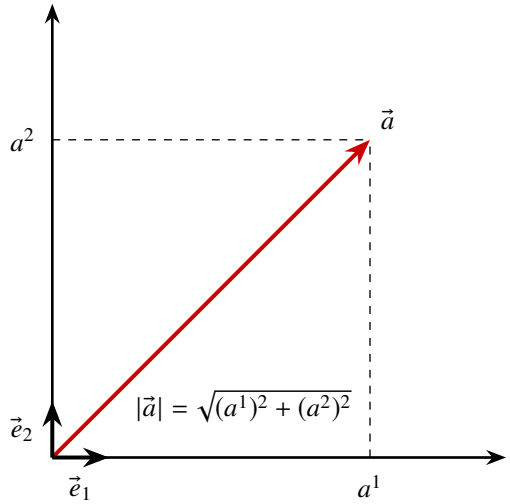


Figura 1.26
Comprimento de um vetor considerando uma base canónica.

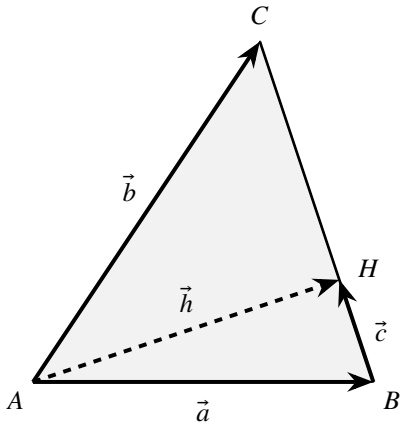


Figura 1.27
Triângulo ABC e a sua altura AH.

A resolução de um problema deste tipo começa pela escolha de uma base em $\mathbb{V}^2$. Podemos escolher, por exemplo, a base formada pelos vetores $\vec{a} = \overrightarrow{AB}$ e $\vec{b} = \overrightarrow{AC}$. Então temos $\vec{c} = \overrightarrow{BH} = \lambda(\vec{b} - \vec{a})$, onde λ é um número. Como o vetor $\vec{h}$ é ortogonal ao vetor $\overrightarrow{BC} = \vec{b} - \vec{a}$, temos

$$0 = \langle \vec{h}, \vec{b} - \vec{a} \rangle = \langle \vec{a} + \vec{c}, \vec{b} - \vec{a} \rangle = \langle \vec{a} + \lambda(\vec{b} - \vec{a}), \vec{b} - \vec{a} \rangle.$$

Daqui encontramos

$$\lambda = \frac{|\vec{a}|^2 - \langle \vec{a}, \vec{b} \rangle}{|\vec{b} - \vec{a}|^2}.$$

Logo, temos

$$\vec{h} = \vec{a} + \vec{c} = \vec{a} + \frac{|\vec{a}|^2 - \langle \vec{a}, \vec{b} \rangle}{|\vec{b} - \vec{a}|^2}(\vec{b} - \vec{a}).$$

Agora é necessário calcular o produto escalar $\langle \vec{a}, \vec{b} \rangle$ em termos dos comprimentos dos lados do triângulo. Para fazer isso notemos que

$$|\vec{b} - \vec{a}|^2 = \langle \vec{b} - \vec{a}, \vec{b} - \vec{a} \rangle = |\vec{a}|^2 + |\vec{b}|^2 - 2\langle \vec{a}, \vec{b} \rangle.$$

Portanto, temos

$$\langle \vec{a}, \vec{b} \rangle = \frac{1}{2}(|\vec{a}|^2 + |\vec{b}|^2 - |\vec{b} - \vec{a}|^2)$$

e

$$|\vec{a}|^2 - \langle \vec{a}, \vec{b} \rangle = \frac{1}{2}(|\vec{a}|^2 - |\vec{b}|^2 + |\vec{b} - \vec{a}|^2).$$

Utilizando esta igualdade encontramos

$$|\vec{h}|^2 = |\vec{a}|^2 + \left(\frac{|\vec{a}|^2 - \langle \vec{a}, \vec{b} \rangle}{|\vec{b} - \vec{a}|^2} \right)^2 |\vec{b} - \vec{a}|^2 - 2\frac{(|\vec{a}|^2 - \langle \vec{a}, \vec{b} \rangle)^2}{|\vec{b} - \vec{a}|^2}$$

$$= |\vec{a}|^2 - \frac{(|\vec{a}|^2 - \langle \vec{a}, \vec{b} \rangle)^2}{|\vec{b} - \vec{a}|^2} = |\vec{a}|^2 - \frac{(|\vec{a}|^2 - |\vec{b}|^2 + |\vec{b} - \vec{a}|^2)^2}{4|\vec{b} - \vec{a}|^2}.$$

□

Exemplo 1.3. Vamos, agora, demonstrar que as medianas de um triângulo têm um ponto comum. Consideremos um triângulo ABC (veja a Fig. 1.28).

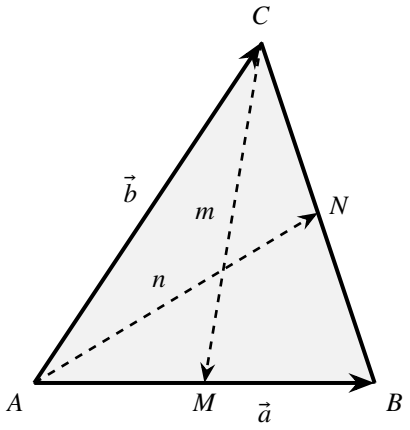


Figura 1.28
Medianas.

Sejam $\vec{a} = \overrightarrow{AB}$ e $\vec{b} = \overrightarrow{AC}$. Estes vetores formam uma base em $\mathbb{V}^2$. Designemos por M e N os pontos de interseção das medianas com os lados do triângulo AB e BC, respetivamente. Usando as notações $\vec{m} = \overrightarrow{CM}$ e $\vec{n} = \overrightarrow{AN}$, temos

$$\vec{m} = \frac{1}{2}\vec{a} - \vec{b} \quad \text{e} \quad \vec{n} = \vec{a} + \frac{1}{2}(\vec{b} - \vec{a}).$$

O ponto de interseção das medianas é $\vec{b} + \tau\vec{m} = t\vec{n}$, onde t e τ são parâmetros (números reais). Portanto, temos

$$\vec{b} + \tau\left(\frac{1}{2}\vec{a} - \vec{b}\right) = t\left(\vec{a} + \frac{1}{2}(\vec{b} - \vec{a})\right)$$

ou, na forma equivalente,

$$\frac{1}{2}\tau\vec{a} + (1-\tau)\vec{b} = \frac{1}{2}t\vec{a} + \frac{1}{2}t\vec{b}.$$

Os coeficientes dos vetores da base no lado esquerdo e no lado direito desta igualdade têm de ser os mesmos. Logo, temos o sistema linear $\frac{1}{2}\tau = \frac{1}{2}t$ e $1 - \tau = \frac{1}{2}t$. Isto significa que $t = \tau = \frac{2}{3}$. Portanto, o ponto de interseção das medianas divide as medianas na proporção $2 : 1$. Isto significa que todas as medianas têm um ponto de interseção comum.

□

1.6. Método de Gauss para sistemas lineares com duas equações e com duas incógnitas

É útil termos uma ferramenta para fazer cálculos com coordenadas. Essa ferramenta é um algoritmo de resolução de sistemas de equações lineares e é conhecido como Método de Gauss (conhecido também por eliminação de Gauss). Com a ajuda deste algoritmo, na secção seguinte, conseguiremos, por exemplo, calcular as coordenadas do ponto de interseção de duas retas.

Consideremos um sistema linear com duas equações e com duas incógnitas x^1 e x^2

$$\begin{cases} ax^1 + bx^2 = c, \\ dx^1 + ex^2 = f, \end{cases}$$

em que a, b, c, d, e e f são números reais. O que pretendemos é determinar os números reais $\bar{x}^1$ e $\bar{x}^2$ tais que

$$\begin{cases} a\bar{x}^1 + b\bar{x}^2 = c, \\ d\bar{x}^1 + e\bar{x}^2 = f. \end{cases}$$

Se existirem, dizemos que o par ordenado $(\bar{x}^1, \bar{x}^2)$ é uma solução do sistema linear. Veremos que há sistemas de equações lineares para os quais só existe um par ordenado que é solução e neste caso dizemos que é um sistema possível e determinado. Há sistemas de equações lineares para os quais existe uma infinidade de pares ordenados que são solução e neste caso dizemos que é um sistema possível e indeterminado, e há sistemas de equações lineares para os quais não existe nenhum par ordenado que é solução e dizemos neste caso que é um sistema impossível. Estas são as únicas situações possíveis, não podendo existir, por exemplo, um sistema linear para o qual existem só dois pares ordenados que são sua solução. No caso de $c = f = 0$, dizemos que o sistema de equações lineares é um sistema homogéneo, que admite sempre a solução trivial $\bar{x}^1 = \bar{x}^2 = 0$.

Existem muitas maneiras para determinar a solução de um sistema linear. Neste texto vamos estudar o Método de Gauss.

O Método de Gauss é constituído por duas fases. O objetivo da primeira é transformar o sistema dado num novo sistema linear mais simples mas equivalente, isto é com a mesma solução. No final da primeira fase já é possível verificar se o sistema dado é possível e determinado, se é possível e indeterminado, ou se é impossível. A segunda fase só se aplica aos dois primeiros casos e é constituída pela obtenção da sua solução.

Vamos agora ilustrar a aplicação do Método de Gauss a sistemas lineares com duas equações e com duas incógnitas considerando os três tipos de sistemas de equações lineares.

Comecemos pelo sistema linear

$$\begin{cases} x^1 + x^2 = 1, \\ 2x^1 - 2x^2 = 1. \end{cases}$$

Se subtrairmos à segunda equação o dobro da primeira, obtemos

$$\begin{cases} x^1 + x^2 = 1 \\ \quad -4x^2 = -1. \end{cases}$$

Com esta operação, termina a primeira fase do Método de Gauss para este exemplo. A segunda fase consiste em determinar a sua solução, que agora é simples de calcular:

da segunda equação determinamos o valor de x^2, que é neste caso $\frac{1}{4}$. Se substituirmos agora este valor na primeira equação, obtemos o valor de x^1, que é neste caso $\frac{3}{4}$. Este é um exemplo de um sistema possível e determinado.

Consideremos, agora, o sistema linear

$$\begin{cases} x^1 + x^2 = 1, \\ -3x^1 - 3x^2 = -3. \end{cases}$$

Se adicionarmos à segunda equação o triplo da primeira, obtemos

$$\begin{cases} x^1 + x^2 = 1, \\ 0x^2 = 0, \end{cases}$$

o que implica que a segunda equação é sempre satisfeita qualquer que seja o valor de x^2. Por este motivo, dizemos que x^2 é uma incógnita livre e que vai ser substituída pelo parâmetro real $x^2 = \alpha \in \mathbb{R}$. O valor da outra incógnita, x^1, que se diz incógnita pivô, terá necessariamente que vir em função daquela, ou seja, $x^1 = 1 - \alpha$. Este é um exemplo de um sistema possível e indeterminado.

Seja, finalmente, o sistema linear

$$\begin{cases} x^1 - x^2 = 1, \\ 2x^1 - 2x^2 = 1. \end{cases}$$

Se subtrairmos à segunda equação o dobro da primeira, obtemos

$$\begin{cases} x^1 - x^2 = 1, \\ 0x^2 = -1, \end{cases}$$

o que implica que a segunda equação nunca é satisfeita, isto é, não existe nenhum número que multiplicado por 0 dê -1. Concluímos, então, que este é um exemplo de um sistema impossível.

Vamos agora mostrar como o Método de Gauss ajuda na resolução de problemas de cálculo vetorial envolvendo coordenadas.

Exemplo 1.4. Consideremos vetores $\vec{a}$, $\vec{b}$ e $\vec{r}$ de $\mathbb{V}^2$ que numa certa base têm coordenadas $(2,1)$, $(3,2)$ e $(1,2)$, respetivamente. Vamos mostrar que $\mathcal{B} = \{\vec{a}, \vec{b}\}$ é uma base em $\mathbb{V}^2$ e determinar as coordenadas do vetor $\vec{r}$ nesta base.

Para mostrar que $\mathcal{B}$ é uma base em $\mathbb{V}^2$ basta demonstrar que os vetores $\vec{a}$ e $\vec{b}$ são linearmente independentes, ou seja, que a igualdade

$$\alpha(2,1) + \beta(3,2) = (0,0) \tag{1.3}$$

implica que $\alpha = \beta = 0$. A igualdade (1.3) é equivalente ao sistema

$$\begin{cases} 2\alpha + 3\beta = 0, \\ \alpha + 2\beta = 0. \end{cases}$$

Resolvendo-o através do Método de Gauss, obtemos $\alpha = \beta = 0$. Portanto, os vetores $\vec{a}$ e $\vec{b}$ são linearmente independentes e formam uma base em $\mathbb{V}^2$.

Para encontrar as coordenadas do vetor $\vec{r}$ nesta base resolvemos a equação $\alpha(2,1) + \beta(3,2) = (1,2)$ que é equivalente ao sistema

$$\begin{cases} 2\alpha + 3\beta = 1, \\ \alpha + 2\beta = 2. \end{cases}$$

Resolvendo-o através do Método de Gauss, obtemos $\alpha = -4$ e $\beta = 3$. Assim, temos que $\vec{r}_{\mathcal{B}} = (-4, 3)$.

□

1.7. Equação geral de uma reta

Vamos agora deduzir a equação geral de uma reta no plano usando as fórmulas, obtidas na Secção 1.4, para a mudança de sistemas de coordenadas.

Consideremos as bases $\mathcal{B} = \{\vec{v}_1, \vec{v}_2\}$ e $\mathcal{B}' = \{\vec{v}'_1, \vec{v}'_2\}$ em $\mathbb{V}^2$ e os sistemas de coordenadas $(O, \mathcal{B})$ e $(O', \mathcal{B}')$. Sejam, ainda, (x^1, x^2) as coordenadas do vetor $\vec{v}$ no sistema de coordenadas $(O, \mathcal{B})$. Atendendo à mudança de coordenadas, vemos que no sistema de coordenadas $(O', \mathcal{B}')$ as coordenadas do mesmo vetor são

$$\begin{aligned} (x^1)' &= a_1^1 x^1 + a_1^2 x^2 + a_1^0, \\ (x^2)' &= a_2^1 x^1 + a_2^2 x^2 + a_2^0. \end{aligned}$$

Vamos usar estas fórmulas para deduzir a equação geral de uma reta no plano. Seja r uma reta. Escolhemos o ponto O' e a base $\{\vec{v}'_1, \vec{v}'_2\}$ como está indicado na Fig. 1.29, em que o vetor $\vec{v}'_1$ tem a direção da reta r. Neste sistema de coordenadas a equação da reta tem forma muito simples: $(x^2)' = 0$. Portanto, a equação da reta r no sistema de coordenadas $(O, \mathcal{B})$ toma a forma

$$a_2^1 x^1 + a_2^2 x^2 + a_2^0 = 0.$$

Assim, a equação geral da reta no plano é

$$Ax^1 + Bx^2 + C = 0.$$

Podemos agora dar uma interpretação geométrica muito simples para os sistemas lineares com duas equações e duas incógnitas: as duas equações definem duas retas do plano, sendo a solução do sistema linear formada pelos pontos comuns destas duas retas. Retomando os três exemplos apresentados na secção anterior, o primeiro sistema

$$\begin{cases} x^1 + x^2 = 1, \\ 2x^1 - 2x^2 = 1, \end{cases}$$

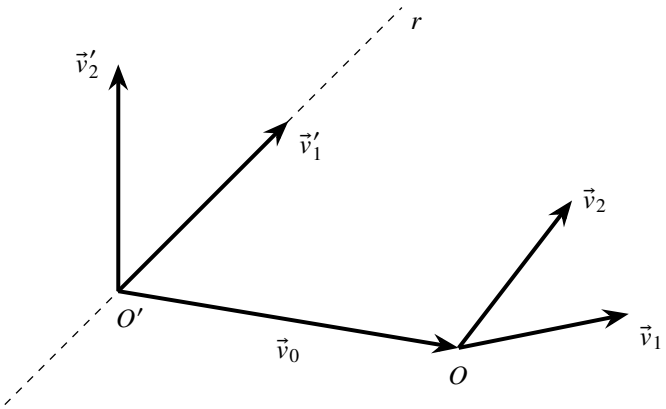


Figura 1.29
Equação de uma reta.

cuja solução é $x^1 = \frac{3}{4}$ e $x^2 = \frac{1}{4}$, corresponde a duas retas concorrentes, isto é, com um único ponto comum (veja a Fig. 1.30), o segundo sistema

$$\begin{cases} x^1 + x^2 = 1, \\ -3x^1 - 3x^2 = -3, \end{cases}$$

cuja solução é $x^1 = 1 - \alpha$ e $= \alpha \in \mathbb{R}$, corresponde a duas retas coincidentes, isto é, com uma infinidade de pontos em comum (veja a Fig. 1.31), e o terceiro sistema

$$\begin{cases} x^1 - x^2 = 1, \\ 2x^1 - 2x^2 = 1, \end{cases}$$

que é um sistema impossível, corresponde a duas retas paralelas, isto é, retas sem pontos em comum (veja a Fig. 1.32).

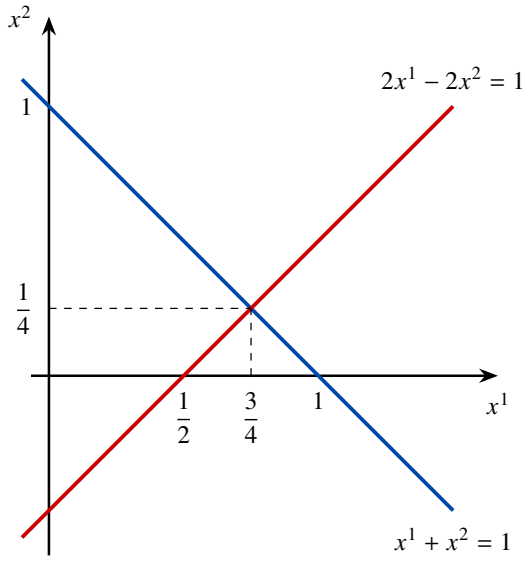


Figura 1.30
Sistema possível e determinado.

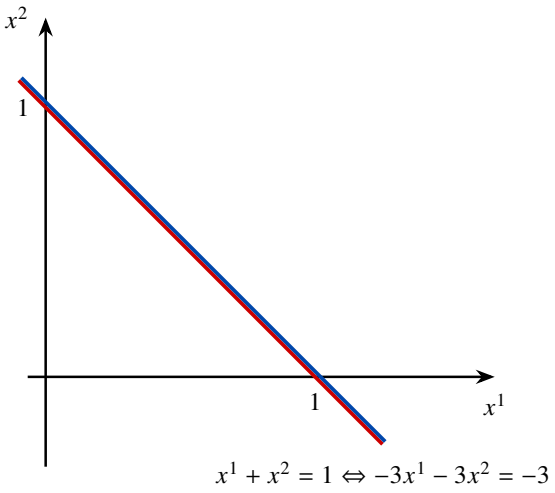


Figura 1.31
Sistema possível e indeterminado.

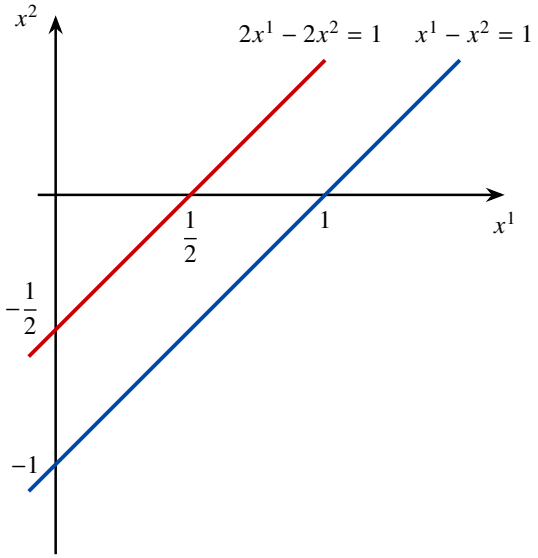


Figura 1.32
Sistema impossível.

2. Espaço de dimensão três

2.1. Introdução

Todas as definições e propriedades que apresentámos para o plano, tais como o Teorema 1.1, reta orientada, projeção de um vetor sobre uma reta orientada, produto escalar e todas as propriedades deste produto, conceito de ortogonalidade e bases canónicas são imediatamente generalizáveis para o caso do espaço de dimensão 3.

Representamos por $\mathbb{V}^3$ o conjunto de todos os vetores no espaço, que continuamos a chamar-lhes apenas vetores.

2.2. Sistemas de coordenadas

Nesta secção vamos deduzir as fórmulas para a mudança de sistemas de coordenadas em $\mathbb{V}^3$. Consideremos, em $\mathbb{V}^3$, dois sistemas de coordenadas $(O, \mathcal{B})$ e $(O', \mathcal{B}')$, onde $\mathcal{B} = \{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ e $\mathcal{B}' = \{\vec{v}'_1, \vec{v}'_2, \vec{v}'_3\}$. Consideremos um vetor $\vec{v}$ que no sistema de coordenadas $(O, \mathcal{B})$ tem a forma

$$\vec{v} = x^1\vec{v}_1 + x^2\vec{v}_2 + x^3\vec{v}_3. \tag{2.1}$$

Vamos ver como escrever este vetor no sistema de coordenadas $(O', \mathcal{B}')$. Suponhamos que na base $\mathcal{B}'$ temos as representações

$$\begin{aligned}
\vec{v}_0 &= a_1^0\vec{v}'_1 + a_2^0\vec{v}'_2 + a_3^0\vec{v}'_3,\\
\vec{v}_1 &= a_1^1\vec{v}'_1 + a_2^1\vec{v}'_2 + a_3^1\vec{v}'_3,\\
\vec{v}_2 &= a_1^2\vec{v}'_1 + a_2^2\vec{v}'_2 + a_3^2\vec{v}'_3,\\
\vec{v}_3 &= a_1^3\vec{v}'_1 + a_2^3\vec{v}'_2 + a_3^3\vec{v}'_3.
\end{aligned}$$

No sistema de coordenadas $(O', \mathcal{B}')$, em vez de (2.1), temos

$$\begin{aligned}
\vec{v} &= x^1\vec{v}_1 + x^2\vec{v}_2 + x^3\vec{v}_3 + \vec{v}_0\\
&= x^1(a_1^1\vec{v}'_1 + a_2^1\vec{v}'_2 + a_3^1\vec{v}'_3) + x^2(a_1^2\vec{v}'_1 + a_2^2\vec{v}'_2 + a_3^2\vec{v}'_3)\\
&\quad + x^3(a_1^3\vec{v}'_1 + a_2^3\vec{v}'_2 + a_3^3\vec{v}'_3) + a_1^0\vec{v}'_1 + a_2^0\vec{v}'_2 + a_3^0\vec{v}'_3\\
&= (a_1^1x^1 + a_1^2x^2 + a_1^3x^3 + a_1^0)\vec{v}'_1 + (a_2^1x^1 + a_2^2x^2 + a_2^3x^3 + a_2^0)\vec{v}'_2\\
&\quad + (a_3^1x^1 + a_3^2x^2 + a_3^3x^3 + a_3^0)\vec{v}'_3.
\end{aligned}$$

Então, analogamente ao caso de $\mathbb{V}^2$, as coordenadas do vetor $\vec{v}$ no sistema $(O', \mathcal{B}')$ são

$$\begin{aligned}
(x^1)' &= a_1^1x^1 + a_1^2x^2 + a_1^3x^3 + a_1^0,\\
(x^2)' &= a_2^1x^1 + a_2^2x^2 + a_2^3x^3 + a_2^0,\\
(x^3)' &= a_3^1x^1 + a_3^2x^2 + a_3^3x^3 + a_3^0.
\end{aligned}$$

2.3. Produto vetorial

Vamos definir agora uma operação em que dados dois vetores dá como resultado um novo vetor. Esta operação é muito importante e posteriormente vai ser usada para definir o produto vetorial nos espaços de dimensão superior.

Definição 2.1. Produto vetorial

Sejam $\vec{a}$ e $\vec{b}$ vetores de $\mathbb{V}^3$. Chamamos *produto vetorial* ou *produto externo* dos vetores $\vec{a}$ e $\vec{b}$, que representamos por $[\vec{a},\vec{b}]$ ou por $\vec{a}\times\vec{b}$, a um vetor de $\mathbb{V}^3$ definido assim:

1. se os vetores $\vec{a}$ ou $\vec{b}$ são vetores nulos ou paralelos, é o vetor nulo do espaço, ou seja, $[\vec{a},\vec{b}]=\vec{0}_3$;
2. caso contrário (veja a Fig. 2.1), é o vetor ortogonal aos vetores $\vec{a}$ e $\vec{b}$, tal que
   - **(a)** o seu comprimento é igual à área do paralelogramo formado pelos vetores $\vec{a}$ e $\vec{b}$
   - **(b)** e o seu sentido é dado pela regra "pés e cabeça", ou seja, colocamos o pé direito alinhado com o vetor $\vec{a}$ e o pé esquerdo alinhado com o vetor $\vec{b}$, então o produto vetorial de $\vec{a}$ e $\vec{b}$ tem o sentido da cabeça.

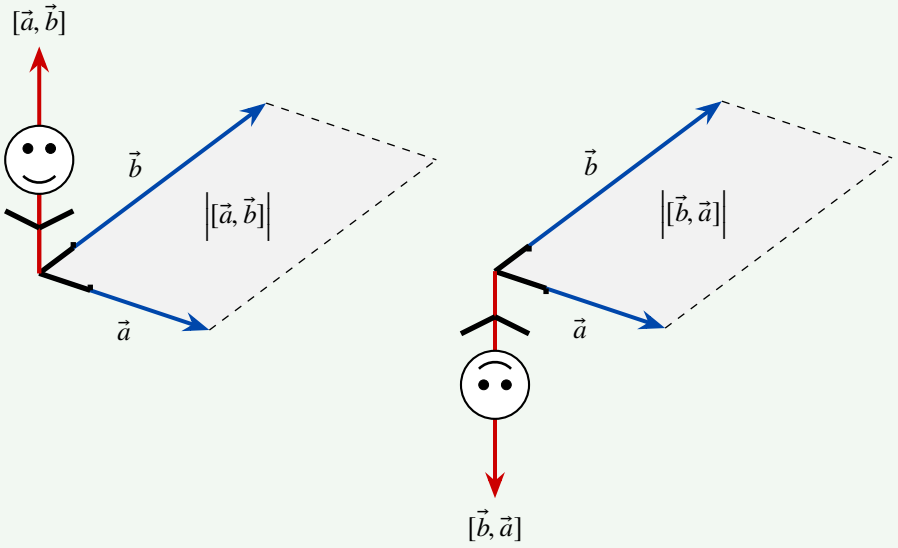


Figura 2.1
Produto vetorial.

A partir da definição podemos concluir que os vetores de uma base canónica verificam as seguintes igualdades:

$$[\vec{e}_1,\vec{e}_1]=\vec{0}_3,\quad [\vec{e}_1,\vec{e}_2]=\vec{e}_3,\quad [\vec{e}_1,\vec{e}_3]=-\vec{e}_2,$$
$$[\vec{e}_2,\vec{e}_1]=-\vec{e}_3,\quad [\vec{e}_2,\vec{e}_2]=\vec{0}_3,\quad [\vec{e}_2,\vec{e}_3]=\vec{e}_1,$$
$$[\vec{e}_3,\vec{e}_1]=\vec{e}_2,\quad [\vec{e}_3,\vec{e}_2]=-\vec{e}_1,\quad [\vec{e}_3,\vec{e}_3]=\vec{0}_3.$$

Na geometria do espaço $\mathbb{V}^3$, a seguinte definição assume um papel fundamental.

Definição 2.2. Projeção de um vetor num plano

Sejam π um plano do espaço $\mathbb{V}^3$ e $\vec{a}$ um vetor. Dizemos que um vetor $\vec{b}$ de π é a *projeção ortogonal* de $\vec{a}$ no plano π se o vetor $\vec{c} = \vec{a} - \vec{b}$ é ortogonal a todos os vetores de π.

Apresentamos, agora, algumas propriedades do produto vetorial, começando com um lema que vai útil numa das demonstrações.

Lema 2.1

Sejam $\vec{p}$ e $\vec{q}$ vetores não nulos de $\mathbb{V}^3$. Então, $[\vec{p}, \vec{q}] = |\vec{q}|\vec{p}_2$, em que $\vec{p}_2$ é o vetor que se obtém rodando a projeção do vetor $\vec{p}$ no plano ortogonal a $\vec{q}$ na direção de $[\vec{p}, \vec{q}]$.

Demonstração. Sejam π o plano ortogonal a $\vec{q}$ e $\vec{p}_1$ a projeção de $\vec{p}$ no plano π. O vetor $[\vec{p}, \vec{q}]$ é ortogonal aos vetores $\vec{p}$ e $\vec{q}$ e portanto também a $\vec{p}_1$ (veja a Fig. 2.2). Seja $\vec{p}_2$ a rotação de $\vec{p}_1$ na direção de $[\vec{p}, \vec{q}]$. Seja, ainda, φ o ângulo entre $\vec{p}$ e $\vec{q}$. Como $|\vec{p}_1| = |\vec{p}| \operatorname{sen} \varphi$ e $|\vec{p}_2| = |\vec{p}_1|$, temos que $|\vec{p}_2| = |\vec{p}| \operatorname{sen} \varphi$, pelo que $|[\vec{p}, \vec{q}]| = |\vec{p}_2||\vec{q}|$, uma vez que $|[\vec{p}, \vec{q}]| = |\vec{p}||\vec{q}| \operatorname{sen} \varphi$ (que é igual à área do paralelogramo formado por $\vec{p}$ e $\vec{q}$). Temos, então, que $[\vec{p}, \vec{q}] = |\vec{q}|\vec{p}_2$.

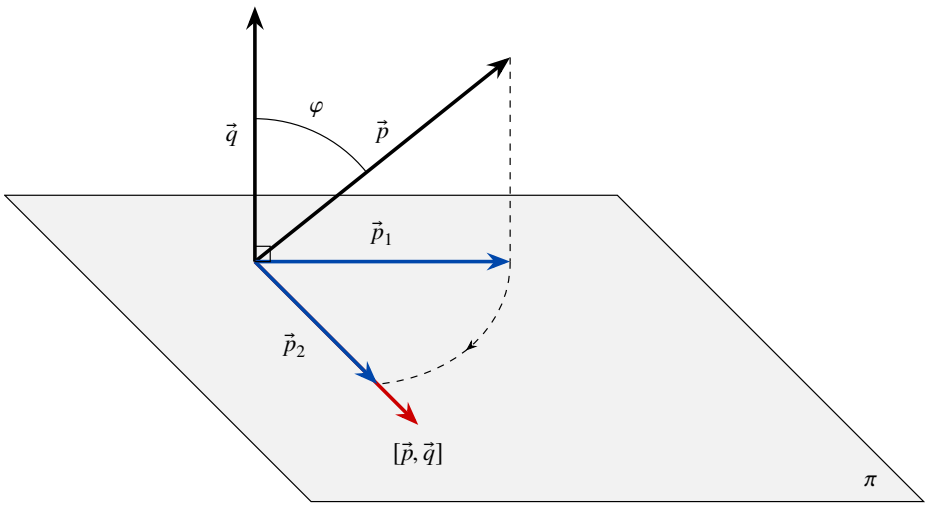


Figura 2.2
Construção geométrica do produto vetorial.

■

Teorema 2.1

Sejam $\vec{a}$, $\vec{b}$ e $\vec{c}$ vetores de $\mathbb{V}^3$ e α um número. Então:

1. $[\vec{a}, \vec{b}] = -[\vec{b}, \vec{a}]$;
2. $[\alpha\vec{a}, \vec{b}] = \alpha[\vec{a}, \vec{b}]$;
3. $[\vec{a} + \vec{b}, \vec{c}] = [\vec{a}, \vec{c}] + [\vec{b}, \vec{c}]$.

Demonstração.

1. Esta propriedade é uma consequência imediata da definição de produto vetorial.

2. Obviamente as normas dos vetores $[\alpha\vec{a},\vec{b}]$ e $\alpha[\vec{a},\vec{b}]$ coincidem. O sentido destes vetores coincide com o sentido do vetor $[\vec{a},\vec{b}]$ quando $\alpha > 0$ e é oposto quando $\alpha < 0$.

3. Seja π o plano ortogonal ao vetor $\vec{c}$. Seja o triângulo OA_1B_1 a projeção do triângulo OAB no plano π e seja o triângulo OA_2B_2 o resultado da rotação com amplitude $\frac{\pi}{2}$ do triângulo OA_1B_1 no sentido do vetor $[\vec{a}+\vec{b},\vec{c}]$ (veja a Fig. 2.3). Se a este último triângulo aplicarmos uma homotetia de centro O e coeficiente $|\vec{c}|$, obtemos o triângulo OA_3B_3. Então, pelo Lema 2.1, temos as três seguintes igualdades:

$$[\vec{a},\vec{c}] = |\vec{c}|\overrightarrow{OA_2} = \overrightarrow{OA_3},$$

$$[\vec{b},\vec{c}] = |\vec{c}|\overrightarrow{A_2B_2} = \overrightarrow{A_3B_3},$$

$$[\vec{a}+\vec{b},\vec{c}] = |\vec{c}|\overrightarrow{OB_2} = \overrightarrow{OB_3}.$$

Como $\overrightarrow{OB_3} = \overrightarrow{OA_3} + \overrightarrow{A_3B_3}$, temos que $[\vec{a}+\vec{b},\vec{c}] = [\vec{a},\vec{c}] + [\vec{b},\vec{c}]$.

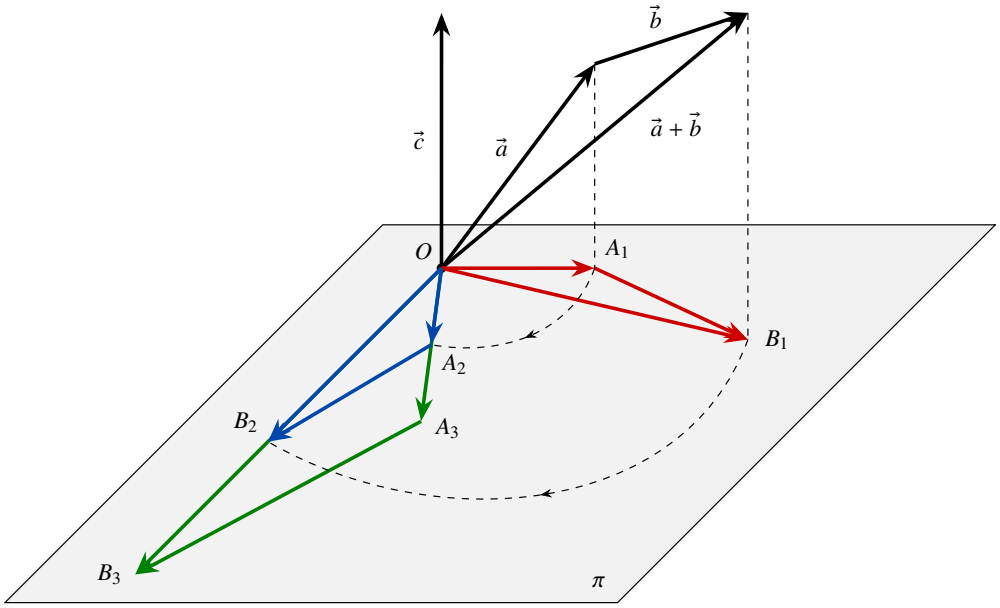


Figura 2.3
Igualdade
$[\vec{a}+\vec{b},\vec{c}] = [\vec{a},\vec{c}] + [\vec{b},\vec{c}]$.

■

Teorema 2.2

Consideremos, numa base canónica em $\mathbb{V}^3$, os vetores $\vec{a} = (a^1,a^2,a^3)$ e $\vec{b} = (b^1,b^2,b^3)$. Então temos

$$[\vec{a},\vec{b}] = (a^2b^3 - a^3b^2, a^3b^1 - a^1b^3, a^1b^2 - a^2b^1).$$

Demonstração. Pelo Teorema 2.1, temos:

$$\begin{aligned}
[\vec{a},\vec{b}] &= [a^1\vec{e}_1 + a^2\vec{e}_2 + a^3\vec{e}_3, b^1\vec{e}_1 + b^2\vec{e}_2 + b^3\vec{e}_3] \\
&= a^1b^1[\vec{e}_1,\vec{e}_1] + a^1b^2[\vec{e}_1,\vec{e}_2] + a^1b^3[\vec{e}_1,\vec{e}_3] \\
&\quad + a^2b^1[\vec{e}_2,\vec{e}_1] + a^2b^2[\vec{e}_2,\vec{e}_2] + a^2b^3[\vec{e}_2,\vec{e}_3] \\
&\quad + a^3b^1[\vec{e}_3,\vec{e}_1] + a^3b^2[\vec{e}_3,\vec{e}_2] + a^3b^3[\vec{e}_3,\vec{e}_3] \\
&= a^1b^1\vec{0}_3 + a^1b^2\vec{e}_3 + a^1b^3(-\vec{e}_2) + a^2b^1(-\vec{e}_3) + a^2b^2\vec{0}_3 + a^2b^3\vec{e}_1 \\
&\quad + a^3b^1\vec{e}_2 + a^3b^2(-\vec{e}_1) + a^3b^3\vec{0}_3 \\
&= (a^2b^3 - a^3b^2, a^3b^1 - a^1b^3, a^1b^2 - a^2b^1).
\end{aligned}$$

■

Notemos que, no caso de vetores $\vec{a} = (a^1, a^2, 0)$ e $\vec{b} = (b^1, b^2, 0)$, temos

$$[\vec{a},\vec{b}] = (0, 0, a^1b^2 - a^2b^1).$$

Portanto, a área do paralelogramo formado pelos vetores (a^1, a^2) e (b^1, b^2) de $\mathbb{V}^2$ é igual a $a^1b^2 - a^2b^1$.

Exemplo 2.1. Consideremos os vetores $\vec{a} = (2, 1, 2)$ e $\vec{b} = (3, 2, -1)$ numa base canónica. Pelo Teorema 2.2 temos que

$$[\vec{a},\vec{b}] = (1\times(-1) - 2\times 2, 3\times 2 - 2\times(-1), 2\times 2 - 3\times 1) = (-5, 8, 1).$$

□

Teorema 2.3

Sejam $\vec{a}$, $\vec{b}$ e $\vec{c}$ vetores de $\mathbb{V}^3$. Então temos

$$[[\vec{a},\vec{b}],\vec{c}] = \langle\vec{a},\vec{c}\rangle\,\vec{b} - \langle\vec{b},\vec{c}\rangle\,\vec{a}. \tag{2.2}$$

Demonstração. Se pelo menos um dos vetores $\vec{a}$, $\vec{b}$ ou $\vec{c}$ for o vetor nulo, a propriedade é imediata. Consideremos, agora, que os três vetores são não nulos. O vetor $\vec{R} = [[\vec{a},\vec{b}],\vec{c}]$ é perpendicular ao vetor $[\vec{a},\vec{b}]$ (veja a Fig. 2.4). Portanto, está contido no plano gerado pelos vetores $\vec{a}$ e $\vec{b}$ e admite a representação $\vec{R} = \alpha\vec{a} + \beta\vec{b}$. Como $\vec{R}$ é também ortogonal ao vetor $\vec{c}$, temos

$$0 = \alpha\langle\vec{a},\vec{c}\rangle + \beta\langle\vec{b},\vec{c}\rangle,$$

pelo que

$$\frac{\beta}{\langle\vec{a},\vec{c}\rangle} = -\frac{\alpha}{\langle\vec{b},\vec{c}\rangle}.$$

Fazendo

$$\sigma = \frac{\beta}{\langle\vec{a},\vec{c}\rangle} = -\frac{\alpha}{\langle\vec{b},\vec{c}\rangle},$$

obtemos

$$\alpha = -\sigma\langle\vec{b},\vec{c}\rangle, \quad \beta = \sigma\langle\vec{a},\vec{c}\rangle.$$

Logo,

$$\vec{R} = \sigma\left(\langle\vec{a},\vec{c}\rangle\,\vec{b} - \langle\vec{b},\vec{c}\rangle\,\vec{a}\right).$$

Para encontrar σ vamos usar a seguinte base ortonormal: o vetor $\vec{k}$ tem o sentido do vetor $[\vec{a},\vec{b}]$, o vetor $\vec{i}$ tem o sentido do vetor $\vec{a}$, e finalmente o vetor $\vec{j}$ é o produto vetorial de vetores $\vec{k}$ e $\vec{i}$. Nesta base temos as representações

$$\vec{a} = a^1\,\vec{i},$$

$$\vec{b} = b^1\,\vec{i} + b^2\,\vec{j},$$

$$\vec{c} = c^1\,\vec{i} + c^2\,\vec{j} + c^3\,\vec{k}.$$

Utilizando a representação do produto vetorial em coordenadas, pelo Teorema 2.2 obtemos (veja a Fig. 2.4)

$$[\vec{a},\vec{b}] = a^1b^2\,\vec{k}$$

e

$$\vec{R} = -a^1b^2c^2\,\vec{i} + a^1b^2c^1\,\vec{j}. \tag{2.3}$$

Por outro lado temos

$$\langle\vec{a},\vec{c}\rangle = a^1c^1 \quad \text{e} \quad \langle\vec{b},\vec{c}\rangle = b^1c^1 + b^2c^2.$$

Logo, obtemos

$$\vec{R} = \sigma(\langle\vec{a},\vec{c}\rangle\,\vec{b} - \langle\vec{b},\vec{c}\rangle\,\vec{a}) = \sigma((b^1\vec{i} + b^2\vec{j})a^1c^1 - a^1\vec{i}(b^1c^1 + b^2c^2))$$
$$= \sigma(-a^1b^2c^2\vec{i} + a^1b^2c^1\vec{j}).$$

Comparando esta última igualdade com a igualdade (2.3), encontramos $\sigma = 1$. Isto implica (2.2).

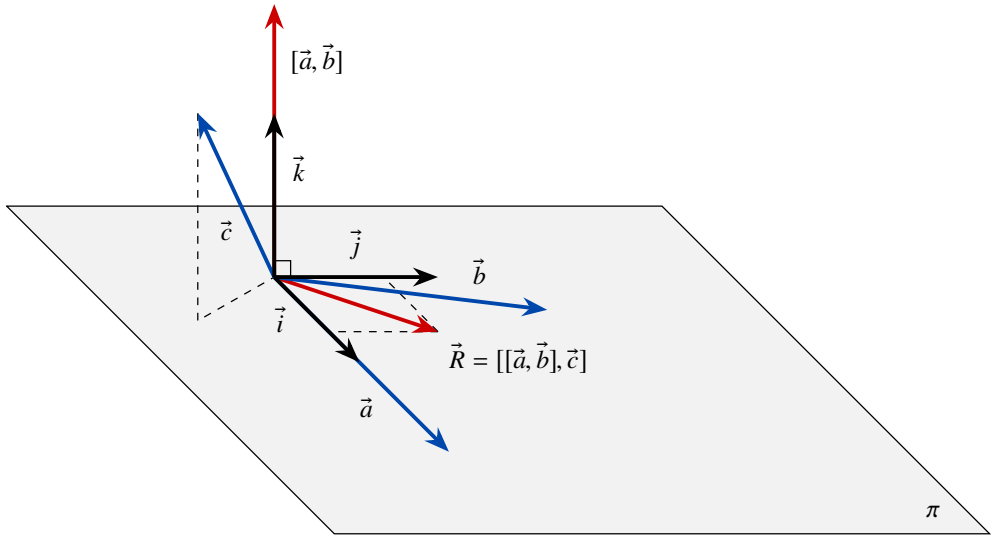


Figura 2.4
Igualdade $[[\vec{a},\vec{b}],\vec{c}] = \langle\vec{a},\vec{c}\rangle\,\vec{b} - \langle\vec{b},\vec{c}\rangle\,\vec{a}$.

■

2.4. Método de Gauss para sistemas lineares com três equações e três incógnitas

Nesta secção vamos representar o Método de Gauss, mas agora para sistemas lineares de três equações e com três incógnitas.

Seja o sistema linear com três equações e com três incógnitas x^1, x^2 e x^3 dado por

$$\begin{cases} ax^1 + bx^2 + cx^3 = d, \\ ex^1 + fx^2 + gx^3 = h, \\ ix^1 + jx^2 + kx^3 = l, \end{cases}$$

em que a, b, c, d, e, f, g, h, i, j, k e l são números reais. O que pretendemos é determinar os números reais $\bar{x}^1$, $\bar{x}^2$ e $\bar{x}^3$ tais que

$$\begin{cases} a\bar{x}^1 + b\bar{x}^2 + c\bar{x}^3 = d, \\ e\bar{x}^1 + f\bar{x}^2 + g\bar{x}^3 = h, \\ i\bar{x}^1 + j\bar{x}^2 + k\bar{x}^3 = l. \end{cases}$$

Se existirem, dizemos que o triplo ordenado $(\bar{x}^1, \bar{x}^2, \bar{x}^3)$ é uma solução do sistema linear. Novamente podemos obter sistemas de equações lineares que só têm uma solução, sistemas de equações lineares que têm uma infinidade de soluções e sistemas de equações lineares que não têm solução. Voltamos a dizer que o sistema é homogéneo se $d = h = l = 0$.

Vamos agora ilustrar a aplicação do Método de Gauss a sistemas lineares com três equações e com três incógnitas considerando os três tipos de sistemas de equações lineares.

Comecemos pelo sistema linear

$$\begin{cases} 2x^3 = 2, \\ 2x^1 + x^2 - x^3 = 1, \\ 3x^1 + 5x^2 + x^3 = 4. \end{cases}$$

É imediato concluir que $x^3 = 1$ e que com este valor o sistema linear dado reduz-se a um sistema linear de duas equações a duas incógnitas. Vamos, no entanto, aplicar o Método de Gauss. Encontramos, no entanto, um obstáculo ao procedimento habitual, que é começar por anular os coeficientes da incógnita x^1 das segunda e terceira equações, pois o facto de o coeficiente da incógnita x^1 da primeira equação ser 0 impossibilita esta tarefa. A solução, no entanto, é simples: devemos trocar linhas por forma a que o coeficiente da incógnita x^1 da nova primeira equação seja diferente de 0. Neste caso podíamos trocar a linha 1 com a linha 2 ou a linha 1 com a linha 3. Optamos, sem nenhum motivo especial, pela primeira hipótese, vindo

$$\begin{cases} 2x^1 + x^2 - x^3 = 1, \\ 2x^3 = 2, \\ 3x^1 + 5x^2 + x^3 = 4. \end{cases}$$

Reparemos que um dos coeficientes que se tem que anular (o coeficiente da incógnita x^1 da nova segunda equação), já é nulo. Basta, pois, anular o coeficiente da incógnita x^1 da terceira equação. Para tal, se subtrairmos à terceira equação, a equação que se obtém multiplicando a primeira por $\frac{3}{2}$, obtemos

$$\begin{cases} 2x^1 + x^2 - x^3 = 1, \\ 2x^3 = 2, \\ \frac{7}{2}x^2 + \frac{5}{2}x^3 = \frac{5}{2}. \end{cases}$$

O problema que detetamos no início da aplicação do Método de Gauss a este exemplo, volta a surgir. A solução é a mesma: trocar a segunda equação pela terceira equação (notamos que aqui é a única hipótese, pois trocar a segunda equação pela primeira equação seria andar para trás no Método de Gauss, dado que o coeficiente da incógnita da nova terceira equação seria não nulo, o que não se pretende). Temos, então

$$\begin{cases} 2x^1 + x^2 - x^3 = 1, \\ \frac{7}{2}x^2 + \frac{5}{2}x^3 = \frac{5}{2}, \\ 2x^3 = 2. \end{cases}$$

Como o coeficiente da incógnita x^2 já é 0 na terceira equação, a primeira fase do Método de Gauss está terminada. Passemos agora à segunda fase do Método de Gauss. A partir da última equação temos que $x^3 = 1$. Substituindo este valor na equação do meio, temos que $x^2 = 0$. Substituindo estes valores na primeira equação, temos que $x^1 = 1$. Este é um exemplo de um sistema possível e determinado.

Consideremos, agora, o sistema linear

$$\begin{cases} x^1 + x^2 + 3x^3 = 2, \\ 3x^1 + 4x^2 + 2x^3 = 3, \\ 2x^1 + 3x^2 - x^3 = 1. \end{cases}$$

Para anular os coeficientes da incógnita x^1 das segunda e terceira equações, subtraímos à segunda equação o triplo da primeira equação e subtraímos à terceira equação o dobro da primeira equação, vindo

$$\begin{cases} x^1 + x^2 + 3x^3 = 2, \\ x^2 - 7x^3 = -3, \\ x^2 - 7x^3 = -3. \end{cases}$$

Para anular o coeficiente da incógnita x^2 da terceira equação, subtraímos à terceira equação a segunda equação, vindo

$$\begin{cases} x^1 + x^2 + 3x^3 = 2, \\ x^2 - 7x^3 = -3, \\ 0x^2 + 0x^3 = 0, \end{cases}$$

o que implica que a última equação é sempre satisfeita quaisquer que sejam os valores de x^2 e x^3. Assim, podemos escrever a solução em função da incógnita x^3, que

por este motivo se diz uma incógnita livre e que vamos substituir pelo parâmetro real $x^3 = \alpha \in \mathbb{R}$. Aplicamos agora a segunda fase do Método de Gauss: a partir da segunda equação temos que $x^2 = 7\alpha - 3$ e a partir da primeira equação temos que $x^1 = 5 - 10\alpha$. Este é um exemplo de um sistema possível e indeterminado.

Consideremos, finalmente, o sistema linear

$$\begin{cases} -x^1 \quad\quad + x^3 = 1, \\ x^1 - 2x^2 \quad\quad = 0, \\ 2x^1 - 2x^2 - x^3 = 2. \end{cases}$$

Para anular os coeficientes da incógnita x^1 das segunda e terceira equações, substituímos a segunda equação pela sua soma com a primeira equação e substituímos a terceira equação pela sua soma com o dobro da primeira equação, vindo

$$\begin{cases} -x^1 \quad\quad + x^3 = 1, \\ \quad -2x^2 + x^3 = 1, \\ \quad -2x^2 + x^3 = 4. \end{cases}$$

Para anular o coeficiente da incógnita x^2 da nova terceira equação, substituímos a segunda equação subtraindo-a à nova segunda equação, vindo

$$\begin{cases} -x^1 \quad\quad + x^3 = 1, \\ \quad -2x^2 + x^3 = 1, \\ \quad\quad 0x^2 + 0x^3 = 3. \end{cases}$$

Concluímos, então, que o sistema linear é impossível.

Consideremos um exemplo que ilustra como o Método de Gauss ajuda na resolução de problemas de cálculo vetorial envolvendo coordenadas.

Exemplo 2.2. Consideremos os vetores $\vec{a} = (2, 1, 2)$, $\vec{b} = (3, 2, -1)$, $\vec{c} = (3, -2, 1)$ e $\vec{d} = (2, -7, 6)$ numa base canónica. Vamos verificar que os vetores $\vec{a}$, $\vec{b}$ e $\vec{c}$ formam uma base e determinar as coordenadas do vetor $\vec{d}$ nessa base.

Resolvendo o sistema

$$\begin{cases} 2x^1 + 3x^2 + 3x^3 = 2, \\ x^1 + 2x^2 - 2x^3 = -7, \\ 2x^1 - x^2 + x^3 = 6, \end{cases}$$

utilizando o Método de Gauss, vemos que existe uma única solução $x^1 = 1$, $x^2 = -2$, $x^3 = 2$. Portanto, os vetores $\vec{a}$, $\vec{b}$ e $\vec{c}$ formam uma base e o vetor $\vec{d}$ nesta base tem coordenadas $(1, -2, 2)$.

□

2.5. Equação geral de um plano

Seja $\vec{n}$ um vetor não nulo de $\mathbb{V}^3$. Então, é possível descrever um plano no espaço $\mathbb{V}^3$ que atravessa a origem como o conjunto de vetores ortogonais a $\vec{n}$, ou seja,

$$\{\vec{v} \in \mathbb{V}^3 : \langle \vec{v}, \vec{n} \rangle = 0\}.$$

Seja A um ponto. Designemos por $\vec{v}_0 \in \mathbb{V}^3$ o vetor $\overrightarrow{OA}$. Então, o plano que atravessa o ponto A é o conjunto

$$\{\vec{v} \in \mathbb{V}^3 : \langle \vec{v} - \vec{v}_0, \vec{n} \rangle = 0\}.$$

Portanto, a equação geral de um plano é

$$\{\vec{v} \in \mathbb{V}^3 : \langle \vec{v}, \vec{n} \rangle = h\},$$

onde h é um número real. Numa base canónica, considerando $\vec{v} = (x^1, x^2, x^3)$ e $\vec{n} = (n^1, n^2, n^3)$, temos

$$n^1 x^1 + n^2 x^2 + n^3 x^3 = h.$$

Vamos ver como se escreve a equação de um plano num sistema de coordenadas qualquer.

Consideremos bases $\mathcal{B} = \{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ e $\mathcal{B}' = \{\vec{v}'_1, \vec{v}'_2, \vec{v}'_3\}$ em $\mathbb{V}^3$ e os sistemas de coordenadas $(O, \mathcal{B})$ e $(O', \mathcal{B}')$. Sejam, ainda, (x^1, x^2, x^3) as coordenadas de um vetor $\vec{v}$ no sistema de coordenadas $(O, \mathcal{B})$. Atendendo à mudança de coordenadas, as coordenadas do mesmo vetor em $(O', \mathcal{B}')$ são

$$\begin{aligned}
(x^1)' &= a_1^1 x^1 + a_1^2 x^2 + a_1^3 x^3 + a_1^0, \\
(x^2)' &= a_2^1 x^1 + a_2^2 x^2 + a_2^3 x^3 + a_2^0, \\
(x^3)' &= a_3^1 x^1 + a_3^2 x^2 + a_3^3 x^3 + a_3^0.
\end{aligned}$$

A partir destas fórmulas, vamos deduzir a equação geral de um plano em $\mathbb{V}^3$. Consideremos um plano em $\mathbb{V}^3$. Escolhemos neste plano um ponto O' e dois vetores $\vec{v}'_1$ e $\vec{v}'_2$ linearmente independentes com origem em O'. Os vetores $\vec{v}'_1$, $\vec{v}'_2$ e $\vec{v}'_3 = [\vec{v}'_1, \vec{v}'_2]$ com origem em O' formam o sistema de coordenadas $(O', \mathcal{B}')$. Neste sistema de coordenadas a equação desse plano tem a forma $x'_3 = 0$. Logo, em $(O, \mathcal{B})$, temos $a_3^1 x^1 + a_3^2 x^2 + a_3^3 x^3 + a_3^0 = 0$. Portanto, em qualquer base a equação do plano é $Ax^1 + Bx^2 + Cx^3 = D$.

Tal como em $\mathbb{V}^2$, também no espaço $\mathbb{V}^3$ existe uma interpretação geométrica muito simples para os sistemas lineares com três equações e três incógnitas: as três equações definem três planos do espaço, sendo a solução do sistema linear formado pelos pontos comuns destes três planos. Existem oito situações possíveis:

1. três planos que possuem um e apenas um ponto em comum, cortando-se dois a dois segundo três retas concorrentes — sistema possível e determinado;

2. três planos paralelos coincidentes — sistema possível e indeterminado;

3. três planos secantes, isto é, com uma reta comum — sistema possível e indeterminado;

4. dois planos paralelos coincidentes secantes a um terceiro plano — sistema possível e indeterminado;

5. três planos estritamente paralelos — sistema impossível;

6. dois planos coincidentes estritamente paralelos a um terceiro plano — sistema impossível;

7. dois planos estritamente paralelos e um terceiro secante aos dois — sistema impossível;

8. três planos secantes dois a dois segundo retas paralelas — sistema impossível.

3. Espaços lineares de dimensão finita

3.1. Introdução

Nos espaços $\mathbb{V}^2$ e $\mathbb{V}^3$ estudados nos capítulos anteriores, definimos os vetores como classes de equivalência de segmentos orientados. Estes objetos são muito cómodos para descrever sistemas físicos. Por exemplo, considerando as velocidades de moléculas de ar, convém colocar os pontos iniciais dos vetores das velocidades nos centros das moléculas. Nestes espaços podemos escolher diferentes sistemas de coordenadas em função do problema que estamos a resolver. Os espaços $\mathbb{V}^2$ e $\mathbb{V}^3$ são dois exemplos de espaços afins, que vamos estudar na sua generalidade no Capítulo 8.

Neste capítulo vamos considerar objetos mais simples, conhecidos como espaços lineares ou espaços vetoriais. Nestes espaços fixamos uma única origem e de todas as classes de segmentos orientados consideramos só representantes destas classes que têm ponto inicial na origem. Desta maneira podemos estabelecer uma correspondência biunívoca entre os vetores e os pontos do espaço. Mais precisamente, fazemos corresponder a um ponto do espaço P o vetor $\overrightarrow{OP}$ que tem ponto inicial na origem e ponto final em P. Aos elementos de um espaço linear vamos chamar vetores ou pontos dependendo do contexto. Vamos, por exemplo, dizer "soma de dois vetores", mas, por outro lado, diremos "função contínua num ponto".

Nos espaços lineares estão definidas as operações de adição de vetores e multiplicação de vetores por um número e no Capítulo 5 vamos desenvolver a teoria de aplicações lineares definidas nos espaços vetoriais, que tem um papel importante na resolução de muitos problemas e que é o objeto principal de estudo deste livro.

3.2. Espaços lineares: definição e exemplos

Num espaço linear é possível adicionar os vetores e multiplicá-los por um elemento de um conjunto de números $\mathbb{K}$, que pode ser o conjunto dos números reais $\mathbb{R}$ ou o conjunto dos números complexos $\mathbb{C}$. Dizemos que espaço linear é *real* se $\mathbb{K} = \mathbb{R}$ e *complexo* se $\mathbb{K} = \mathbb{C}$. No estudo dos espaços lineares, a natureza dos vetores e do conjunto de números $\mathbb{K}$ não é importante. As operações nesta teoria estudamos de modo abstrato. Os elementos de um espaço linear podem ser, por exemplo, polinómios algébricos ou trigonométricos ou outras funções. Esta abstração dá a possibilidade de aplicar os resultados da teoria ao estudo de vários tipos de problemas. Vamos então começar por dar a definição de espaço linear.

Definição 3.1. Espaço linear

Seja $\mathbb{K}$ um conjunto de números. Dizemos que um conjunto $\mathbb{X}$ é um *espaço linear* sobre $\mathbb{K}$, se em $\mathbb{X}$ estão definidas as operações:

1. de adição, que faz corresponder a cada par $(x, y) \in \mathbb{X} \times \mathbb{X}$ um vetor $z = x + y \in \mathbb{X}$, e

continua na página seguinte

continuação da página anterior

2. de multiplicação por números $\lambda \in \mathbb{K}$, que faz corresponder ao par $(x, \lambda) \in \mathbb{X} \times \mathbb{K}$ um vetor $z = \lambda x \in \mathbb{X}$,

e estas operações têm as seguintes propriedades:

1. $x + y = y + x$, para quaisquer $x, y \in \mathbb{X}$;
2. $(x + y) + z = x + (y + z)$, para quaisquer $x, y, z \in \mathbb{X}$;
3. existe zero, isto é, um elemento $0 \in \mathbb{X}$ tal que $x + 0 = x$, para qualquer $x \in \mathbb{X}$;
4. existe elemento inverso, isto é, um $y \in \mathbb{X}$ tal que $x + y = 0$, para todo o $x \in \mathbb{X}$;
5. $1x = x$, para qualquer $x \in \mathbb{X}$;
6. $\alpha(\beta x) = (\alpha\beta)x$, para qualquer $x \in \mathbb{X}$ e para quaisquer $\alpha, \beta \in \mathbb{K}$;
7. $(\alpha + \beta)x = \alpha x + \beta x$, para qualquer $x \in \mathbb{X}$ e para quaisquer $\alpha, \beta \in \mathbb{K}$;
8. $\alpha(x + y) = \alpha x + \alpha y$, para quaisquer $x, y \in \mathbb{X}$ e $\alpha \in \mathbb{K}$.

Esta definição implica as seguintes propriedades.

Teorema 3.1

Seja $\mathbb{X}$ um espaço linear. Então:

1. existe só um zero;
2. para todo o $x \in \mathbb{X}$ existe só um elemento inverso;
3. $0x = 0$ (notemos que o caractere "0" do lado esquerdo da igualdade representa o número zero enquanto que o caractere "0" do lado direito da igualdade representa o elemento zero do espaço linear $\mathbb{X}$);
4. o inverso de $x \in \mathbb{X}$ é $y = (-1)x \in \mathbb{X}$.

Demonstração.

1. Sejam $0_1 \in \mathbb{X}$ e $0_2 \in \mathbb{X}$ dois zeros. Então temos $0_1 + 0_2 = 0_1$ e $0_2 + 0_1 = 0_2$. Logo, $0_1 = 0_2$.

2. Sejam y_1 e y_2 elementos inversos de $x \in \mathbb{X}$. Então, é fácil ver que $y_2 + (x + y_1) = y_2 + 0 = y_2$. Por outro lado, temos $y_2 + (x + y_1) = (y_2 + x) + y_1 = 0 + y_1 = y_1$. Portanto, concluímos que $y_1 = y_2$.

3. Com efeito, notemos que $0x + x = 0x + 1x = (0 + 1)x = 1x = x$. Seja y o elemento inverso de x. Então temos $0x = 0x + 0 = 0x + x + y = x + y = 0$.

4. Utilizando a propriedade 3, temos $x+(-1)x = 1x+(-1)x = (1-1)x = 0x = 0$.
■

Vamos designar o elemento inverso de x por $-x$. Do Teorema 3.1 temos $-x = (-1)x$. Definimos a diferença de x e y como a soma de x e $-y$, ou seja, $x - y = x + (-y)$.

Consideremos alguns exemplos de espaços lineares.

Exemplo 3.1. Espaço real das sequências ordenadas com n componentes reais:

$$\mathbb{R}^n = \{x = (x^1, x^2, \ldots, x^n) : x^k \in \mathbb{R}, k = \overline{1,n}\}.$$

Sejam $x = (x^1, x^2, \ldots, x^n), y = (y^1, y^2, \ldots, y^n) \in \mathbb{R}^n$ e $\lambda \in \mathbb{R}$. Definimos a adição como

$$x + y = (x^1 + y^1, x^2 + y^2, \ldots, x^n + y^n)$$

e a multiplicação por um número real como

$$\lambda x = (\lambda x^1, \lambda x^2, \ldots, \lambda x^n).$$

□

Exemplo 3.2. Espaço complexo das sequências ordenadas com n componentes complexas:

$$\mathbb{C}^n = \{z = (z^1, z^2, \ldots, z^n) : z^k \in \mathbb{C}, k = \overline{1,n}\}.$$

Sejam $z = (z^1, z^2, \ldots, z^n), w = (w^1, w^2, \ldots, w^n) \in \mathbb{C}^n$ e $\zeta \in \mathbb{C}$. Definimos a adição e a multiplicação por um número como no exemplo anterior.

□

Exemplo 3.3. Espaço de polinómios de grau menor ou igual a n com coeficientes reais:

$$\mathcal{P}^n = \{p(x) = a_0x^n + a_1x^{n-1} + \cdots + a_n : a_i \in \mathbb{R}, i = \overline{0,n}\}.$$

Sejam $p(x) = a_0x^n + a_1x^{n-1} + \cdots + a_n, q(x) = b_0x^n + b_1x^{n-1} + \cdots + b_n \in \mathcal{P}^n$ e $\alpha \in \mathbb{R}$. Definimos a adição como

$$(p + q)(x) = p(x) + q(x) = (a_0 + b_0)x^n + (a_1 + b_1)x^{n-1} + \cdots + (a_n + b_n)$$

e a multiplicação por um número real como

$$(\alpha p)(x) = \alpha p(x) = (\alpha a_0)x^n + (\alpha a_1)x^{n-1} + \cdots + (\alpha a_n).$$

□

Exemplo 3.4. Espaço de polinómios de grau menor ou igual a n com coeficientes complexos, que representamos utilizando o mesmo símbolo:

$$\mathcal{P}^n = \{p(x) = a_0x^n + a_1x^{n-1} + \cdots + a_n : a_i \in \mathbb{C}, i = \overline{0,n}\}.$$

Aqui, definimos a adição e a multiplicação por um número complexo como no exemplo anterior.

□

3.3. Método de Gauss para sistemas lineares com m equações e n incógnitas

O Método de Gauss pode ser generalizado para resolver agora sistemas lineares com m equações e com n incógnitas. Consideremos o sistema linear com m equações e com n incógnitas $x^1, \ldots, x^n$ dado por

$$\begin{cases} a_1^1 x^1 + a_1^2 x^2 + a_1^3 x^3 + \cdots + a_1^n x^n = b_1, \\ a_2^1 x^1 + a_2^2 x^2 + a_2^3 x^3 + \cdots + a_2^n x^n = b_2, \\ \vdots \quad\quad \vdots \quad\quad \vdots \quad\quad \ddots \quad\quad \vdots \quad\quad \vdots \\ a_m^1 x^1 + a_m^2 x^2 + a_m^3 x^3 + \cdots + a_m^n x^n = b_m, \end{cases}$$

em que todos os coeficientes $a_1^1, \ldots, a_m^n$ e todos os termos independentes $b_1, \ldots, b_m$ são números reais (complexos). O que pretendemos é determinar os números reais (complexos) $\bar{x}^1, \ldots, \bar{x}^n$ tais que

$$\begin{cases} a_1^1 \bar{x}^1 + a_1^2 \bar{x}^2 + a_1^3 \bar{x}^3 + \cdots + a_1^n \bar{x}^n = b_1, \\ a_2^1 \bar{x}^1 + a_2^2 \bar{x}^2 + a_2^3 \bar{x}^3 + \cdots + a_2^n \bar{x}^n = b_2, \\ \vdots \quad\quad \vdots \quad\quad \vdots \quad\quad \ddots \quad\quad \vdots \quad\quad \vdots \\ a_m^1 \bar{x}^1 + a_m^2 \bar{x}^2 + a_m^3 \bar{x}^3 + \cdots + a_m^n \bar{x}^n = b_m. \end{cases}$$

Novamente podemos obter sistemas de equações lineares que só têm uma solução, sistemas de equações lineares que têm uma infinidade de soluções e sistemas de equações lineares que não têm solução.

A primeira fase do Método de Gauss consiste em transformar o sistema num sistema linear equivalente mais simples, que tem *forma em escada* ou *forma escalonada*

$$\begin{cases} a_1^{j_1} x^1 + a_1^2 x^2 + \cdots + a_1^{j_2} x^{j_2} + \cdots + a_1^{j_i} x^{j_i} + \cdots + a_1^n x^n = b_1, \\ 0 + 0 + \cdots + a_2^{j_2(1)} x^{j_2} + \cdots + a_2^{j_i(1)} x^{j_i} + \cdots + a_2^{n(1)} x^n = b_2^{(1)}, \\ \vdots \quad \vdots \quad \ddots \quad \vdots \quad \ddots \quad \vdots \quad \ddots \quad \vdots \quad \vdots \\ 0 + 0 + \cdots + 0 + \cdots + a_i^{j_i(i-1)} x^{j_i} + \cdots + a_i^{n(i-1)} x^n = b_i^{(i-1)}, \\ \vdots \quad \vdots \quad \ddots \quad \vdots \quad \ddots \quad \vdots \quad \ddots \quad \vdots \quad \vdots \\ 0 + 0 + \cdots + 0 + \cdots + 0 + \cdots + 0 = 0. \end{cases}$$

(Aqui, $j_1 = 1$.) Para obtermos um sistema desta forma anulamos primeiro os coeficientes da incógnita x^1 da segunda até à m-ésima linha, vindo

$$\begin{cases} a_1^1 x^1 + a_1^2 x^2 + \cdots + a_1^n x^n = b_1, \\ 0 + a_2^{2(1)} x^2 + \cdots + a_2^{n(1)} x^n = b_2^{(1)}, \\ \vdots \quad\quad \vdots \quad\quad\quad\quad \vdots \quad\quad \vdots \\ 0 + a_m^{2(1)} x^2 + \cdots + a_m^{n(1)} x^n = b_m^{(1)}, \end{cases}$$

onde $a_i^{j(1)} = a_i^j - \frac{a_i^1}{a_1^1} a_1^j$ e $b_i^{(1)} = b_i - \frac{a_i^1}{a_1^1} b_1$, para $i = \overline{2, m}$ e $j = \overline{2, n}$. No caso $a_1^1 = 0$ devemos trocar linhas por forma a que o coeficiente da incógnita x^1 da nova

primeira equação seja diferente de zero. Pode acontecer que no novo sistema todos os coeficientes $a_i^{2(1)}$, $i = \overline{2,m}$, são iguais a zero. Neste caso consideramos o sistema que tem só incógnitas $x^3, \ldots, x^n$. Continuando este procedimento chegamos a um sistema equivalente que tem forma escalonada. Os coeficientes não nulos mais à esquerda em cada equação da forma escalonada são designados por *elementos pivô*. As incógnitas correspondentes chamamos incógnitas pivô. As restantes incógnitas são designadas por incógnitas livres. Se como resultado desta transformação do sistema se obtém a igualdade $0 = b_i^{i-1} \neq 0$, o sistema é impossível. Caso contrário o sistema tem soluções. A segunda fase consiste em determinar a sua solução tal como foi feito nos capítulos anteriores. Se o sistema não tem incógnitas livres a solução é única. Quando há incógnitas livres essas são substituídas por parâmetros reais (complexos), obtemos no fim uma infinidade de soluções, portanto, um sistema possível e indeterminado.

No caso de $b_1 = b_2 = \cdots = b_m = 0$, voltamos a dizer que o sistema é homogéneo, que pelo menos admite sempre a solução trivial. Se $n > m$, isto é, o número de incógnitas é superior ao número de equações, o sistema admite soluções não triviais.

O algoritmo do Método de Gauss só envolve os coeficientes e os termos independentes, não sendo necessário estar sempre a escrever as incógnitas do sistema. Estes coeficientes e termos independentes podemos escrever na forma de uma tabela:

$$\begin{pmatrix} a_1^1 & a_1^2 & \cdots & a_1^n & b_1 \\ a_2^1 & a_2^2 & \cdots & a_2^n & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_m^1 & a_m^2 & \cdots & a_m^n & b_m \end{pmatrix}. \tag{3.1}$$

A uma tabela de números

$$\begin{pmatrix} a_1^1 & a_1^2 & \cdots & a_1^n \\ a_2^1 & a_2^2 & \cdots & a_2^n \\ \vdots & \vdots & \ddots & \vdots \\ a_m^1 & a_m^2 & \cdots & a_m^n \end{pmatrix}$$

com m linhas e n colunas damos o nome de *matriz*. Este objeto é muito importante e vamos encontrá-lo ao longo de todo o livro. Vamos usar a notação a_i^j para os elementos de uma matriz. O índice inferior i corresponde ao número da linha e o índice superior j ao número da coluna.

Dizemos que a matriz é *quadrada* se o número de linhas for igual ao número de colunas. Caso contrário, dizemos que a matriz é *retangular*. Por exemplo, a matriz (3.1) tem m linhas e $n+1$ colunas e se $m \neq n+1$ a matriz é retangular. Dizemos que uma matriz é *triangular superior* se é uma matriz quadrada em que todos os elementos abaixo da diagonal (formada pelos elementos $a_1^1, \ldots, a_n^n$) são zeros. Analogamente definimos matriz *triangular inferior*. Dizemos que a matriz é *diagonal* se todos os elementos a_i^j com $i \neq j$ são nulos.

Vamos voltar ao exemplo do sistema linear considerado no capítulo anterior

$$\begin{cases} 2x^3 = 2, \\ 2x^1 + x^2 - x^3 = 1, \\ 3x^1 + 5x^2 + x^3 = 4, \end{cases}$$

e mostrar como a linguagem das matrizes simplifica o funcionamento do Método de Gauss. A matriz dos coeficientes do sistema e dos termos independentes é

$$\begin{pmatrix} 0 & 0 & 2 & 2 \\ 2 & 1 & -1 & 1 \\ 3 & 5 & 1 & 4 \end{pmatrix}.$$

Segundo o Método de Gauss, a primeira operação que temos que efetuar é trocar linhas por forma a colocar um elemento não nulo na primeira linha e primeira coluna. A troca da i-ésima linha com a j-ésima linha vamos denotar por $l_i \leftrightarrow l_j$. Uma das hipóteses é trocar a primeira linha com a segunda, ou seja, efetuar a operação $l_1 \leftrightarrow l_2$, vindo a matriz

$$\begin{pmatrix} 2 & 1 & -1 & 1 \\ 0 & 0 & 2 & 2 \\ 3 & 5 & 1 & 4 \end{pmatrix}.$$

A substituição da i-ésima linha pela soma da j-ésima linha multiplicada por a e da k-ésima linha vamos denotar por $l_i \leftarrow al_j + l_k$. Portanto, a operação $l_3 \leftarrow l_3 - \frac{3}{2}l_1$ dá a matriz

$$\begin{pmatrix} 2 & 1 & -1 & 1 \\ 0 & 0 & 2 & 2 \\ 0 & \frac{7}{2} & \frac{5}{2} & \frac{5}{2} \end{pmatrix}.$$

Temos agora que efetuar a operação $l_2 \leftrightarrow l_3$. Obtemos, então, a matriz

$$\begin{pmatrix} 2 & 1 & -1 & 1 \\ 0 & \frac{7}{2} & \frac{5}{2} & \frac{5}{2} \\ 0 & 0 & 2 & 2 \end{pmatrix},$$

que já está na forma em escada. Podemos concluir que o sistema é possível e determinado, e que a aplicação da segunda fase do Método de Gauss leva à solução do sistema $x^3 = 1$, $x^2 = 0$ e $x^1 = 1$.

Com o Método de Gauss estamos em condições de determinar se o sistema é impossível ou resolver qualquer sistema de equações lineares que tem uma solução ou uma infinidade de soluções. No entanto, existe uma variação deste método, conhecida por Método de Gauss-Jordan (ou Método de eliminação de Gauss-Jordan), que consiste em continuar a transformar a matriz em escada, reduzindo-a a uma matriz bastante mais simples de modo a que no fim a solução pode ser quase lida diretamente a partir da matriz. Nessa transformação, as operações que vamos aplicar às linhas da matriz

correspondem no sistema a divisões de equações por números não nulos e a somas de equações com múltiplos de outras equações. Por isso, o novo sistema linear vai ser equivalente ao sistema inicial no sentido de ter a mesma solução.

Assim, dada a matriz em escada associada ao sistema linear geral considerado anteriormente

$$\begin{pmatrix} a_1^{j_1} & a_1^2 & \cdots & a_1^{j_2} & \cdots & a_1^{j_i} & \cdots & a_1^n & b_1 \\ 0 & 0 & \cdots & a_2^{j_2(1)} & \cdots & a_2^{j_i(1)} & \cdots & a_2^{n(1)} & b_2^{(1)} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & a_i^{j_i(i-1)} & \cdots & a_i^{n(i-1)} & b_i^{(i-1)} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & 0 \end{pmatrix},$$

se dividirmos a última linha não nula da matriz pelo elemento pivô dessa linha, $a_i^{j_i(i-1)}$, obtemos

$$\begin{pmatrix} a_1^{j_1} & a_1^2 & \cdots & a_1^{j_2} & \cdots & a_1^{j_i} & \cdots & a_1^n & b_1 \\ 0 & 0 & \cdots & a_2^{j_2(1)} & \cdots & a_2^{j_i(1)} & \cdots & a_2^{n(1)} & b_2^{(1)} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 1 & \cdots & a_i^{n(i)} & b_i^{(i)} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & 0 \end{pmatrix},$$

onde todos os elementos da linha i são o resultado da divisão por $a_i^{j_i(i-1)}$. Temos, por exemplo,

$$a_i^{n(i)} = \frac{a_i^{n(i-1)}}{a_i^{j_i(i-1)}}, \quad b_i^{(i)} = \frac{b_i^{(i-1)}}{a_i^{j_i(i-1)}}.$$

Se anularmos agora os elementos que estão na mesma coluna acima do pivô, ou seja, os elementos, $a_h^{j_i(h-1)}$, $h = \overline{1, i-1}$ (onde $a_1^{j_i(0)} = a_1^{j_i}$), com a operação $l_h \leftarrow l_h - a_h^{j_i(h-1)} l_i$, $h = \overline{0, i-1}$, ou seja, subtraindo a cada uma dessas linhas a linha i multiplicada pelo elemento correspondente da linha por cima do pivô, obtemos

$$\begin{pmatrix} a_1^{j_1} & a_1^2 & \cdots & a_1^{j_2} & \cdots & 0 & \cdots & a_1^{n(i+1)} & b_1^{(i+1)} \\ 0 & 0 & \cdots & a_2^{j_2(1)} & \cdots & 0 & \cdots & a_2^{n(i+1)} & b_2^{(i+1)} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 1 & \cdots & a_i^{n(i)} & b_i^{(i)} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & 0 \end{pmatrix},$$

onde

$$a_h^{n(i+1)} = a_h^{n(h-1)} - a_i^{n(i)} a_h^{j_i(h-1)}, \quad b_h^{(i+1)} = b_h^{(h-1)} - b_i^{(i)} a_h^{j_i(h-1)}$$

para $h = \overline{1, i-1}$ (com $b_1^{(0)} = b_1$). Vemos que nesta última matriz, o pivô da última linha é 1 e na coluna correspondente os restantes elementos são todos nulos.

Se aplicarmos sucessivamente este processo (divisão de uma linha pelo pivô e anulamento dos elementos que estão por cima do pivô) a cada linha, subindo na matriz até alcançarmos a primeira linha, obtemos a matriz mais simples, que designamos por *matriz em escada reduzida*,

$$\begin{pmatrix} 1 & a_1^{2(2i-1)} & \cdots & 0 & \cdots & 0 & \cdots & a_1^{n(2i-1)} & b_1^{(2i-1)} \\ 0 & 0 & \cdots & 1 & \cdots & 0 & \cdots & a_2^{n(2i-2)} & b_2^{(2i-2)} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 1 & \cdots & a_i^{n(i)} & b_i^{(i)} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & 0 \end{pmatrix}.$$

A solução pode ser determinada a partir desta matriz tal como no Método de Gauss. A diferença é que agora é mais fácil resolver o sistema e, em certos casos, como vamos ver no exemplo apresentado a seguir, a solução pode ser lida a partir da última coluna.

Para ilustrar a aplicação do Método de Gauss-Jordan, vamos considerar a matriz em escada obtida anteriormente

$$\begin{pmatrix} 2 & 1 & -1 & 1 \\ 0 & \frac{7}{2} & \frac{5}{2} & \frac{5}{2} \\ 0 & 0 & 2 & 2 \end{pmatrix}.$$

Dividindo agora a última linha pelo seu pivô, 2, usando a operação $l_3 \leftarrow \frac{1}{2} l_3$, e anulando de seguida os elementos que estão por cima desse pivô na mesma coluna usando as operações $l_2 \leftarrow l_2 - \frac{5}{2} l_3$ e $l_1 \leftarrow l_1 + l_3$, obtemos

$$\begin{pmatrix} 2 & 1 & 0 & 2 \\ 0 & \frac{7}{2} & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}.$$

Subindo agora para a segunda linha, dividindo esta pelo seu pivô, $\frac{7}{2}$, usando a operação $l_2 \leftarrow \frac{2}{7} l_2$, e anulando depois o elemento que está por cima deste pivô na primeira linha com a operação $l_1 \leftarrow l_1 - l_2$, obtemos a matriz

$$\begin{pmatrix} 2 & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}.$$

Finalmente, dividindo a primeira linha da matriz pelo seu pivô, $l_1 \leftarrow \frac{1}{2} l_1$, transformamos a matriz na forma em escada reduzida

$$\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix},$$

em que a solução $x^1 = 1$, $x^2 = 0$ e $x^3 = 1$ está na última coluna.

Relativamente ao Método de Gauss, o Método de Gauss-Jordan nem tem vantagens computacionais, uma vez que o número de operações aritméticas é normalmente maior, nem teóricas. Porém, nas resoluções manuais de exercícios tem, na nossa opinião, maior elegância e clareza, pois nos casos dos sistemas possíveis e determinados os valores das incógnitas aparecem explicitamente na última coluna da matriz final.

3.4. Combinação linear, base, coordenadas e dimensão

Vamos agora introduzir várias definições importantes relativas a conjuntos de vetores nos espaços lineares.

Definição 3.2. Combinação linear

Sejam $\mathbb{X}$ um espaço linear sobre um conjunto de números $\mathbb{K}$, x_j, $j = \overline{1,m}$, vetores de $\mathbb{X}$ e α^j, $j = \overline{1,m}$, números de $\mathbb{K}$. Dizemos que o vetor

$$y = \sum_{j=1}^{m} \alpha^j x_j$$

é uma *combinação linear* dos vetores x_j, $j = \overline{1,m}$.
O conjunto de todas as combinações lineares destes vetores vamos designar por $\text{Lin}\{x_1, x_2, \ldots, x_m\}$.

Definição 3.3. Independência linear e dependência linear

Sejam $\mathbb{X}$ um espaço linear sobre um conjunto de números $\mathbb{K}$, x_j, $j = \overline{1,m}$, vetores de $\mathbb{X}$ e α^j, $j = \overline{1,m}$, números de $\mathbb{K}$. Dizemos que os vetores x_j, $j = \overline{1,m}$, são *linearmente independentes* se a igualdade

$$\sum_{j=1}^{m} \alpha^j x_j = 0$$

implica $\alpha^j = 0$, $j = \overline{1,m}$. Caso contrário, ou seja, se é possível representar zero na forma de uma combinação linear dos vetores x_j, $j = \overline{1,m}$, onde nem todos os coeficientes nesta combinação são zeros, dizemos que os vetores são *linearmente dependentes*.

Consideremos dois exemplos.

Exemplo 3.5. Os vetores $e_1 = (1, 0, \ldots, 0)$, $e_2 = (0, 1, \ldots, 0)$, ..., e $e_n = (0, 0, \ldots, 1)$ em $\mathbb{R}^n$ são linearmente independentes. □

Exemplo 3.6. Os vetores $v_1 = (2,0,0)$, $v_2 = (4,0,0)$ e $v_3 = (0,0,3)$ em $\mathbb{R}^3$ são linearmente dependentes. De facto, $v_1 - \frac{1}{2}v_2 + 0v_3 = 0$. □

Notemos que, se alguns dos vetores x_j, $j = \overline{1,m}$, são linearmente dependentes, então todos estes vetores são linearmente dependentes e se os m vetores são linearmente dependentes, então pelo menos um destes vetores pode ser representado como uma combinação linear dos restantes.

Vamos introduzir a noção fundamental de base num espaço linear.

Definição 3.4. Base num espaço linear e coordenadas de um vetor numa base

Seja $\mathbb{X}$ um espaço linear sobre um conjunto de números $\mathbb{K}$. Consideremos os vetores linearmente independentes $e_1, e_2, \ldots, e_n \in \mathbb{X}$. A este conjunto damos o nome de *base* se para todo $x \in \mathbb{X}$ existem $\xi^j \in \mathbb{K}$, $j = \overline{1,n}$, tais que

$$x = \sum_{j=1}^{n} \xi^j e_j.$$

Aos números $\xi^j \in \mathbb{K}$, $j = \overline{1,n}$, chamamos *coordenadas* do vetor x na base $\mathcal{B} = \{e_1, e_2, \ldots, e_n\}$ e escrevemos $x_{\mathcal{B}} = (\xi^1, \ldots, \xi^n)$.

Consideremos dois exemplos.

Exemplo 3.7. Os vetores $e_1 = (1,0,\ldots,0)$, $e_2 = (0,1,\ldots,0)$, …, e $e_n = (0,0,\ldots,1)$ formam uma base em $\mathbb{R}^n$. A esta base chamamos *base canónica em* $\mathbb{R}^n$. Por defeito, é esta base que consideramos em $\mathbb{R}^n$. □

Exemplo 3.8. O conjunto dos monómios x^k, $k = \overline{0,n}$, forma uma base no espaço de polinómios $\mathcal{P}^n$. □

A pertinência do método das coordenadas fica explícita no seguinte teorema.

Teorema 3.2

As coordenadas de um qualquer elemento de um espaço linear associadas a uma dada base estão definidas de modo único.

Demonstração. Sejam $\mathbb{X}$ um espaço linear sobre um conjunto de números $\mathbb{K}$ e $\{e_1, e_2, \ldots, e_n\}$ uma base em $\mathbb{X}$ e $x \in \mathbb{X}$. Suponhamos que ξ^j e η^j, $j = \overline{1,n}$, são coordenadas de x na base dada. Então temos

$$x = \sum_{j=1}^{n} \xi^j e_j = \sum_{j=1}^{n} \eta^j e_j,$$

pelo que

$$0 = \sum_{j=1}^{n} (\xi^j - \eta^j) e_j.$$

Como os vetores e_j, $j = \overline{1,n}$, são linearmente independentes, temos $\xi^j = \eta^j$, $j = \overline{1,n}$, ou seja, as coordenadas estão definidas de modo único como queríamos mostrar. ■

É fácil ver que as coordenadas de uma soma de vetores são as somas das coordenadas e as coordenadas de um produto de um número e um vetor são os resultados da multiplicação das coordenadas pelo número. De facto, seja $\{e_1, e_2, \ldots, e_n\}$ uma base em $\mathbb{X}$. Consideremos $\alpha \in \mathbb{K}$ e $x, y \in \mathbb{X}$ representados por

$$x = \sum_{j=1}^{n} \xi^j e_j, \quad y = \sum_{j=1}^{n} \eta^j e_j.$$

Então,

$$x + y = \sum_{j=1}^{n} \xi^j e_j + \sum_{j=1}^{n} \eta^j e_j = \sum_{j=1}^{n} (\xi^j + \eta^j) e_j$$

e

$$\alpha x = \alpha \sum_{j=1}^{n} \xi^j e_j = \sum_{j=1}^{n} (\alpha \xi^j) e_j.$$

Consideremos o seguinte exemplo de cálculo com coordenadas.

Exemplo 3.9. Sejam $x = (2,2)$, $y = (1,2)$ e $\alpha = -2$. Consideremos a base $\mathcal{B} = \{(1,0),(1,1)\}$ em $\mathbb{R}^2$. Como $x = 0(1,0) + 2(1,1)$, então $x_{\mathcal{B}} = (0,2)$ e como $y = -1(1,0) + 2(1,1)$, então $y_{\mathcal{B}} = (-1,2)$. A soma dos vetores $x + y = (3,4)$ tem coordenadas $(x+y)_{\mathcal{B}} = (-1,4)$. De facto, $(3,4) = -1(1,0) + 4(1,1)$. O produto de y por α, $\alpha y = (-2,-4)$, tem coordenadas $(\alpha y)_{\mathcal{B}} = (2,-4)$. De facto, $(-2,-4) = 2(1,0) - 4(1,1)$. □

Para resolver exercícios deste tipo pode ser muito útil o Método de Gauss.

Exemplo 3.10. Vamos estudar a dependência linear dos vetores $(1,2,1)$, $(-1,-1,-3)$ e $(1,3,2)$ através do Método de Gauss.

A dependência linear é equivalente à existência de uma solução não trivial do sistema homogéneo

$$\begin{cases} \alpha^1 - \alpha^2 + \alpha^3 = 0, \\ 2\alpha^1 - \alpha^2 + 3\alpha^3 = 0, \\ \alpha^1 - 3\alpha^2 + 2\alpha^3 = 0. \end{cases}$$

Resolvendo este sistema com ajuda do Método de Gauss encontramos que a única solução do sistema é $\alpha^1 = \alpha^2 = \alpha^3 = 0$. Portanto, os vetores são linearmente independentes. □

Exemplo 3.11. Vamos verificar que os vetores $(1,-2,1)$, $(1,-1,3)$, $(1,-2,3)$ formam uma base e depois determinar as coordenadas do vetor $(1,-4,1)$ nesta base.

Resolvendo o sistema

$$\begin{cases} \alpha^1 + \alpha^2 + \alpha^3 = 0, \\ -2\alpha^1 - \alpha^2 - 2\alpha^3 = 0, \\ \alpha^1 + 3\alpha^2 + 3\alpha^3 = 0, \end{cases}$$

através do Método de Gauss encontramos que a única solução do sistema é $\alpha^1 = \alpha^2 = \alpha^3 = 0$. Portanto, os vetores são linearmente independentes. Logo, formam uma base no espaço de dimensão três. Agora, resolvendo o sistema

$$\begin{cases} \alpha^1 + \alpha^2 + \alpha^3 = 1, \\ -2\alpha^1 - \alpha^2 - 2\alpha^3 = -4, \\ \alpha^1 + 3\alpha^2 + 3\alpha^3 = 1, \end{cases}$$

encontramos as coordenadas do vetor $(1, -4, 1)$: $\alpha^1 = 1$, $\alpha^2 = -2$, $\alpha^3 = 2$. □

O seguinte teorema ajuda a esclarecer o conceito de dimensão de um espaço linear.

Teorema 3.3

Seja $\mathbb{X}$ um espaço linear sobre um conjunto de números $\mathbb{K}$ tal que admite uma base com n elementos. Então, cada $n+1$ vetores de $\mathbb{X}$ são linearmente dependentes e cada conjunto de n vetores linearmente independentes forma uma base.

Demonstração. Suponhamos que os vetores e_j, $j = \overline{1,n}$, formam uma base em $\mathbb{X}$. Consideremos um conjunto de $n+1$ vetores $\{x_1, x_2, \ldots, x_{n+1}\} \subset \mathbb{X}$. Cada um destes vetores pode ser representado na forma

$$x_k = \sum_{j=1}^{n} \xi_k^j e_j, k = \overline{1, n+1}.$$

Consideremos o sistema de equações lineares

$$\sum_{k=1}^{n+1} \xi_k^j \alpha^k = 0, j = \overline{1,n}.$$

Como o sistema é homogéneo e o número de incógnitas é superior ao número de equações, o Método de Gauss garante a existência de uma solução não trivial, $\alpha^1, \alpha^2, \ldots, \alpha^{n+1}$, deste sistema. É fácil ver que

$$\sum_{k=1}^{n+1} \alpha^k x_k = \sum_{k=1}^{n+1} \alpha^k \sum_{j=1}^{n} \xi_k^j e_j = \sum_{j=1}^{n} \left(\sum_{k=1}^{n+1} \xi_k^j \alpha^k \right) e_j = 0.$$

Portanto, os vetores $x_1, x_2, \ldots, x_{n+1}$ são linearmente dependentes.

Consideremos agora n vetores linearmente independentes $h_1, h_2, \ldots, h_n$. Então, para todo $x \in \mathbb{X}$, os vetores $x, h_1, h_2, \ldots, h_n$ são linearmente dependentes, isto é, existem números $\alpha^0, \alpha^1, \ldots, \alpha^n$ tais que

$$\alpha^0 x + \sum_{j=1}^{n} \alpha^j h_j = 0$$

e

$$\sum_{j=0}^{n} |\alpha^j| > 0.$$

Mostremos que $\alpha^0 \neq 0$. Com efeito, caso contrário, teríamos

$$\sum_{j=1}^{n} \alpha^j h_j = 0 \quad \text{e} \quad \sum_{j=1}^{n} |\alpha^j| > 0,$$

o que contradiz a independência linear dos vetores h_j, $j = \overline{1,n}$. Assim, temos que

$$x = -\sum_{j=1}^{n} \frac{\alpha^j}{\alpha^0} h_j,$$

ou seja, qualquer elemento do espaço linear $\mathbb{X}$ pode ser escrito de maneira única como uma combinação linear dos vetores linearmente independentes h_j, $j = \overline{1,n}$, que formam então uma base em $\mathbb{X}$. ■

Este teorema leva-nos à definição seguinte.

Definição 3.5. Dimensão de um espaço linear

Seja $\mathbb{X}$ um espaço linear sobre um conjunto de números $\mathbb{K}$ tal que admite uma base com n elementos. A este número n chamamos *dimensão do espaço linear* $\mathbb{X}$ e escrevemos $\dim(\mathbb{X}) = n$.

Consideremos dois exemplos.

Exemplo 3.12. $\dim(\mathbb{R}^n) = n$. □

Exemplo 3.13. $\dim(\mathbb{C}^n) = n$. □

O Método de Gauss leva-nos à definição da *característica de uma matriz*. Seja

$$A = \begin{pmatrix} a_1^1 & a_1^2 & \cdots & a_1^n \\ a_2^1 & a_2^2 & \cdots & a_2^n \\ \vdots & \vdots & \ddots & \vdots \\ a_m^1 & a_m^2 & \cdots & a_m^n \end{pmatrix}$$

uma matriz com m linhas e n colunas. Após a transformação desta matriz em forma em escada

$$\begin{pmatrix} a_1^{j_1} & a_1^2 & \cdots & a_1^{j_2} & \cdots & a_1^{j_i} & \cdots & a_1^n \\ 0 & 0 & \cdots & a_2^{j_2} & \cdots & a_2^{j_i} & \cdots & a_2^n \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & a_i^{j_i} & \cdots & a_i^n \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \end{pmatrix},$$

as linhas da matriz original e da matriz em escada podem ser consideradas como coordenadas de vetores numa certa base que podem ser obtidos uns dos outros como resultado de combinações lineares. Obviamente as primeiras i linhas são linearmente independentes. A este número i chamamos *característica da matriz* A. Portanto, a característica da matriz é o número de linhas linearmente independentes. Escolhendo as colunas com os pivôs $a_k^{j_k}$, obtemos i colunas linearmente independentes. Logo, a característica é também igual ao número de colunas linearmente independentes.

3.5. Subespaços

Na teoria dos espaços lineares existe o conceito muito importante de subespaço. Este conceito generaliza os objetos tais como retas em $\mathbb{V}^2$ ou retas e planos em $\mathbb{V}^3$ que atravessam a origem.

Começamos por introduzir a noção de subespaço de um espaço linear.

Definição 3.6. Subespaço de um espaço linear

Dizemos que o conjunto $\mathbb{L}$ do espaço linear $\mathbb{X}$ é um *subespaço* de $\mathbb{X}$ se $\mathbb{L}$ é um espaço linear.

Imediatamente da definição de espaço linear, temos o seguinte teorema.

Teorema 3.4

O conjunto $\mathbb{L}$ do espaço linear $\mathbb{X}$ é um subespaço de $\mathbb{X}$ se e só se

$$\alpha x + \beta y \in \mathbb{L},$$

para todos os $x, y \in \mathbb{L}$ e $\alpha, \beta \in \mathbb{K}$.

Eis alguns exemplos de subespaços, a seguir.

Exemplo 3.14. O vetor 0 forma o subespaço de $\mathbb{X}$ mínimo no sentido de inclusão. □

Exemplo 3.15. O espaço $\mathbb{X}$ é ele mesmo o subespaço de $\mathbb{X}$ maior no sentido de inclusão. □

Exemplo 3.16. O conjunto formado pelas soluções de um sistema de equações lineares homogéneo com m equações e n incógnitas é um subespaço de $\mathbb{R}^n$. □

Exemplo 3.17. O conjunto $\mathcal{P}^m$ dos polinómios de grau menor ou igual a m é um subespaço de $\mathcal{P}^n$ se $m \leqslant n$. □

É fácil de ver que a interseção de dois subespaços de um espaço linear $\mathbb{X}$ é um subespaço de $\mathbb{X}$ e que a dimensão de um subespaço não supera a dimensão do espaço.

Na resolução de muitos problemas temos uma base num subespaço e é necessário expandir esta base para todo o espaço linear. Para esse fim, vamos mostrar que cada conjunto de vetores linearmente independentes num espaço de dimensão finita pode ser completado até uma base.

Teorema 3.5

Sejam $\mathbb{X}$ um espaço linear de dimensão n e $e_1, e_2, \ldots, e_m$, $m < n$, vetores linearmente independentes. Então, existem vetores $e_{m+1}, \ldots, e_n$ tais que $\{e_1, \ldots, e_m, e_{m+1}, \ldots, e_n\}$ é uma base em $\mathbb{X}$.

Demonstração. Existe um vetor e_{m+1} que não pode ser representado na forma de uma combinação linear dos vetores $e_1, e_2, \ldots, e_m$, pois, caso contrário, os vetores e_k, $k = \overline{1, m}$, formariam uma base em $\mathbb{X}$ e a dimensão de $\mathbb{X}$ seria igual a m, o que não é possível. Os vetores $e_1, e_2, \ldots, e_{m+1}$ são linearmente independentes porque as condições

$$\sum_{i=1}^{m+1} \alpha^i e_i = 0 \quad \text{e} \quad \sum_{i=1}^{m+1} |\alpha^i| > 0$$

implicam que $\alpha^{m+1} \neq 0$ (os vetores $e_1, e_2, \ldots, e_m$ são linearmente independentes), isto é, que e_{m+1} é uma combinação linear dos vetores $e_1, e_2, \ldots, e_m$, o que é uma contradição. Se cada vetor de $\mathbb{X}$ pode ser representado na forma de uma combinação linear dos vetores $e_1, e_2, \ldots, e_{m+1}$, então $\dim(\mathbb{X}) = m + 1$ e os vetores $e_1, e_2, \ldots, e_{m+1}$ formam uma base. Caso contrário, existe e_{m+2} que não pode ser escrito como uma combinação linear dos vetores $e_1, e_2, \ldots, e_{m+1}$, etc. Continuando este procedimento, depois de um número finito de passos construimos uma base $\{e_1, \ldots, e_m, e_{m+1}, \ldots, e_n\}$ em $\mathbb{X}$. ■

Vamos utilizar este teorema várias vezes.

4. Espaços normados e espaços Euclidianos

4.1. Introdução

Tal como nos primeiros dois capítulos, onde estudamos espaços de dimensão dois e três, os conceitos de norma e de produto escalar (produto interno) também são muito importantes na resolução de problemas em espaços de dimensão superior. Neste capítulo vamos definir e estudar as propriedades de norma e de produto escalar em espaços lineares. Aos espaços lineares nos quais está definida uma norma damos o nome de *espaços normados* e aos espaços lineares nos quais está definido um produto escalar damos o nome de *espaços Euclidianos*.

4.2. Espaços normados

A noção de comprimento de um vetor no plano ou no espaço de dimensão três é generalizada para o caso dos espaços lineares através da definição da função chamada *norma*, que faz corresponder a um vetor um número real não negativo.

Definição 4.1. Norma e espaço normado

Seja $\mathbb{X}$ um espaço linear sobre o conjunto de números $\mathbb{K}$. Chamamos *norma* a uma função $p : \mathbb{X} \to \mathbb{R}$ se a função verifica as seguintes condições:

1. $p(x) \geqslant 0$, $p(x) = 0 \Leftrightarrow x = 0$ (não negatividade);
2. $p(\alpha x) = |\alpha| p(x)$, para quaisquer $x \in \mathbb{X}, \alpha \in \mathbb{K}$ (homogeneidade positiva);
3. $p(x+y) \leqslant p(x)+p(y)$, para quaisquer $x, y \in \mathbb{X}$ (desigualdade triangular).

Dizemos que $\mathbb{X}$ é um espaço *normado* se nele está definida uma norma. A norma de um vetor $x \in \mathbb{X}$ representamos por $|x|$.

É fácil verificar que os seguintes espaços lineares são normados.

Exemplo 4.1. Espaço $\mathbb{K}^n = \{(x^1, \ldots, x^n) : x^k \in \mathbb{K},\ k = \overline{1,n}\}$ com a norma

$$|x|_1 = \sum_{k=1}^{n} |x^k|.$$

□

Exemplo 4.2. Espaço $\mathbb{K}^n = \{(x^1, \ldots, x^n) : x^k \in \mathbb{K},\ k = \overline{1,n}\}$ com a norma

$$|x|_\infty = \max\{|x^k| : k = \overline{1,n}\}.$$

□

Num espaço linear pode ser útil definir a distância entre dois pontos, isto é, uma função que faz corresponder a dois pontos um número não negativo e que tem as propriedades naturais que tem a distância entre dois pontos de $\mathbb{V}^2$.

Definição 4.2. Distância

Seja $\mathbb{X}$ um espaço linear. Chamamos *distância* a uma função $d : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ que verifica as seguintes condições:

1. $d(x, y) \geqslant 0$, $d(x, y) = 0 \Leftrightarrow x = y$ (não negatividade);
2. $d(x, y) = d(y, x)$, para quaisquer $x, y \in \mathbb{X}$ (simetria);
3. $d(x, y) \leqslant d(x, z) + d(z, y)$, para quaisquer $x, y, z \in \mathbb{X}$ (desigualdade triangular).

A norma dá a possibilidade de definir a distância entre dois pontos de um espaço normado. É fácil verificar diretamente da definição de norma que a função definida por

$$d(x, y) = |x - y|$$

é uma distância.

Demonstremos uma consequência útil da desigualdade triangular.

Teorema 4.1

Sejam $\mathbb{X}$ um espaço normado e x e y dois vetores de $\mathbb{X}$. Então temos a desigualdade

$$||x| - |y|| \leqslant |x - y|. \tag{4.1}$$

Demonstração. Da desigualdade triangular (veja a Definição 4.1) obtemos

$$|x| = |x - y + y| \leqslant |x - y| + |y|, \quad |y| = |y - x + x| \leqslant |y - x| + |x|.$$

Assim, temos

$$|x| - |y| \leqslant |x - y|, \quad |y| - |x| \leqslant |y - x|,$$

que implicam (4.1). ■

Uma das consequências importantes desta desigualdade é a continuidade da função norma no sentido que se a distância entre dois pontos é pequena, então a diferença entre as respetivas normas também o é.

4.3. Espaços Euclidianos reais

Ao introduzir o produto escalar nos espaços lineares vamos conseguir, tal como nos dois primeiros capítulos, obter uma geometria muito mais interessante e rica do que no caso dos espaços normados. O conceito de produto escalar num espaço linear é formalizado através da seguinte definição.

Definição 4.3. Espaço Euclidiano real

Seja $\mathbb{X}$ um espaço linear sobre o conjunto dos números reais. Dizemos que este espaço é *Euclidiano* (Euclidiano real) se está definida uma função $\langle \cdot, \cdot \rangle : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ que verifica as seguintes condições:

1. $\langle x, y \rangle = \langle y, x \rangle$, para quaisquer $x, y \in \mathbb{X}$;
2. $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$, para quaisquer $x, y, z \in \mathbb{X}$;
3. $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$, para quaisquer $x, y \in \mathbb{X}$ e qualquer $\alpha \in \mathbb{R}$;
4. $\langle x, x \rangle \geqslant 0$, para qualquer $x \in \mathbb{X}$, e $\langle x, x \rangle = 0 \Leftrightarrow x = 0$.

À função $\langle \cdot, \cdot \rangle$ damos o nome de *produto escalar* ou *produto interno*.

Apresentamos três exemplos de espaços Euclidianos.

Exemplo 4.3. O espaço $\mathbb{R}^n$. O produto escalar dos vetores x e y de $\mathbb{R}^n$ pode ser definido por

$$\langle x, y \rangle = x^1 y^1 + \cdots + x^n y^n.$$

□

No mesmo espaço linear é possível definir produtos escalares diferentes. Eis mais um exemplo de produto escalar em $\mathbb{R}^n$.

Exemplo 4.4. A fórmula

$$\langle x, y \rangle = a^1 x^1 y^1 + \cdots + a^n x^n y^n,$$

com $a^k > 0$, $k = \overline{1, n}$, define outro produto escalar em $\mathbb{R}^n$. □

Exemplo 4.5. O espaço $\mathcal{P}^n([a, b])$ de polinómios de grau menor ou igual a n considerados no intervalo $[a, b]$. Definimos o produto escalar dos polinómios p e q de $\mathcal{P}^n([a, b])$ por

$$\langle p, q \rangle = \int_a^b p(x) q(x) \, \mathrm{d}x.$$

□

Neste exemplo, a única propriedade do produto escalar que não é óbvia é que a igualdade

$$\int_a^b p^2(x) \, \mathrm{d}x = 0$$

implica que $p(x) \equiv 0$. Para mostrar isto, precisamos do seguinte lema.

Lema 4.1

Seja $f : [a, b] \to \mathbb{R}$ uma função contínua não negativa que verifica a igualdade

$$\int_a^b f(x)\,\mathrm{d}x = 0.$$

Então, $f(x) \equiv 0$.

Demonstração. Suponhamos que existe $x_0 \in]a, b[$ tal que $f(x_0) = c > 0$. Como a função f é contínua, existe $\delta > 0$ tal que $[x_0 - \delta, x_0 + \delta] \subset [a, b]$ e $f(x) > \frac{c}{2}$ sempre que $x \in [x_0 - \delta, x_0 + \delta]$. Como a função f é não negativa, logo temos

$$0 = \int_a^b f(x)\,\mathrm{d}x \geqslant \int_{x_0-\delta}^{x_0+\delta} f(x)\,\mathrm{d}x \geqslant \int_{x_0-\delta}^{x_0+\delta} \frac{c}{2}\,\mathrm{d}x = c\delta > 0,$$

o que é uma contradição. Portanto, $f(x) = 0$ nos pontos interiores do intervalo $[a, b]$. Pela continuidade de f temos que $f(a) = f(b) = 0$. ■

Num espaço Euclidiano o produto escalar dá a possibilidade de definir uma norma, conhecida por *norma Euclidiana*. Para verificar as propriedades de norma, precisamos do seguinte lema.

Lema 4.2. Desigualdade de Cauchy-Schwarz

Sejam $\mathbb{X}$ um espaço Euclidiano e $x, y \in \mathbb{X}$. Então, $|\langle x, y\rangle| \leqslant |x||y|$. A igualdade $|\langle x, y\rangle| = |x||y|$ verifica-se se e só se x e y são linearmente dependentes.

Demonstração. Temos $0 \leqslant \langle tx + y, tx + y\rangle$, para todo $t \in \mathbb{R}$, ou, na forma equivalente, $0 \leqslant \langle x, x\rangle t^2 + 2\langle x, y\rangle t + \langle y, y\rangle$, para todo $t \in \mathbb{R}$. Um polinómio de segundo grau é maior ou igual a zero para cada t se e só se o seu discriminante é não positivo, isto é, $\langle x, y\rangle^2 - \langle x, x\rangle\langle y, y\rangle \leqslant 0$, ou $\langle x, y\rangle^2 \leqslant |x|^2|y|^2$, sendo o discriminante igual a zero, isto é, $\langle x, y\rangle^2 = |x|^2|y|^2$, se e só se existe t tal que $0 = \langle tx + y, tx + y\rangle$, ou seja, $tx + y = 0$, isto é x e y serem linearmente dependentes. ■

A definição de norma Euclidiana baseia-se no seguinte teorema.

Teorema 4.2

Seja $\mathbb{X}$ um espaço Euclidiano real. Então,

$$|x| = \sqrt{\langle x, x\rangle}$$

é uma norma, que chamamos *norma Euclidiana*.

Demonstração. Sejam $x, y \in \mathbb{X}$ e $\alpha \in \mathbb{R}$. Então temos:

1. $|x| = \sqrt{\langle x, x\rangle} \geqslant 0$, onde a igualdade é válida se e só se $x = 0$ pois $\langle x, x\rangle = 0 \Leftrightarrow x = 0$;

2. $|\alpha x| = \sqrt{\langle \alpha x, \alpha x\rangle} = \sqrt{\alpha^2\langle x, x\rangle} = |\alpha||x|$;

3. Usando a desigualdade de Cauchy-Schwarz, temos

$$|x+y|^2 = |x|^2 + 2\langle x, y\rangle + |y|^2 \leqslant |x|^2 + 2|x||y| + |y|^2 = (|x| + |y|)^2,$$

o que é equivalente a $|x+y| \leqslant |x| + |y|$.

■

No conjunto dos espaços normados, os espaços Euclidianos são poucos, no sentido que a propriedade da norma ser Euclidiana, isto é, definida através do produto escalar, é um acontecimento raro. Com efeito, se $\mathbb{X}$ é um espaço Euclidiano e $x, y \in \mathbb{X}$, o leitor facilmente pode demonstrar a *igualdade do paralelogramo*

$$|x+y|^2 + |x-y|^2 = 2(|x|^2 + |y|^2),$$

que para a maioria das normas não se verifica. É fácil ver, por exemplo, que no espaço $\mathbb{R}^2$ a norma $|\cdot|_\infty$ não é Euclidiana. De facto, sejam $e_1 = (1,0)$ e $e_2 = (0,1)$. Então temos $|e_1|_\infty = 1$, $|e_2|_\infty = 1$, $|e_1 + e_2|_\infty = 1$ e $|e_1 - e_2|_\infty = 1$. Portanto,

$$|e_1 + e_2|_\infty^2 + |e_1 - e_2|_\infty^2 = 2 \quad \text{mas} \quad 2(|e_1|_\infty^2 + |e_2|_\infty^2) = 4.$$

Na geometria dos espaços Euclidianos, a seguinte definição assume um papel fundamental.

Definição 4.4. Projeção de um vetor num subespaço

Sejam $\mathbb{L}$ um subespaço de um espaço Euclidiano $\mathbb{X}$ e x um vetor de $\mathbb{X}$. Dizemos que um vetor z de $\mathbb{L}$ é a *projeção ortogonal* de x no subespaço $\mathbb{L}$ se o vetor $y = x - z$ é ortogonal a todos os vetores de $\mathbb{L}$.

O seguinte teorema pode ser considerada como uma definição alternativa da definição de projeção ortogonal.

Teorema 4.3

A projeção ortogonal de um vetor x no subespaço $\mathbb{L}$ é o ponto $z \in \mathbb{L}$ mais próximo de x no sentido da norma Euclidiana de $\mathbb{X}$.

Demonstração. Seja $v \neq z$ um vetor de $\mathbb{L}$. Então o vetor $z - v \in \mathbb{L}$ é ortogonal ao vetor $y = x - z$. Logo, temos

$$\begin{aligned} |x-v|^2 &= \langle x-v, x-v\rangle = \langle y+z-v, y+z-v\rangle = \langle y, y\rangle + \langle z-v, z-v\rangle \\ &= |y|^2 + |z-v|^2 > |y|^2. \end{aligned}$$

■

Há nos espaços Euclidianos uma classe de subespaços conhecidos por *complementos ortogonais*, que aí assumem um papel muito importante. Seja $\mathbb{L}$ um subespaço de um espaço Euclidiano $\mathbb{X}$.

Definição 4.5. Complemento ortogonal

Chamamos *complemento ortogonal* de $\mathbb{L}$, que representamos por $\mathbb{L}^\perp$, ao conjunto dos vetores que lhe são ortogonais.

Teorema 4.4

O complemento ortogonal, $\mathbb{L}^\perp$, é um subespaço de $\mathbb{X}$.

Demonstração. Sejam $y, z \in \mathbb{L}^\perp$ e $a, b \in \mathbb{R}$. Como $\mathbb{L}^\perp = \{y \in \mathbb{X} : \langle y, x\rangle = 0, x \in \mathbb{L}\}$, temos

$$\langle ay + bz, x\rangle = a\langle y, x\rangle + b\langle z, x\rangle = 0,$$

para todo $x \in \mathbb{L}$, ou seja, $ay + bz \in \mathbb{L}^\perp$. ∎

4.4. Geometria do Método de Gauss

Nesta secção vamos dar uma interpretação geométrica do Método de Gauss.

Consideremos um vetor $e \in \mathbb{R}^n$, o subespaço $\mathbb{L} = \{z \in \mathbb{R}^n : \langle z, e\rangle = 0\}$, um vetor $v \in \mathbb{R}^n$ e uma reta $\mathbb{T} = \{tv : t \in \mathbb{R}\}$. O problema da construção da projeção ortogonal da reta $\mathbb{T}$ no subespaço $\mathbb{L}$ tem uma solução óbvia: a reta gerada pelo vetor $w = v - \langle v, e\rangle e \in \mathbb{L}$.

O mesmo problema é bastante mais complicado quando a reta é definida como um sistema de equações $\mathbb{T} = \{v \in \mathbb{R}^n : \langle c_i, v\rangle = 0,\ i = \overline{1, n-1}\}$, onde $c_i \in \mathbb{R}^n$ são vetores linearmente independentes. Neste caso, a solução do problema tem de ter a forma de um sistema de equações $\{w \in \mathbb{L} : \langle d_j, w\rangle = 0,\ j = \overline{2, n-1}\}$, onde $d_j \in \mathbb{L}$ são vetores linearmente independentes. Vamos procurar os vetores $d_j \in \mathbb{L}$ na forma de combinações lineares dos vetores $c_j - \langle c_j, e\rangle e \in \mathbb{L}$. Como os vetores $c_j \in \mathbb{R}^n$ são linearmente independentes, pelo menos um dos produtos escalares $\langle c_j, e\rangle$ é diferente de zero. Seja, então, sem perda de generalidade, $\langle c_1, e\rangle \neq 0$, e façamos

$$d_j = \begin{cases} c_j, & \text{se } \langle c_j, e\rangle = 0, \\ \frac{1}{\langle c_j, e\rangle}(c_j - \langle c_j, e\rangle e) - \frac{1}{\langle c_1, e\rangle}(c_1 - \langle c_1, e\rangle e), & \text{se } \langle c_j, e\rangle \neq 0. \end{cases}$$

Como

$$\langle v - \langle v, e\rangle e, c_j - \langle c_j, e\rangle e\rangle = -\langle v, e\rangle\langle c_j, e\rangle, \tag{4.2}$$

temos que

$$\langle d_j, w\rangle = \langle d_j, v - \langle v, e\rangle e\rangle = \left\{ \begin{array}{ll} -\langle v, e\rangle\langle c_j, e\rangle, & \text{se } \langle c_j, e\rangle = 0 \\ \langle v, e\rangle - \langle v, e\rangle & \text{se } \langle c_j, e\rangle \neq 0 \end{array} \right\} = 0.$$

É fácil ver que os vetores $d_j \in \mathbb{L}$, $j = \overline{2, n-1}$, são linearmente independentes. De facto, a igualdade

$$0 = \sum_{j=2}^{n-1} \alpha_j d_j$$

implica a igualdade

$$0 = \sum_{j \in J_0} \alpha_j c_j + \sum_{j \in J_\#} \alpha_j \left(\frac{c_j}{\langle c_j, e \rangle} - \frac{c_1}{\langle c_1, e \rangle} \right),$$

onde $J_0 = \{j \in \{2, \ldots, n-1\} : \langle c_j, e \rangle = 0\}$ e $J_\# = \{j \in \{2, \ldots, n-1\} : \langle c_j, e \rangle \neq 0\}$. Como os vetores $c_i \in \mathbb{R}^n$, $i = \overline{1, n-1}$, são linearmente independentes, vemos que $\alpha_j = 0$, $j = \overline{2, n-1}$.

Notemos que, se $v \in \mathbb{T}$ e $w \in \mathbb{L}$ é a sua projeção, então é fácil, dado w, reconstruir v. De facto, como para j tais que $\langle c_j, e \rangle \neq 0$ (veja (4.2)) temos

$$\langle v, e \rangle = -\frac{\langle w, c_j - \langle c_j, e \rangle e \rangle}{\langle c_j, e \rangle} = -\frac{\langle w, c_j \rangle}{\langle c_j, e \rangle},$$

fazendo

$$v = w + \langle v, e \rangle e = w - \frac{\langle w, c_j \rangle}{\langle c_j, e \rangle} e,$$

temos

$$\langle v, c_j \rangle = 0.$$

No caso $\langle c_j, e \rangle = 0$, temos

$$\langle v, c_j \rangle = \langle w, c_j \rangle = \langle w, d_j \rangle = 0.$$

Portanto, como $\langle c_1, e \rangle \neq 0$, a reconstrução de v é obtida através da fórmula

$$v = w - \frac{\langle w, c_1 \rangle}{\langle c_1, e \rangle} e. \tag{4.3}$$

Consideremos o sistema linear possível e determinado com n equações e com n incógnitas $x^1, \ldots, x^n$ dado por

$$\begin{cases} a_1^1 x^1 + a_1^2 x^2 + a_1^3 x^3 + \cdots + a_1^n x^n = b_1 \\ a_2^1 x^1 + a_2^2 x^2 + a_2^3 x^3 + \cdots + a_2^n x^n = b_2 \\ \quad \vdots \qquad \vdots \qquad \vdots \qquad \ddots \qquad \vdots \qquad \vdots \\ a_n^1 x^1 + a_n^2 x^2 + a_n^3 x^3 + \cdots + a_n^n x^n = b_n. \end{cases}$$

Introduzamos mais uma variável x^{n+1} e consideremos o sistema homogéneo

$$\begin{cases} a_1^1 x^1 + a_1^2 x^2 + a_1^3 x^3 + \cdots + a_1^n x^n = b_1 x^{n+1} \\ a_2^1 x^1 + a_2^2 x^2 + a_2^3 x^3 + \cdots + a_2^n x^n = b_2 x^{n+1} \\ \quad \vdots \qquad \vdots \qquad \vdots \qquad \ddots \qquad \vdots \qquad \vdots \\ a_n^1 x^1 + a_n^2 x^2 + a_n^3 x^3 + \cdots + a_n^n x^n = b_n x^{n+1}. \end{cases}$$

Este sistema determina uma reta $\mathbb{T} \subset \mathbb{R}^{n+1}$. Fazendo

$$e = e_1 = (1, 0, 0, \ldots, 0)$$

e

$$c_i = (a_i^1, a_i^2, a_i^3, \ldots, a_i^n, -b_i), \quad i = \overline{1, n},$$

e aplicando o raciocínio anterior efetuamos o primeiro passo do Método de Gauss e obtemos sistema de equações que determina uma reta no subespaço $\{x \in \mathbb{R}^n : x^1 = 0\}$. Isto corresponde à operação elementar

$$l_j \leftarrow \frac{1}{a_j^1} l_j - \frac{1}{a_1^1} l_1,$$

que é equivalente à operação habitual

$$l_j \leftarrow l_j - \frac{a_j^1}{a_1^1} l_1.$$

Continuando esta construção chegamos à equação de uma reta no plano $\{x \in \mathbb{R}^n : x^i = 0,\ i = \overline{1, n-1}\}$.

Passamos agora à segunda fase do Método de Gauss. Fazendo $x^{n+1} = 1$ e aplicando (4.3), determinamos x^n. Determinamos todas as outras incógnitas aplicando repetidamente este processo.

Geometricamente falando, em cada passo da segunda fase do Método de Gauss temos um subespaço $\mathbb{L}$ e um seu ponto w, que é a projeção ortogonal do ponto $v \in \mathbb{L}'$ no subespaço $\mathbb{L}$, onde $\mathbb{L}'$ é um subespaço tal que $\dim(\mathbb{L}') = \dim(\mathbb{L}) + 1$. A fórmula (4.3) proporciona o método de reconstrução.

4.5. Bases ortogonais e ortogonalização

Nos espaços Euclidianos é possível e prático considerar bases ortogonais tal como fizemos nos dois primeiros capítulos.

Definição 4.6. Conjunto ortogonal

Seja $\{x_1, \ldots, x_n\}$ um conjunto de vetores de um espaço Euclidiano $\mathbb{X}$. Dizemos que este conjunto é *ortogonal* se $\langle x_k, x_{k'} \rangle = 0$, $k \neq k'$.

Como podemos ver do seguinte teorema, num espaço Euclidiano $\mathbb{X}$ de dimensão n, um conjunto de n vetores ortogonais não nulos forma uma base.

Teorema 4.5

Seja $\{e_1, \ldots, e_n\}$ um conjunto de vetores ortogonais de um espaço Euclidiano $\mathbb{X}$. Se $\dim(\mathbb{X}) = n$ e $|e_k| = 1$, $k = \overline{1, n}$, então o conjunto é uma base, que chamamos *ortonormal*.

Demonstração. Graças ao Teorema 3.3, basta mostrar que os vetores e_k, $k = \overline{1,n}$, são linearmente independentes. Suponhamos que estes vetores são linearmente dependentes, isto é, que existem números α_k nem todos nulos e tais que

$$\sum_{k=1}^{n} \alpha_k e_k = 0.$$

Suponhamos que $\alpha_m \neq 0$. Como o conjunto dos vetores é ortogonal, o produto escalar desta soma e do vetor e_m, implica

$$\alpha_m = \alpha_m \langle e_m, e_m \rangle = 0,$$

o que é uma contradição. ■

Um exemplo clássico de uma base ortonormal no espaço $\mathbb{R}^n$ é o seguinte.

Exemplo 4.6. Os vetores $e_1 = (1, 0, \ldots, 0)$, $e_2 = (0, 1, \ldots, 0)$, ..., e $e_n = (0, 0, \ldots, 1)$ formam uma base ortonormal no espaço $\mathbb{R}^n$. □

Um conjunto ortogonal de vetores pode ser obtido de um conjunto de vetores linearmente independentes utilizando o seguinte *processo de ortogonalização*.

Teorema 4.6. (Gram-Schmidt)

Sejam $\mathbb{X}$ um espaço Euclidiano e $x_1, \ldots, x_n$ vetores linearmente independentes de $\mathbb{X}$. Então, existem vetores $y_1, \ldots, y_n$ de $\mathbb{X}$ tais que

1. $\{y_1, \ldots, y_n\}$ é um conjunto ortogonal e
2. $\text{Lin}\{y_1, \ldots, y_m\} = \text{Lin}\{x_1, \ldots, x_m\}$, $m = \overline{1,n}$.

Demonstração. Façamos $y_1 = x_1$. Suponhamos que os vetores $y_1, y_2, \ldots, y_{m-1}$ já estão construídos. Definamos o vetor y_m por

$$y_m = x_m - s_{m-1}, \quad \text{com} \quad s_{m-1} = \sum_{k=1}^{m-1} \frac{\langle x_m, y_k \rangle}{|y_k|^2} y_k. \tag{4.4}$$

Obviamente $\langle y_m, y_k \rangle = 0$, $k = \overline{1, m-1}$, e $\text{Lin}\{y_1, y_2, \ldots, y_m\} = \text{Lin}\{x_1, x_2, \ldots, x_m\}$. ■

Notemos que a soma (4.4) que aparece no processo de ortogonalização de Gram-Schmidt, é a projeção ortogonal do vetor x_m no subespaço $\text{Lin}\{y_1, y_2, \ldots, y_{m-1}\}$ e $y_m \in (\text{Lin}\{y_1, y_2, \ldots, y_{m-1}\})^{\perp}$ (veja a Fig. 4.1).

A partir da relação entre os vetores linearmente independentes $\{x_1, \ldots, x_m\}$ e os vetores ortogonais correspondentes $\{y_1, \ldots, y_m\}$ obtidos pelo processo de ortogonalização podemos estabelecer uma desigualdade útil. Os pontos 0, x_m e s_{m-1} formam um triângulo retângulo pois o vetor $x_m - s_{m-1}$ é ortogonal ao vetor s_{m-1}. Obviamente a hipotenusa x_m não pode ser mais curta que os lados do triângulo. Este facto leva à desigualdade

$$|x_m|^2 \geqslant |y_m|^2, \quad m = \overline{1,n}. \tag{4.5}$$

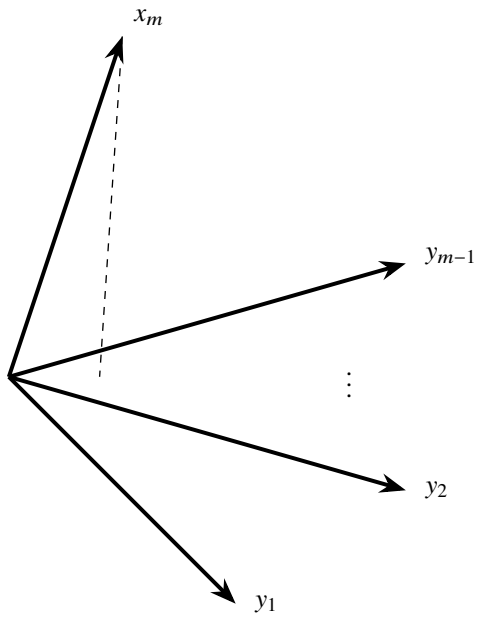


Figura 4.1
Projeção ortogonal do vetor x_m no subespaço $\text{Lin}\{y_1, y_2, \ldots, y_{m-1}\}$.

Esta desigualdade também pode ser obtida analiticamente. De facto, temos

$$|x_m|^2 = |y_m + s_{m-1}|^2 = \langle y_m + s_{m-1}, y_m + s_{m-1}\rangle = |y_m|^2 + |s_{m-1}|^2.$$

Vamos usar esta desigualdade no Capítulo 6.

Consideremos dois exemplos de aplicação do Método de ortogonalização de Gram-Schmidt.

Exemplo 4.7. Consideremos a base $\{(1,-3),(2,2)\}$ de $\mathbb{R}^2$. Vamos transformar esta base numa base ortonormal. Sejam $x_1 = (1,-3)$ e $x_2 = (2,2)$. Normalizando x_1 obtemos

$$w_1 = \frac{x_1}{|x_1|} = \frac{1}{\sqrt{10}}(1,-3).$$

Aplicando o processo de ortogonalização de Gram-Schmidt, usando o vetor normalizado w_1, obtemos

$$\begin{aligned} y_2 &= x_2 - \frac{\langle x_2, w_1\rangle}{|w_1|^2} w_1 = (2,2) - \left\langle (2,2), \frac{1}{\sqrt{10}}(1,-3)\right\rangle \frac{1}{\sqrt{10}}(1,-3) \\ &= (2,2) + \frac{2}{5}(1,-3) = \frac{1}{5}(12,4), \\ w_2 &= \frac{y_2}{|y_2|} = \frac{1}{\sqrt{10}}(3,1). \end{aligned}$$

Assim, concluímos que a base ortonormal $\{w_1, w_2\}$ obtida a partir dos vetores x_1, x_2 é dada por

$$\left\{\frac{1}{\sqrt{10}}(1,-3), \frac{1}{\sqrt{10}}(3,1)\right\}.$$

□

Exemplo 4.8. Consideremos a base $\{(1,1,1),(-1,1,0),(1,2,1)\}$ de $\mathbb{R}^3$. Vamos aplicar o processo de ortogonalização de Gram-Schmidt para transformarmos esta base numa base ortonormal. Sejam $x_1 = (1,1,1)$, $x_2 = (-1,1,0)$ e $x_3 = (1,2,1)$. Estes vetores são linearmente independentes. Aplicando o processo de ortogonalização, normalizando primeiro o vetor x_1, vem

$$
\begin{aligned}
w_1 &= \frac{x_1}{|x_1|} = \frac{\sqrt{3}}{3}(1,1,1),\\
y_2 &= x_2 - w_1\langle x_2, w_1\rangle = (-1,1,0) - \frac{1}{\sqrt{3}}(1,1,1)\left\langle (-1,1,0)\cdot\frac{1}{\sqrt{3}}(1,1,1)\right\rangle\\
&= (-1,1,0),\\
w_2 &= \frac{y_2}{|y_2|} = \frac{\sqrt{2}}{2}(-1,1,0),\\
y_3 &= x_3 - w_1\langle x_3, w_1\rangle - w_2\langle x_3, w_2\rangle\\
&= (1,2,1) - \frac{1}{\sqrt{3}}(1,1,1)\left\langle (1,2,1), \frac{1}{\sqrt{3}}(1,1,1)\right\rangle\\
&\quad - \frac{\sqrt{2}}{2}(-1,1,0)\left\langle (1,2,1), \frac{\sqrt{2}}{2}(-1,1,0)\right\rangle\\
&= (1,2,1) - \frac{4}{3}(1,1,1) - \frac{1}{2}(-1,1,0) = \frac{1}{6}(1,1,-2),\\
w_3 &= \frac{z_3}{|z_3|} = \frac{\sqrt{6}}{6}(1,1,-2).
\end{aligned}
$$

Assim, concluímos que a base ortonormal $\{w_1, w_2, w_3\}$ obtida a partir dos vetores x_1, x_2, x_3 é dada por

$$\left\{\frac{\sqrt{3}}{3}(1,1,1), \frac{\sqrt{2}}{2}(-1,1,0), \frac{\sqrt{6}}{6}(1,1,-2)\right\}.$$

□

4.6. Base no espaço de polinómios trigonométricos e coeficientes de Fourier

Um exemplo interessante é o do espaço linear de polinómios trigonométricos. Neste espaço, da mesma maneira como foi feito no caso dos polinómios algébricos, podemos definir um produto escalar. O que é interessante é que neste espaço com esse produto escalar existe uma base ortogonal natural e esse facto tem uma influência marcante em vários ramos da Matemática, da Teoria dos Números até às equações da Física Matemática (equações em derivadas parciais).

Os polinómios trigonométricos

$$f(x) = \frac{a_0}{2} + a_1\cos x + b_1\,\mathrm{sen}\,x + \cdots + a_n\cos(nx) + b_n\,\mathrm{sen}(nx)$$

são combinações lineares das funções trigonométricas

$$1, \cos x, \mathrm{sen}\,x, \ldots, \cos(nx), \mathrm{sen}(nx). \tag{4.6}$$

Obviamente que os polinómios trigonométricos $\mathcal{T}^n$, definidos no intervalo $[-\pi, \pi]$, formam um espaço linear. Neste espaço podemos definir o produto escalar

$$\langle f, g\rangle = \int_{-\pi}^{\pi} f(x)g(x)\,\mathrm{d}x.$$

Verificamos exatamente como no caso do espaço de polinómios algébricos $\mathcal{P}^n$ as propriedades do produto escalar.

O seguinte lema estabelece uma propriedade importante das funções trigonométricas.

Lema 4.3

As funções (4.6) são ortogonais, ou seja,

$$\int_{-\pi}^{\pi} \cos(nx)\cos(mx)\,\mathrm{d}x = 0, \quad n \neq m,$$

$$\int_{-\pi}^{\pi} \operatorname{sen}(nx)\operatorname{sen}(mx)\,\mathrm{d}x = 0, \quad n \neq m, \quad \text{e}$$

$$\int_{-\pi}^{\pi} \cos(nx)\operatorname{sen}(mx)\,\mathrm{d}x = 0, \quad n, m = 0, 1, 2, \ldots.$$

Além disso, temos as igualdades

$$\int_{-\pi}^{\pi} 1\,\mathrm{d}x = 2\pi,$$

$$\int_{-\pi}^{\pi} \cos^2(nx)\,\mathrm{d}x = \pi \quad \text{e}$$

$$\int_{-\pi}^{\pi} \operatorname{sen}^2(nx)\,\mathrm{d}x = \pi, n = 1, 2, \ldots.$$

Demonstração. Seja $n \neq m$. Temos, por exemplo,

$$\int_{-\pi}^{\pi} \operatorname{sen}(nx)\operatorname{sen}(mx)\,\mathrm{d}x = \frac{1}{2}\int_{-\pi}^{\pi} (\cos(n-m)x - \cos(n+m)x)\,\mathrm{d}x$$
$$= \frac{\operatorname{sen}(n-m)x}{2(n-m)}\Big|_{-\pi}^{\pi} - \frac{\operatorname{sen}(n+m)x}{2(n+m)}\Big|_{-\pi}^{\pi} = 0.$$

Analogamente mostramos que

$$\int_{-\pi}^{\pi} \cos(nx)\cos(mx)\,\mathrm{d}x = 0, \quad n \neq m,$$

e

$$\int_{-\pi}^{\pi} \cos(nx)\operatorname{sen}(mx)\,\mathrm{d}x = 0, \quad n, m = 0, 1, 2, \ldots$$

Os integrais dos quadrados das funções calculamos assim:

$$\int_{-\pi}^{\pi} \cos^2(nx)\,\mathrm{d}x = \frac{1}{2}\int_{-\pi}^{\pi} (1 + \cos(2nx))\,\mathrm{d}x = \pi$$

e

$$\int_{-\pi}^{\pi} \operatorname{sen}^2(nx)\,\mathrm{d}x = \frac{1}{2}\int_{-\pi}^{\pi}(1-\cos(2nx))\,\mathrm{d}x = \pi.$$

■

Seja f um polinómio trigonométrico. Então, os respetivos coeficientes podem ser expressos em termos de f da seguinte maneira.

Teorema 4.7

Seja o polinómio trigonométrico

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{n}(a_k\cos(kx) + b_k\operatorname{sen}(kx)). \tag{4.7}$$

Então, os seus coeficientes têm as seguintes formas:

$$a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\,\mathrm{d}x,$$

$$a_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos(kx)\,\mathrm{d}x \quad \text{e}$$

$$b_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\operatorname{sen}(kx)\,\mathrm{d}x.$$

Demonstração. Como

$$\int_{-\pi}^{\pi} f(x)\,\mathrm{d}x = \int_{-\pi}^{\pi}\left(\frac{a_0}{2} + \sum_{k=1}^{n} a_k\cos(kx) + b_k\operatorname{sen}(kx)\right)\mathrm{d}x$$

$$= \frac{a_0}{2}\int_{-\pi}^{\pi}\mathrm{d}x + \sum_{k=1}^{n} a_k\int_{-\pi}^{\pi}\cos(kx)\,\mathrm{d}x + b_k\int_{-\pi}^{\pi}\operatorname{sen}(kx)\,\mathrm{d}x = \pi a_0,$$

temos a fórmula para a_0.

Multiplicando (4.7) termo a termo por $\cos(kx)$ ou $\operatorname{sen}(kx)$, integrando e usando a propriedade de ortogonalidade (veja o Lema 4.3), temos

$$\int_{-\pi}^{\pi} f(x)\cos(kx)\,\mathrm{d}x = \int_{-\pi}^{\pi} a_k\cos^2(kx)\,\mathrm{d}x = \pi a_k$$

e

$$\int_{-\pi}^{\pi} f(x)\operatorname{sen}(kx)\,\mathrm{d}x = \int_{-\pi}^{\pi} b_k\operatorname{sen}^2(kx)\,\mathrm{d}x = \pi b_k.$$

■

Deste teorema vemos que os coeficientes de um polinómio trigonométrico são definidos de maneira única, isto é, se

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{n}(a_k\cos(kx) + b_k\operatorname{sen}(kx))$$

e

$$F(x) = \frac{A_0}{2} + \sum_{k=1}^{n}(A_k\cos(kx) + B_k\operatorname{sen}(kx))$$

são dois polinómios trigonométricos, tais que $f(x) \equiv F(x)$, então $a_k = A_k$ e $b_k = B_k$ para $k = \overline{0,n}$. Portanto, as funções (4.6) formam uma base ortogonal no espaço $\mathcal{T}^n$, donde $\dim(\mathcal{T}^n) = 2n+1$.

Os coeficientes de um polinómio trigonométrico na forma apresentada no Teorema 4.7 são conhecidos como coeficientes de Fourier. Estes coeficientes são as coordenadas de um polinómio trigonométrico na base ortogonal formada pelas funções (4.6). Quando trabalhamos num espaço Euclidiano com uma base ortonormal, as coordenadas de um vetor relativamente a essa base chamamos também, por analogia, *coeficientes de Fourier*.

Definição 4.7. Coeficientes de Fourier

Sejam $\mathbb{X}$ um espaço Euclidiano, $\{e_1, \ldots, e_n\}$ um conjunto ortonormal e $x \in \mathbb{X}$. Então, os coeficientes

$$c_k = \langle x, e_k \rangle, \quad k = \overline{1,n},$$

são conhecidos como *coeficientes de Fourier* do vetor x relativamente ao conjunto ortonormal $\{e_1, \ldots, e_n\}$.

Exemplo 4.9. No espaço dos polinómios trigonométricos $\mathcal{T}^n$, o conjunto de funções

$$\left\{\frac{1}{\sqrt{2\pi}}, \frac{\cos x}{\sqrt{\pi}}, \frac{\operatorname{sen} x}{\sqrt{\pi}}, \ldots, \frac{\cos(nx)}{\sqrt{\pi}}, \frac{\operatorname{sen}(nx)}{\sqrt{\pi}}\right\}$$

é ortonormal (veja Lema 4.3). □

Seja $\mathbb{L} = \operatorname{Lin}\{e_1, \ldots, e_n\}$ com $\{e_1, \ldots, e_n\}$ um conjunto ortonormal num espaço Euclidiano $\mathbb{X}$. Acontece que o ponto mais próximo em $\mathbb{L}$ de um vetor $x \in \mathbb{X}$ é a combinação linear dos vetores $\{e_1, \ldots, e_n\}$ com coeficientes de Fourier de x.

Teorema 4.8. Propriedade mínima dos coeficientes de Fourier

Sejam $\mathbb{X}$ um espaço Euclidiano e $x \in \mathbb{X}$. Então

$$\left| x - \sum_{k=1}^{n} \langle x, e_k \rangle e_k \right|^2 \leqslant \left| x - \sum_{k=1}^{n} \alpha_k e_k \right|^2$$

para todos os números $\alpha_k \in \mathbb{R}$, $k = \overline{1,n}$.

Demonstração. Como

$$\begin{aligned}
\left| x - \sum_{k=1}^{n} \alpha_k e_k \right|^2 &= \left\langle x - \sum_{k=1}^{n} \alpha_k e_k, x - \sum_{k=1}^{n} \alpha_k e_k \right\rangle \\
&= \langle x, x \rangle - 2\left\langle x, \sum_{k=1}^{n} \alpha_k e_k \right\rangle + \left\langle \sum_{k=1}^{n} \alpha_k e_k, \sum_{k=1}^{n} \alpha_k e_k \right\rangle \\
&= |x|^2 - 2\sum_{k=1}^{n} \alpha_k c_k + \sum_{k=1}^{n} \alpha_k^2 = |x|^2 - \sum_{k=1}^{n} c_k^2 + \sum_{k=1}^{n} (\alpha_k - c_k)^2,
\end{aligned}$$

obtemos

$$\left|x-\sum_{k=1}^{n}\alpha_k e_k\right|^2 \geqslant |x|^2-\sum_{k=1}^{n}c_k^2=\left|x-\sum_{k=1}^{n}c_k e_k\right|^2=\left|x-\sum_{k=1}^{n}\langle x,e_k\rangle e_k\right|^2.$$

■

É fácil de ver que a soma

$$\sum_{k=1}^{n}\langle x,e_k\rangle e_k$$

é a projeção ortogonal do vetor x no subespaço $\mathrm{Lin}\{e_1,\ldots,e_n\}$.

Obviamente que a norma do vetor mais próximo de x no subespaço $\mathbb{L}$ não pode superar a norma do vetor x. Temos, então, o seguinte corolário desta última fórmula.

Corolário 4.1. (Desigualdade de Bessel)

Seja $\mathbb{X}$ um espaço Euclidiano. Para cada vetor $x\in\mathbb{X}$ verifica-se a desigualdade de Bessel

$$\sum_{k=1}^{n}c_k^2\leqslant|x|^2.$$

Demonstração. Como

$$|x|^2-\sum_{k=1}^{n}c_k^2=\left|x-\sum_{k=1}^{n}c_k e_k\right|^2\geqslant 0,$$

temos

$$|x|^2\geqslant\sum_{k=1}^{n}c_k^2.$$

■

4.7. Espaços Euclidianos complexos

Também é possível introduzir espaços Euclidianos complexos, mas o conjunto de propriedades do produto escalar tem de ser modificado dado que, no caso complexo, as propriedades $\langle x,y\rangle=\langle y,x\rangle$ e $\langle\alpha x,y\rangle=\alpha\langle x,y\rangle$ não são compatíveis com a propriedade $\langle x,x\rangle\geqslant 0$. Com efeito,

$$\langle \mathtt{i}x,\mathtt{i}x\rangle=\mathtt{i}^2\langle x,x\rangle=-\langle x,x\rangle,$$

isto é, os quadrados das normas dos vetores x e $\mathtt{i}x$ não podem ser positivos ao mesmo tempo. Para evitar esta contradição, vamos definir o espaço Euclidiano complexo $\mathbb{X}$ da seguinte maneira. (Recordemos que o número conjugado de um número complexo $z=a+\mathtt{i}b$ é $\overline{z}=a-\mathtt{i}b$.)

Definição 4.8. Espaço Euclidiano complexo

Seja $\mathbb{X}$ um espaço linear sobre o conjunto dos números complexos. Dizemos que este espaço é *Euclidiano* (Euclidiano complexo) se está definida uma função $\langle \cdot, \cdot \rangle : \mathbb{X} \times \mathbb{X} \to \mathbb{C}$ que verifica as seguintes condições:

1. $\langle x, y \rangle = \overline{\langle y, x \rangle}$, para quaisquer $x, y \in \mathbb{X}$;
2. $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$, para quaisquer $x, y, z \in \mathbb{X}$;
3. $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$, para quaisquer $x, y \in \mathbb{X}$ e qualquer $\alpha \in \mathbb{C}$;
4. $\langle x, x \rangle \geqslant 0$, para qualquer $x \in \mathbb{X}$, e $\langle x, x \rangle = 0 \Leftrightarrow x = 0$.

Consideremos dois exemplos de espaços Euclidianos complexos.

Exemplo 4.10. O espaço $\mathbb{C}^n$ de vetores complexos $x = (x^1, x^2, \ldots, x^n)$ com o produto escalar

$$\langle x, y \rangle = \sum_{k=1}^{n} x^k \bar{y}^k$$

é um espaço Euclidiano complexo. □

Exemplo 4.11. O espaço de polinómios $p : [a, b] \to \mathbb{C}$, $p(z) = a_0 z^n + a_1 z^{n-1} + \cdots + a_n$, de grau menor ou igual a n com coeficientes complexos, é um espaço Euclidiano complexo com o produto escalar

$$\langle p, q \rangle = \int_a^b p(z) \overline{q(z)} \, \mathrm{d}z.$$

□

Tal como no caso de espaços Euclidianos reais, para definir a norma Euclidiana precisamos da versão complexa da desigualdade de Cauchy-Schwarz

$$|\langle x, y \rangle|^2 \leqslant \langle x, x \rangle \langle y, y \rangle. \tag{4.8}$$

Se $\langle x, y \rangle = 0$, a desigualdade é óbvia. Suponhamos que $\langle x, y \rangle \neq 0$. Para todo $t \in \mathbb{C}$ temos

$$\langle tx - y, tx - y \rangle \geqslant 0,$$

ou seja,

$$|t|^2 \langle x, x \rangle - t \langle x, y \rangle - \overline{t \langle x, y \rangle} + \langle y, y \rangle \geqslant 0.$$

Fazendo $t = \frac{\tau \overline{\langle x, y \rangle}}{|\langle x, y \rangle|}$, onde $\tau \in \mathbb{R}$, obtemos $t \langle x, y \rangle = \tau |\langle x, y \rangle|$. Logo, $\overline{t \langle x, y \rangle} = t \langle x, y \rangle$. Daqui obtemos

$$\tau^2 \langle x, x \rangle - 2\tau |\langle x, y \rangle| + \langle y, y \rangle \geqslant 0$$

para qualquer $\tau \in \mathbb{R}$, e portanto, como no caso real, temos (4.8).

A norma Euclidiana de $x \in \mathbb{X}$ definimos por

$$|x| = \sqrt{\langle x, x\rangle}$$

(notemos que $\langle x, x\rangle \in \mathbb{R}$).

As duas primeiras propriedades da norma são consequências imediatas da definição. A única propriedade que precisa de demonstração é a desigualdade triangular que deduzimos da desigualdade de Cauchy-Schwarz:

$$\begin{aligned} |x+y|^2 &= \langle x+y, x+y\rangle = \langle x, x\rangle + \langle x, y\rangle + \overline{\langle x, y\rangle} + \langle y, y\rangle \\ &\leqslant \langle x, x\rangle + 2|\langle x, y\rangle| + \langle y, y\rangle \leqslant (|x| + |y|)^2. \end{aligned}$$

5. Aplicações lineares

5.1. Introdução

Neste capítulo introduzimos o objeto principal que estudamos neste livro — as aplicações lineares. As aplicações lineares transformam combinações lineares de vetores de um espaço linear em combinações lineares de vetores de outro espaço linear com os mesmos coeficientes. As aplicações lineares são muito importantes em Matemática. Por exemplo, as derivadas e integrais são aplicações lineares.

5.2. Aplicações lineares: definição e exemplos

Começamos por introduzir a definição de aplicação linear seguida da apresentação de alguns exemplos.

Definição 5.1. Aplicação linear

Sejam $\mathbb{X}$ e $\mathbb{Y}$ espaços lineares sobre um conjunto de números $\mathbb{K}$. A uma aplicação $A : \mathbb{X} \to \mathbb{Y}$ damos o nome de *aplicação linear* se

$$A(\alpha_1 x_1 + \alpha_2 x_2) = \alpha_1 A(x_1) + \alpha_2 A(x_2)$$

para todos os $x_1, x_2 \in \mathbb{X}$ e $\alpha_1, \alpha_2 \in \mathbb{K}$.

Vamos designar por $\mathcal{L}(\mathbb{X}, \mathbb{Y})$ o conjunto de todas as aplicações lineares de $\mathbb{X}$ em $\mathbb{Y}$ e designar também por Ax a imagem de x, isto é, $A(x) = Ax$.

Por indução, podemos imediatamente deduzir o seguinte teorema.

Teorema 5.1

Seja $A \in \mathcal{L}(\mathbb{X}, \mathbb{Y})$. Então

$$A\left(\sum_{j=1}^{m} \alpha_j x_j\right) = \sum_{j=1}^{m} \alpha_j A x_j,$$

para todos os $x_j \in \mathbb{X}$ e todos os $\alpha_j \in \mathbb{K}$, $j = \overline{1, m}$.

Eis alguns exemplos de aplicações lineares.

Exemplo 5.1. A aplicação identidade $I \in \mathcal{L}(\mathbb{X}, \mathbb{X})$, que faz corresponder a cada $x \in \mathbb{X}$ o mesmo vetor x.

□

Exemplo 5.2. A aplicação zero $0 \in \mathcal{L}(\mathbb{X}, \mathbb{Y})$, que transforma todo o $x \in \mathbb{X}$ no ponto zero do espaço $\mathbb{Y}$.

□

Exemplo 5.3. A derivação é uma aplicação linear no espaço $\mathcal{P}^n$ de todos os polinómios de grau menor ou igual a n: a cada polinómio

$$p(t) = a_0 t^n + a_1 t^{n-1} + \cdots + a_{n-1} t + a_n \in \mathcal{P}^n,$$

a aplicação $\frac{\mathrm{d}}{\mathrm{d}t}$ faz corresponder o polinómio

$$p'(t) = n a_0 t^{n-1} + (n-1) a_1 t^{n-2} + \cdots + a_{n-1}.$$

□

Vamos agora introduzir duas operações no conjunto das aplicações lineares.

Definição 5.2. Soma de aplicações lineares

Sejam $A, B \in \mathcal{L}(\mathbb{X}, \mathbb{Y})$. Então, definimos a sua *soma* por

$$(A + B)x = Ax + Bx,$$

para todo o $x \in \mathbb{X}$.

Definição 5.3. Multiplicação de uma aplicação linear por um número

Sejam $A \in \mathcal{L}(\mathbb{X}, \mathbb{Y})$ e $\lambda \in \mathbb{K}$. Então, definimos a *multiplicação* de A por λ por

$$(\lambda A)x = \lambda(Ax),$$

para todo o $x \in \mathbb{X}$.

Deixamos a demonstração do seguinte teorema para o leitor como um simples exercício (basta verificar que as operações de soma e de multiplicação por um número têm como resultado uma aplicação de $\mathcal{L}(\mathbb{X}, \mathbb{Y})$, e a seguir verificar todas as propriedades de espaço linear).

Teorema 5.2

O conjunto $\mathcal{L}(\mathbb{X}, \mathbb{Y})$ com as operações de soma e de multiplicação por um número dadas nas Definições 5.2 e 5.3, respetivamente, é um espaço linear.

Em muitos problemas é necessário trabalhar com mais do que dois espaços e com aplicações entre eles. Isto naturalmente leva à necessidade da definição seguinte.

Definição 5.4. Composição de aplicações lineares

Sejam $A \in \mathcal{L}(\mathbb{X}, \mathbb{Y})$ e $B \in \mathcal{L}(\mathbb{Y}, \mathbb{Z})$. Então, definimos a *composição de duas aplicações lineares* por

$$(BA)x = B(Ax).$$

A demonstração do seguinte teorema também deixamos para o leitor como um simples exercício (basta utilizar a definição de aplicação linear).

Teorema 5.3

A composição de duas aplicações lineares $A \in \mathcal{L}(\mathbb{X}, \mathbb{Y})$ e $B \in \mathcal{L}(\mathbb{Y}, \mathbb{Z})$ é uma aplicação linear de $\mathbb{X}$ em $\mathbb{Z}$.

5.3. Aplicações lineares e matrizes

A resolução de um problema que envolve uma aplicação linear começa pela introdução de bases nos respetivos espaços. Como veremos, definidas as duas bases, podemos estabelecer uma correspondência biunívoca entre a aplicação linear e uma matriz. Uma vez estabelecida esta correspondência, vamos desenvolver um cálculo matricial, que é uma ferramenta muito importante no estudo das propriedades das aplicações lineares.

Consideremos dois espaços lineares $\mathbb{X}$ e $\mathbb{Y}$ de dimensões n e m, respetivamente, e uma aplicação linear $A \in \mathcal{L}(\mathbb{X}, \mathbb{Y})$. Seja $\mathcal{E} = \{e_1, e_2, \ldots, e_n\}$ uma base em $\mathbb{X}$ e $\mathcal{F} = \{f_1, f_2, \ldots, f_m\}$ uma base em $\mathbb{Y}$. Cada vetor $Ae_k \in \mathbb{Y}$ pode ser representado na forma

$$Ae_k = \sum_{j=1}^{m} a_j^k f_j, \quad k = \overline{1,n}. \tag{5.1}$$

Os coeficientes a_j^k escrevemos na forma da matriz

$$A_{(\mathcal{E},\mathcal{F})} = \begin{pmatrix} a_1^1 & a_1^2 & \cdots & a_1^n \\ a_2^1 & a_2^2 & \cdots & a_2^n \\ \vdots & \vdots & \ddots & \vdots \\ a_m^1 & a_m^2 & \cdots & a_m^n \end{pmatrix}.$$

A esta matriz damos o nome de *matriz da aplicação linear* A relativamente às bases $\mathcal{E}$ e $\mathcal{F}$. As colunas desta matriz são formadas pelas coordenadas dos vetores $Ae_1, Ae_2, \ldots, Ae_n$ na base $\mathcal{F}$. Vamos designar a matriz $A_{(\mathcal{E},\mathcal{F})}$ simplesmente por A quando as bases estão fixas e vamos usar a notação $A = (a_i^j)$ para indicar as componentes que a matriz tem. Sendo A uma matriz com m linhas e n colunas, vamos passar a usar a notação $m \times n$ (que se lê "m por n") para representar as dimensões da matriz e denotar por $\mathcal{M}_{m\times n}$ o conjunto das matrizes com m linhas e n colunas.

Exemplo 5.4. Consideremos o espaço de polinómios $\mathcal{P}^n$ com coeficientes reais. Consideremos a base formada pelos monómios x^k, $k = \overline{0,n}$. Nesta base o polinómio

$$p(x) = a_0 x^n + a_1 x^{n-1} + \cdots + a_n$$

tem as coordenadas $(a_0, a_1, \ldots, a_n) \in \mathbb{R}^{n+1}$. Determinemos a matriz D da aplicação linear $\frac{\mathrm{d}}{\mathrm{d}x} : \mathcal{P}^n \to \mathcal{P}^n$. Como

$$\frac{\mathrm{d}}{\mathrm{d}x} p(x) = n a_0 x^{n-1} + (n-1) a_1 x^{n-2} + \cdots + a_{n-1},$$

então obtemos

$$D = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ n & 0 & \cdots & 0 & 0 \\ 0 & n-1 & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}.$$

□

Determinemos a ligação entre as coordenadas ξ^k, $k = \overline{1,n}$, do vetor x na base $\mathcal{E}$ e as coordenadas η^j, $j = \overline{1,m}$, do vetor $y = Ax$ na base $\mathcal{F}$.

Teorema 5.4

Sejam $A \in \mathcal{L}(\mathbb{X}, \mathbb{Y})$, $\mathcal{E} = \{e_1, e_2, \ldots, e_n\}$ uma base em $\mathbb{X}$ e $\mathcal{F} = \{f_1, f_2, \ldots, f_m\}$ uma base em $\mathbb{Y}$. Sejam, ainda, $x \in \mathbb{X}$ e $y = Ax \in \mathbb{Y}$ tais que

$$x = \sum_{k=1}^{n} \xi^k e_k, \quad y = \sum_{j=1}^{m} \eta^j f_j.$$

Então temos

$$\eta^j = \sum_{k=1}^{n} a_j^k \xi^k, \quad j = \overline{1,m}. \tag{5.2}$$

Demonstração. Como A é uma aplicação linear, pela fórmula (5.1) temos

$$\sum_{j=1}^{m} \eta^j f_j = y = A\left(\sum_{k=1}^{n} \xi^k e_k\right) = \sum_{k=1}^{n} \xi^k A e_k = \sum_{k=1}^{n} \xi^k \sum_{j=1}^{m} a_j^k f_j$$

$$= \sum_{j=1}^{m} \left(\sum_{k=1}^{n} a_j^k \xi^k\right) f_j.$$

Comparando as partes esquerda e direita desta igualdade, chegamos a (5.2). ■

Vamos agora aprofundar a relação entre aplicações lineares e matrizes associadas às operações lineares soma e multiplicação por um número.

Teorema 5.5

Sejam $\mathbb{X}$ e $\mathbb{Y}$ espaços lineares com bases $\mathcal{E} = \{e_1, e_2, \ldots, e_n\}$ e $\mathcal{F} = \{f_1, f_2, \ldots, f_m\}$, respetivamente. Então, a matriz $C = (c_i^j)$ associada à soma de duas aplicações lineares A e B de $\mathcal{L}(\mathbb{X}, \mathbb{Y})$ tem componentes

$$c_i^j = a_i^j + b_i^j \tag{5.3}$$

continua na página seguinte

continuação da página anterior

e a matriz $C = (c_i^j)$ associada ao produto de um número λ e uma aplicação linear A tem componentes

$$c_i^j = \lambda a_i^j, \tag{5.4}$$

onde a_i^j e b_i^j são as componentes das matrizes das aplicações A e B relativamente às bases $\mathcal{E}$ e $\mathcal{F}$.

Demonstração. Esta demonstração baseia-se na fórmula (5.1). A fórmula (5.3) é uma consequência das igualdades

$$(A+B)e_j = Ae_j + Be_j = \sum_{i=1}^{m} a_i^j f_i + \sum_{i=1}^{m} b_i^j f_i = \sum_{i=1}^{m} (a_i^j + b_i^j) f_i = \sum_{i=1}^{m} c_i^j f_i,$$

e a fórmula (5.4) é uma consequência das igualdades

$$(\lambda A)e_j = \lambda(Ae_j) = \lambda \sum_{i=1}^{m} a_i^j f_i = \sum_{i=1}^{m} \lambda a_i^j f_i = \sum_{i=1}^{m} c_i^j f_i.$$

■

Este último teorema leva naturalmente às definições das seguintes operações com matrizes.

Definição 5.5. Soma de duas matrizes

A *soma* de duas matrizes $A = (a_i^j)$ e $B = (b_i^j)$ de $\mathcal{M}_{m\times n}$, que denotamos por $A + B$, é a matriz $C = (c_i^j) \in \mathcal{M}_{m\times n}$ com $c_i^j = a_i^j + b_i^j$.

Queremos reforçar que só podemos somar matrizes com as mesmas dimensões.

Exemplo 5.5. Sejam $A = \begin{pmatrix} 1 & 2 \\ -1 & -1 \\ 5 & 0 \end{pmatrix}$ e $B = \begin{pmatrix} 4 & -2 \\ 3 & 7 \\ 2 & -1 \end{pmatrix}$. Então, $A + B = \begin{pmatrix} 5 & 0 \\ 2 & 6 \\ 7 & -1 \end{pmatrix}$. □

Definição 5.6. Multiplicação de uma matriz por um número

O *produto* de uma matriz $A = (a_i^j) \in \mathcal{M}_{m\times n}$ por um número λ, que denotamos por λA, é a matriz $C = (c_i^j) \in \mathcal{M}_{m\times n}$ com $c_i^j = \lambda a_i^j$.

Exemplo 5.6. Seja $A = \begin{pmatrix} 1 & 2 & 4 \\ -2 & 1 & 8 \end{pmatrix}$. Então, $3A = \begin{pmatrix} 3 & 6 & 12 \\ -6 & 3 & 24 \end{pmatrix}$. □

Com a ajuda destas operações é, por exemplo, muito fácil calcular as coordenadas do vetor $A_{(\mathcal{E},\mathcal{F})}x$ na base $\mathcal{F}$.

O leitor facilmente pode verificar as propriedades de espaço linear e desta maneira demonstrar o seguinte teorema.

Teorema 5.6

O conjunto $\mathcal{M}_{m\times n}$ com as operações de soma e multiplicação por um número é um espaço linear.

Vamos agora introduzir o conceito de *isomorfismo de espaços lineares* e que é muito útil na resolução de vários problemas.

Definição 5.7. Isomorfismo de espaços lineares

Sejam $\mathbb{X}$ e $\mathbb{Y}$ espaços lineares sobre o conjunto de números $\mathbb{K}$. Dizemos que os espaços $\mathbb{X}$ e $\mathbb{Y}$ são *isomorfos* se existe uma aplicação biunívoca $\phi : \mathbb{X} \to \mathbb{Y}$ que preserva as combinações lineares, ou seja, tal que

$$\phi(\alpha_1 x_1 + \alpha_2 x_2) = \alpha_1 \phi(x_1) + \alpha_2 \phi(x_2),$$

para todos os $x_1, x_2 \in \mathbb{X}$ e $\alpha_1, \alpha_2 \in \mathbb{K}$. À aplicação ϕ damos o nome de *isomorfismo*.

O conceito de isomorfismo é muito importante. Com efeito, podemos estar a trabalhar com espaços cujos elementos são de natureza muito diferentes mas, se os espaços forem isomorfos, do ponto de vista da teoria que estamos a desenvolver, são idênticos. Este facto permite-nos aplicar as mesmas técnicas matemáticas ao estudo de objetos muito distintos. Por exemplo, do Teorema 5.5, imediatamente temos o teorema seguinte.

Teorema 5.7

Os espaços lineares $\mathcal{L}(\mathbb{X}, \mathbb{Y})$ e $\mathcal{M}_{m\times n}$ são isomorfos.

A diferença entre $\mathcal{L}(\mathbb{X}, \mathbb{Y})$ e $\mathcal{M}_{m\times n}$ é muito grande. Com efeito, as aplicações lineares são funções que transformam vetores em vetores e as matrizes são tabelas de números. No entanto, o facto de serem isomorfos, vai ser muito útil.

Em muitos problemas é necessário considerar composições de aplicações lineares. Consideremos espaços lineares de dimensão finita $\mathbb{X}$, $\mathbb{Y}$ e $\mathbb{Z}$. Sejam $\mathcal{E} = \{e_1, \ldots, e_n\}$, $\mathcal{F} = \{f_1, \ldots, f_m\}$ e $\mathcal{G} = \{g_1, \ldots, g_l\}$ bases em $\mathbb{X}$, $\mathbb{Y}$ e $\mathbb{Z}$, respetivamente. Consideremos duas aplicações lineares $B \in \mathcal{L}(\mathbb{X}, \mathbb{Y})$ e $A \in \mathcal{L}(\mathbb{Y}, \mathbb{Z})$. À aplicação B corresponde a matriz com as componentes b_j^k, $j = \overline{1,m}$, $k = \overline{1,n}$, e à aplicação A corresponde a matriz com as componentes a_i^j, $i = \overline{1,l}$, $j = \overline{1,m}$. O leitor facilmente pode verificar que a aplicação $C : \mathbb{X} \to \mathbb{Z}$ definida como a composição das aplicações A e B, ou seja,

$$z = Cx = (AB)x = A(Bx), x \in \mathbb{X},$$

é uma aplicação linear.

Teorema 5.8

A matriz $C = (c_i^k)$ associada à composição AB das aplicações lineares A e B tem componentes

$$c_i^k = \sum_{j=1}^{m} a_i^j b_j^k. \tag{5.5}$$

Demonstração. A fórmula (5.5) é uma consequência das seguintes igualdades:

$$(AB)e_k = A(Be_k) = A\left(\sum_{j=1}^{m} b_j^k f_j\right) = \sum_{j=1}^{m} b_j^k A f_j = \sum_{j=1}^{m} b_j^k \sum_{i=1}^{l} a_i^j g_i$$

$$= \sum_{i=1}^{l}\left(\sum_{j=1}^{m} a_i^j b_j^k\right) g_i = \sum_{i=1}^{l} c_i^k g_i.$$

■

Este teorema leva naturalmente à definição da operação muito importante de multiplicação de matrizes.

Definição 5.8. Produto de duas matrizes

O *produto* de duas matrizes $A = (a_i^j) \in \mathcal{M}_{m\times n}$ e $B = (b_j^k) \in \mathcal{M}_{n\times l}$, que denotamos por AB, é a matriz $C = (c_i^k) \in \mathcal{M}_{m\times l}$ com $c_i^k = \sum_{j=1}^{n} a_i^j b_j^k$.

Queremos reforçar que só podemos multiplicar duas matrizes quando o número de colunas da primeira é igual ao número de linhas da segunda. Por exemplo, na fórmula (5.2) calculamos o produto da matriz $A_{(\mathcal{E},\mathcal{F})} \in \mathcal{M}_{m\times n}$ e do vetor-coluna com as coordenadas $(\xi^1, \xi^2, \ldots, \xi^n)$, que é uma matriz $n \times 1$:

$$\begin{pmatrix} \eta^1 \\ \eta^2 \\ \vdots \\ \eta^m \end{pmatrix} = \begin{pmatrix} a_1^1 & a_1^2 & \cdots & a_1^n \\ a_2^1 & a_2^2 & \cdots & a_2^n \\ \vdots & \vdots & \ddots & \vdots \\ a_m^1 & a_m^2 & \cdots & a_m^n \end{pmatrix} \begin{pmatrix} \xi^1 \\ \xi^2 \\ \vdots \\ \xi^n \end{pmatrix}.$$

O resultado é o vetor-coluna com as coordenadas $(\eta^1, \eta^2, \ldots, \eta^m)$, ou seja, uma matriz $m \times 1$.

Exemplo 5.7. Sejam $A = \begin{pmatrix} 1 & 2 \\ -1 & -1 \\ 5 & 0 \end{pmatrix}$ e $B = \begin{pmatrix} 2 & -1 \\ 3 & 2 \end{pmatrix}$. Então, $AB = \begin{pmatrix} 8 & 3 \\ -5 & -1 \\ 10 & -5 \end{pmatrix}$. □

A partir deste momento, vamos utilizar os termos *vetor-linha* e *vetor-coluna* para dizer que são matrizes com dimensões $1\times n$ e $n\times 1$, respetivamente. Estes objetos podem ser considerados como elementos do conjunto $\mathbb{K}^n$, mas do ponto de vista de operações com matrizes a diferença entre estes dois objetos é muito grande. Por exemplo, o produto de um vetor-linha por um vetor-coluna com n componentes é um número e o produto de um vetor-coluna por um vetor-linha é uma matriz com dimensões $n \times n$.

O esquema geral de resolução dos problemas que envolvem aplicações lineares pode ser apresentado por

$$\mathcal{L} \to \mathcal{M} \to \mathcal{L}.$$

De facto, graças ao Teorema 5.7, estando definidas as bases, às aplicações lineares podemos fazer corresponder as respetivas matrizes e utilizando o cálculo matricial fazer as contas necessárias. Podemos depois regressar às aplicações lineares para terminar a resolução do problema.

Como ilustração desta ideia, vamos ver um exemplo muito simples.

Exemplo 5.8. Consideremos num espaço Euclidiano de dimensão dois uma base ortonormal $\mathcal{E} = \{e_1, e_2\}$. Seja A a rotação no sentido anti-horário do espaço que tem centro na origem e amplitude $\frac{\pi}{2}$. Suponhamos que a base está orientada de maneira que $Ae_1 = e_2$. Queremos, agora, determinar o resultado da rotação do vetor $e_1 + 3e_2$. Na base introduzida a matriz da aplicação A é

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Então, recorrendo ao cálculo matricial, vemos que as coordenadas do vetor $A(e_1 + 3e_2)$ são

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} -3 \\ 1 \end{pmatrix}.$$

Logo, $A(e_1 + 3e_2) = -3e_1 + e_2$. □

5.4. Aplicações lineares de $\mathbb{X}$ em $\mathbb{X}$

Sejam $\mathbb{X}$ um espaço linear de dimensão n e $\mathcal{E} = \{e_1, e_2, \ldots, e_n\}$ uma base em $\mathbb{X}$. Consideremos, agora, aplicações lineares em que o conjunto de chegada é igual ao seu domínio. Seja, então, $A \in \mathcal{L}(\mathbb{X}, \mathbb{X})$. A matriz que corresponde a esta aplicação

$$A = \begin{pmatrix} a_1^1 & a_1^2 & \cdots & a_1^n \\ a_2^1 & a_2^2 & \cdots & a_2^n \\ \vdots & \vdots & \ddots & \vdots \\ a_n^1 & a_n^2 & \cdots & a_n^n \end{pmatrix}$$

está determinada pelas igualdades

$$Ae_k = \sum_{j=1}^{n} a_j^k e_j, \quad k = \overline{1,n}.$$

Para as coordenadas do vetor $y = Ax$ relativamente à base $\mathcal{E}$ ($y = \sum_{j=1}^{n} \eta^j e_j$), temos

$$\eta^j = \sum_{k=1}^{n} a_j^k \xi^k, \quad j = \overline{1,n},$$

(veja (5.2)) onde ξ^k, $k = \overline{1,n}$, são as coordenadas do vetor x relativamente à mesma base $\mathcal{E}$ ($x = \sum_{k=1}^{n} \xi^k e_k$).

Apresentemos agora alguns exemplos.

Exemplo 5.9. Sejam $\{e_1, e_2, \ldots, e_n\}$ uma base do espaço linear $\mathbb{X}$ e os números $\lambda_i \in \mathbb{K}$, $i = \overline{1,n}$. Consideremos a aplicação linear $A \in \mathcal{L}(\mathbb{X}, \mathbb{X})$ definida por

$$Ae_k = \lambda_k e_k, \quad k = \overline{1,n}.$$

A matriz desta aplicação tem a forma

$$A = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix}.$$

A uma aplicação (e à matriz correspondente) deste tipo chamamos *diagonal*. □

Exemplo 5.10. Seja $\mathbb{X}$ um espaço linear. Vamos designar por I a *aplicação identidade* definida como

$$Ix = x,$$

para todo $x \in \mathbb{X}$. Obviamente

$$AI = IA = A,$$

para qualquer $A \in \mathcal{L}(\mathbb{X}, \mathbb{X})$. A esta aplicação corresponde a *matriz identidade*

$$I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}.$$

□

Exemplo 5.11. Seja $A \in \mathcal{L}(\mathbb{X}, \mathbb{X})$. À aplicação $P \in \mathcal{L}(\mathbb{X}, \mathbb{X})$ damos o nome de *aplicação inversa esquerda* da aplicação A se

$$PA = I.$$

À aplicação $Q \in \mathcal{L}(\mathbb{X}, \mathbb{X})$ damos o nome de *aplicação inversa direita* da aplicação A se

$$AQ = I.$$

□

O seguinte teorema estabelece a propriedade mais importante destas duas últimas aplicações.

Teorema 5.9

Sejam $A \in \mathcal{L}(\mathbb{X},\mathbb{X})$, $P \in \mathcal{L}(\mathbb{X},\mathbb{X})$ uma aplicação inversa esquerda de A e $Q \in \mathcal{L}(\mathbb{X},\mathbb{X})$ uma aplicação inversa direita de A. Então, $P = Q$ e estas aplicações estão definidas de modo único.

Demonstração. Com efeito, temos

$$P = PI = P(AQ) = (PA)Q = IQ = Q.$$

Sejam P_1 e P_2 duas aplicações inversas esquerdas de A e Q uma aplicação inversa direita. Então, como já demonstramos, $P_1 = Q$ e $P_2 = Q$. Portanto, $P_1 = P_2$. A unicidade da aplicação inversa direita demonstra-se da mesma maneira. ■

Deste teorema vemos que a existência de aplicações inversas à direita e à esquerda implica a existência de uma única aplicação $A^{-1} = P = Q$. Temos então a seguinte definição.

Definição 5.9. Aplicação inversa e aplicação linear invertível

Seja $A \in \mathcal{L}(\mathbb{X},\mathbb{X})$ tal que admite aplicações inversas à direita e à esquerda. Então, à aplicação linear A^{-1} chamamos *aplicação inversa* de A. A aplicação A^{-1} verifica as condições

$$A^{-1}A = AA^{-1} = I.$$

A uma aplicação linear que admite inversa chamamos *aplicação linear invertível*.

Exemplo 5.12. A aplicação $A : \mathbb{R}^n \to \mathbb{R}^n$ com a matriz diagonal

$$A = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix}$$

com $\lambda_k \neq 0$, $k = \overline{1,n}$, é invertível e a matriz da aplicação inversa é

$$A^{-1} = \begin{pmatrix} \lambda_1^{-1} & 0 & \cdots & 0 \\ 0 & \lambda_2^{-1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n^{-1} \end{pmatrix}.$$

□

Sejam A e B aplicações lineares de $\mathbb{X}$ em $\mathbb{X}$. Então, a composição destas aplicações, AB, também é uma aplicação linear de $\mathbb{X}$ em $\mathbb{X}$. A composição de duas aplicações pode ser vista como um produto que tem as propriedades de associatividade $A(BC) = (AB)C$ e de distributividade $(A+B)C = AC+BC$ e $C(A+B) = CA+CB$.

(O leitor facilmente pode deduzir estas propriedades diretamente das definições.) A um espaço linear onde está definido o produto dos elementos com estas duas propriedades chamamos *álgebra*. Obviamente $\mathcal{L}(\mathbb{X},\mathbb{X})$ e $\mathcal{M}_{n\times n}$ são álgebras. Duas álgebras são isomorfas se o isomorfismo ϕ dos respetivos espaços lineares preserva também o produto:

$$\phi(x_1x_2)=\phi(x_1)\phi(x_2).$$

Por exemplo, do Teorema 5.7 e do Teorema 5.8 imediatamente temos o teorema seguinte.

Teorema 5.10

As álgebras $\mathcal{L}(\mathbb{X},\mathbb{X})$ e $\mathcal{M}_{n\times n}$ são isomorfas sempre que $\dim(\mathbb{X})=n$.

Notemos que as álgebras $\mathcal{L}(\mathbb{X},\mathbb{X})$ e $\mathcal{M}_{n\times n}$ não são comutativas, isto é, no caso geral $AB\neq BA$, para $A,B\in\mathcal{L}(\mathbb{X},\mathbb{X})$ ou $A,B\in\mathcal{M}_{n\times n}$. De facto, consideremos num espaço Euclidiano de dimensão dois uma base ortonormal $\mathcal{E}=\{e_1,e_2\}$. Sejam A a simetria de reflexão em relação à bissetriz do ângulo formado pelos vetores da base e B a rotação no sentido anti-horário do espaço que tem centro na origem e amplitude $\frac{\pi}{2}$. Suponhamos que a base está orientada de maneira que $Be_1=e_2$. Então temos

$$ABe_1=e_1 \quad \text{e} \quad BAe_1=-e_1.$$

Obviamente as respetivas matrizes são

$$A=\begin{pmatrix}0&1\\1&0\end{pmatrix} \quad \text{e} \quad B=\begin{pmatrix}0&-1\\1&0\end{pmatrix},$$

pelo que

$$AB=\begin{pmatrix}1&0\\0&-1\end{pmatrix} \quad \text{e} \quad BA=\begin{pmatrix}-1&0\\0&1\end{pmatrix}.$$

Pretendemos agora determinar a matriz da aplicação inversa no caso geral. Seja A uma aplicação linear invertível cujos domínio e conjunto de chegada são o mesmo espaço linear de dimensão n. Vamos designar a matriz desta aplicação e a matriz da sua inversa, numa certa base, por A e A^{-1}, respetivamente. A existência da aplicação inversa implica que a equação matricial $AX=I$ tem uma única solução, a matriz $X\in\mathcal{M}_{n\times n}$. Designemos as colunas da matriz X por x_k, $k=\overline{1,n}$. Então, a equação $AX=I$ é equivalente ao conjunto de equações $Ax_k=e_k$, $k=\overline{1,n}$, onde e_k é a k-ésima coluna da matriz I. A cada uma destas equações podemos aplicar o Método de Gauss. As transformações que fazemos com a matriz A não dependem do vetor e_k. Portanto, podemos fazer os cálculos simultaneamente para todos os $k=\overline{1,n}$. Para fazer isso começamos por construir a matriz de dimensões $n\times 2n$, cujas n primeiras colunas são as colunas da matriz A, sendo as restantes as colunas da matriz identidade. Vamos denotar esta matriz por $(A|I)$. Aplicamos de seguida a primeira fase do Método de Gauss por forma a transformar a matriz $(A|I)$ numa matriz em escada.

Para determinar as colunas da matriz inversa temos que aplicar a segunda fase do Método de Gauss n vezes.

Vamos agora ilustrar o procedimento acabado de descrever. Seja A uma aplicação linear cuja matriz em certa base é

$$A = \begin{pmatrix} 1 & 0 & 2 \\ 2 & -1 & 3 \\ 4 & 1 & 8 \end{pmatrix}.$$

Construimos então a matriz $(A|I)$:

$$\begin{pmatrix} 1 & 0 & 2 & 1 & 0 & 0 \\ 2 & -1 & 3 & 0 & 1 & 0 \\ 4 & 1 & 8 & 0 & 0 & 1 \end{pmatrix}.$$

Depois das transformações $l_2 \leftarrow l_2 - 2l_1$ e $l_3 \leftarrow l_3 - 4l_1$ obtemos a matriz

$$\begin{pmatrix} 1 & 0 & 2 & 1 & 0 & 0 \\ 0 & -1 & -1 & -2 & 1 & 0 \\ 0 & 1 & 0 & -4 & 0 & 1 \end{pmatrix}.$$

Aplicando a transformação $l_3 \leftarrow l_3 + l_2$ obtemos

$$\begin{pmatrix} 1 & 0 & 2 & 1 & 0 & 0 \\ 0 & -1 & -1 & -2 & 1 & 0 \\ 0 & 0 & -1 & -6 & 1 & 1 \end{pmatrix}.$$

A primeira fase do Método de Gauss está terminada uma vez que a matriz $(A|I)$ já está na forma em escada. Estamos então em condições de determinar as colunas da matriz inversa. Os elementos da primeira coluna da matriz inversa $(a^{-1})^1_1$, $(a^{-1})^1_2$ e $(a^{-1})^1_3$ são a solução do sistema linear

$$\begin{pmatrix} 1 & 0 & 2 & 1 \\ 0 & -1 & -1 & -2 \\ 0 & 0 & -1 & -6 \end{pmatrix},$$

vindo $(a^{-1})^1_3 = 6$, $(a^{-1})^1_2 = -4$ e $(a^{-1})^1_1 = -11$. Os elementos da segunda coluna da matriz inversa $(a^{-1})^2_1$, $(a^{-1})^2_2$ e $(a^{-1})^2_3$ são a solução do sistema linear

$$\begin{pmatrix} 1 & 0 & 2 & 0 \\ 0 & -1 & -1 & 1 \\ 0 & 0 & -1 & 1 \end{pmatrix},$$

vindo $(a^{-1})^2_3 = -1$, $(a^{-1})^2_2 = 0$ e $(a^{-1})^2_1 = 2$. Os elementos da terceira coluna da matriz inversa $(a^{-1})^3_1$, $(a^{-1})^3_2$ e $(a^{-1})^3_3$ são a solução do sistema linear

$$\begin{pmatrix} 1 & 0 & 2 & 0 \\ 0 & -1 & -1 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix}.$$

vindo $(a^{-1})_3^3 = -1$, $(a^{-1})_2^3 = 1$ e $(a^{-1})_1^3 = 2$. Temos, então, que

$$A^{-1} = \begin{pmatrix} -11 & 2 & 2 \\ -4 & 0 & 1 \\ 6 & -1 & -1 \end{pmatrix}.$$

Uma resolução alternativa para determinar a matriz inversa é recorrer ao Método de Gauss-Jordan onde, como sabemos da Secção 3.3, a matriz $(A|b)$, correspondente ao sistema $Ax = b$, é transformada na matriz $(I|A^{-1}b)$. Portanto, aplicando este método à matriz $(A|I)$ obtemos a matriz $(I|A^{-1})$.

Consideremos, então, novamente a matriz

$$A = \begin{pmatrix} 1 & 0 & 2 \\ 2 & -1 & 3 \\ 4 & 1 & 8 \end{pmatrix}.$$

Construimos a matriz $(A|I)$:

$$\begin{pmatrix} 1 & 0 & 2 & 1 & 0 & 0 \\ 2 & -1 & 3 & 0 & 1 & 0 \\ 4 & 1 & 8 & 0 & 0 & 1 \end{pmatrix}.$$

Depois das transformações $l_2 \leftarrow l_2 - 2l_1$ e $l_3 \leftarrow l_3 - 4l_1$ obtemos a matriz

$$\begin{pmatrix} 1 & 0 & 2 & 1 & 0 & 0 \\ 0 & -1 & -1 & -2 & 1 & 0 \\ 0 & 1 & 0 & -4 & 0 & 1 \end{pmatrix}.$$

Aplicando a transformação $l_3 \leftarrow l_3 + l_2$ obtemos

$$\begin{pmatrix} 1 & 0 & 2 & 1 & 0 & 0 \\ 0 & -1 & -1 & -2 & 1 & 0 \\ 0 & 0 & -1 & -6 & 1 & 1 \end{pmatrix}.$$

A matriz A já está na forma em escada. Pretendemos agora transformar a matriz

$$\begin{pmatrix} 1 & 0 & 2 \\ 0 & -1 & -1 \\ 0 & 0 & -1 \end{pmatrix}$$

na matriz identidade. Para tal, continuamos a recorrer às operações com linhas da matriz obtida. Fazendo $l_2 \leftarrow -l_2$ e $l_3 \leftarrow -l_3$ chegamos à matriz

$$\begin{pmatrix} 1 & 0 & 2 & 1 & 0 & 0 \\ 0 & 1 & 1 & 2 & -1 & 0 \\ 0 & 0 & 1 & 6 & -1 & -1 \end{pmatrix}.$$

Finalmente depois das transformações $l_1 \leftarrow l_1 - 2l_3$ e $l_2 \leftarrow l_2 - l_3$ obtemos

$$\begin{pmatrix} 1 & 0 & 0 & -11 & 2 & 2 \\ 0 & 1 & 0 & -4 & 0 & 1 \\ 0 & 0 & 1 & 6 & -1 & -1 \end{pmatrix}.$$

Portanto,

$$A^{-1} = \begin{pmatrix} -11 & 2 & 2 \\ -4 & 0 & 1 \\ 6 & -1 & -1 \end{pmatrix}.$$

Como já dissemos na Secção 3.3, este método não tem nenhuma vantagem fundamental quando comparado com a abordagem baseada no Método de Gauss. No entanto, em vez da abordagem sequencial da segunda fase do Método de Gauss, em que obtemos uma a uma as colunas da matriz inversa, na segunda fase do Método de Gauss-Jordan fazemos as contas em simultâneo, onde usamos o mesmo tipo de operações sobre as linhas como na primeira fase do método, até obter a matriz inversa. Ao aplicarmos sempre o mesmo tipo de operações, as resoluções deste tipo de exercícios criam, aos olhos humanos, uma boa ilusão de ser uma abordagem mais simples.

5.5. Operações elementares com matrizes

No Método de Gauss-Jordan fazemos com as linhas da matriz $(A|I)$ as seguintes operações elementares:

1. Multiplicamos uma linha por um número;
2. Adicionamos uma linha a outra linha;
3. Trocamos duas linhas.

Todas estas operações são possíveis de escrever em termos de multiplicações de matrizes:

1. A multiplicação da linha i da matriz A por um número λ é equivalente à multiplicação à esquerda da matriz A pela matriz $M(i,\lambda)$, que é a matriz I que na linha i e coluna i em vez de 1 tem λ:

$$M(i,\lambda) = \begin{array}{c} 1 \\ \vdots \\ i-1 \\ i \\ i+1 \\ \vdots \\ n \end{array} \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & 1 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \lambda & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & 1 \end{pmatrix};$$

2. A adição da linha j à linha i é equivalente à multiplicação à esquerda da matriz A pela matriz $S(i,j)$, que é a matriz I que na linha i e coluna j em vez de 0 tem 1:

$$S(i,j) = \begin{array}{c} \\ 1 \\ \vdots \\ i-1 \\ i \\ i+1 \\ \vdots \\ \vdots \\ n \end{array} \begin{array}{c} \begin{array}{cccccccc} & & & & j & & & \end{array} \\ \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & 1 \end{pmatrix} \end{array}.$$

3. A troca das linhas i e j é equivalente à multiplicação à esquerda da matriz A pela matriz $T(i,j)$, que é a matriz I com as linhas i e j trocadas:

$$T(i,j) = \begin{array}{c} \\ 1 \\ \vdots \\ \vdots \\ i \\ \vdots \\ \vdots \\ \vdots \\ j \\ \vdots \\ \vdots \\ n \end{array} \begin{array}{c} \begin{array}{ccccccccccc} & & & & i & & & j & & & \end{array} \\ \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & 1 & \ddots & \ddots & \ddots & 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 & \ddots & \ddots & \ddots & 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 1 & \ddots & \ddots & \ddots & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & 0 & \ddots & \ddots & \ddots & 1 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 & \ddots & \ddots & \ddots & 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 & \ddots & \ddots & \ddots & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & 1 \end{pmatrix} \end{array}.$$

A multiplicação à direita da matriz A pelas matrizes $M(i,\lambda)$, $S(i,j)$ e $T(i,j)$, dá, respetivamente, a multiplicação da coluna i pelo número λ, a adição da coluna i à coluna j e a troca das colunas i e j.

Estas observações são muito importantes em termos teóricos e práticos.

5.6. Mudança de coordenadas

Sejam $\mathcal{E} = \{e_1, e_2, \ldots, e_n\}$ e $\mathcal{F} = \{f_1, f_2, \ldots, f_n\}$ duas bases do espaço linear $\mathbb{X}$. Para os vetores f_k, $k = \overline{1,n}$, temos a representação

$$f_k = \sum_{j=1}^{n} p_j^k e_j, \quad k = \overline{1,n}. \tag{5.6}$$

À matriz P de componentes p_j^k, $j,k = \overline{1,n}$, isto é, a matriz cujas colunas contêm as coordenadas dos vetores f_k na base $\mathcal{E}$, chamamos *matriz de transição* da base $\mathcal{E}$ para a base $\mathcal{F}$. Esta matriz é invertível, pois caso contrário existiria uma dependência linear entre as suas colunas e portanto entre os vetores f_k, $k = \overline{1,n}$. À aplicação linear P definida por

$$f_k = Pe_k, \quad k = \overline{1,n},$$

damos o nome de *aplicação de transição* da base $\mathcal{E}$ para a base $\mathcal{F}$. A sua matriz coincide com a matriz P.

Seja P uma matriz invertível. Construamos vetores f_k, $k = \overline{1,n}$, utilizando as fórmulas (5.6). Obviamente estes vetores são linearmente independentes e logo formam uma base em $\mathbb{X}$. Portanto, cada matriz invertível define segundo as fórmulas (5.6) uma transição de uma base para outra.

Vamos agora ver como se transformam as coordenadas de um vetor quando passamos da base $\mathcal{E}$ para a base $\mathcal{F}$. Para todo o vetor $x \in \mathbb{X}$ temos

$$x = \sum_{k=1}^{n} \xi^k e_k = \sum_{j=1}^{n} \eta^j f_j. \tag{5.7}$$

Suponhamos que as coordenadas $(\xi^1, \xi^2, \ldots, \xi^n)$ do vetor x na base $\mathcal{E}$ são dadas. Pretendemos encontrar as coordenadas $(\eta^1, \eta^2, \ldots, \eta^n)$ do vetor x na base $\mathcal{F}$. Designemos por q_i^j, $i,j = \overline{1,n}$, as componentes da matriz P^{-1}. Então temos

$$e_k = \sum_{j=1}^{n} q_j^k f_j, \quad k = \overline{1,n}.$$

Substituindo estas expressões na fórmula (5.7) temos

$$x = \sum_{k=1}^{n} \xi^k e_k = \sum_{k=1}^{n} \xi^k \sum_{j=1}^{n} q_j^k f_j = \sum_{j=1}^{n} \left(\sum_{k=1}^{n} q_j^k \xi^k \right) f_j = \sum_{j=1}^{n} \eta^j f_j.$$

Logo, obtemos

$$\eta^j = \sum_{k=1}^{n} q_j^k \xi^k, \quad j = \overline{1,n}.$$

Podemos reescrever esta fórmula na forma matricial

$$\eta = P^{-1}\xi.$$

Portanto, o vetor-coluna η composto pelas coordenadas $(\eta^1, \eta^2, \ldots, \eta^n)$ é o produto da matriz P^{-1} e do vetor-coluna ξ composto pelas coordenadas $(\xi^1, \xi^2, \ldots, \xi^n)$.

Vamos agora ver como se transformam os elementos da matriz de uma aplicação linear quando passamos de uma base para uma outra base. Seja $A \in \mathcal{L}(\mathbb{X}, \mathbb{X})$ uma aplicação linear de um espaço de dimensão n em si mesmo. Consideremos duas bases $\mathcal{E} = \{e_1, e_2, \ldots, e_n\}$ e $\mathcal{F} = \{f_1, f_2, \ldots, f_n\}$ em $\mathbb{X}$. Designemos por P a matriz de transição da base $\mathcal{E}$ para a base $\mathcal{F}$. A matriz da aplicação A na base $\mathcal{E}$ é designada por

$$A_{\mathcal{E}} = A_{(\mathcal{E},\mathcal{E})} = \begin{pmatrix} (a_{\mathcal{E}})_1^1 & \cdots & (a_{\mathcal{E}})_1^n \\ \vdots & \ddots & \vdots \\ (a_{\mathcal{E}})_n^1 & \cdots & (a_{\mathcal{E}})_n^n \end{pmatrix}$$

e a matriz na base $\mathcal{F}$ por

$$A_{\mathcal{F}} = A_{(\mathcal{F},\mathcal{F})} = \begin{pmatrix} (a_{\mathcal{F}})_1^1 & \cdots & (a_{\mathcal{F}})_1^n \\ \vdots & \ddots & \vdots \\ (a_{\mathcal{F}})_n^1 & \cdots & (a_{\mathcal{F}})_n^n \end{pmatrix}.$$

Encontremos a relação entre $A_{\mathcal{E}}$, $A_{\mathcal{F}}$ e P. As componentes das matrizes $A_{\mathcal{E}}$ e $A_{\mathcal{F}}$ determinam-se pelas igualdades

$$Ae_i = \sum_{j=1}^{n} (a_{\mathcal{E}})_j^i e_j, \quad i = \overline{1,n}, \tag{5.8}$$

e

$$Af_m = \sum_{k=1}^{n} (a_{\mathcal{F}})_k^m f_k, \quad m = \overline{1,n}, \tag{5.9}$$

respetivamente. Substituindo f_k em (5.9) por (5.6), obtemos

$$Af_m = \sum_{k=1}^{n} (a_{\mathcal{F}})_k^m \sum_{j=1}^{n} p_j^k e_j. \tag{5.10}$$

Por outro lado, pela fórmula (5.6), onde mudamos o índice do somatório de j para i, temos

$$Af_m = A \sum_{i=1}^{n} p_i^m e_i = \sum_{i=1}^{n} p_i^m Ae_i.$$

Combinando esta igualdade com (5.8) obtemos

$$Af_m = \sum_{i=1}^{n} p_i^m \sum_{j=1}^{n} (a_{\mathcal{E}})_j^i e_j. \tag{5.11}$$

Comparando (5.10) e (5.11) vemos que

$$\sum_{k=1}^{n} (a_{\mathcal{F}})_k^m p_j^k = \sum_{i=1}^{n} p_i^m (a_{\mathcal{E}})_j^i,$$

quaisquer que sejam $m, j = \overline{1,n}$, ou na forma matricial

$$PA_{\mathcal{F}} = A_{\mathcal{E}} P. \tag{5.12}$$

Logo, temos

$$A_{\mathcal{F}} = P^{-1} A_{\mathcal{E}} P.$$

Consideremos um exemplo de mudança de base.

Exemplo 5.13. Seja $A \in \mathcal{L}(\mathbb{R}^3, \mathbb{R}^3)$ tal que a sua matriz na base $e_1 = (-2, 1, 2)$, $e_2 = (1, 2, 1)$, $e_3 = (1, -1, 1)$ é

$$A_{\mathcal{E}} = \begin{pmatrix} 1 & -3 & 2 \\ 1 & 0 & 4 \\ 2 & -1 & 2 \end{pmatrix}.$$

Vamos calcular a matriz $A_{\mathcal{F}}$ de A na base $f_1 = (1, 4, 5)$, $f_2 = (4, 0, 0)$, $f_3 = (-6, 0, 6)$. Sejam E e F as matrizes cujas colunas são os vetores das bases $\mathcal{E}$ e $\mathcal{F}$, ou seja,

$$E = \begin{pmatrix} -2 & 1 & 1 \\ 1 & 2 & -1 \\ 2 & 1 & 1 \end{pmatrix}, \quad F = \begin{pmatrix} 1 & 4 & -6 \\ 4 & 0 & 0 \\ 5 & 0 & 6 \end{pmatrix}.$$

Encontrando a matriz de transição da base $\mathcal{E}$ para a base $\mathcal{F}$

$$P = E^{-1} F = \begin{pmatrix} 1 & -1 & 3 \\ 2 & 1 & -1 \\ 1 & 1 & 1 \end{pmatrix},$$

obtemos

$$A_{\mathcal{F}} = P^{-1} A_{\mathcal{E}} P = \frac{1}{8} \begin{pmatrix} 10 & 10 & 26 \\ 13 & -7 & 25 \\ -7 & -11 & 21 \end{pmatrix}.$$

□

5.7. Forma geral de uma função linear

Sejam $\mathbb{X}$ um espaço linear de dimensão n e $L \in \mathcal{L}(\mathbb{X}, \mathbb{R})$. A esta aplicação linear chamamos *função linear*. Consideremos uma base $\{e_1, \ldots, e_n\}$ em $\mathbb{X}$. Designemos os valores $L(e_k)$ por l^k, $k = \overline{1, n}$. Então, para cada $x = \sum_{k=1}^{n} \xi^k e_k \in \mathbb{X}$ temos

$$L(x) = L\left(\sum_{k=1}^{n} \xi^k e_k\right) = \sum_{k=1}^{n} \xi^k L(e_k) = \sum_{k=1}^{n} \xi^k l^k.$$

Portanto, num espaço Euclidiano existe uma correspondência biunívoca entre os vetores do espaço, x^*, e as funções lineares, $L(x)$, dada pelo produto escalar

$$L(x) = \langle x^*, x \rangle.$$

5.8. Aplicações adjuntas e autoadjuntas (simétricas)

Sejam $\mathbb{X}$ e $\mathbb{Y}$ espaços Euclidianos reais de dimensões n e m, respetivamente, e $A \in \mathcal{L}(\mathbb{X}, \mathbb{Y})$ uma aplicação linear. Consideremos nos espaços $\mathbb{X}$ e $\mathbb{Y}$ bases ortonormais. Seja $y^* \in \mathbb{Y}$. Obviamente a aplicação $x \to \langle y^*, Ax\rangle$ é uma função linear. Logo, (veja a secção anterior) existe $x^* \in \mathbb{X}$ tal que

$$\langle x^*, x\rangle = \langle y^*, Ax\rangle. \tag{5.13}$$

Assim, a cada $y^* \in \mathbb{Y}$ corresponde um vetor $x^* \in \mathbb{X}$, pelo que temos uma aplicação A^* que faz corresponder a um vetor y^* um vetor x^*. Obviamente esta aplicação $A^* : \mathbb{Y} \to \mathbb{X}$ é linear. Esta observação leva-nos à seguinte definição.

Definição 5.10. Aplicação adjunta

À aplicação linear $A^* \in \mathcal{L}(\mathbb{Y}, \mathbb{X})$ que faz corresponder a cada $y^* \in \mathbb{Y}$ um vetor $x^* = A^*y^* \in \mathbb{X}$, tal que para cada $x \in \mathbb{X}$ se verifica (5.13), chamamos *aplicação adjunta* da aplicação A.

Vamos agora definir a matriz transposta, que tem um papel muito importante na teoria das aplicações lineares como veremos a seguir e que está relacionada com a aplicação adjunta.

Definição 5.11. Matriz transposta

A *transposta* da matriz $X = (x_i^j)$ de $\mathcal{M}_{m\times n}$, que denotamos por X^T, é a matriz $Y = (y_i^j)$ de $\mathcal{M}_{n\times m}$ com $y_i^j = x_j^i$, ou seja a matriz que obtemos a partir de X trocando as linhas com as colunas.

A operação de transposição de matrizes tem a seguinte propriedade relevante.

Teorema 5.11

Sejam $A \in \mathcal{M}_{m\times n}$ e $B \in \mathcal{M}_{n\times l}$ duas matrizes. Então temos $(AB)^\mathsf{T} = B^\mathsf{T}A^\mathsf{T}$.

Demonstração. Seja $C = AB$. Então

$$c_i^k = \sum_{j=1}^{n} a_i^j b_j^k.$$

Designemos por $(a^\mathsf{T})_i^j$, $(b^\mathsf{T})_j^k$ e $(c^\mathsf{T})_i^k$ os elementos das matrizes transpostas A^T, B^T e C^T, respetivamente. Como $(a^\mathsf{T})_i^j = a_j^i$, $(b^\mathsf{T})_j^k = b_k^j$ e $(c^\mathsf{T})_i^k = c_k^i$, temos

$$(c^\mathsf{T})_i^k = c_k^i = \sum_{j=1}^{n} a_k^j b_j^i = \sum_{j=1}^{n} (b^\mathsf{T})_i^j (a^\mathsf{T})_j^k.$$

A última soma é o elemento da matriz $B^\mathsf{T}A^\mathsf{T}$ que está na linha i e na coluna k. ■

Vamos encontrar a matriz que corresponde à aplicação adjunta.

Teorema 5.12

Seja $A \in \mathcal{L}(\mathbb{X}, \mathbb{Y})$. A matriz que corresponde à aplicação A^* é a matriz transposta da matriz da aplicação A.

Demonstração. Sejam a_i^j os elementos da matriz correspondente à aplicação $A \in \mathcal{L}(\mathbb{X}, \mathbb{Y})$. Então, a igualdade $y = Ax$ é equivalente ao sistema

$$y^i = \sum_{j=1}^{n} a_i^j x^j, \quad i = \overline{1, m}.$$

A função linear associada ao vetor x^* tem a forma

$$\langle x^*, x \rangle = \sum_{j=1}^{n} (x^*)^j x^j.$$

A igualdade $\langle x^*, x \rangle = \langle y^*, Ax \rangle$ toma a forma

$$\sum_{i=1}^{m} (x^*)^i x^i = \sum_{i=1}^{m} (y^*)^i y^i = \sum_{i=1}^{m} \sum_{j=1}^{n} (y^*)^i a_i^j x^j = \sum_{j=1}^{n} x^j \sum_{i=1}^{m} a_i^j (y^*)^i.$$

Logo, obtemos

$$(x^*)^j = \sum_{i=1}^{m} a_i^j (y^*)^i,$$

isto é, a matriz que corresponde à aplicação A^* é a matriz transposta da matriz da aplicação A.

■

O seguinte teorema, cuja demonstração vem diretamente da definição de aplicação adjunta, dá-nos as propriedades principais da aplicação adjunta.

Teorema 5.13

Sejam $\mathbb{X}$ um espaço Euclidiano, $A, B \in \mathcal{L}(\mathbb{X}, \mathbb{X})$ e α um número. Então:

1. $(A + B)^* = A^* + B^*$;
2. $(\alpha A)^* = \alpha A^*$;
3. $(AB)^* = B^* A^*$.

A igualdade (5.13) pode ser reescrita como

$$\langle A^* y^*, x \rangle = \langle y^*, Ax \rangle. \tag{5.14}$$

Em coordenadas, utilizando matrizes correspondentes às aplicações lineares A e A^*, esta igualdade toma a forma

$$(A^{\mathsf{T}} y)^{\mathsf{T}} x = y^{\mathsf{T}} A x$$

ou

$$\langle A^{\mathsf{T}} y, x\rangle = \langle y, Ax\rangle.$$

Vamos agora definir uma classe de aplicações que, como iremos ver, têm propriedades muito interessantes.

Definição 5.12. Aplicação autoadjunta ou simétrica

Seja $\mathbb{X}$ um espaço Euclidiano. Dizemos que uma aplicação linear $A \in \mathcal{L}(\mathbb{X}, \mathbb{X})$ é *autoadjunta ou simétrica* se $A^* = A$.

Às aplicações autoadjuntas, segundo o Teorema 5.12, correspondem *matrizes simétricas*, ou seja, as matrizes que verificam a condição $A^{\mathsf{T}} = A$.

Consideremos alguns exemplos.

Exemplo 5.14. Voltemos ao Exemplo 5.4 onde a base no espaço $\mathcal{P}^n$ é formada pelos monómios x^k, $k = \overline{0,n}$. Definamos o produto escalar de dois polinómios

$$p(x) = a_0 x^n + a_1 x^{n-1} + \cdots + a_n$$

e

$$q(x) = b_0 x^n + b_1 x^{n-1} + \cdots + b_n$$

como

$$\sum_{i=0}^{n} a_i b_i.$$

A matriz D da aplicação linear $\frac{\mathrm{d}}{\mathrm{d}x}$ é

$$D = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ n & 0 & \cdots & 0 & 0 \\ 0 & n-1 & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix};$$

À aplicação adjunta corresponde a matriz transposta

$$D^{\mathsf{T}} = \begin{pmatrix} 0 & n & 0 & \cdots & 0 \\ 0 & 0 & n-1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}.$$

Logo, temos

$$\left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^* p(x) = n a_1 x^n + (n-1) a_2 x^{n-1} + \cdots + a_n x.$$

Esta aplicação não é autoadjunta. □

Exemplo 5.15. Em $\mathbb{R}^2$ a aplicação de simetria na bissetriz do primeiro e terceiro quadrantes tem matriz simétrica

$$S = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Esta aplicação é autoadjuna. □

5.9. Aplicações e matrizes ortogonais

Seja $\mathbb{X}$ um espaço Euclidiano. Nesta secção vamos introduzir uma classe importante de aplicações lineares de $\mathcal{L}(\mathbb{X}, \mathbb{X})$ conhecidas como aplicações ortogonais.

Definição 5.13. Aplicação ortogonal

Dizemos que $P \in \mathcal{L}(\mathbb{X}, \mathbb{X})$ é uma *aplicação ortogonal* se $P^*P = I$.

A matriz da aplicação P^*, a aplicação adjunta de P, é a matriz transposta da matriz correspondente à aplicação P. Vemos que a matriz que corresponde à aplicação inversa P^{-1} coincide com a matriz transposta correspondente à aplicação P. Obviamente, as colunas da matriz P formam um conjunto de vetores ortonormais.

Exemplo 5.16. Consideremos em $\mathbb{R}^2$ a aplicação de simetria na bissetriz do primeiro e terceiro quadrantes, cuja matriz é

$$S = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Como $S^{\mathsf{T}}S = I$, esta aplicação é ortogonal. □

Exemplo 5.17. À aplicação de rotação em torno da origem com amplitude $\frac{\pi}{2}$ no sentido anti-horário corresponde a matriz

$$R = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Esta matriz também verifica a condição $R^{\mathsf{T}}R = I$, pelo que também é uma aplicação ortogonal. □

As matrizes correspondentes às aplicações autoadjuntas são muito importantes e merecem um nome especial.

Definição 5.14. Matriz ortogonal

Dizemos que P é uma *matriz ortogonal* se $P^{\mathsf{T}}P = I$.

A propriedade fundamental das aplicações ortogonais está no seguinte teorema.

Teorema 5.14

As aplicações ortogonais preservam o produto escalar no sentido que

$$\langle x, y\rangle = \langle Px, Py\rangle.$$

Demonstração. De facto, atendendo à definição de aplicação adjunta (Definição 5.10) e à igualdade (5.14), temos

$$\langle Px, Py\rangle = \langle x, P^*Py\rangle = \langle x, y\rangle.$$

■

5.10. Espaço de aplicações lineares como um espaço normado

Sejam $\mathbb{X}$ e $\mathbb{Y}$ espaços normados de dimensão finita com normas $|\cdot|_{\mathbb{X}}$ e $|\cdot|_{\mathbb{Y}}$, respetivamente. Existe uma maneira natural de introduzir no espaço linear $\mathcal{L}(\mathbb{X}, \mathbb{Y})$ uma norma, transformando-o num espaço normado. Esta norma depende das normas $|\cdot|_{\mathbb{X}}$ e $|\cdot|_{\mathbb{Y}}$. Consideremos uma aplicação linear $A \in \mathcal{L}(\mathbb{X}, \mathbb{Y})$. Como A é uma aplicação linear, ela é contínua e o valor

$$|A|_{\mathcal{L}} = \max_{\{x\in\mathbb{X}:|x|_{\mathbb{X}}=1\}} |Ax|_{\mathbb{Y}}$$

está corretamente definido e é finito. De facto, pelo Teorema de Weierstrass, uma função contínua definida num conjunto limitado e fechado atinge o seu máximo.

Teorema 5.15

A função $|A|_{\mathcal{L}}$, que faz corresponder a uma aplicação linear $A \in \mathcal{L}(\mathbb{X}, \mathbb{Y})$ um número real, é uma norma no espaço $\mathcal{L}(\mathbb{X}, \mathbb{Y})$.

Demonstração. É óbvio que $|A|_{\mathcal{L}} \geqslant 0$ e $|A|_{\mathcal{L}} = 0$ se e só se $Ax = 0$ para todos os vetores $x \in \mathbb{X}$, isto é, $A = 0$.

Seja $\alpha \in \mathbb{R}$. Então temos

$$|\alpha A|_{\mathcal{L}} = \max_{\{x\in\mathbb{X}:|x|_{\mathbb{X}}=1\}} |\alpha Ax|_{\mathbb{Y}} = |\alpha| \max_{\{x\in\mathbb{X}:|x|_{\mathbb{X}}=1\}} |Ax|_{\mathbb{Y}} = |\alpha||A|_{\mathcal{L}}.$$

Finalmente, temos

$$|A_1 + A_2|_{\mathcal{L}} = \max_{\{x\in\mathbb{X}:|x|_{\mathbb{X}}=1\}} |A_1x + A_2x|_{\mathbb{Y}} \leqslant \max_{\{x\in\mathbb{X}:|x|_{\mathbb{X}}=1\}} (|A_1x|_{\mathbb{Y}} + |A_2x|_{\mathbb{Y}})$$

$$\leqslant \max_{\{x\in\mathbb{X}:|x|_{\mathbb{X}}=1\}} |A_1x|_{\mathbb{Y}} + \max_{\{x\in\mathbb{X}:|x|_{\mathbb{X}}=1\}} |A_2x|_{\mathbb{Y}} = |A_1|_{\mathcal{L}} + |A_2|_{\mathcal{L}}.$$

■

Obter uma fórmula simples para a norma de uma aplicação linear nem sempre é possível. Apresentemos um caso em que o é. Suponhamos que o espaço $\mathbb{R}^n$ é equipado com a norma $|x|_1$ e o espaço $\mathbb{R}^m$ com a norma $|y|_\infty$. Consideremos uma aplicação $A \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, cuja matriz tem componentes a_i^j. Demonstremos que $|A|_{\mathcal{L}} = \max_{ij} |a_i^j|$. Com efeito,

$$|Ax|_\infty = \max_{i=\overline{1,m}} \left| \sum_{j=1}^{n} a_i^j x^j \right| \leqslant a \sum_{j=1}^{n} |x^j| = a|x|_1,$$

onde $a = \max_{ij} |a_i^j|$. Seja $a = |a_p^q|$. Então, escolhendo $\hat{x}$ com componentes $x^j = 0$, $j \neq q$, $x^q = \operatorname{sgn}(a_p^q)$, temos $|A\hat{x}|_\infty = a$.

6. Determinantes

6.1. Introdução

Nesta secção introduzimos e estudamos o conceito de determinante. É possível escrever os resultados de muitos problemas de geometria dos espaços de dimensão finita em termos de determinantes. Como vamos dar ênfase à interpretação geométrica deste conceito, definimos em primeiro lugar os determinantes em $\mathbb{V}^2$ e $\mathbb{V}^3$, onde têm um claro significado geométrico. Generalizamos depois estas ideias para o caso dos espaços lineares de dimensão finita.

6.2. Determinante em $\mathbb{V}^2$

Consideremos o espaço $\mathbb{V}^2$ com uma base canónica fixa. Sejam $\vec{x} = (x^1, x^2)$ e $\vec{y} = (y^1, y^2)$ dois vetores de $\mathbb{V}^2$. O espaço $\mathbb{V}^2$ é natural considerar como um subespaço do espaço $\mathbb{V}^3$. Então, aos vetores $\vec{x} = (x^1, x^2)$ e $\vec{y} = (y^1, y^2)$ correspondem os vetores $\vec{x}' = (x^1, x^2, 0)$ e $\vec{y}' = (y^1, y^2, 0)$ em $\mathbb{V}^3$. O produto vetorial destes vetores é um vetor do espaço $\mathbb{V}^3$ que é ortogonal ao plano formado por $\vec{x}'$ e $\vec{y}'$ e que tem coordenadas

$$[\vec{x}', \vec{y}'] = (0, 0, x^1y^2 - x^2y^1)$$

(veja o Teorema 2.2). O vetor $[\vec{x}', \vec{y}']$ é conhecido como *vetor da área* do paralelogramo formado pelos vetores $\vec{x}' = (x^1, x^2, 0)$ e $\vec{y}' = (y^1, y^2, 0)$, visto que a norma deste vetor é a área do paralelogramo. O sentido deste vetor depende da orientação dos vetores $\vec{x}'$ e $\vec{y}'$ (regra "pés e cabeça").

Para que é necessário transformar a área, que é um número, num vetor? Acontece que a consideração da área como um vetor tem muitas vantagens. Este conceito é muito utilizado em Física e na parte do Cálculo Integral que tem a ver com os integrais de superfície. Por exemplo, suponhamos que no espaço $\mathbb{V}^3$ existe um fluxo homogéneo de um líquido. A velocidade deste fluxo é $\vec{v}$. Que quantidade de líquido atravessa o paralelogramo formado pelos vetores $\vec{x}' = (x^1, x^2, 0)$ e $\vec{y}' = (y^1, y^2, 0)$ durante uma unidade de tempo? A noção da área vetorial permite responder a esta pergunta imediatamente. A quantidade é o produto da área do paralelogramo e da projeção do vetor da velocidade sobre o vetor da normal ao plano que contém o paralelogramo, ou seja, $\langle \vec{v}, [\vec{x}', \vec{y}'] \rangle$. É muito importante que o vetor da área tenha sentido. Logo, a quantidade de líquido que atravessa o paralelogramo pode ser positiva ou negativa.

Em muitas situações o conceito da área vetorial pode ser reduzido ao conceito de *área orientada*, isto é, a área (um número) que tem sinal positivo ou negativo. Este sinal depende do sentido do vetor da área. A ideia de considerar áreas orientadas leva à definição do determinante de dois vetores.

Definição 6.1. Determinante em $\mathbb{V}^2$

O *determinante* dos vetores $\vec{x}$ e $\vec{y}$ de $\mathbb{V}^2$, que representamos por $\det(\vec{x}, \vec{y})$, é definido por

$$\det(\vec{x}, \vec{y}) = x^1y^2 - x^2y^1.$$

Tradicionalmente o determinante de vetores é definido como o determinante da matriz cujas linhas são as coordenadas dos vetores $\vec{x}$ e $\vec{y}$. Neste caso é utilizada a notação

$$\det\begin{pmatrix} x^1 & x^2 \\ y^1 & y^2 \end{pmatrix} = x^1y^2 - x^2y^1.$$

Para manter esta tradição, vamos também por vezes usar esta notação e a expressão "determinante de uma matriz".

6.3. Determinante em $\mathbb{V}^3$

Na construção da teoria de determinantes vamos seguir a ideia geométrica apresentada na secção anterior. Sejam $\vec{x} = (x^1, x^2, x^3)$, $\vec{y} = (y^1, y^2, y^3)$ e $\vec{z} = (z^1, z^2, z^3)$ vetores do espaço $\mathbb{V}^3$, onde está fixa uma base canónica. Então, por analogia com a secção anterior, podemos definir o determinante destes vetores $\det(\vec{x}, \vec{y}, \vec{z})$ como um volume orientado de um paralelepípedo formado pelos três vetores $\vec{x}$, $\vec{y}$ e $\vec{z}$. Para implementar esta ideia precisamos em primeiro lugar de definir e calcular o volume orientado do paralelepípedo formado por três vetores. Para o calcular é útil o assim dito *produto misto*.

Definição 6.2. Produto misto

O *produto misto* dos vetores $\vec{x}$, $\vec{y}$ e $\vec{z}$ de $\mathbb{V}^3$, que representamos por $(\vec{x}, \vec{y}, \vec{z})$, é definido por

$$(\vec{x}, \vec{y}, \vec{z}) = \langle \vec{x}, [\vec{y}, \vec{z}] \rangle.$$

O produto misto e o volume do paralelepípedo formado por três vetores estão ligados entre si através do seguinte teorema.

Teorema 6.1

Sejam $\vec{x}$, $\vec{y}$ e $\vec{z}$ vetores de $\mathbb{V}^3$ que formam um paralelepípedo de volume V (veja a Fig. 6.1). Então,

$$V = |(\vec{x}, \vec{y}, \vec{z})|.$$

Demonstração. O volume de um paralelepípedo pode ser calculado através da fórmula $V = Ah$, onde A é a área da base e h a altura do paralelepípedo. Então, para o paralelepípedo da Fig. 6.1, temos

$$V = Ah = |[\vec{y}, \vec{z}]|h, \quad h = |\vec{x}||\cos\varphi|,$$

onde φ é o ângulo que formam os vetores $\vec{x}$ e $[\vec{y}, \vec{z}]$, sendo este último ortogonal ao plano gerado por $\vec{y}$ e $\vec{z}$. Daí resulta que

$$V = |[\vec{y}, \vec{z}]||\vec{x}||\cos\varphi| = |\langle \vec{x}, [\vec{y}, \vec{z}] \rangle| = |(\vec{x}, \vec{y}, \vec{z})|$$

pelas definições do produto escalar e do produto misto.

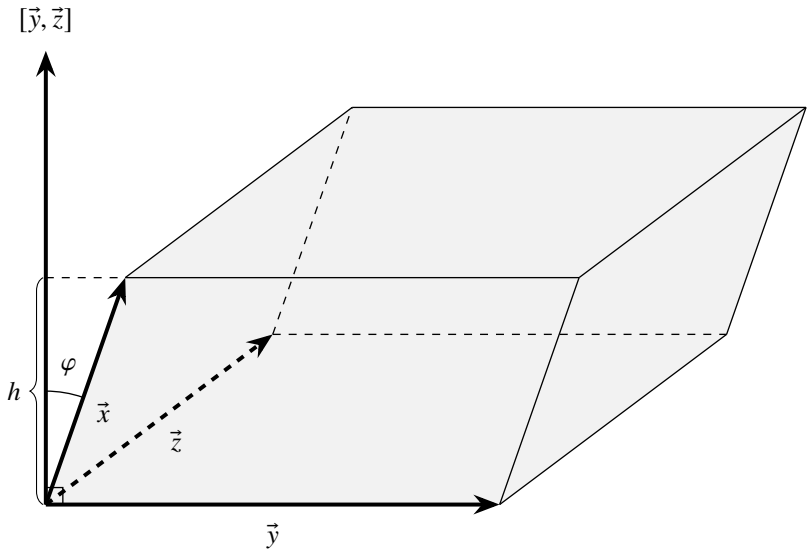


Figura 6.1
Paralelepípedo formado por $\vec{x}$, $\vec{y}$ e $\vec{z}$.

■

A fórmula do produto misto em termos de coordenadas dos três vetores do espaço $\mathbb{V}^3$ considerando uma base canónica de $\mathbb{V}^3$ é dada no seguinte teorema.

Teorema 6.2

Sejam $\vec{x} = (x^1, x^2, x^3)$, $\vec{y} = (y^1, y^2, y^3)$ e $\vec{z} = (z^1, z^2, z^3)$ vetores de $\mathbb{V}^3$ com coordenadas numa base canónica $\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$. Então:

$$(\vec{x},\vec{y},\vec{z}) = x^1y^2z^3 - x^1y^3z^2 + x^2y^3z^1 - x^2y^1z^3 + x^3y^1z^2 - x^3y^2z^1.$$

Demonstração. Utilizando as propriedades do produto vetorial e do produto escalar, obtemos

$$\begin{aligned}(\vec{x},\vec{y},\vec{z}) &= \langle \vec{x}, [\vec{y},\vec{z}]\rangle \\ &= \langle x^1\vec{e}_1 + x^2\vec{e}_2 + x^3\vec{e}_3, [y^1\vec{e}_1 + y^2\vec{e}_2 + y^3\vec{e}_3, z^1\vec{e}_1 + z^2\vec{e}_2 + z^3\vec{e}_3]\rangle \\ &= \langle x^1\vec{e}_1 + x^2\vec{e}_2 + x^3\vec{e}_3, (y^2z^3 - y^3z^2)\vec{e}_1 + (y^3z^1 - y^1z^3)\vec{e}_2 + (y^1z^2 - y^2z^1)\vec{e}_3\rangle \\ &= x^1y^2z^3 - x^1y^3z^2 + x^2y^3z^1 - x^2y^1z^3 + x^3y^1z^2 - x^3y^2z^1.\end{aligned}$$

■

Como podemos ver do teorema seguinte, o produto misto é um volume orientado, pois o sinal depende da ordem dos vetores no produto.

Teorema 6.3

Para todos os vetores $\vec{x}, \vec{y}, \vec{z}$ de $\mathbb{V}^3$ temos

$$(\vec{x},\vec{y},\vec{z}) = (\vec{y},\vec{z},\vec{x}) = (\vec{z},\vec{x},\vec{y}) = -(\vec{z},\vec{y},\vec{x}) = -(\vec{y},\vec{x},\vec{z}) = -(\vec{x},\vec{z},\vec{y}).$$

Demonstração. Sejam $\vec{x} = (x^1, x^2, x^3)$, $\vec{y} = (y^1, y^2, y^3)$ e $\vec{z} = (z^1, z^2, z^3)$ vetores de $\mathbb{V}^3$ com coordenadas numa base canónica $\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$. Provamos apenas a primeira

igualdade $(\vec{x},\vec{y},\vec{z}) = (\vec{y},\vec{z},\vec{x})$, deixando a demonstração das outras igualdades como exercício. Utilizando o Teorema 6.2, temos

$$(\vec{x},\vec{y},\vec{z}) = x^1y^2z^3 - x^1y^3z^2 + x^2y^3z^1 - x^2y^1z^3 + x^3y^1z^2 - x^3y^2z^1.$$

Por outro lado, o produto

$$(\vec{y},\vec{z},\vec{x}) = y^1z^2x^3 - y^1z^3x^2 + y^2z^3x^1 - y^2z^1x^3 + y^3z^1x^2 - y^3z^2x^1,$$

dá o mesmo resultado.

■

Notemos que, se $\vec{x}$, $\vec{y}$ e $\vec{z}$ são vetores linearmente dependentes de $\mathbb{V}^3$, então $(\vec{x},\vec{y},\vec{z}) = 0$, pois o volume do paralelepípedo formado por estes vetores é zero.

Vamos agora definir o determinante de três vetores $\vec{x} = (x^1,x^2,x^3)$, $\vec{y} = (y^1,y^2,y^3)$ e $\vec{z} = (z^1,z^2,z^3)$ de $\mathbb{V}^3$.

Definição 6.3. Determinante em $\mathbb{V}^3$

O *determinante* dos vetores $\vec{x}$, $\vec{y}$ e $\vec{z}$ de $\mathbb{V}^3$, que representamos por $\det(\vec{x},\vec{y},\vec{z})$, é definido por

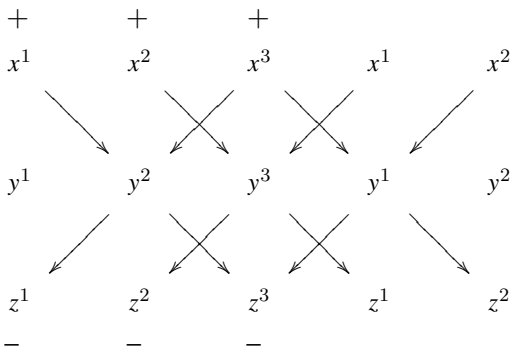
$$\begin{aligned}\det(\vec{x},\vec{y},\vec{z}) &= (\vec{x},\vec{y},\vec{z})\\ &= x^1y^2z^3 - x^1y^3z^2 + x^2y^3z^1 - x^2y^1z^3 + x^3y^1z^2 - x^3y^2z^1.\end{aligned} \tag{6.1}$$

Desta definição vemos que o volume do paralelepípedo gerado por três vetores é o módulo do determinante destes vetores.

Tal como no caso de dois vetores vamos também definir o determinante da matriz cujas linhas são as coordenadas dos vetores $\vec{x}$, $\vec{y}$ e $\vec{z}$ em certa base. Denotamos este determinante por

$$\det\begin{pmatrix} x^1 & x^2 & x^3\\ y^1 & y^2 & y^3\\ z^1 & z^2 & z^3\end{pmatrix},$$

que representa $\det(\vec{x},\vec{y},\vec{z})$. O determinante em $\mathbb{V}^3$ também pode ser calculado com a ajuda do seguinte método, conhecido por "Regra de Sarrus":



Para calcular o determinante recorrendo a esta regra, repetimos do lado direito da matriz as primeiras duas colunas, como no esquema acima. De seguida calculamos os produtos dos elementos das diagonais indicadas, em que os produtos dos termos das diagonais com início em x^1, x^2 e x^3 recebem o sinal $+$ e os produtos dos termos das diagonais terminando em z^1, z^2 e z^3 recebem o sinal $-$, obtemos, como é fácil ver, o valor do determinante (6.1).

6.4. Determinante em $\mathbb{R}^n$

Nesta secção vamos apresentar o determinante em $\mathbb{R}^n$ como um objeto geométrico a partir da generalização dos dois casos anteriores. Para isso definimos primeiro a aplicação projeção.

Seja $\{e_1, \ldots, e_n\}$ uma base ortonormal em $\mathbb{R}^n$.

Definição 6.4. Aplicação projeção

Seja

$$\begin{aligned} x &= x_1 e_1 + \cdots + x_{k-1} e_{k-1} + x_k e_k + x_{k+1} e_{k+1} + \cdots + x_n e_n \\ &= (x_1, \ldots, x_{k-1}, x_k, x_{k+1}, \ldots, x_n) \end{aligned}$$

um vetor de $\mathbb{R}^n$, $n \geqslant 2$. A *aplicação projeção* π_k, $k = \overline{1,n}$, é definida por

$$\pi_k : \mathbb{R}^n \longrightarrow \mathbb{R}^{n-1}$$

$$\pi_k(x) = \pi_k(x_1, \ldots, x_{k-1}, x_k, x_{k+1}, \ldots, x_n) = (x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_n).$$

A aplicação π_k elimina a coordenada k do vetor $x \in \mathbb{R}^n$ projetando-o na direção de e_k no espaço $\mathbb{R}^{n-1}$ de base $\{e_1, \ldots, e_{k-1}, e_{k+1}, \ldots, e_n\}$ (veja o exemplo na Fig. 6.2).

Definição 6.5. Determinante em $\mathbb{R}^n$

O *determinante* dos vetores $x_1, \ldots, x_n \in \mathbb{R}^n$, $n \geqslant 2$, que representamos por $\det(x_1, \ldots, x_n)$, é definido por

$$\det(x_1, \ldots, x_n) = \langle x_1, [x_2, \ldots, x_n] \rangle, \tag{6.2}$$

onde

$$[x_2, \ldots, x_n] = \sum_{k=1}^{n} (-1)^{k-1} \det(\pi_k(x_2), \ldots, \pi_k(x_n)) e_k.$$

Se $x \in \mathbb{R}$ definimos

$$\det(x) = x. \tag{6.3}$$

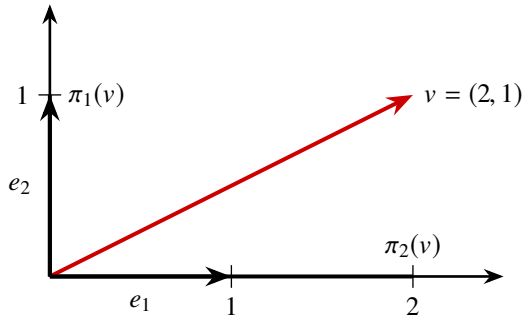


Figura 6.2
Projeções $\pi_1(v) = 1$ e $\pi_2(v) = 2$ do vetor $v = (2, 1) \in \mathbb{R}^2$.

Nesta definição, $[x_2, \ldots, x_n]$ é uma generalização do produto vetorial de $n-1$ vetores de $\mathbb{R}^n$, ou seja, *volume vetorial* do paralelepípedo $(n-1)$-dimensional gerado pelos vetores $x_2, \ldots, x_n$ (veja a este respeito a Secção 6.8). Se $n = 3$, esta definição coincide com o produto $[x_2, x_3]$.

Definição 6.6. Determinante de uma matriz

Sejam

$$\begin{aligned} x_1 &= (x_1^1, x_1^2, \ldots, x_1^n), \\ x_2 &= (x_2^1, x_2^2, \ldots, x_2^n), \\ &\vdots \\ x_n &= (x_n^1, x_n^2, \ldots, x_n^n), \end{aligned}$$

vetores de $\mathbb{R}^n$. Então, o *determinante* da matriz

$$X = \begin{pmatrix} x_1^1 & x_1^2 & \cdots & x_1^n \\ x_2^1 & x_2^2 & \cdots & x_2^n \\ \vdots & \vdots & \ddots & \vdots \\ x_n^1 & x_n^2 & \cdots & x_n^n \end{pmatrix},$$

que representamos por $\det(X)$, é definido por $\det(X) = \det(x_1, \ldots, x_n)$.

Para desenvolver o cálculo de determinantes precisamos do conceito de permutação.

Definição 6.7. Permutação

Uma *permutação* p dos números naturais $(1, \ldots, n)$ é uma sua reordenação, que é designada por

$$p = (i_1, i_2, \ldots, i_n) = \begin{pmatrix} 1 & 2 & \cdots & n \\ i_1 & i_2 & \cdots & i_n \end{pmatrix}.$$

O conjunto de todas as permutações de n elementos denotamos por Π_n.

Exemplo 6.1. O conjunto de todas as permutações dos números $(1,2,3)$ é

$$\Pi_3 = \{(1,2,3),(2,1,3),(3,2,1),(1,3,2),(2,3,1),(3,1,2)\}.$$

□

Dadas duas permutações $(i_1, i_2, \ldots, i_n)$ e $(j_1, j_2, \ldots, j_n)$, podemos definir a permutação onde i_k é substituído por j_k, $k = \overline{1,n}$. É prático escrever esta permutação na forma

$$\begin{pmatrix} i_1 & i_2 & \cdots & i_n \\ j_1 & j_2 & \cdots & j_n \end{pmatrix}.$$

Definição 6.8. Inversão

Seja $p \in \Pi_n$. Uma *inversão* em $p = (i_1, i_2, \ldots, i_n)$ é um par de números k e m entre 1 e n com $k < m$ e $i_k > i_m$.

Exemplo 6.2. Consideremos o conjunto Π_4 de todas as permutações dos números $(1,2,3,4)$.

1. A permutação

$$p = (1,2,3,4) = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix}$$

tem zero inversões.

2. O número de inversões da permutação

$$p = (3,1,4,2) = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 4 & 2 \end{pmatrix}$$

é três, porque $i_1 > i_2$, $i_1 > i_4$ e $i_3 > i_4$.

□

Definição 6.9. Sinal de uma permutação

Seja N o número de inversões da permutação p. Definimos por

$$\operatorname{sgn}(p) = (-1)^N$$

o *sinal da permutação* p, ou seja, $\operatorname{sgn}(p) = 1$ se o número de inversões for par e $\operatorname{sgn}(p) = -1$ se o número de inversões for ímpar.

Exemplo 6.3. Consideremos o conjunto Π_3.

1. A permutação

$$p = (2,1,3) = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} \in \Pi_3$$

tem sinal $\operatorname{sgn}(p) = -1$, porque o número de inversões é um: $i_1 > i_2$.

2. A permutação

$$p = (3,1,2) = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} \in \Pi_3,$$

tem sinal $\operatorname{sgn}(p) = 1$, porque neste caso o número de inversões é dois: $i_1 > i_2$ e $i_1 > i_3$.

□

Lema 6.1

Sejam j e k números naturais tais que $1 < j < k$. Então,

$$\operatorname{sgn}(i_1, \ldots, i_{j-1}, i_j, i_{j+1}, \ldots, i_k) = (-1)^{j-1} \operatorname{sgn}(i_j, i_1, \ldots, i_{j-1}, i_{j+1}, \ldots, i_k).$$

Demonstração. Permutando i_j com i_{j-1} na permutação $(i_1, \ldots, i_{j-1}, i_j, i_{j+1}, \ldots, i_k)$, obtemos

$$\operatorname{sgn}(i_1, \ldots, i_{j-1}, i_j, i_{j+1}, \ldots, i_k) = (-1) \operatorname{sgn}(i_1, \ldots, i_{j-2}, i_j, i_{j-1}, i_{j+1}, \ldots, i_k).$$

Para colocar i_j na primeira posição, partindo da sua posição inicial, são necessárias $j-1$ permutações. Logo, temos

$$\operatorname{sgn}(i_1, \ldots, i_{j-1}, i_j, i_{j+1}, \ldots, i_k) = (-1)^{j-1} \operatorname{sgn}(i_j, i_1, \ldots, i_{j-1}, i_{j+1}, \ldots, i_k).$$

■

Estamos agora em condições de apresentar o teorema que nos permite calcular determinantes utilizando o conceito de permutação. Este teorema ajuda, como veremos, a estabelecer várias propriedades importantes de determinantes.

Teorema 6.4

Sejam $x_1 = (x_1^1, \ldots, x_1^n), \ldots, x_n = (x_n^1, \ldots, x_n^n) \in \mathbb{R}^n$. Então,

$$\det(x_1, \ldots, x_n) = \sum_{p=(i_1,i_2,\ldots,i_n)\in\Pi_n} \operatorname{sgn}(p) x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n}. \tag{6.4}$$

Demonstração. Para $n = 1$, a demonstração é trivial. Considerando $x \in \mathbb{R}$, temos pela fórmula (6.4) que

$$\det(x) = \sum_{p=(1)\in\Pi_1} \operatorname{sgn}(p) x = x,$$

que coincide com (6.3).

Para o leitor perceber melhor a demonstração que vamos fazer por indução, consideremos ainda o caso $n = 2$. Sejam $x_1 = (x_1^1, x_1^2)$ e $x_2 = (x_2^1, x_2^2) \in \mathbb{R}^2$. Pela fórmula (6.4), sendo $\Pi_2 = \{(1,2),(2,1)\}$, temos

$$\begin{aligned} \det(x_1, x_2) &= \sum_{p=(i_1,i_2)\in\Pi_2} \operatorname{sgn}(p) x_1^{i_1} x_2^{i_2} = \operatorname{sgn}(1,2) x_1^1 x_2^2 + \operatorname{sgn}(2,1) x_1^2 x_2^1 \\ &= x_1^1 x_2^2 - x_1^2 x_2^1. \end{aligned} \tag{6.5}$$

Da Definição 6.5, obtemos

$$\det(x_1, x_2) = \langle x_1, [x_2] \rangle,$$

em que

$$[x_2] = \det(\pi_1(x_2))e_1 - \det(\pi_2(x_2))e_2 = x_2^2 e_1 - x_2^1 e_2.$$

Logo, temos

$$\det(x_1, x_2) = \langle x_1^1 e_1 + x_1^2 e_2, x_2^2 e_1 - x_2^1 e_2 \rangle = x_1^1 x_2^2 - x_1^2 x_2^1,$$

que coincide com (6.5) (e também com a Definição 6.1).

Supondo agora que as fórmulas (6.2) e (6.4) são equivalentes para $n = k$, $k \geqslant 2$, provamos a seguir que elas também são equivalentes para $n = k + 1$. Pela hipótese de indução, no caso de k vetores, é válida a igualdade

$$\left\langle x_1, \sum_{j=1}^{k} (-1)^{j-1} \det(\pi_j(x_2), \ldots, \pi_j(x_k)) e_j \right\rangle = \sum_{p=(i_1,\ldots,i_k)\in\Pi_k} \operatorname{sgn}(p) x_1^{i_1} x_2^{i_2} \cdots x_k^{i_k},$$

onde $x_i \in \mathbb{R}^k, i = \overline{1,k}$. Pela Definição 6.5, temos

$$\begin{aligned}
\det(x_1, \ldots, x_{k+1}) &= \left\langle x_1, \sum_{j=1}^{k+1} (-1)^{j-1} \det(\pi_j(x_2), \ldots, \pi_j(x_{k+1})) e_j \right\rangle \\
&= x_1^1 \det(\pi_1(x_2), \ldots, \pi_1(x_{k+1})) \\
&\quad - x_1^2 \det(\pi_2(x_2), \ldots, \pi_2(x_{k+1})) \\
&\quad + \cdots \\
&\quad + (-1)^{k-1} x_1^k \det(\pi_k(x_2), \ldots, \pi_k(x_{k+1})) \\
&\quad + (-1)^k x_1^{k+1} \det(\pi_{k+1}(x_2), \ldots, \pi_{k+1}(x_{k+1})).
\end{aligned}$$

Como os k vetores $\pi_j(x_2), \ldots, \pi_j(x_{k+1})$, para cada $j = \overline{1, k+1}$, têm k componentes, pela hipótese de indução vemos que

$$\begin{aligned}
\det(x_1, \ldots, x_{k+1}) &= x_1^1 \sum_{p=(i_1,\ldots,i_{k+1})\in\Pi_{k+1}^1} \operatorname{sgn}(p) \left(\pi_1(x_2)\right)_{i_1} \cdots \left(\pi_1(x_{k+1})\right)_{i_{k+1}} \\
&\quad - x_1^2 \sum_{p=(i_1,\ldots,i_{k+1})\in\Pi_{k+1}^2} \operatorname{sgn}(p) \left(\pi_2(x_2)\right)_{i_1} \cdots \left(\pi_2(x_{k+1})\right)_{i_{k+1}} \\
&\quad + \cdots \\
&\quad + (-1)^{k-1} x_1^k \sum_{p=(i_1,\ldots,i_{k+1})\in\Pi_{k+1}^k} \operatorname{sgn}(p) \left(\pi_k(x_2)\right)_{i_1} \cdots \left(\pi_k(x_{k+1})\right)_{i_{k+1}} \\
&\quad + (-1)^k x_1^{k+1} \sum_{p=(i_1,\ldots,i_{k+1})\in\Pi_{k+1}^{k+1}} \operatorname{sgn}(p) \left(\pi_{k+1}(x_2)\right)_{i_1} \cdots \left(\pi_{k+1}(x_{k+1})\right)_{i_{k+1}},
\end{aligned} \tag{6.6}$$

onde Π_n^j é o conjunto de todas as permutações dos números $(1,\ldots,j-1,j+1,\ldots,n)$. Como a projeção $\pi_j(x)$ não contém a coordenada x^j e a permutação $p \in \Pi_{k+1}^j$ não contém o número j, de (6.6) obtemos

$$\begin{aligned}\det(x_1,\ldots,x_{k+1}) = {} & x_1^1 \sum_{p=(i_1,\ldots,i_{k+1})\in\Pi_{k+1}^1} \operatorname{sgn}(p) x_2^{i_1}\cdots x_{k+1}^{i_{k+1}} \\ & - x_1^2 \sum_{p=(i_1,\ldots,i_{k+1})\in\Pi_{k+1}^2} \operatorname{sgn}(p) x_2^{i_1}\cdots x_{k+1}^{i_{k+1}} \\ & + \cdots \\ & + (-1)^{k-1} x_1^k \sum_{p=(i_1,\ldots,i_{k+1})\in\Pi_{k+1}^k} \operatorname{sgn}(p) x_2^{i_1}\cdots x_{k+1}^{i_{k+1}} \\ & + (-1)^k x_1^{k+1} \sum_{p=(i_1,\ldots,i_{k+1})\in\Pi_{k+1}^{k+1}} \operatorname{sgn}(p) x_2^{i_1}\cdots x_{k+1}^{i_{k+1}}.\end{aligned} \tag{6.7}$$

Seja $p = (i_1,\ldots,i_{k+1}) \in \Pi_{k+1}^j$. Do Lema 6.1 vemos que $\operatorname{sgn}(p^j)$, o sinal da permutação $p^j = (i_1,\ldots,i_{j-1},j,i_{j+1},\ldots,i_{k+1}) \in \Pi_{k+1}$, obtida da permutação p colocando o número j na posição j, é igual a $(-1)^{j-1}\operatorname{sgn}(p)$. Logo, (6.7) implica

$$\begin{aligned}\det(x_1,\ldots,x_{k+1}) = {} & x_1^1 \sum_{p^1=(1,i_1,\ldots,i_{k+1})\in\Pi_{k+1}} \operatorname{sgn}(p^1) x_2^{i_1}\cdots x_{k+1}^{i_{k+1}} \\ & + x_1^2 \sum_{p^2=(i_1,2,\ldots,i_{k+1})\in\Pi_{k+1}} \operatorname{sgn}(p^2) x_2^{i_1}\cdots x_{k+1}^{i_{k+1}} \\ & + \cdots \\ & + x_1^k \sum_{p^k=(i_1,\ldots,k,i_{k+1})\in\Pi_{k+1}} \operatorname{sgn}(p^k) x_2^{i_1}\cdots x_{k+1}^{i_{k+1}} \\ & + x_1^{k+1} \sum_{p^{k+1}=(i_1,\ldots,i_{k+1},k+1)\in\Pi_{k+1}} \operatorname{sgn}(p^{k+1}) x_2^{i_1}\cdots x_{k+1}^{i_{k+1}}.\end{aligned}$$

Isto implica

$$\det(x_1,\ldots,x_{k+1}) = \sum_{p=(i_1,\ldots,i_{k+1})\in\Pi_{k+1}} \operatorname{sgn}(p) x_1^{i_1} x_2^{i_2}\cdots x_{k+1}^{i_{k+1}}.$$

■

Vamos ver algumas aplicações deste teorema.

Corolário 6.1

Consideremos os vetores $x_1, \ldots, x_k, \ldots, x_n$, y_k, $z_k \in \mathbb{R}^n$. Se $x_k = \alpha y_k + \beta z_k$, onde α e β são números, então

$$\det(x_1, \ldots, x_k, \ldots, x_n) = \alpha \det(x_1, \ldots, y_k, \ldots, x_n) + \beta \det(x_1, \ldots, z_k, \ldots, x_n).$$

Demonstração. Pelo Teorema 6.4, temos

$$\begin{aligned}
\det(x_1, \ldots, x_k, \ldots, x_n) &= \sum_{p=(i_1,\ldots,i_n)\in\Pi_n} \operatorname{sgn}(p) x_1^{i_1} \cdots x_k^{i_k} \cdots x_n^{i_n} \\
&= \sum_{p=(i_1,\ldots,i_n)\in\Pi_n} \operatorname{sgn}(p) x_1^{i_1} \cdots (\alpha y_k^{i_k} + \beta z_k^{i_k}) \cdots x_n^{i_n} \\
&= \alpha \sum_{p=(i_1,\ldots,i_n)\in\Pi_n} \operatorname{sgn}(p) x_1^{i_1} \cdots y_k^{i_k} \cdots x_n^{i_n} \\
&\quad + \beta \sum_{p=(i_1,\ldots,i_n)\in\Pi_n} \operatorname{sgn}(p) x_1^{i_1} \cdots z_k^{i_k} \cdots x_n^{i_n} \\
&= \alpha \det(x_1, \ldots, y_k, \ldots, x_n) + \beta \det(x_1, \ldots, z_k, \ldots, x_n).
\end{aligned}$$

■

Corolário 6.2

Consideremos os vetores $x_1, \ldots, x_j, x_k, \ldots, x_n, \in \mathbb{R}^n$. Então,

$$\det(x_1, \ldots, x_j, x_k, \ldots, x_n) = -\det(x_1, \ldots, x_k, x_j, \ldots, x_n).$$

Demonstração. Aplicando o Teorema 6.4, temos

$$\begin{aligned}
\det(x_1, \ldots, x_j, x_k, \ldots, x_n) &= \sum_{p=(i_1,\ldots,i_j,i_k,\ldots,i_n)\in\Pi_n} \operatorname{sgn}(p) x_1^{i_1} \cdots x_j^{i_j} x_k^{i_k} \cdots x_n^{i_n} \\
&= -\sum_{p=(i_1,\ldots,i_k,i_j,\ldots,i_n)\in\Pi_n} \operatorname{sgn}(p) x_1^{i_1} \cdots x_k^{i_k} x_j^{i_j} \cdots x_n^{i_n} \\
&= -\det(x_1, \ldots, x_k, x_j, \ldots, x_n).
\end{aligned}$$

A segunda igualdade é consequência do primeiro passo na demonstração do Lema 6.1.

■

Corolário 6.3

Consideremos os vetores $x_1, \ldots, x_n, \in \mathbb{R}^n$, com $n \geqslant 2$, e dois vetores x_i, x_j, com $i \neq j$. Se $x_i = x_j$, então $\det(x_1, \ldots, x_n) = 0$.

Demonstração. Este resultado é consequência imediata do Corolário 6.2. ■

Dos Corolários 6.1 e 6.3 temos o seguinte resultado.

Corolário 6.4

Consideremos os vetores $x_1, \ldots, x_n, \in \mathbb{R}^n$, com $n \geqslant 2$, dois vetores x_i, x_j, com $i \neq j$, e α um número. Então,

$$\det(x_1, \ldots, x_{i-1}, x_i + \alpha x_j, x_{i+1}, \ldots, x_n) = \det(x_1, \ldots, x_n).$$

Deste corolário vemos que o determinante de n vetores tem o mesmo valor quando a um dos vetores adicionamos um outro multiplicado por um número. Daqui temos a seguinte propriedade.

Corolário 6.5

Se os vetores $x_1, \ldots, x_n \in \mathbb{R}^n$ são linearmente dependentes, então

$$\det(x_1, \ldots, x_n) = 0.$$

Demonstração. Como os vetores $x_1, \ldots, x_n$ são linearmente dependentes, um deles, por exemplo, x_1, é uma combinação linear dos restantes vetores:

$$x_1 = \sum_{i=2}^{n} \alpha_i x_i.$$

Portanto, pelo Corolário 6.4 temos

$$\det(x_1, x_2, \ldots, x_n) = \det\left(x_1 - \sum_{i=2}^{n} \alpha_i x_i, x_2, \ldots, x_n\right) = \det(0, x_2, \ldots, x_n) = 0.$$

■

Recordemos que o determinante de uma matriz quadrada foi definido como o determinante dos vetores cujas coordenadas são as linhas da matriz (veja a Definição 6.6). A operação de transposição de uma matriz quadrada não altera o valor do determinante.

Teorema 6.5

Seja X uma matriz quadrada. Então,

$$\det(X) = \det(X^\top),$$

ou seja, o determinante da matriz X também é o determinante dos vetores cujas coordenadas formam as colunas da matriz X.

Demonstração. Pelo Teorema 6.4 temos que

$$\det(X) = \det(x_1, \ldots, x_n) = \sum_{p=(i_1,\ldots,i_n)\in\Pi_n} \operatorname{sgn}(p) x_1^{i_1} \cdots x_n^{i_n},$$

onde $x_1, \ldots, x_n$ representam as linhas da matriz X. Cada termo de $\det(X)$ é da forma

$$\operatorname{sgn}(i_1, \ldots, i_n) x_1^{i_1} \cdots x_n^{i_n}$$

e todos os termos deste produto pertencem a linhas e colunas diferentes. Portanto, os mesmos termos fazem parte de $\det(X^\intercal)$. Como o sinal das permutações

$$\begin{pmatrix} 1 & \cdots & n \\ i_1 & \cdots & i_n \end{pmatrix} \quad \text{e} \quad \begin{pmatrix} i_1 & \cdots & i_n \\ 1 & \cdots & n \end{pmatrix}$$

é o mesmo, temos $\det(X) = \det(X^\intercal)$. ■

Agora vamos demonstrar uma das propriedades mais importantes dos determinantes.

Teorema 6.6

Sejam A e B matrizes quadradas da mesma ordem. Então,

$$\det(AB) = \det(A)\det(B).$$

Demonstração. Sejam $a_i = (a_i^1, \ldots, a_i^n)$ e $b_i = (b_i^1, \ldots, b_i^n)$, $i = \overline{1,n}$, as linhas das matrizes A e B de $\mathcal{M}_{n\times n}$, respetivamente. Pelo Corolário 6.1 temos

$$\begin{aligned}
\det(AB) &= \det\left(\sum_j a_1^j b_j, \sum_j a_2^j b_j, \ldots, \sum_j a_n^j b_j\right) \\
&= \sum_{k_1} a_1^{k_1} \det\left(b_{k_1}, \sum_j a_2^j b_j, \ldots, \sum_j a_n^j b_j\right) \\
&= \sum_{k_1 \neq k_2} a_1^{k_1} a_2^{k_2} \det\left(b_{k_1}, b_{k_2}, \ldots, \sum_j a_n^j b_j\right) \\
&= \sum_{(k_1,\ldots,k_n)} a_1^{k_1} \cdots a_n^{k_n} \det\left(b_{k_1}, \ldots, b_{k_n}\right) \\
&= \sum_{p=(k_1,\ldots,k_n)} \operatorname{sgn}(p) a_1^{k_1} \cdots a_n^{k_n} \det(B) = \det(A)\det(B).
\end{aligned}$$

Na terceira igualdade, o somatório é aplicado independentemente aos índices k_1 e k_2, que devem ser diferentes (ver Corolário 6.3). Na quarta igualdade, o somatório é aplicado independentemente, mas por ordem, aos índices distintos $k_1, \ldots, k_n$. Finalmente a última igualdade é uma consequência do Teorema 6.4. ■

Do Corolário 6.4 vemos que o Método de Gauss pode ser aplicado também para o cálculo do determinante de uma matriz uma vez que a operação $l_i \leftarrow l_i + \alpha l_j$ não altera o valor do determinante. Portanto, o determinante da matriz original é igual ao determinante da matriz em escada obtida aplicando o Método de Gauss. Este último é igual ao produto dos elementos que estão na diagonal principal da matriz. De facto, como o determinante é a soma de produtos de elementos de uma matriz que todos pertencem a linhas e colunas diferentes, obviamente o único produto não nulo é o produto dos elementos diagonais da matriz. O sinal da respetiva permutação é positivo. Para ilustrar este processo, consideremos o seguinte exemplo.

Exemplo 6.4. Seja a matriz

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 5 & 1 & 4 \\ 3 & 2 & 5 \end{pmatrix}.$$

Depois das operações $l_2 \leftarrow l_2 - 5l_1$ e $l_3 \leftarrow l_3 - 3l_1$, obtemos

$$\det A = \det \begin{pmatrix} 1 & 2 & 3 \\ 0 & -9 & -11 \\ 0 & -4 & -4 \end{pmatrix}.$$

Fazendo $l_3 \leftarrow l_3 - \frac{4}{9}l_2$, temos

$$\det A = \det \begin{pmatrix} 1 & 2 & 3 \\ 0 & -9 & -11 \\ 0 & 0 & \frac{8}{9} \end{pmatrix} = -8.$$

□

O determinante de uma matriz quadrada foi definido como o determinante dos vetores cujas coordenadas formam as linhas da matriz. Podíamos, como mais habitualmente é feito, ter definido o determinante de uma matriz $X = (x_i^j)$ através da fórmula (6.4). No entanto, seguindo esta última abordagem, o significado geométrico de determinante não aparece de maneira clara. Este é o motivo da opção de introduzir uma definição de cariz geométrico neste livro.

6.5. Cálculo do determinante pela fórmula de Laplace

Além dos métodos e fórmulas apresentados nas secções anteriores, o determinante de uma matriz pode ser calculado de outra maneira.

Definição 6.10. Co-fator de um elemento de uma matriz

Seja $A = (a_i^j)$ uma matriz quadrada de ordem n. Seja $\tilde{A}(i,j)$ a submatriz $(n-1)\times(n-1)$ de A obtida eliminando a linha i e a coluna j de A. O *co-fator* (ou *complemento algébrico*) do elemento a_i^j da matriz A, que representamos por A_i^j, é definido por

$$A_i^j = (-1)^{i+j} \det(\tilde{A}(i,j)).$$

O seguinte teorema dá a possibilidade de calcular o determinante na forma de soma de produtos de elementos de uma linha ou coluna e os respetivos co-fatores. Em termos computacionais, este método não é competitivo com o Método de Gauss apresentado na secção anterior, mas tem uma grande importância teórica.

Teorema 6.7. (Laplace)

Seja $A = (a_i^j)$ uma matriz quadrada de ordem n. Então,

$$\det(A) = \sum_{\substack{j=1 \\ i \text{ fixo}}}^{n} a_i^j A_i^j = \sum_{\substack{i=1 \\ j \text{ fixo}}}^{n} a_i^j A_i^j.$$

Demonstração. Designemos por a_i a i-ésima linha de A. Então,

$$\det(A) = \det(a_1, \ldots, a_n) = (-1)^{i-1} \det(a_i, a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_n).$$

Pela definição do determinante (Definição 6.5) temos

$$\begin{aligned}\det(A) &= (-1)^{i-1} \det(a_i, a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_n) \\ &= (-1)^{i-1} \sum_{j=1}^{n} (-1)^{j-1} a_i^j \det(\tilde{A}(i,j)) = \sum_{j=1}^{n} a_i^j A_i^j.\end{aligned}$$

Como $\det(A^{\mathsf{T}}) = \det(A)$, temos também a fórmula

$$\det(A) = \sum_{\substack{i=1 \\ j \text{ fixo}}}^{n} a_i^j A_i^j.$$

■

As fórmulas deste teorema são conhecidas como o desenvolvimento do determinante ao longo de uma linha e de uma coluna.

Exemplo 6.5. Vamos calcular o determinante Δ da matriz $\left(\begin{smallmatrix} 9 & 2 & 4 \\ 3 & 1 & 6 \\ 7 & 6 & 5 \end{smallmatrix}\right)$. Utilizando o Teorema de Laplace e desenvolvendo o determinante pela primeira linha temos

$$\begin{aligned}\Delta &= 9 \det \begin{pmatrix} 1 & 6 \\ 6 & 5 \end{pmatrix} - 2 \det \begin{pmatrix} 3 & 6 \\ 7 & 5 \end{pmatrix} + 4 \det \begin{pmatrix} 3 & 1 \\ 7 & 6 \end{pmatrix} \\ &= 9(5-36) - 2(15-42) + 4(18-7) = -181.\end{aligned}$$

Podemos também calcular o determinante utilizando as operações elementares com matrizes. A adição a uma linha de uma outra multiplicada por um número não altera o valor do determinante. A troca de duas linhas altera o sinal do determinante. Aplicando as transformações $\ell_2 \leftarrow \ell_2 - \frac{1}{3}\ell_1$ e $\ell_3 \leftarrow \ell_3 - \frac{7}{9}\ell_1$, obtemos

$$\Delta = \det \begin{pmatrix} 9 & 2 & 4 \\ 0 & \frac{1}{3} & \frac{14}{3} \\ 0 & \frac{40}{9} & \frac{17}{9} \end{pmatrix}.$$

Aplicando agora a transformação $\ell_3 \leftarrow \ell_3 - \frac{40}{3}\ell_2$, vemos que

$$\Delta = \det \begin{pmatrix} 9 & 1 & 6 \\ 0 & \frac{1}{3} & \frac{14}{3} \\ 0 & 0 & -\frac{181}{3} \end{pmatrix}.$$

Obviamente o determinante de uma matriz triangular é o produto dos elementos diagonais. Portanto, novamente temos $\Delta = -181$.

□

6.6. Aplicações da teoria dos determinantes

Nesta secção vamos mostrar como os determinantes ajudam na resolução de alguns problemas.

O primeiro exemplo é a Regra de Cramer, que dá a resolução de um sistema de n equações lineares e n incógnitas $Ax = b$, desde que $\det(A) \neq 0$. O processo de resolução consiste em calcular cada incógnita x_i a partir de uma fórmula explícita que envolve determinantes.

As colunas da matriz A vamos designar por a_j, $j = \overline{1,n}$.

Teorema 6.8. (Regra de Cramer)

Consideremos o sistema de n equações lineares e n incógnitas $Ax = b$, onde $A = (a_i^j)$, $x = (x_1, \ldots, x_n)$ e $b = (b_1, \ldots, b_n)$ são vetores coluna. Se $\det(A) \neq 0$, então

$$x_i = \frac{\det(a_1, \ldots, a_{i-1}, b, a_{i+1}, \ldots, a_n)}{\det(A)}, i = \overline{1,n}. \tag{6.8}$$

A matriz que aparece no numerador é obtida a partir da matriz A substituindo a sua coluna a_i pelo vetor b.

Demonstração. O sistema de equações lineares $Ax = b$ pode ser escrito na forma

$$x_1 a_1 + \cdots + x_i a_i + \cdots + x_n a_n = b,$$

que é equivalente a

$$x_1 a_1 + \cdots + (x_i a_i - b) + \cdots + x_n a_n = 0.$$

Isto significa que as colunas $a_1, \ldots, x_i a_i - b, \ldots, a_n$ da matriz

$$C = \begin{pmatrix} a_1^1 & \cdots & x_i a_1^i - b_1 & \cdots & a_1^n \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_n^1 & \cdots & x_i a_n^i - b_n & \cdots & a_n^n \end{pmatrix}$$

são linearmente dependentes, logo, $\det(C) = 0$ (veja o Corolário 6.5). Pelo Corolário 6.1, temos

$$\det(C) = x_i \det(A) - \det(a_1, \ldots, a_{i-1}, b, a_{i+1} \ldots, a_n) = 0$$

donde concluímos que

$$x_i = \frac{\det(a_1, \ldots, a_{i-1}, b, a_{i+1}, \ldots, a_n)}{\det(A)}, i = \overline{1, n}.$$

■

Exemplo 6.6. Utilizando a Regra de Cramer para resolver o sistema

$$\begin{cases} x^1 + 2x^2 = 4, \\ 3x^1 + x^2 = 7, \end{cases}$$

temos

$$x^1 = \frac{\det\begin{pmatrix} 4 & 2 \\ 7 & 1 \end{pmatrix}}{\det\begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix}} = \frac{4-14}{1-6} = 2, \quad x^2 = \frac{\det\begin{pmatrix} 1 & 4 \\ 3 & 7 \end{pmatrix}}{\det\begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix}} = \frac{7-12}{1-6} = 1.$$

□

Apesar da Regra de Cramer não ser muito interessante do ponto de vista prático, é importante do ponto de vista teórico. Apresentamos a seguir uma sua aplicação.

Sejam A uma matriz com $\det(A) \neq 0$ e A_k e b_k, $k = 1, 2, \ldots$, sucessões de matrizes e vetores cujos elementos convergem para os respetivos elementos de A e b. Obviamente, $\det(A_k) \neq 0$ para k's suficientemente grandes, o que significa que para esses k's os sistemas $A_k x = b_k$ são possíveis e determinados.

Corolário 6.6

A sucessão das soluções x_k dos sistemas $A_k x = b_k$ converge para a solução do sistema $Ax = b$.

Demonstração. Este resultado é uma consequência imediata das fórmulas (6.8). ■

Um outro exemplo de aplicação da teoria de determinantes é a fórmula para a matriz inversa.

Teorema 6.9

Sejam A uma aplicação linear invertível e $B = A^{-1}$ a sua aplicação inversa. Então, $\det(A) \neq 0$ e as componentes da matriz B são

$$b_i^j = \frac{A_j^i}{\det A}, \tag{6.9}$$

onde A_j^i são os complementos algébricos dos elementos a_j^i (atenção à troca dos índices em A_j^i).

Demonstração. Sejam a_i^j e b_i^j, $i, j = \overline{1,n}$, as componentes das matrizes A e B, respetivamente. Como $AB = I$, temos que

$$1 = \det(I) = \det(AB) = \det(A)\det(B).$$

Daqui vemos que $\det(A) \neq 0$. Pela fórmula para o produto das matrizes vemos que o elemento da matriz AB que está na linha k e na coluna i é igual a

$$\sum_{j=1}^{n} a_k^j b_j^i = \begin{cases} 0, & \text{se } k \neq i, \\ 1, & \text{se } k = i. \end{cases}$$

Fixando i, temos um sistema linear para encontrar $b_1^i, b_2^i, \ldots, b_n^i$. Como $\det(A) \neq 0$, pela Regra de Cramer, existe uma única solução do sistema (6.9). ■

Mais uma aplicação das propriedades dos determinantes é a obtenção do valor do *determinante de Vandermonde*

$$D_n(x_1, x_2, \ldots, x_n) = \det \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ \vdots & \vdots & \ddots & \vdots \\ (x_1)^{n-1} & (x_2)^{n-1} & \cdots & (x_n)^{n-1} \end{pmatrix},$$

onde x_i, $i = \overline{1,n}$, são números diferentes. Este determinante aparece em vários ramos da Matemática. O seu valor, como veremos, é diferente de zero. Isto, por exemplo, implica que as funções

$$1, x, x^2, \ldots, x^{n-1}$$

são linearmente independentes, donde podemos concluir que estas funções formam uma base no espaço $\mathcal{P}^{n-1}$ e, portanto, que $\dim(\mathcal{P}^n) = n + 1$.

Para calcular o valor do determinante, vamos usar o seguinte método, que, até um certo ponto, é parecido com o Método de Gauss. Nomeadamente, efetuando as operações

$$l_n \leftarrow l_n - x_1 l_{n-1},$$

$$l_{n-1} \leftarrow l_{n-1} - x_1 l_{n-2},$$

$$\vdots$$

$$l_2 \leftarrow l_2 - x_1 l_1,$$

obtemos

$$D_n(x_1, x_2, \ldots, x_n) = \det \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 0 & x_2 - x_1 & \cdots & x_n - x_1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & (x_2)^{n-1} - x_1(x_2)^{n-2} & \cdots & (x_n)^{n-1} - x_1(x_n)^{n-2} \end{pmatrix}$$

$$= \det \begin{pmatrix} x_2 - x_1 & \cdots & x_n - x_1 \\ \vdots & \ddots & \vdots \\ (x_2)^{n-1} - x_1(x_2)^{n-2} & \cdots & (x_n)^{n-1} - x_1(x_n)^{n-2} \end{pmatrix}$$

$$= (x_2 - x_1) \ldots (x_n - x_1) D_{n-1}(x_2, x_3, \ldots, x_n).$$

Continuando este procedimento, obtemos

$$D_n(x_1, x_2, \ldots, x_n) = \prod_{1 \leqslant j < i \leqslant n} (x_i - x_j). \tag{6.10}$$

Como todos os números x_i são diferentes, vemos que $D_n(x_1, x_2, \ldots, x_n) \neq 0$.

Qual é a utilidade do determinante de Vandermonde? Uma, das suas muitas aplicações, é a seguinte. Consideremos um problema de interpolação polinomial. Temos um conjunto finito de pares de pontos

$$(x_0, y_0), (x_1, y_1), \ldots, (x_n, y_n),$$

que podem ser o resultado de medição (numa experiência física ou química) dos valores $y_k = f(x_k)$ de uma função f nos pontos $x_0 < x_1 < \ldots < x_n$. De alguma maneira é necessário encontrar uma forma analítica da função f. Vamos tentar construir um polinómio

$$P(x) = c_0 + c_1 x + c_2 x^2 + \cdots + c_n x^n$$

que verifica as condições $P(x_k) = y_k$, $k = \overline{0, n}$. Este problema é equivalente à resolução do sistema linear

$$\begin{cases} c_0 + c_1 x_0 + c_2 x_0^2 + \cdots + c_n x_0^n = y_0, \\ c_0 + c_1 x_1 + c_2 x_1^2 + \cdots + c_n x_1^n = y_1, \\ \vdots \quad \vdots \quad \vdots \quad \ddots \quad \vdots \quad \vdots \\ c_0 + c_1 x_n + c_2 x_n^2 + \cdots + c_n x_n^n = y_n, \end{cases}$$

onde as incógnitas são os coeficientes do polinómio P. O determinante deste sistema é o determinante de Vandermonde

$$D_{n+1}(x_0, x_1, \ldots, x_n) \neq 0.$$

Portanto, o sistema é possível e determinado.

Para encontrar os coeficientes c_k analiticamente podemos usar a Regra de Cramer, mas este caminho implica fazer muitos cálculos. Existe uma maneira mais simples e

elegante para resolver o problema. Fixemos um $x \neq x_k$, $k = \overline{0,n}$, e consideremos um sistema de equações lineares alargado com mais uma incógnita c_{-1}

$$\begin{cases} c_{-1}P(x) = 0, \\ c_{-1}y_0 \ = 0, \\ c_{-1}y_1 \ + c_0 + c_1x_1 + c_2x_1^2 + \cdots + c_nx_1^n = 0, \\ \quad \vdots \qquad \vdots \quad \vdots \qquad \vdots \quad \ddots \quad \vdots \quad \vdots \\ c_{-1}y_n \ + c_0 + c_1x_n + c_2x_n^2 + \cdots + c_nx_n^n = 0. \end{cases}$$

Este sistema é homogéneo e tem uma solução não trivial $(-1, c_0, c_1, \ldots, c_n)$. Isto significa que o determinante deste sistema é zero, ou seja,

$$\det \begin{pmatrix} P(x) & 1 & x & x^2 & \cdots & x^n \\ y_0 & 1 & x_0 & x_0^2 & \cdots & x_0^n \\ y_1 & 1 & x_1 & x_1^2 & \cdots & x_1^n \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ y_n & 1 & x_n & x_n^2 & \cdots & x_n^n \end{pmatrix} = 0.$$

Desenvolvendo este determinante ao longo da primeira coluna obtemos

$$\begin{aligned} 0 = P(x)&D_{n+1}(x_0, x_1, x_2, \ldots, x_n) \\ &- y_0 D_{n+1}(x, x_1, x_2, \ldots, x_n) \\ &+ y_1 D_{n+1}(x, x_0, x_2, \ldots, x_n) \\ &+ \cdots \\ &+ (-1)^{(n+1)} y_n D_{n+1}(x, x_0, x_1, \ldots, x_{n-1}). \end{aligned}$$

Trocando as colunas no determinante de Vandermonde temos

$$P(x) = \sum_{k=0}^{n} y_k \frac{D_{n+1}(x_0, x_1, \ldots, x_{k-1}, x, x_{k+1}, \ldots, x_n)}{D_{n+1}(x_0, x_1, \ldots, x_n)}.$$

Agora, aplicando a fórmula (6.10), chegamos ao resultado final

$$P(x) = \sum_{k=0}^{n} y_k \frac{(x - x_0) \ldots (x - x_{k-1})(x - x_{k+1}) \ldots (x - x_n)}{(x_k - x_0) \ldots (x_k - x_{k-1})(x_k - x_{k+1}) \ldots (x_k - x_n)}.$$

Este polinómio é conhecido como *polinómio de interpolação de Lagrange*.

6.7. Problema geral de construção de uma projeção ortogonal e determinante de Gram

Sejam $\mathbb{X}$ um espaço Euclidiano de dimensão n e $x \in \mathbb{X}$. Consideremos um conjunto ortonormal $\{e_1, \ldots, e_m\}$, $m < n$. Como sabemos, o vetor $x - \sum_{k=1}^{m} \langle x, e_k \rangle e_k$ é ortogonal ao subespaço $\mathbb{L} = \text{Lin}\{e_1, \ldots, e_m\}$ e o vetor $\sum_{k=1}^{m} \langle x, e_k \rangle e_k$ é a projeção ortogonal do

vetor x sobre o subespaço $\mathbb{L}$. Mas como construir uma projeção ortogonal sobre um subespaço quando neste subespaço não há uma base ortonormal? Este problema, que evidentemente é mais complicado, vamos resolver nesta secção.

Consideremos $x \in \mathbb{X}$ e um seu subespaço com base $\{b_1, \ldots, b_m\}$. A projeção que estamos a procurar pode ser representada como uma combinação linear dos vetores $\{b_1, \ldots, b_m\}$, ou seja,

$$y = \beta_1 b_1 + \cdots + \beta_m b_m. \tag{6.11}$$

O vetor $h = x - y$ tem de ser ortogonal a todos os vetores $\{b_1, \ldots, b_m\}$. Esta condição dá-nos o sistema de equações lineares

$$\begin{cases} \langle x - y, b_1 \rangle = 0, \\ \quad \vdots \qquad \vdots \\ \langle x - y, b_m \rangle = 0. \end{cases}$$

Daqui, utilizando (6.11), obtemos

$$\begin{cases} \beta_1 \langle b_1, b_1 \rangle + \beta_2 \langle b_2, b_1 \rangle + \cdots + \beta_m \langle b_m, b_1 \rangle = \langle x, b_1 \rangle, \\ \beta_1 \langle b_1, b_2 \rangle + \beta_2 \langle b_2, b_2 \rangle + \cdots + \beta_m \langle b_m, b_2 \rangle = \langle x, b_2 \rangle, \\ \quad \vdots \qquad\qquad \vdots \qquad \ddots \qquad \vdots \qquad\qquad \vdots \\ \beta_1 \langle b_1, b_m \rangle + \beta_2 \langle b_2, b_m \rangle + \cdots + \beta_m \langle b_m, b_m \rangle = \langle x, b_m \rangle. \end{cases}$$

O determinante da matriz dos coeficientes deste sistema é

$$D = \det \begin{pmatrix} \langle b_1, b_1 \rangle & \langle b_2, b_1 \rangle & \cdots & \langle b_m, b_1 \rangle \\ \langle b_1, b_2 \rangle & \langle b_2, b_2 \rangle & \cdots & \langle b_m, b_2 \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle b_1, b_m \rangle & \langle b_2, b_b \rangle & \cdots & \langle b_m, b_m \rangle \end{pmatrix}.$$

Aplicando a Regra de Cramer, encontramos

$$\beta_j = \frac{1}{D} \det \begin{pmatrix} \langle b_1, b_1 \rangle & \cdots & \langle b_{j-1}, b_1 \rangle & \langle x, b_1 \rangle & \langle b_{j+1}, b_1 \rangle & \cdots & \langle b_m, b_1 \rangle \\ \langle b_1, b_2 \rangle & \cdots & \langle b_{j-1}, b_2 \rangle & \langle x, b_2 \rangle & \langle b_{j+1}, b_2 \rangle & \cdots & \langle b_m, b_2 \rangle \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \langle b_1, b_m \rangle & \cdots & \langle b_{j-1}, b_m \rangle & \langle x, b_m \rangle & \langle b_{j+1}, b_m \rangle & \cdots & \langle b_m, b_m \rangle \end{pmatrix},$$

para $j = \overline{1, m}$.

O determinante D aparece na resolução de muitos outros problemas.

Definição 6.11. Determinante de Gram

Chamamos *determinante de Gram* dos vetores $x_1, \ldots, x_k \in \mathbb{X}$, que representamos por $G(x_1, \ldots, x_k)$, ao determinante da matriz dos produtos escalares $\langle x_i, x_j \rangle$, $i, j = \overline{1,k}$, ou seja,

$$G(x_1, \ldots, x_k) = \det \begin{pmatrix} \langle x_1, x_1 \rangle & \langle x_1, x_2 \rangle & \cdots & \langle x_1, x_k \rangle \\ \langle x_2, x_1 \rangle & \langle x_2, x_2 \rangle & \cdots & \langle x_2, x_k \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle x_k, x_1 \rangle & \langle x_k, x_2 \rangle & \cdots & \langle x_k, x_k \rangle \end{pmatrix}.$$

Vamos estabelecer uma propriedade importante dos determinantes de Gram relacionada com o processo de ortogonalização de Gram-Schmidt.

Teorema 6.10

Sejam $x_1, \ldots, x_k$ vetores de $\mathbb{X}$ e $G(x_1, \ldots, x_k)$ o determinante de Gram. Consideremos os vetores $y_1, \ldots, y_k$ de $\mathbb{X}$, ortogonais dois a dois, ou seja, $\langle y_i, y_j \rangle = 0$, $i \neq j = \overline{1,k}$, construídos através do processo de ortogonalização de Gram-Schmidt (veja o Teorema 4.6):

$$\begin{aligned} y_1 &= x_1, \\ y_2 &= \beta_1^{(2)} y_1 + x_2 \in (\mathrm{Lin}\{y_1\})^\perp, \\ y_3 &= \beta_1^{(3)} y_1 + \beta_2^{(3)} y_2 + x_3 \in (\mathrm{Lin}\{y_1, y_2\})^\perp \\ &\vdots \\ y_k &= \beta_1^{(k)} y_1 + \beta_2^{(k)} y_2 + \cdots + \beta_{k-1}^{(k)} y_{k-1} + x_k \in (\mathrm{Lin}\{y_1, \ldots, y_{k-1}\})^\perp. \end{aligned}$$

Então,

$$G(x_1, \ldots, x_k) = \langle y_1, y_1 \rangle \langle y_2, y_2 \rangle \cdots \langle y_k, y_k \rangle.$$

Demonstração. Substituindo $x_1 = y_1$ no determinante de Gram, obtemos

$$G(x_1, \ldots, x_k) = \det \begin{pmatrix} \langle y_1, y_1 \rangle & \langle y_1, x_2 \rangle & \cdots & \langle y_1, x_k \rangle \\ \langle x_2, y_1 \rangle & \langle x_2, x_2 \rangle & \cdots & \langle x_2, x_k \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle x_k, y_1 \rangle & \langle x_k, x_2 \rangle & \cdots & \langle x_k, x_k \rangle \end{pmatrix}.$$

Como $y_2 = \beta_1^{(2)} y_1 + x_2$, aplicando as operações elementares $l_2 \leftarrow l_2 + \beta_1^{(2)} l_1$ e $c_2 \leftarrow c_2 + \beta_1^{(2)} c_1$, obtemos a matriz com o mesmo valor do determinante

$$G(x_1, \ldots, x_k) = \det \begin{pmatrix} \langle y_1, y_1 \rangle & \langle y_1, y_2 \rangle & \langle y_1, x_3 \rangle & \cdots & \langle y_1, x_k \rangle \\ \langle y_2, y_1 \rangle & \langle y_2, y_2 \rangle & \langle y_2, x_3 \rangle & \cdots & \langle y_2, x_k \rangle \\ \langle x_3, y_1 \rangle & \langle x_3, y_2 \rangle & \langle x_3, y_3 \rangle & \cdots & \langle x_3, x_k \rangle \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \langle x_k, y_1 \rangle & \langle x_k, y_2 \rangle & \langle x_k, y_3 \rangle & \cdots & \langle x_k, x_k \rangle \end{pmatrix}.$$

Prosseguindo do mesmo modo, transformamos sucessivamente as restantes linhas e colunas da matriz anterior usando as operações elementares $l_i \leftarrow l_i + \beta_1^{(i)} l_1 + \cdots + \beta_{i-1}^{(i)} l_{i-1}$ e $c_i \leftarrow c_i + \beta_1^{(i)} c_1 + \cdots + \beta_{i-1}^{(i)} c_{i-1}$ para $i = 3, \ldots, k$, obtendo no final

$$G(x_1, \ldots, x_k) = \det \begin{pmatrix} \langle y_1, y_1 \rangle & \langle y_1, y_2 \rangle & \langle y_1, y_3 \rangle & \cdots & \langle y_1, y_k \rangle \\ \langle y_2, y_1 \rangle & \langle y_2, y_2 \rangle & \langle y_2, y_3 \rangle & \cdots & \langle y_2, y_k \rangle \\ \langle y_3, y_1 \rangle & \langle y_3, y_2 \rangle & \langle y_3, y_3 \rangle & \cdots & \langle y_3, y_k \rangle \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \langle y_k, y_1 \rangle & \langle y_k, y_2 \rangle & \langle y_k, y_3 \rangle & \cdots & \langle y_k, y_k \rangle \end{pmatrix}.$$

Logo, temos

$$G(x_1, \ldots, x_k) = \langle y_1, y_1 \rangle \langle y_2, y_2 \rangle \cdots \langle y_k, y_k \rangle,$$

pois a matriz anterior é diagonal.

■

6.8. Determinantes de Gram e volumes de paralelepípedos

Consideremos um espaço Euclidiano $\mathbb{X}$ e uma base ortonormal de $\mathbb{X}$. Recorrendo ao determinante de Gram podemos, de outra maneira, mostrar que os determinantes permitem calcular os volumes $\text{Vol}(x_1, \ldots, x_k)$ dos paralelepípedos gerados pelos vetores $x_1, \ldots, x_k$.

Consideremos um vetor x_1. Então, o volume $\text{Vol}(x_1)$ do paralelepípedo gerado pelo vetor x_1 é a norma do vetor x_1 (veja a Fig. 6.3). Sendo assim, temos

$$\text{Vol}(x_1) = |x_1|.$$

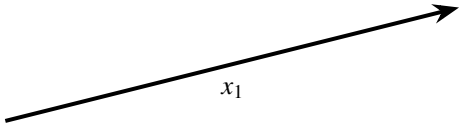


Figura 6.3
Vetor x_1.

Dois vetores x_1 e x_2 definem um paralelepípedo, que é um paralelogramo, cujo volume $\text{Vol}(x_1, x_2)$ é a área do paralelogramo, que pode ser calculada multiplicando o comprimento da base $|x_1|$ pela altura h_2 (veja a Fig. 6.4). Sendo assim, temos

$$\text{Vol}(x_1, x_2) = \text{Vol}(x_1) h_2 = |x_1| h_2.$$

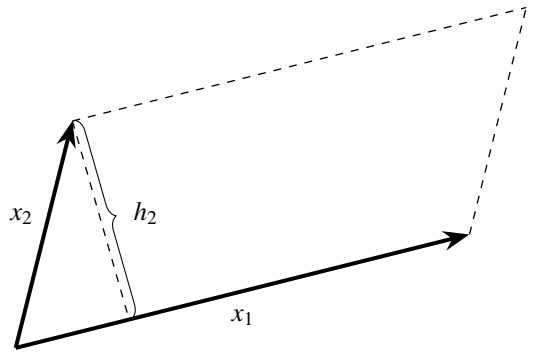


Figura 6.4
Paralelogramo gerado pelos vetores x_1 e x_2.

Voltando ao Teorema 6.10, em que foi ortogonalizado o conjunto de vetores $\{x_1, \ldots, x_k\}$, vemos que a altura h_2 é igual à norma do vetor y_2. Este vetor é ortogonal ao vetor $x_1 = y_1$. Assim, temos

$$\text{Vol}(x_1, x_2) = \text{Vol}(x_1)|y_2| = |x_1||y_2| = |y_1||y_2|.$$

Consideremos um paralelepípedo gerado por três vetores x_1, x_2 e x_3. O seu volume $\text{Vol}(x_1, x_2, x_3)$ é calculado multiplicando a área da base, ou seja, o volume $\text{Vol}(x_1, x_2)$ do paralelepípedo definido por x_1 e x_2, pela altura h_3, que coincide com a norma do vetor y_3. Este vetor é ortogonal aos vetores x_1 e x_2 (veja a Fig. 6.5). Sendo assim, temos

$$\text{Vol}(x_1, x_2, x_3) = \text{Vol}(x_1, x_2)|y_3| = |y_1||y_2||y_3|.$$

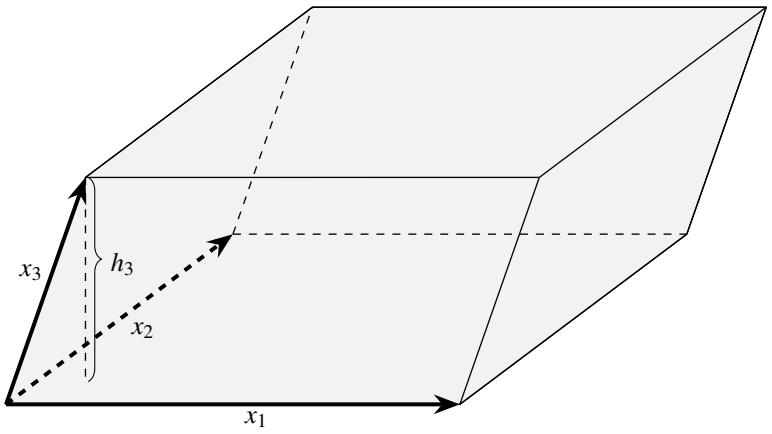


Figura 6.5
Paralelepípedo gerado pelos vetores x_1, x_2 e x_3.

O volume $\text{Vol}(x_1, \ldots, x_k)$ de um paralelepípedo gerado pelos vetores $x_1, \ldots, x_k$ é o produto do volume $\text{Vol}(x_1, \ldots, x_{k-1})$ e da respetiva altura h_k

$$\text{Vol}(x_1, \ldots, x_k) = \text{Vol}(x_1, \ldots, x_{k-1})h_k = \text{Vol}(x_1, \ldots, x_{k-1})|y_k|$$

(veja a Fig. 6.6). Portanto, por indução, temos que

$$\text{Vol}(x_1, \ldots, x_k) = |y_1| \cdots |y_k|.$$

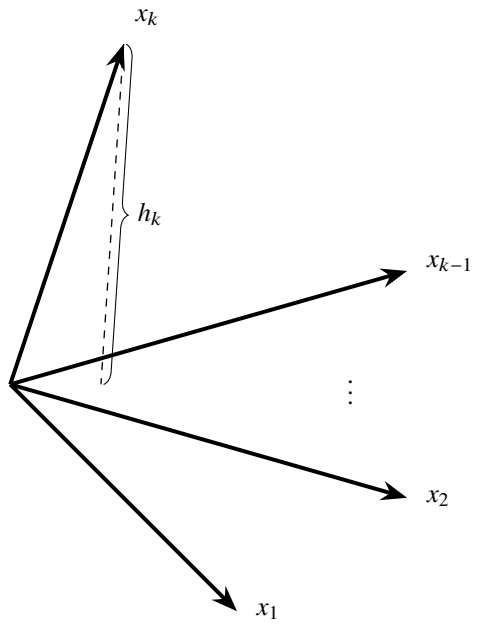


Figura 6.6
Paralelepípedo gerado pelos vetores $x_1, \ldots, x_k$.

Esta fórmula dá a possibilidade de estabelecer a relação entre o determinante de um conjunto de vetores, e portanto de uma certa matriz, e o volume de um paralelepípedo gerado por esses vetores.

Teorema 6.11

Seja $\text{Vol}(x_1, \ldots, x_k)$ o volume do paralelepípedo gerado pelos vetores $x_1, \ldots, x_k \in \mathbb{X}$. Então,

$$\text{Vol}(x_1, \ldots, x_k) = |\det(X)|,$$

onde X é a matriz cujas linhas contêm as coordenadas dos vetores x_1, $x_2, \ldots, x_k$, ou seja,

$$X = \begin{pmatrix} x_1^1 & x_1^2 & \cdots & x_1^k \\ x_2^1 & x_2^2 & \cdots & x_2^k \\ \vdots & \vdots & \ddots & \vdots \\ x_k^1 & x_k^2 & \cdots & x_k^k \end{pmatrix}.$$

Demonstração. Como

$$\begin{aligned} \text{Vol}(x_1, \ldots, x_k)^2 &= |y_1|^2 |y_2|^2 \cdots |y_k|^2 \\ &= \langle y_1, y_1 \rangle \langle y_2, y_2 \rangle \cdots \langle y_k, y_k \rangle = G(x_1, \ldots, x_k) \\ &= \det \begin{pmatrix} \langle x_1, x_1 \rangle & \langle x_1, x_2 \rangle & \cdots & \langle x_1, x_k \rangle \\ \langle x_2, x_1 \rangle & \langle x_2, x_2 \rangle & \cdots & \langle x_2, x_k \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle x_k, x_1 \rangle & \langle x_k, x_2 \rangle & \cdots & \langle x_k, x_k \rangle \end{pmatrix}. \end{aligned}$$

Tendo em conta que $x_1 = (x_1^1, x_1^2, \ldots, x_1^k)$,..., $x_k = (x_k^1, x_k^2, \ldots, x_k^k)$ e calculando os produtos escalares na matriz anterior, obtemos

$$\begin{aligned}&\mathrm{Vol}(x_1, \ldots, x_k)^2 \\ &\quad = \det \begin{pmatrix} x_1^1 x_1^1 + \cdots + x_1^k x_1^k & x_1^1 x_2^1 + \cdots + x_1^k x_2^k & \cdots & x_1^1 x_k^1 + \cdots + x_1^k x_k^k \\ x_2^1 x_1^1 + \cdots + x_2^k x_1^k & x_2^1 x_2^1 + \cdots + x_2^k x_2^k & \cdots & x_2^1 x_k^1 + \cdots + x_2^k x_k^k \\ \vdots & \vdots & \ddots & \vdots \\ x_k^1 x_1^1 + \cdots + x_k^k x_1^k & x_k^1 x_2^1 + \cdots + x_k^k x_2^k & \cdots & x_k^1 x_k^1 + \cdots + x_k^k x_k^k \end{pmatrix}.\end{aligned}$$

A matriz da expressão anterior pode ser escrita como produto de duas matrizes:

$$\begin{aligned}\mathrm{Vol}(x_1, \ldots, x_k)^2 &= \det \left(\begin{pmatrix} x_1^1 & x_1^2 & \cdots & x_1^k \\ x_2^1 & x_2^2 & \cdots & x_2^k \\ \vdots & \vdots & \ddots & \vdots \\ x_k^1 & x_k^2 & \cdots & x_k^k \end{pmatrix} \begin{pmatrix} x_1^1 & x_2^1 & \cdots & x_k^1 \\ x_1^2 & x_2^2 & \cdots & x_k^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^k & x_2^k & \cdots & x_k^k \end{pmatrix} \right) \\ &= \det(XX^{\mathsf{T}}) = \det(X) \det(X^{\mathsf{T}}).\end{aligned}$$

Como $\det(X) = \det(X^{\mathsf{T}})$, então temos

$$\mathrm{Vol}(x_1, \ldots, x_k)^2 = \det(X)^2$$

e daí resulta que

$$\mathrm{Vol}(x_1, \ldots, x_k) = |\det(X)|.$$

■

Recordemos que $\langle y_j, y_j \rangle \leqslant \langle x_j, x_j \rangle$ (veja a desigualdade (4.5)). Logo, temos a *desigualdade de Hadamard*:

$$\det(X)^2 = G(x_1, \ldots, x_k) = \langle y_1, y_1 \rangle \ldots \langle y_k, y_k \rangle \leqslant \langle x_1, x_1 \rangle \ldots \langle x_k, x_k \rangle.$$

Obtemos a igualdade se o conjunto $\{x_1, \ldots, x_k\}$ for ortogonal. Isto significa que o volume do paralelepípedo gerado pelas linhas de X não pode exceder o volume de um paralelepípedo retangular com arestas de comprimentos $|x_1|, \ldots, |x_k|$.

6.9. Fatorização QR

O processo de ortogonalização de Gram-Schmidt (Teorema 4.6) pode ser aplicado à construção da *fatorização QR* (ou *decomposição QR*) de matrizes, que é utilizada em vários métodos numéricos da Álgebra Linear.

Consideremos uma matriz A cujos vetores-coluna são linearmente independentes. Através do processo de ortogonalização de Gram-Schmidt podemos transformar o conjunto dos vetores-coluna num conjunto ortonormal. Vamos chamar à matriz formada pelos vetores-coluna ortonormais matriz Q. Como veremos, existe uma matriz R triangular superior tal que $A = QR$. O que é interessante é que o determinante

de R está relacionado com o determinante de Gram dos vetores-coluna de A. Mais precisamente, temos o seguinte resultado.

Teorema 6.12

Seja A uma matriz com m linhas e com n colunas linearmente independentes. Então, A admite uma única fatorização QR (ou decomposição QR) na forma

$$A = QR,$$

onde a matriz Q tem as mesmas dimensões da matriz A. Os vetores-coluna da matriz Q formam um conjunto ortonormal. A matriz R é uma matriz com n linhas e n colunas, triangular superior, invertível e cujos elementos na diagonal são números positivos.

Demonstração. Seja $\{p_1, \ldots, p_n\}$ um conjunto linearmente independente do espaço Euclidiano $\mathbb{X}$ de dimensão finita. Consideremos a base ortonormal $\{q_1, \ldots, q_n\}$ obtida aplicando o processo de Gram-Schmidt a $\{p_1, \ldots, p_n\}$, isto é,

$$q_1 = \frac{p_1}{|p_1|},$$

$$q_m = \frac{p_m - s_{m-1}}{|p_m - s_{m-1}|}, \quad m = \overline{2, n},$$

onde

$$s_{m-1} = \sum_{k=1}^{m-1} \langle p_m, q_k \rangle q_k.$$

Recordemos que

$$p_m \notin \mathbb{L}_{m-1} = \text{Lin}\{q_1, \ldots, q_{m-1}\} = \text{Lin}\{p_1, \ldots, p_{m-1}\}$$

e que a soma s_{m-1} é a projeção ortogonal do vetor p_m no subespaço $\mathbb{L}_{m-1}$ e $q_m \in \mathbb{L}_{m-1}^{\perp}$.

Sejam A a matriz cujas colunas são os vetores $p_1, \ldots, p_n$ e Q a matriz cujas colunas são os vetores $q_1, \ldots, q_n$. Então, como a base $\{q_1, \ldots, q_n\}$ é ortonormal, temos

$$p_1 = \langle p_1, q_1 \rangle q_1 + \langle p_1, q_2 \rangle q_2 + \cdots + \langle p_1, q_n \rangle q_n,$$

$$p_2 = \langle p_2, q_1 \rangle q_1 + \langle p_2, q_2 \rangle q_2 + \cdots + \langle p_2, q_n \rangle q_n,$$

$$\vdots$$

$$p_n = \langle p_n, q_1 \rangle q_1 + \langle p_n, q_2 \rangle q_2 + \cdots + \langle p_n, q_n \rangle q_n.$$

Fazendo

$$R = \begin{pmatrix} \langle p_1, q_1 \rangle & \langle p_2, q_1 \rangle & \cdots & \langle p_n, q_1 \rangle \\ \langle p_1, q_2 \rangle & \langle p_2, q_2 \rangle & \cdots & \langle p_n, q_2 \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle p_1, q_n \rangle & \langle p_2, q_n \rangle & \cdots & \langle p_n, q_n \rangle \end{pmatrix},$$

vemos que as equações anteriores são equivalentes à igualdade matricial

$$A = QR.$$

Como o vetor q_j é ortogonal a todos os vetores $p_1, \ldots, p_{j-1}$ para $j \geqslant 2$, temos

$$R = \begin{pmatrix} \langle p_1, q_1\rangle & \langle p_2, q_1\rangle & \cdots & \langle p_{n-1}, q_1\rangle & \langle p_n, q_1\rangle \\ 0 & \langle p_2, q_2\rangle & \cdots & \langle p_{n-1}, q_2\rangle & \langle p_n, q_2\rangle \\ \vdots & 0 & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \langle p_{n-1}, q_{n-1}\rangle & \langle p_n, q_{n-1}\rangle \\ 0 & 0 & \cdots & 0 & \langle p_n, q_n\rangle \end{pmatrix}.$$

A matriz R é invertível porque os elementos da diagonal de R são números positivos. De facto, como a norma do vetor p_m é estritamente maior do que a norma da sua projeção ortogonal s_{m-1}, temos

$$\langle p_m, q_m\rangle = \left\langle p_m, \frac{p_m - \sum_{k=1}^{m-1}\langle p_m, q_k\rangle q_k}{|p_m - s_{m-1}|}\right\rangle = \frac{|p_m|^2 - |s_{m-1}|^2}{|p_m - s_{m-1}|} > 0.$$

■

Notemos que

$$\langle p_m, q_m\rangle = \langle p_m - s_{m-1} + s_{m-1}, q_m\rangle = \langle p_m - s_{m-1}, q_m\rangle = |p_m - s_{m-1}|.$$

Portanto, o determinante de Gram $G(p_1, \ldots, p_n)$ está relacionado com o determinante da matriz R da forma

$$\det(R)^2 = \langle p_1, q_1\rangle^2 \cdots \langle p_n, q_n\rangle^2 = |p_1|^2|p_2 - s_1|^2 \cdots |p_n - s_{n-1}|^2 = G(p_1, \ldots, p_n)$$

(veja o Teorema 6.10). Logo, o volume do paralelepípedo gerado pelos vetores $p_1, \ldots, p_n$ é igual ao produto dos elementos da diagonal de R

$$\text{Vol}(p_1, \ldots, p_k) = \det(R).$$

Exemplo 6.7. Determinemos a fatorização QR da matriz $A = \left(\begin{smallmatrix} 1 & 2 \\ -3 & 2 \end{smallmatrix}\right)$. Aplicando o processo de ortogonalização de Gram-Schmidt aos vetores-coluna $e_1 = \left(\begin{smallmatrix} 1 \\ -3 \end{smallmatrix}\right)$ e $e_2 = \left(\begin{smallmatrix} 2 \\ 2 \end{smallmatrix}\right)$, utilizando o vetor normalizado

$$q_1 = \frac{e_1}{|e_1|} = \frac{1}{\sqrt{10}}\begin{pmatrix} 1 \\ -3 \end{pmatrix},$$

obtemos

$$\begin{aligned} y_2 = e_2 - \frac{\langle e_2, q_1\rangle}{|q_1|^2}q_1 &= \begin{pmatrix} 2 \\ 2 \end{pmatrix} - \left\langle \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \frac{1}{\sqrt{10}}\begin{pmatrix} 1 \\ -3 \end{pmatrix}\right\rangle \frac{1}{\sqrt{10}}\begin{pmatrix} 1 \\ -3 \end{pmatrix} \\ &= \begin{pmatrix} 2 \\ 2 \end{pmatrix} + \frac{2}{5}\begin{pmatrix} 1 \\ -3 \end{pmatrix} = \frac{1}{5}\begin{pmatrix} 12 \\ 4 \end{pmatrix}, \end{aligned}$$

$$q_2 = \frac{y_2}{|y_2|} = \frac{1}{\sqrt{10}}\begin{pmatrix} 3 \\ 1 \end{pmatrix}.$$

Assim, $A = QR$, onde

$$Q = \begin{pmatrix} \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \\ -\frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \end{pmatrix}$$

e

$$R = \begin{pmatrix} \langle e_1, q_1 \rangle & \langle e_2, q_1 \rangle \\ 0 & \langle e_2, q_2 \rangle \end{pmatrix} = \begin{pmatrix} \sqrt{10} & -\frac{2\sqrt{10}}{5} \\ 0 & \frac{4\sqrt{10}}{5} \end{pmatrix}.$$

□

6.10. Método dos mínimos quadrados

Os determinantes de Gram e a fatorização QR são úteis na resolução de vários problemas práticos. Consideremos, por exemplo, um sistema de equações lineares impossível. Como definir a sua "solução"? Responder a esta questão é necessário em muitos problemas que encontramos na vida real onde o sistema é obtido como resultado de experiências físicas ou químicas e pode não ter solução no sentido clássico.

Consideremos uma matriz A com m linhas e n colunas $a_1, \ldots, a_n$ e um vetor b de $\mathbb{R}^m$. O sistema de equações

$$Ax = b \tag{6.12}$$

pode não ter solução. Nesta situação é natural procurar um vetor x que minimiza a norma Euclidiana do vetor $r = Ax - b$. Este vetor x é a "solução" do problema, chamada *solução generalizada*. Obviamente, quando o sistema tem soluções, elas coincidem com as soluções generalizadas. O problema de minimização do quadrado da norma Euclidiana $|Ax - b|$ no caso do sistema (6.12) ser impossível é conhecido como o *problema de mínimos quadrados*. Se x é o vetor que minimiza $|r|^2 = |Ax - b|^2$, então Ax é a projeção ortogonal de b no subespaço $\text{Lin}\{a_1, \ldots, a_n\}$ (veja a Fig. 6.7 e o Teorema 4.3).

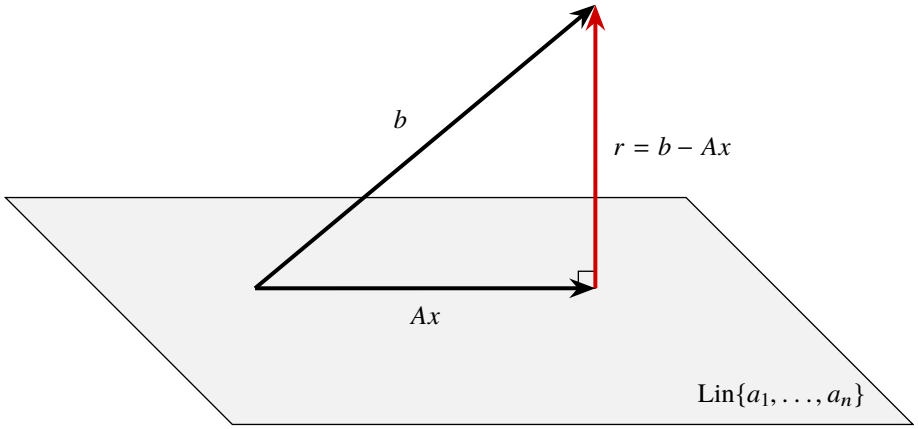


Figura 6.7
Projeção de b em $\text{Lin}\{a_1, \ldots, a_n\}$ e erro r.

Sabemos que através da construção da projeção ortogonal de b no subespaço $\text{Lin}\{a_1, \ldots, a_n\}$ temos $Ax - b \in (\text{Lin}\{a_1, \ldots, a_n\})^\perp$, donde obtemos o sistema de equações

lineares

$$\begin{cases} \langle Ax, a_1\rangle = \langle a_1, b\rangle, \\ \quad \vdots \qquad\quad \vdots \\ \langle Ax, a_n\rangle = \langle a_n, b\rangle. \end{cases}$$

Este sistema é designado por *sistema normal* associado ao sistema (6.12), sendo mais cómodo escrevê-lo na forma matricial

$$A^{\mathsf{T}}Ax = A^{\mathsf{T}}b. \tag{6.13}$$

Se as colunas da matriz A forem linearmente independentes, então

$$D = G(a_1, \ldots, a_n) = \det(A^{\mathsf{T}}A) \neq 0,$$

e as incógnitas do sistema (6.13) x_j, $j = \overline{1,n}$, podem ser obtidas pela Regra de Cramer

$$x_j = \frac{1}{D}\det\begin{pmatrix} \langle a_1, a_1\rangle & \cdots & \langle a_{j-1}, a_1\rangle & \langle x, a_1\rangle & \langle a_{j+1}, a_1\rangle & \cdots & \langle a_n, a_1\rangle \\ \langle a_1, a_2\rangle & \cdots & \langle a_{j-1}, a_2\rangle & \langle x, a_2\rangle & \langle a_{j+1}, a_2\rangle & \cdots & \langle a_n, a_2\rangle \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \langle a_1, a_n\rangle & \cdots & \langle a_{j-1}, a_n\rangle & \langle x, a_n\rangle & \langle a_{j+1}, a_n\rangle & \cdots & \langle a_n, a_n\rangle \end{pmatrix}.$$

É claro que ninguém utiliza a Regra de Cramer para a resolução de sistemas lineares pois este processo necessita um número astronómico de operações mesmo para sistemas de dimensões modestas. A estrutura dos sistemas do tipo (6.13) de grande dimensão também desaconselha a utilização do Método de Gauss pois os erros de arredondamento que o computador faz podem ter efeitos catastróficos. Este problema pode ser resolvido usando a fatorização QR da matriz A. Fazendo a substituição $A = QR$ no sistema (6.13), obtemos

$$(QR)^{\mathsf{T}}QRx = (QR)^{\mathsf{T}}b.$$

Utilizando o Teorema 5.11, obtemos

$$R^{\mathsf{T}}Q^{\mathsf{T}}QRx = R^{\mathsf{T}}Q^{\mathsf{T}}b.$$

Como os vetores-coluna de Q formam uma base ortonormal, temos $Q^{\mathsf{T}}Q = I$. Portanto, o sistema toma a forma

$$R^{\mathsf{T}}Rx = R^{\mathsf{T}}Q^{\mathsf{T}}b.$$

Como R é uma matriz invertível, R^{T} também o é. Logo, o sistema (6.13) é equivalente ao sistema

$$Rx = Q^{\mathsf{T}}b.$$

Como R é uma matriz triangular superior, é muito simples determinar a sua solução. A dificuldade principal na utilização deste método é a determinação da matriz Q. Nós demonstramos a existência desta matriz através do processo de ortogonalização de

Gram-Schmidt. Este algoritmo não pode ser usado nos métodos numéricos pois pode levar a grandes erros de arredondamento que o computador faz. Existem, no entanto, algoritmos que não sofrem deste problema.

Usando os determinantes de Gram, podemos também encontrar o valor

$$\delta = \min_x |Ax - b|.$$

De facto, δ é a altura do paralelepípedo gerado pelos vetores $a_1, \ldots, a_n, b$ e, portanto, é igual a

$$\delta = \frac{\text{Vol}(a_1, \ldots, a_n, b)}{\text{Vol}(a_1, \ldots, a_n)},$$

ou, em termos de determinantes de Gram,

$$\delta^2 = \frac{G(a_1, \ldots, a_n, b)}{G(a_1, \ldots, a_n)}.$$

O método dos mínimos quadrados tem muitas aplicações. Uma das mais importantes é a aproximação de funções por polinómios. Por exemplo, nós conhecemos a função f em alguns pontos. Estes seus valores são, por exemplo, resultado de uma experiência física. Como reconstruir a função f analiticamente? Este problema já foi considerado na Secção 6.6, onde foi construído o polinómio de interpolação de Lagrange. O grau deste polinómio é igual ao número de pontos obtidos como resultado da experiência menos um. Quando o número dos resultados é muito grande convém procurar um polinómio de um grau mais pequeno. Fazemos isto usualmente através do método dos mínimos quadrados.

Consideremos uma função $f : \mathbb{R} \to \mathbb{R}$ cujos valores nos pontos t_j, $j = \overline{1, m}$, são dados. O problema consiste em determinar o polinómio $P(t) = \xi_n t^{n-1} + \cdots + \xi_2 t + \xi_1$ de grau $n - 1$ que minimiza o valor

$$\delta(f, P) = \sum_{j=1}^{m} (f(t_j) - P(t_j))^2.$$

Nos problemas práticos temos sempre $m > n$, porque o número de medições m costuma ser grande e o objetivo é encontrar um polinómio de grau bastante pequeno que aproxima a função de uma maneira razoável.

Introduzamos o espaço Euclidiano de funções $\mathcal{F}^m$, que têm como domínio o conjunto finito de pontos t_j, $j = \overline{1, m}$, com o produto escalar

$$\langle p, q \rangle = \sum_{j=1}^{m} p(t_j) q(t_j).$$

Por exemplo, o produto escalar de t^α e t^β é

$$\langle t^\alpha, t^\beta \rangle = \sum_{j=1}^{m} (t_j)^\alpha (t_j)^\beta.$$

Então, o problema fica reduzido ao problema de projeção da função f no subespaço gerado pelos polinómios de grau menor ou igual a $n - 1$. O valor de δ satisfaz

$$\delta^2 = \frac{G(1, t, \ldots, t^{n-1}, f)}{G(1, t, \ldots, t^{n-1})}.$$

7. Vetores e valores próprios

7.1. Introdução

Este capítulo é dedicado ao estudo da estrutura de algumas classes de aplicações lineares do espaço $\mathcal{L}(\mathbb{X},\mathbb{X})$. Nesse estudo, os subespaços invariantes desempenham um papel fundamental. O estudo completo da estrutura de uma aplicação linear é feito no último capítulo do livro enquanto neste capítulo só vamos considerar classes de aplicações lineares de estrutura muito simples. Estas aplicações possuem um conjunto de subespaços invariantes de dimensão um e os vetores que geram estes subespaços formam uma base no espaço $\mathbb{X}$. Vamos elaborar técnicas para construir estes subespaços baseadas nas propriedades de determinantes. Apesar da sua simplicidade, esta classe de aplicações lineares é muito importante na resolução de vários problemas práticos.

7.2. Definições

A estrutura de uma aplicação linear descreve-se em termos do conjunto de subespaços que a aplicação transforma neles mesmos. Formalmente definimos estes subespaços assim.

Definição 7.1. Subespaço invariante

Seja $\mathbb{X}$ um espaço linear. Dizemos que o subespaço $\mathbb{L} \subset \mathbb{X}$ é um *subespaço invariante* da aplicação $A \in \mathcal{L}(\mathbb{X},\mathbb{X})$ se para todo $x \in \mathbb{L}$ temos $Ax \in \mathbb{L}$, ou seja, $A\mathbb{L} \subset \mathbb{L}$.

Os subespaços $\{0\}$ e $\mathbb{X}$, que são subespaços invariantes de todas as aplicações lineares de $\mathbb{X}$ em $\mathbb{X}$, são dois exemplos triviais. Um exemplo muito importante e não trivial é dado na seguinte definição.

Definição 7.2. Vetores e valores próprios

Seja $\mathbb{X}$ um espaço linear. Dizemos que $x \in \mathbb{X}$ é um *vetor próprio* da aplicação $A \in \mathcal{L}(\mathbb{X},\mathbb{X})$ se $x \neq 0$ e

$$Ax = \lambda x, \tag{7.1}$$

onde λ é um número. A este número damos o nome de *valor próprio* da aplicação A correspondente ao vetor próprio x.

Obviamente se x é um vetor próprio da aplicação A, então A transforma $\text{Lin}\{x\}$ em $\text{Lin}\{x\}$, ou seja, cada vetor próprio gera um subespaço invariante da aplicação A de dimensão um.

Seja x um vetor diferente de zero que verifica a equação (7.1). Então, qualquer vetor $y = \alpha x$ com $\alpha \neq 0$ também verifica a mesma equação e portanto todos eles são

vetores próprios correspondentes ao valor próprio λ. Estes vetores juntamente com o vetor nulo formam um subespaço de dimensão um. Este subespaço pode ser representado por qualquer vetor não nulo que lhe pertence. A partir de agora, quando dissermos "o vetor próprio da aplicação linear" queremos dizer "um qualquer vetor não nulo deste subespaço".

As propriedades mais importantes dos vetores próprios são dadas no seguinte teorema.

Teorema 7.1

Os vetores próprios correspondentes ao mesmo valor próprio λ juntamente com o vetor nulo formam um subespaço em $\mathbb{X}$. Os vetores próprios que correspondem aos valores próprios diferentes são linearmente independentes.

Demonstração. Se $Ax_1 = \lambda x_1$ e $Ax_2 = \lambda x_2$, então temos

$$A(\alpha x_1 + \beta x_2) = \alpha A x_1 + \beta A x_2 = \lambda(\alpha x_1 + \beta x_2),$$

quaisquer que sejam os números α e β, o que demonstra a primeira parte do teorema.

A segunda parte vamos demonstrar por indução. Quando temos um vetor próprio, não há nada para demonstrar. Suponhamos que o teorema é verdadeira para $s-1$ vetores próprios de A. Consideremos s vetores próprios $x_1, x_2, \ldots, x_s$ da aplicação A que correspondem a s valores próprios diferentes $\lambda_1, \lambda_2, \ldots, \lambda_s$. Suponhamos que existem $\alpha_1, \alpha_2, \ldots, \alpha_s \in \mathbb{K}$ tais que

$$\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_s x_s = 0 \tag{7.2}$$

e, por exemplo, $\alpha_k \neq 0$, $k \in \{1, \ldots, s\}$. Aplicando à igualdade (7.2) a aplicação A obtemos

$$\alpha_1 \lambda_1 x_1 + \alpha_2 \lambda_2 x_2 + \cdots + \alpha_s \lambda_s x_s = 0. \tag{7.3}$$

Multiplicando (7.2) por λ_m, $m \in \{1, \ldots, s\}$, $m \neq k$, e subtraindo a igualdade obtida da igualdade (7.3), vem

$$\begin{aligned}\alpha_1(\lambda_1 - \lambda_m)x_1 + \cdots + \alpha_{m-1}(\lambda_{m-1} - \lambda_m)x_{m-1} \\ + \alpha_{m+1}(\lambda_{m+1} - \lambda_m)x_{m+1} + \cdots + \alpha_s(\lambda_s - \lambda_m)x_s = 0.\end{aligned}$$

Pela hipótese de indução, todos os coeficientes nesta igualdade são nulos. Em particular $\alpha_k(\lambda_k - \lambda_m) = 0$. Logo, $\lambda_k = \lambda_m$, o que é uma contradição. Portanto, os vetores $x_1, x_2, \ldots, x_s$ são linearmente independentes. ■

No caso particular em que a aplicação linear possui um subespaço gerado pelos vetores próprios correspondentes ao valor próprio 0, a este subespaço damos o nome de *núcleo* da aplicação linear A, que denotamos por $\ker A$. Neste livro precisamos só do núcleo de uma aplicação de $\mathbb{X}$ em $\mathbb{X}$. Notemos que o conceito de núcleo pode ser introduzido para qualquer aplicação linear como o conjunto de vetores que a aplicação transforma em zero.

Consideremos agora a questão como encontrar os valores e vetores próprios. Sejam $\mathbb{X}$ um espaço linear de dimensão n sobre o conjunto de números $\mathbb{K}$, $\{e_1, e_2, \ldots, e_n\}$ uma base em $\mathbb{X}$ e $A \in \mathcal{L}(\mathbb{X}, \mathbb{X})$ uma aplicação linear à qual corresponde a matriz $A = (a^i_j)$. Suponhamos que $x = \sum_{i=1}^n \xi^i e_i$ é um vetor próprio de A, isto é, $Ax = \lambda x$, onde $\lambda \in \mathbb{K}$. Então, em coordenadas, temos o sistema

$$\sum_{i=1}^{n} a^i_j \xi^i = \lambda \xi^j, \quad j = \overline{1,n},$$

ou seja,

$$\begin{pmatrix} a^1_1 - \lambda & a^2_1 & \cdots & a^n_1 \\ a^1_2 & a^2_2 - \lambda & \cdots & a^n_2 \\ \vdots & \vdots & \ddots & \vdots \\ a^1_n & a^2_n & \cdots & a^n_n - \lambda \end{pmatrix} \begin{pmatrix} \xi^1 \\ \xi^2 \\ \vdots \\ \xi^n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{7.4}$$

Este sistema homogéneo tem uma solução não trivial se e só se

$$\Delta(\lambda) = \det(A - \lambda I) = 0. \tag{7.5}$$

Este determinante é um polinómio de grau n de λ, a que chamamos *polinómio característico* da matriz A e que denotamos por $\Delta(\lambda)$. À equação (7.5) damos o nome de *equação característica*. A cada raiz λ deste polinómio corresponde pelo menos uma solução não trivial do sistema (7.4).

Consideremos alguns exemplos.

Exemplo 7.1. Polinómio característico com n raízes diferentes. Se todas as raízes $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{K}$ do polinómio característico $\Delta(\lambda) = 0$ são diferentes, então resolvendo o sistema (7.4) com $\lambda = \lambda_k$, $k = \overline{1,n}$, encontramos n vetores próprios diferentes $f_1, \ldots, f_n$. Pelo Teorema 7.1 estes vetores são linearmente independentes e portanto formam uma base em $\mathbb{X}$. Como $Af_k = \lambda_k f_k$, $k = \overline{1,n}$, a matriz da aplicação A nesta base tem a forma diagonal:

$$A_{\mathcal{F}} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix}.$$

□

Exemplo 7.2. Polinómio característico sem raízes reais. Consideremos a aplicação linear $A \in \mathcal{L}(\mathbb{R}^2, \mathbb{R}^2)$ de rotação de amplitude $\frac{\pi}{2}$ no sentido anti-horário. A sua matriz na base $\{(1,0), (0,1)\}$ é

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

A equação característica

$$\det \begin{pmatrix} -\lambda & -1 \\ 1 & -\lambda \end{pmatrix} = \lambda^2 + 1 = 0$$

não tem raízes no conjunto dos números reais. Portanto, esta aplicação não tem vetores próprios reais. Notemos que, pelo Teorema Fundamental da Álgebra (veja o Capítulo 12), cada aplicação linear $A \in \mathcal{L}(\mathbb{C}^n, \mathbb{C}^n)$ tem sempre pelo menos um valor próprio e, portanto, pelo menos um vetor próprio. Neste caso a aplicação tem dois valores próprios complexos: $\pm\mathtt{i}$. A cada um deles corresponde um vetor próprio complexo. Considerando o valor próprio $\lambda = \mathtt{i}$, o sistema de equações lineares

$$\begin{pmatrix} -\mathtt{i} & -1 \\ 1 & -\mathtt{i} \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

tem soluções $\begin{pmatrix} \mathtt{i}x^2 \\ x^2 \end{pmatrix}$ e, assim, por exemplo, $\begin{pmatrix} \mathtt{i} \\ 1 \end{pmatrix}$ é um vetor próprio correspondente ao valor próprio $\lambda = \mathtt{i}$. Para $\lambda = -\mathtt{i}$ obtemos

$$\begin{pmatrix} \mathtt{i} & -1 \\ 1 & \mathtt{i} \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

com soluções $\begin{pmatrix} -\mathtt{i}x^2 \\ x^2 \end{pmatrix}$ e, por exemplo, $\begin{pmatrix} -\mathtt{i} \\ 1 \end{pmatrix}$ é um vetor próprio correspondente ao valor próprio $\lambda = -\mathtt{i}$.

□

Exemplo 7.3. Polinómio característico com raízes múltiplas. Consideremos a aplicação linear $A \in \mathcal{L}(\mathbb{R}^2, \mathbb{R}^2)$ cuja matriz tem a forma

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

A equação característica

$$\det \begin{pmatrix} 1-\lambda & 1 \\ 0 & 1-\lambda \end{pmatrix} = (1-\lambda)^2 = 0$$

tem só uma raiz $\lambda = 1$. O sistema de equações lineares

$$\begin{pmatrix} 1-1 & 1 \\ 0 & 1-1 \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

possui apenas uma solução não trivial $\begin{pmatrix} x^1 \\ 0 \end{pmatrix}$. Portanto, existe só um vetor próprio, por exemplo, $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$, pelo que não existe uma base em $\mathbb{R}^2$ formada pelos vetores próprios desta aplicação.

Agora consideremos a aplicação identidade

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Neste caso, a equação característica é

$$\det \begin{pmatrix} 1-\lambda & 0 \\ 0 & 1-\lambda \end{pmatrix} = (1-\lambda)^2 = 0$$

e como no exemplo anterior tem só uma raiz $\lambda = 1$. Mas o sistema de equações lineares

$$\begin{pmatrix} 1-1 & 0 \\ 0 & 1-1 \end{pmatrix}\begin{pmatrix} x^1 \\ x^2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

tem duas soluções não triviais linearmente independentes. Portanto, neste caso, em $\mathbb{R}^2$ existe uma base formada pelos vetores próprios, por exemplo, $\{\binom{1}{0}, \binom{0}{1}\}$. □

Exemplo 7.4. Vamos calcular os valores e vetores próprios da matriz $A = \begin{pmatrix} 2 & -1 & -1 \\ 0 & -1 & 0 \\ 0 & 2 & 1 \end{pmatrix}$. A equação característica da matriz A é

$$0 = \det(A - \lambda I) = \begin{pmatrix} 2-\lambda & -1 & -1 \\ 0 & -1-\lambda & 0 \\ 0 & 2 & 1-\lambda \end{pmatrix} = -(2-\lambda)(1+\lambda)(1-\lambda).$$

Logo, os valores próprios são $\lambda_1 = 2$, $\lambda_2 = -1$, $\lambda_3 = 1$. Vamos determinar o vetor próprio correspondente ao valor próprio $\lambda_1 = 2$. Temos de encontrar uma solução não trivial do sistema

$$\begin{pmatrix} 2 & -1 & -1 \\ 0 & -1 & 0 \\ 0 & 2 & 1 \end{pmatrix}\begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix} = 2\begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix}.$$

A primeira equação implica que $-x^2 - x^3 = 0$. Da segunda equação temos $x^2 = 0$. Portanto, $x^3 = 0$. A terceira equação, naturalmente, não tem nova informação, pois, as equações são linearmente dependentes. Assim, vemos que o vetor próprio correspondente ao valor próprio $\lambda_1 = 2$ é $v_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$. Da mesma maneira calculamos os restantes vetores próprios: $\lambda_2 = -1$, $v_2 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$ e $\lambda_3 = 1$, $v_3 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$. □

Exemplo 7.5. Determinemos os valores e vetores próprios da matriz $A = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 1 & -1 \\ 0 & 2 & 4 \end{pmatrix}$. A equação característica da matriz A é

$$0 = \det(A - \lambda I) = \begin{pmatrix} 2-\lambda & 1 & 0 \\ 0 & 1-\lambda & -1 \\ 0 & 2 & 4-\lambda \end{pmatrix} = -(2-\lambda)^2(3-\lambda).$$

Logo, os valores próprios são $\lambda_1 = 3$, $\lambda_2 = 2$. Raciocinando como no exemplo anterior obtemos o vetor próprio correspondente a $\lambda_1 = 3$, $v_1 = \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}$, e o vetor próprio correspondente a $\lambda_2 = 2$, $v_2 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$. Neste caso, não existe uma base de $\mathbb{R}^3$ constituída por estes vetores próprios. □

7.3. Polinómio característico de uma aplicação linear

Seja $A \in \mathcal{L}(\mathbb{X}, \mathbb{X})$ uma aplicação linear de um espaço de dimensão n em si mesmo. Consideremos duas bases $\mathcal{E} = \{e_1, e_2, \ldots, e_n\}$ e $\mathcal{F} = \{f_1, f_2, \ldots, f_n\}$ em $\mathbb{X}$. Designe-

mos por P a matriz de transição da base $\mathcal{E}$ para a base $\mathcal{F}$. Como sempre, representamos por $A_{\mathcal{E}}$ e $A_{\mathcal{F}}$ as matrizes da aplicação A nas bases $\mathcal{E}$ e $\mathcal{F}$, respetivamente. A relação entre $A_{\mathcal{E}}$, $A_{\mathcal{F}}$ e P é dada pela fórmula (veja (5.12))

$$PA_{\mathcal{F}} = A_{\mathcal{E}}P.$$

Como o determinante do produto de matrizes é igual ao produto dos determinantes, obtemos

$$\det(P)\det(A_{\mathcal{F}}) = \det(A_{\mathcal{E}})\det(P),$$

pelo que

$$\det(A_{\mathcal{F}}) = \det(A_{\mathcal{E}}),$$

pois $\det(P) \neq 0$. Logo, o determinante da matriz de uma aplicação linear não depende da base.

Consideremos agora a aplicação linear $A - \lambda I$, onde $\lambda \in \mathbb{K}$. Sejam $A_{\mathcal{E}} - \lambda I$ a matriz desta aplicação na base $\mathcal{E}$ e $A_{\mathcal{F}} - \lambda I$ a matriz desta aplicação na base $\mathcal{F}$. Então temos

$$\det(A_{\mathcal{F}} - \lambda I) = \det(A_{\mathcal{E}} - \lambda I),$$

qualquer que seja $\lambda \in \mathbb{K}$. Em ambas as partes desta igualdade estão polinómios de grau n. Como eles são iguais para todos os $\lambda \in \mathbb{K}$, os coeficientes destes polinómios são iguais. Portanto, o polinómio característico $\det(A_{\mathcal{E}} - \lambda I)$ não depende da base. Temos, então, a seguinte definição.

Definição 7.3. Polinómio característico de uma aplicação linear

Seja $\lambda \in \mathbb{K}$ e a $A_{\mathcal{E}} - \lambda I$ a matriz da aplicação linear $A - \lambda I$ na base $\mathcal{E}$. Ao polinómio definido por $\det(A_{\mathcal{E}} - \lambda I)$ chamamos *polinómio característico* da aplicação linear A e à equação

$$\det(A_{\mathcal{E}} - \lambda I) = 0$$

chamamos *equação característica* da aplicação linear A.

7.4. Vetores e valores próprios de aplicações simétricas

Num espaço linear real os valores próprios podem ser reais ou complexos. Este facto torna natural a seguinte construção, chamada *complexificação*, que ajuda a perceber quando uma aplicação linear no espaço real tem ou não tem valores próprios complexos. Seja $\mathbb{X}$ um espaço Euclidiano real. Construamos um novo espaço linear complexo $\tilde{\mathbb{X}}$ com base nos elementos de $\mathbb{X}$ da seguinte maneira: os elementos de $\tilde{\mathbb{X}}$ são todas as somas formais do tipo $x_1 + \mathtt{i}x_2$, onde $x_1, x_2 \in \mathbb{X}$ e $\mathtt{i}$ é o número complexo que verifica $\mathtt{i}^2 = -1$. Definamos as operações algébricas como

$$(x_1 + \mathtt{i}x_2) + (y_1 + \mathtt{i}y_2) = (x_1 + y_1) + \mathtt{i}(x_2 + y_2)$$

e

$$\alpha(x_1 + \mathtt{i}x_2) = \alpha x_1 + \mathtt{i}(\alpha x_2)$$

e o produto escalar como

$$\langle x_1 + \mathtt{i}x_2, y_1 + \mathtt{i}y_2\rangle = \langle x_1, y_1\rangle + \langle x_2, y_2\rangle + \mathtt{i}(\langle x_2, y_1\rangle - \langle x_1, y_2\rangle).$$

Recordemos que nos espaços Euclidianos complexos o produto escalar não é comutativo mas quando trocamos a ordem dos multiplicadores temos

$$\langle x_1 + \mathtt{i}x_2, y_1 + \mathtt{i}y_2\rangle = \overline{\langle y_1 + \mathtt{i}y_2, x_1 + \mathtt{i}x_2\rangle}.$$

Com estas definições, é fácil verificar todas as propriedades de espaço Euclidiano complexo. Sejam $A \in \mathcal{L}(\mathbb{X}, \mathbb{X})$ uma aplicação linear e $x_1, x_2 \in \mathbb{X}$. Definamos a aplicação $A : \tilde{\mathbb{X}} \to \tilde{\mathbb{X}}$ por

$$A(x_1 + \mathtt{i}x_2) = Ax_1 + \mathtt{i}Ax_2.$$

Pode acontecer que a aplicação $A : \tilde{\mathbb{X}} \to \tilde{\mathbb{X}}$ tenha vetores próprios que correspondem a valores próprios complexos:

$$A(x_1 + \mathtt{i}x_2) = (\lambda_1 + \mathtt{i}\lambda_2)(x_1 + \mathtt{i}x_2).$$

Esta igualdade é equivalente às igualdades

$$Ax_1 = \lambda_1 x_1 - \lambda_2 x_2, \quad Ax_2 = \lambda_2 x_1 + \lambda_1 x_2. \tag{7.6}$$

Com a ajuda da operação da complexificação vamos agora estabelecer uma propriedade muito importante das aplicações simétricas.

Teorema 7.2

Seja $A \in \mathcal{L}(\mathbb{X}, \mathbb{X})$ uma aplicação simétrica. Então, todos os valores próprios de A são reais, existem vetores próprios reais que correspondem a estes valores próprios, e os vetores próprios que correspondem aos valores próprios diferentes são ortogonais.

Demonstração. Subtraindo à segunda igualdade de (7.6) multiplicada por x_1 a primeira igualdade multiplicada por x_2, temos

$$\langle Ax_2, x_1\rangle - \langle Ax_1, x_2\rangle = \lambda_2(|x_1|^2 + |x_2|^2).$$

Como a aplicação A é simétrica, a parte esquerda desta igualdade é igual a zero. Como o vetor próprio é diferente de zero, temos $|x_1|^2 + |x_2|^2 \neq 0$. Logo, $\lambda_2 = 0$, isto é, o valor próprio é real, e as igualdades (7.6) tomam a forma

$$Ax_1 = \lambda_1 x_1 \quad \text{e} \quad Ax_2 = \lambda_1 x_2,$$

isto é, pode ser escolhido um vetor próprio real que corresponde ao valor próprio λ_1.

Sejam λ_1 e λ_2 valores próprios diferentes de A e x_1 e x_2 os vetores próprios correspondentes. Subtraindo à igualdade $Ax_1 = \lambda_1 x_1$ multiplicada por x_2 a igualdade $Ax_2 = \lambda_2 x_2$ multiplicada por x_1, temos

$$\langle Ax_1, x_2\rangle - \langle Ax_2, x_1\rangle = (\lambda_1 - \lambda_2)\langle x_1, x_2\rangle.$$

Como a aplicação A é simétrica, a parte esquerda desta igualdade é igual a zero. Como $\lambda_1 \neq \lambda_2$ temos $\langle x_1, x_2\rangle = 0$.

■

7.5. Bases de vetores próprios

O seguinte teorema dá-nos a propriedade mais importante das aplicações simétricas definidas num espaço Euclidiano de dimensão finita.

Teorema 7.3

Sejam $\mathbb{X}$ um espaço Euclidiano de dimensão finita e $A \in \mathcal{L}(\mathbb{X}, \mathbb{X})$ uma aplicação linear simétrica. Então, existe em $\mathbb{X}$ uma base ortonormal de vetores próprios.

No caso em que todas as raízes da equação característica da aplicação A são distintas, a demonstração deste teorema é completamente elementar. De facto, pelo Teorema 7.2, todos os n vetores próprios que correspondem aos n valores próprios diferentes são ortogonais. Depois da normalização obtemos uma base ortonormal de vetores próprios. A demonstração no caso geral é mais complicada.

Demonstração. Pelo Teorema Fundamental da Álgebra (Teorema 12.1), existem um vetor e respetivo valor próprio λ. Estes vetor e valor próprios são reais (veja o Teorema 7.2).

Consideremos o subespaço $X \subset \mathbb{X}$ gerado por todos os vetores próprios da aplicação A que correspondem ao valor próprio λ. Este subespaço não é trivial. Escolhemos uma base ortonormal em X. Se $X = \mathbb{X}$, obtemos uma base ortonormal em $\mathbb{X}$ formada por vetores próprios da aplicação A. Se $X \neq \mathbb{X}$, podemos completar a base $e_1, \ldots, e_m$ em X até obter uma base ortonormal em $\mathbb{X}$, $e_1, \ldots, e_m, e_{m+1}, \ldots, e_n$ (veja o Teorema 3.5 e o processo de ortogonalização de Gram-Schmidt). O subespaço gerado pelos vetores $e_{m+1}, \ldots, e_n$ é o complemento ortogonal $X^\perp$. Seja $y \in X^\perp$. Então, para todo $x \in X$ temos $Ax \in X$ e portanto

$$\langle x, Ay\rangle = \langle Ax, y\rangle = 0,$$

isto é, $AX^\perp \subset X^\perp$. Consideremos a restrição da aplicação A ao subespaço $X^\perp$. Esta restrição também tem um vetor próprio e podemos repetir os raciocínios feitos. Assim, depois de um número finito de passos, obtemos uma base ortonormal em $\mathbb{X}$ formada por vetores próprios da aplicação A.

■

7.6. Norma de aplicações lineares nos espaços Euclidianos*

Consideremos uma aplicação linear $A \in \mathcal{L}(\mathbb{X}, \mathbb{Y})$, onde $\mathbb{X}$ e $\mathbb{Y}$ são espaços Euclidianos. Vamos designar os produtos escalares nos espaços $\mathbb{X}$ e $\mathbb{Y}$ por $\langle \cdot, \cdot \rangle_{\mathbb{X}}$ e $\langle \cdot, \cdot \rangle_{\mathbb{Y}}$, respetivamente, e por $|\cdot|_{\mathbb{X}}$ e $|\cdot|_{\mathbb{Y}}$ as respetivas normas Euclidianas.

Teorema 7.4

A norma da aplicação A é a raiz quadrada do valor próprio máximo da aplicação A^*A.

Demonstração. Pela definição da norma de uma aplicação linear, temos

$$|A| = \max_{|x|_{\mathbb{X}}=1} |Ax|_{\mathbb{Y}}.$$

Portanto, basta resolver o problema de maximização

$$\max \langle Ax, Ax \rangle_{\mathbb{Y}}, \text{ sujeito à restrição } \langle x, x \rangle_{\mathbb{X}} = 1.$$

Pelo Teorema de Weierstrass, a função contínua $\langle Ax, Ax \rangle_{\mathbb{Y}}$ atinge o seu máximo no conjunto fechado e limitado $\langle x, x \rangle_{\mathbb{X}} = 1$ num ponto $\hat{x}$. Pela Regra dos multiplicadores de Lagrange, existe um número μ tal que

$$\nabla_x (\langle A\hat{x}, A\hat{x} \rangle_{\mathbb{Y}} + \mu(\langle \hat{x}, \hat{x} \rangle_{\mathbb{X}} - 1)) = 0,$$

que é equivalente à igualdade

$$A^*A\hat{x} + \mu\hat{x} = 0.$$

Logo, $\hat{x}$ é um vetor próprio da aplicação A^*A que corresponde ao valor próprio $-\mu$. Como

$$\langle A\hat{x}, A\hat{x} \rangle_{\mathbb{Y}} = \mu^2 \langle \hat{x}, \hat{x} \rangle_{\mathbb{X}} = \mu^2,$$

temos o pretendido. ■

7.7. Equações diferenciais lineares*

Nesta secção, numa maneira informal, vamos considerar dois problemas — um sobre dinâmica de populações e outro sobre um sistema mecânico simples. O primeiro exemplo serve para explicar ao leitor o que é uma equação diferencial ordinária e como estas equações aparecem. O segundo exemplo mostra como os vetores e valores próprios são necessários para a sua análise. Estes dois exemplos pertencem a uma larga classe de problemas que nós vamos analisar nesta secção.

Exemplo 7.6. Neste exemplo modelamos a evolução do número de indivíduos numa certa população. Seja $x(t)$ o tamanho de uma população de, por exemplo, bactérias,

no instante t. É claro que na vida real $x(t)$ tem que ser um número natural. Todavia se o número de bactérias, N, é muito grande, digamos aproximadamente um milhão, a diferença relativa entre N e $N+1$ é tão pequena $\left(\frac{N}{N+1} \approx 1\right)$ que podemos introduzir, por exemplo, a variável $x(t) = N(t) \times 10^{-6}$, que pode perfeitamente ser considerada como uma variável contínua. Suponhamos que a variação da população num dado intervalo de tempo, Δt, pode ser representada pela igualdade

$$\Delta x = ax\Delta t - bx\Delta t,$$

onde a e b são as taxas de natalidade e mortalidade, respetivamente (número de nascimentos e mortes por indivíduo e por unidade de tempo). A passagem ao limite (quando $\Delta t \to 0$) na igualdade

$$\frac{\Delta x}{\Delta t} = (a-b)x$$

dá a equação diferencial ordinária

$$\frac{\mathrm{d}}{\mathrm{d}t}x(t) = (a-b)x(t).$$

A solução desta equação é dada por

$$x(t) = x(0)\exp((a-b)t).$$

Se $a < b$ a população desaparece com o decorrer do tempo e se $a > b$ temos um crescimento exponencial.

□

Neste exemplo, a determinação da solução da equação não precisa de grandes conhecimentos. No caso de equações mais complicadas precisamos da teoria que foi desenvolvida neste capítulo.

Exemplo 7.7. Este exemplo vem da Física. Na Física é habitual denotar as derivadas em ordem ao tempo por um ponto, ou seja,

$$\frac{\mathrm{d}}{\mathrm{d}t}x(t) = \dot{x}(t), \quad \frac{\mathrm{d}^2}{\mathrm{d}t^2}x(t) = \ddot{x}(t).$$

Assim, o movimento de um oscilador (sistema mecânico composto por um corpo ligado a uma mola fixa (veja a Fig. 7.1) é descrito pela segunda lei de Newton:

$$\ddot{x}(t) = -\beta^2 x(t).$$

Aqui $x(t)$ representa a posição do corpo de massa $m = 1$ no instante t onde $x = 0$ corresponde à posição de equilíbrio. O termo $-\beta^2 x$ representa a força elástica da mola onde o sinal menos é para sublinhar que a força é negativa, isto é, a força faz o corpo voltar para a posição de equilíbrio.

Consideremos agora o movimento de um oscilador num meio denso (veja a Fig. 7.2). Neste caso o oscilador também está sujeito a uma força de resistência que depende da velocidade do movimento. Se a velocidade é bastante pequena, a força é uma função linear da velocidade: $-2\alpha\dot{x}$, α um número positivo. A segunda lei de Newton toma a forma

$$\ddot{x}(t) = -2\alpha\dot{x}(t) - \beta^2 x(t). \tag{7.7}$$

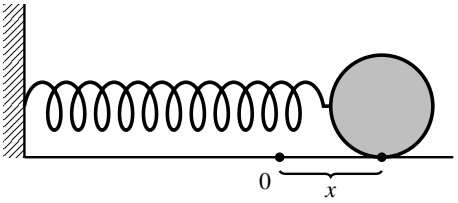


Figura 7.1
Oscilador.

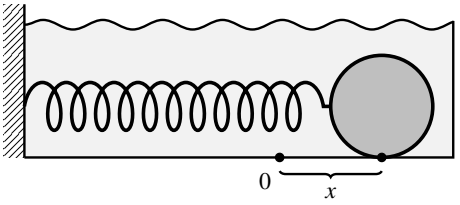


Figura 7.2
Oscilador em meio denso.

□

Notemos que a mesma equação diferencial também descreve a dinâmica da carga elétrica num condensador depois de ligar o interruptor I (veja a Fig. 7.3). O papel do corpo é feito pela bobina de indutância L, a resistência R faz o papel do coeficiente 2α e $\frac{1}{C}$ faz o papel do coeficiente β^2.

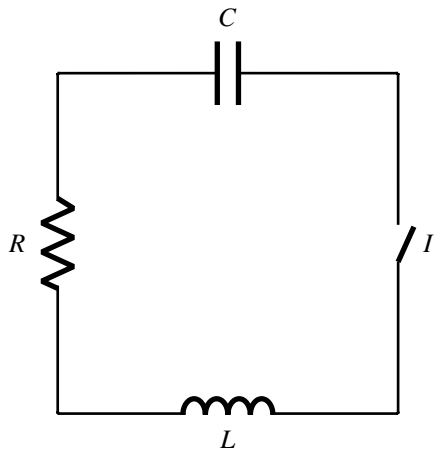


Figura 7.3
Circuito elétrico.

Portanto, resolvendo uma equação diferencial, na realidade resolvemos muitos problemas físicos.

Regressemos à equação (7.7). O seu estudo é mais fácil de efetuar considerando-a escrita na forma de um sistema de equações diferenciais. Para tal, e introduzindo a velocidade do oscilador $v = \dot{x}$, podemos reescrever a equação diferencial de segunda ordem (por conter uma derivada de segunda ordem) como o sistema de equações diferenciais de primeira ordem (por conter apenas derivadas de primeira ordem)

$$\begin{cases} \dot{x} = v, \\ \dot{v} = -2\alpha v - \beta^2 x, \end{cases}$$

ou ainda, equivalentemente, na forma matricial

$$\begin{pmatrix} \dot{x} \\ \dot{v} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\beta^2 & -2\alpha \end{pmatrix} \begin{pmatrix} x \\ v \end{pmatrix}.$$

Este sistema é só um exemplo de um sistema de equações diferenciais de primeira ordem. O caso geral é a equação diferencial

$$\dot{x} = Ax, \tag{7.8}$$

onde A é uma matriz de dimensões $n \times n$. Dizemos que a função $x = x(t)$ é a solução da equação diferencial se ela verifica a equação (7.8) em todos os pontos t. Suponhamos que todos os valores próprios $\lambda_1, \ldots, \lambda_n$ da matriz A são reais e diferentes. Então, os vetores próprios $\{v_1, \ldots, v_n\}$ de A formam uma base no espaço $\mathbb{R}^n$. Obviamente para qualquer escolha das constantes $c_1, \ldots, c_n$, a função

$$x(t) = \sum_{k=1}^{n} c_k \exp(\lambda_k t) v_k \tag{7.9}$$

é uma solução da equação (7.8). De facto, por um lado

$$\dot{x}(t) = \frac{\mathrm{d}}{\mathrm{d}t} \sum_{k=1}^{n} c_k \exp(\lambda_k t) v_k = \sum_{k=1}^{n} c_k \lambda_k \exp(\lambda_k t) v_k$$

e, por outro lado, temos

$$Ax(t) = A\left(\sum_{k=1}^{n} c_k \exp(\lambda_k t) v_k\right) = \sum_{k=1}^{n} c_k \exp(\lambda_k t) A v_k = \sum_{k=1}^{n} c_k \lambda_k \exp(\lambda_k t) v_k.$$

Se todos os valores próprios $\lambda_1, \ldots, \lambda_n$ da matriz A são agora complexos e diferentes, a função (7.9) também é uma solução da equação (7.8). Só é necessário definir corretamente a função exponencial de uma variável complexa. Para o fazer é necessário escolher uma definição onde possa ser substituída uma variável real por uma variável complexa.

Como sabemos

$$\exp(x) = \lim_{N \to \infty} \left(1 + \frac{x}{N}\right)^N.$$

Seja $z_N = x_N + \mathtt{i} y_N$, $N = 1, 2, \ldots$, uma sucessão de números complexos. Vamos dizer que o número complexo $z = x + \mathtt{i} y$ é o limite da sucessão z_N e escrever

$$\lim_{N \to \infty} z_N = z,$$

se

$$\lim_{N \to \infty} x_N = x \quad \text{e} \quad \lim_{N \to \infty} y_N = y,$$

ou seja, se

$$\lim_{N \to \infty} |z_N| = |z| \quad \text{e} \quad \lim_{N \to \infty} \arg z_N = \arg z.$$

Encontremos o limite da sucessão

$$z_N = \left(1 + \frac{z}{N}\right)^N$$

através do módulo e do argumento do termo z_N. Como

$$|z_N| = \left|\left(1 + \frac{z}{N}\right)^N\right| = \left(\left(1 + \frac{x}{N}\right)^2 + \frac{y^2}{N^2}\right)^{\frac{N}{2}} = \left(1 + \frac{2x}{N} + \frac{x^2 + y^2}{N^2}\right)^{\frac{N}{2}}$$

e

$$\arg z_N = N \arg\left(1 + \frac{z}{N}\right) = N \operatorname{arctg} \frac{\frac{y}{N}}{1 + \frac{x}{N}},$$

temos que $|z_N| \to \exp(x)$ e $\arg z_N \to y$, pelo que

$$\lim_{N\to\infty} z_N = \lim_{N\to\infty} \left[|z_N| \left(\cos(\arg z_N) + \mathtt{i} \operatorname{sen}(\arg z_N)\right)\right] = \exp(x)(\cos y + \mathtt{i} \operatorname{sen} y).$$

Portanto, é natural definir a *função exponencial complexa*

$$\exp(z) = \exp(x)(\cos y + \mathtt{i} \operatorname{sen} y).$$

O que acabamos de fazer não é nehuma demonstração. É um raciocínio baseado em bom senso que nos ajudou a chegar a uma definição razoável da função exponencial de um argumento complexo. Agora é necessário verificar as propriedades principais da função exponencial, nomeadamente, a igualdade

$$\exp(z_1 + z_2) = \exp(z_1) \exp(z_2) \tag{7.10}$$

e a sua consequência

$$\frac{\mathrm{d}}{\mathrm{d}z} \exp(z) = \exp(z).$$

Sejam $z_1 = x_1 + \mathtt{i}y_1$ e $z_2 = x_2 + \mathtt{i}y_2$ dois números complexos. Pela definição da função exponencial temos

$$\begin{aligned}
\exp(z_1 + z_2) &= \exp(x_1 + x_2)(\cos(y_1 + y_2) + \mathtt{i} \operatorname{sen}(y_1 + y_2)) \\
&= \exp(x_1) \exp(x_2)((\cos y_1 \cos y_2 - \operatorname{sen} y_1 \operatorname{sen} y_2) \\
&\quad + \mathtt{i}(\operatorname{sen} y_1 \cos y_2 + \cos y_1 \operatorname{sen} y_2)) \\
&= \exp(x_1)(\cos y_1 + \mathtt{i} \operatorname{sen} y_1) \exp(x_2)(\cos y_2 + \mathtt{i} \operatorname{sen} y_2) \\
&= \exp(z_1) \exp(z_2),
\end{aligned}$$

o que demonstra (7.10).

Sejam $z = x + \mathtt{i}y$ e $\Delta z = \Delta x + \mathtt{i}\Delta y$. Utilizando a fórmula de Taylor, de (7.10) obtemos

$$\begin{aligned}
\frac{\exp(z + \Delta z) - \exp(z)}{\Delta z} &= \exp(z) \frac{\exp(\Delta z) - 1}{\Delta z} \\
&= \exp(z) \frac{\exp(\Delta x)(\cos \Delta y + \mathtt{i} \operatorname{sen} \Delta y) - 1}{\Delta x + \mathtt{i}\Delta y} \\
&= \exp(z) \frac{(1 + \Delta x + \alpha(\Delta x)\Delta x)(1 + \beta(\Delta y)\Delta y + \mathtt{i}(\Delta y + \gamma(\Delta y)\Delta y) - 1}{\Delta x + \mathtt{i}\Delta y},
\end{aligned}$$

onde as funções $\alpha(\Delta x)$, $\beta(\Delta y)$ e $\gamma(\Delta y)$ tendem para zero quando Δx e Δy tendem para zero. Portanto, temos

$$\begin{aligned}\frac{\exp(z+\Delta z)-\exp(z)}{\Delta z} &= \exp(z)\frac{\exp(\Delta z)-1}{\Delta z}\\ &= \exp(z)\frac{(1+\Delta x+\alpha(\Delta x)\Delta x)(1+\beta(\Delta y)\Delta y+\mathtt{i}(\Delta y+\gamma(\Delta y)\Delta y)-1}{\Delta x+\mathtt{i}\Delta y}\\ &= \exp(z)\frac{\Delta x+\mathtt{i}\Delta y+\xi(\Delta x,\Delta y)\Delta x+\eta(\Delta x,\Delta y)\Delta y}{\Delta x+\mathtt{i}\Delta y},\end{aligned}$$

onde as funções (complexas) $\xi(\Delta x,\Delta y)$ e $\eta(\Delta x,\Delta y)$ tendem para zero quando Δx e Δy tendem para zero. Como

$$\left|\frac{\Delta x}{\Delta x+\mathtt{i}\Delta y}\right|\leqslant 1 \quad \text{e} \quad \left|\frac{\Delta y}{\Delta x+\mathtt{i}\Delta y}\right|\leqslant 1,$$

vemos que

$$\frac{\mathrm{d}}{\mathrm{d}z}\exp(z)=\lim_{\Delta z\to 0}\frac{\exp(z+\Delta z)-\exp(z)}{\Delta z}=\exp(z).$$

As fórmulas do cálculo diferencial para as funções complexas são idênticas às fórmulas para as funções reais. Por exemplo, da regra da cadeia, temos

$$\frac{\mathrm{d}}{\mathrm{d}t}\exp(\lambda t)=\lambda\exp(\lambda t),$$

onde λ é um número complexo.

Vamos voltar à equação do oscilador e encontrar as soluções do sistema que descreve os seus movimentos. O polinómio característico da matriz

$$A=\begin{pmatrix}0 & 1\\ -\beta^2 & -2\alpha\end{pmatrix}.$$

é

$$\det(A-\lambda I)=\det\begin{pmatrix}-\lambda & 1\\ -\beta^2 & -2\alpha-\lambda\end{pmatrix}=\lambda^2+2\alpha\lambda+\beta^2$$

e tem raízes $\lambda_\pm=-\alpha\pm\sqrt{\alpha^2-\beta^2}$. Os respetivos vetores próprios são

$$v_+=\begin{pmatrix}1\\ -\alpha+\sqrt{\alpha^2-\beta^2}\end{pmatrix} \quad \text{e} \quad v_-=\begin{pmatrix}1\\ -\alpha-\sqrt{\alpha^2-\beta^2}\end{pmatrix}.$$

Portanto, quando $\alpha^2-\beta^2>0$ a solução é

$$\begin{aligned}\begin{pmatrix}x(t)\\ v(t)\end{pmatrix} &= c_+\exp\left(\left(-\alpha+\sqrt{\alpha^2-\beta^2}\right)t\right)\begin{pmatrix}1\\ -\alpha+\sqrt{\alpha^2-\beta^2}\end{pmatrix}\\ &\quad + c_-\exp\left(\left(-\alpha-\sqrt{\alpha^2-\beta^2}\right)t\right)\begin{pmatrix}1\\ -\alpha-\sqrt{\alpha^2-\beta^2}\end{pmatrix},\end{aligned}$$

onde $c_\pm$ são constantes reais. Dadas as condições iniciais $x(0) = x_0$ e $v(0) = v_0$, as constantes $c_\pm$ são determinadas através da resolução do sistema linear

$$\begin{pmatrix} x_0 \\ v_0 \end{pmatrix} = c_+ \begin{pmatrix} 1 \\ -\alpha + \sqrt{\alpha^2 - \beta^2} \end{pmatrix} + c_- \begin{pmatrix} 1 \\ -\alpha - \sqrt{\alpha^2 - \beta^2} \end{pmatrix},$$

uma vez que é um sistema possível e determinado pois os vetores próprios $v_\pm$ formam uma base em $\mathbb{R}^2$.

No caso $\alpha^2 - \beta^2 < 0$ a solução tem a mesma forma mas com constantes complexas $c_\pm$. Como apenas uma solução real tem sentido físico e como os vetores $\exp(\lambda_+ t)v_+$ e $\exp(\lambda_- t)v_-$ são complexos conjugados, é necessário escolher as constantes $c_\pm$ complexas conjugadas. Sejam, então, $c_\pm = \frac{1}{2}(a \mp \mathtt{i}b)$. Com esta escolha temos

$$\begin{aligned}
\begin{pmatrix} x(t) \\ v(t) \end{pmatrix} &= \frac{1}{2}(a - \mathtt{i}b) \begin{pmatrix} 1 \\ -\alpha + \mathtt{i}\sqrt{\beta^2 - \alpha^2} \end{pmatrix} \exp\left(\left(-\alpha + \mathtt{i}\sqrt{\beta^2 - \alpha^2}\right) t\right) \\
&\quad + \frac{1}{2}(a + \mathtt{i}b) \begin{pmatrix} 1 \\ -\alpha - \mathtt{i}\sqrt{\beta^2 - \alpha^2} \end{pmatrix} \exp\left(\left(-\alpha - \mathtt{i}\sqrt{\beta^2 - \alpha^2}\right) t\right) \\
&= \frac{1}{2}(a - \mathtt{i}b) \begin{pmatrix} 1 \\ -\alpha + \mathtt{i}\sqrt{\beta^2 - \alpha^2} \end{pmatrix} \exp(-\alpha t) \left(\cos \sqrt{\beta^2 - \alpha^2} t + \mathtt{i} \operatorname{sen} \sqrt{\beta^2 - \alpha^2} t\right) \\
&\quad + \frac{1}{2}(a + \mathtt{i}b) \begin{pmatrix} 1 \\ -\alpha - \mathtt{i}\sqrt{\beta^2 - \alpha^2} \end{pmatrix} \exp(-\alpha t) \left(\cos \sqrt{\beta^2 - \alpha^2} t - \mathtt{i} \operatorname{sen} \sqrt{\beta^2 - \alpha^2} t\right) \\
&= \exp(-\alpha t) \begin{pmatrix} a \\ -\alpha a + b\sqrt{\beta^2 - \alpha^2} \end{pmatrix} \cos \sqrt{\beta^2 - \alpha^2} t \\
&\quad + \exp(-\alpha t) \begin{pmatrix} b \\ -\alpha b - a\sqrt{\beta^2 - \alpha^2} \end{pmatrix} \operatorname{sen} \sqrt{\beta^2 - \alpha^2} t. \qquad (7.11)
\end{aligned}$$

Para determinar as constantes a e b é necessário resolver o sistema linear

$$\begin{pmatrix} x_0 \\ v_0 \end{pmatrix} = \begin{pmatrix} a \\ -\alpha a + b\sqrt{\beta^2 - \alpha^2} \end{pmatrix}.$$

Quando os valores próprios não são todos distintos, a forma da solução da equação diferencial (7.8) é mais complicada. Consideramos este caso geral no último capítulo do livro.

7.8. Da Álgebra Linear para a Análise Linear*

A Álgebra Linear tem a ver com espaços de dimensão finita. O método de resolução de equações diferenciais que utiliza vetores e valores próprios considerado na secção anterior também pode ser adaptado para espaços de dimensão infinita. A resolução

de um problema da Álgebra Linear começa com a introdução de uma base. No caso de espaços de dimensão infinita só é necessário, em vez da igualdade na forma de somatório

$$x = \sum_{k=1}^{n} \xi^k e_k,$$

conseguirmos escrever uma igualdade na forma de série

$$x = \sum_{k=1}^{\infty} \xi^k e_k.$$

Portanto, já estamos a considerar uma série convergente e que faz parte de um outro ramo de Matemática que é conhecido como Análise Linear (ou Análise Funcional, visto que normalmente os elementos dos espaços lineares de dimensão infinita com que trabalhamos são funções). Se os vetores próprios de uma aplicação linear A formam uma base no espaço de dimensão infinita, então podemos representar a solução da equação diferencial

$$\dot{x} = Ax$$

na forma da série

$$x(t) = \sum_{k=1}^{\infty} c_k v_k \exp(\lambda_k t),$$

onde v_k são os vetores próprios da aplicação linear A correspondentes aos valores próprios λ_k e c_k são constantes.

Consideremos um exemplo deste tipo — um modelo do processo de difusão. Consideremos partículas que se movem ao longo de uma reta e podem ocupar os pontos hk, onde $h > 0$ é um número muito pequeno e $k = 0, \pm 1, \pm 2, \ldots$. O tempo é discreto $t = \tau n$, onde $\tau > 0$ é um número muito pequeno, e $n = 0, 1, 2, \ldots$. Uma partícula que no instante de tempo τn está no ponto kh, vai, com probabilidade $\frac{1}{2}$, no instante de tempo seguinte para o ponto $(k-1)h$ ou para o ponto $(k+1)h$ (veja a Fig. 7.4).

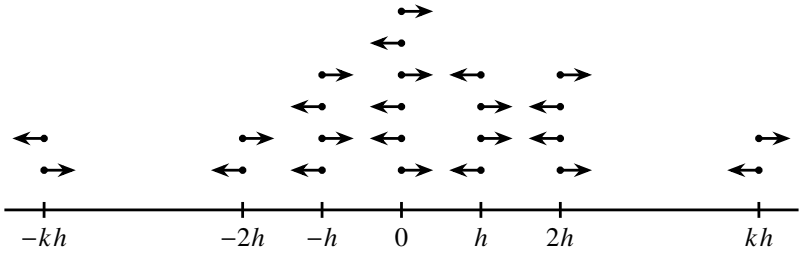


Figura 7.4
Posição e sentido do movimento das partículas.

Seja $N_{k,n}$ o número das partículas no ponto kh no instante de tempo $n\tau$. Então, no instante de tempo $(n+1)\tau$ no ponto kh está metade do número de partículas que saíram do ponto $(k-1)h$ e metade do número de partículas que saíram do ponto $(k+1)h$, logo

$$N_{k,n+1} = \frac{1}{2}(N_{k+1,n} + N_{k-1,n}).$$

Subtraindo $N_{k,n}$ desta igualdade, obtemos

$$N_{k,n+1} - N_{k,n} = \frac{1}{2}(N_{k+1,n} - 2N_{k,n} + N_{k-1,n}).$$

Seja $\sigma = \frac{h^2}{2\tau}$. Então temos

$$\frac{N_{k,n+1} - N_{k,n}}{\tau} = \sigma \frac{N_{k+1,n} - 2N_{k,n} + N_{k-1,n}}{h^2}.$$

Dividindo $N_{k,n}$ pelo número total de partículas, N, obtemos a concentração $u_{k,n}$ das partículas no ponto kh, no instante de tempo $n\tau$, ou seja, $u_{k,n} = \frac{N_{k,n}}{N}$. A concentração verifica a igualdade

$$\frac{u_{k,n+1} - u_{k,n}}{\tau} = \sigma \frac{u_{k+1,n} - 2u_{k,n} + u_{k-1,n}}{h^2}. \tag{7.12}$$

Recordemos que a derivada de uma função $x(t)$ pode ser aproximada por

$$\dot{x}(t) \approx \frac{x(t+\tau) - x(t)}{\tau} \quad \text{e} \quad \dot{x}(t) \approx \frac{x(t) - x(t-\tau)}{\tau}. \tag{7.13}$$

Então, temos a aproximação à segunda derivada

$$\begin{aligned} \ddot{x}(t) &\approx \frac{\dot{x}(t+\tau) - \dot{x}(t)}{\tau} \approx \frac{\frac{x(t+\tau)-x(t)}{\tau} - \frac{x(t)-x(t-\tau)}{\tau}}{\tau} \\ &= \frac{x(t+\tau) - 2x(t) + x(t-\tau)}{\tau^2}. \end{aligned} \tag{7.14}$$

Portanto, o primeiro membro da igualdade (7.12) é um valor aproximado da derivada parcial em ordem ao tempo da função $u(x,t)$, a concentração de partículas no ponto x e no instante de tempo t. O segundo membro é um valor aproximado da segunda derivada parcial em ordem à variável espacial x da função $u(x,t)$ (veja (7.13) e (7.14)). Fixemos a constante $\sigma = \frac{h^2}{2\tau}$. Passando ao limite quando τ e h tendem para zero e mantendo a razão $\frac{h^2}{2\tau}$ igual a σ, obtemos a equação

$$\dot{u}(x,t) = \sigma u''(x,t),$$

conhecida por *equação de difusão*, onde $\dot{u}$ representa novamente a primeira derivada de u em ordem ao tempo t e u'' representa a segunda derivada de u em ordem à variável espacial x. Ao coeficiente σ chamamos coeficiente de *difusão*.

O esquema que consideramos é muito geral. As "partículas" podem ser, por exemplo, moléculas de um gás que difunde num outro ou a quantidade de calor que se propaga num meio. Neste segundo caso, em vez de equação de difusão dizemos *equação do calor*, a função $u(x,t)$ corresponde à temperatura do meio no ponto x no instante de tempo t e ao coeficiente σ chamamos coeficiente de *termodifusividade*. (A temperatura pode ser considerada como uma medida da quantidade de calor.)

Consideremos um exemplo com $\sigma = 1$

$$\dot{u} = u'', \tag{7.15}$$

onde a função $u = u(x,t)$ é a temperatura entre dois planos infinitos localizados a uma distância π um do outro. Suponhamos que a temperatura dos planos é zero,

$u(0,t) = u(\pi,t) = 0$, e que no momento inicial a temperatura entre os planos é $u(x,0) = x(\pi - x)$. A equação (7.15) não é nada mais do que a equação (7.8), onde $Au = u''$. Para resolver esta equação precisamos de encontrar os vetores próprios da aplicação da segunda derivada $A = \frac{\mathrm{d}^2}{\mathrm{d}x^2}$, isto é, as funções $X = X(x) \neq 0$ que verificam a equação

$$X''(x) = \lambda X(x), \tag{7.16}$$

onde λ é um número, e as condições na fronteira $X(0) = X(\pi) = 0$.

Consideremos três casos.

1. Seja $\lambda = 0$. Neste caso temos a equação $X''(x) = 0$. A solução geral desta equação é $X(x) = ax + b$, onde a e b são constantes. Como $X(0) = X(\pi) = 0$, obviamente $a = b = 0$. Portanto, $X(x) \equiv 0$. Isto significa que $\lambda = 0$ não é um valor próprio da aplicação A.

2. Seja $\lambda > 0$. Neste caso a solução geral da equação (7.16) é

$$X(x) = a\exp(\sqrt{\lambda}x) + b\exp(-\sqrt{\lambda}x). \tag{7.17}$$

Para ver isto, escrevemos a equação (7.16) na forma de um sistema equivalente como foi feito no caso do oscilador

$$\begin{cases} X' = V, \\ V' = \lambda X. \end{cases}$$

A matriz deste sistema é

$$\begin{pmatrix} 0 & 1 \\ \lambda & 0 \end{pmatrix}.$$

A respetiva equação característica

$$\det\begin{pmatrix} -\mu & 1 \\ \lambda & -\mu \end{pmatrix} = \mu^2 - \lambda = 0$$

tem raízes reais $\mu = \pm\sqrt{\lambda}$ e os respetivos vetores próprios são

$$\begin{pmatrix} 1 \\ \pm\sqrt{\lambda} \end{pmatrix}.$$

Portanto, a solução geral do sistema é

$$\begin{pmatrix} X(x) \\ V(x) \end{pmatrix} = a\begin{pmatrix} 1 \\ \sqrt{\lambda} \end{pmatrix}\exp(\sqrt{\lambda}x) + b\begin{pmatrix} 1 \\ -\sqrt{\lambda} \end{pmatrix}\exp(-\sqrt{\lambda}x).$$

Daqui temos (7.17).

As condições $X(0) = X(\pi) = 0$ implicam que a e b são soluções do sistema linear

$$\begin{cases} a + b = 0, \\ \exp(\sqrt{\lambda}\pi)a + \exp(-\sqrt{\lambda}\pi)b = 0. \end{cases}$$

O determinante da matriz deste sistema é $\exp(-\sqrt{\lambda}\pi) - \exp(\sqrt{\lambda}\pi) \neq 0$. Logo, $a = b = 0$. Portanto, $X(x) \equiv 0$. Isto significa que $\lambda > 0$ não pode ser um valor próprio da aplicação $A = \frac{\mathrm{d}^2}{\mathrm{d}x^2}$.

3. Finalmente, consideremos o caso $\lambda < 0$. Neste caso a solução geral da equação (7.16) é

$$X(x) = a\cos(\sqrt{|\lambda|}x) + b\,\mathrm{sen}(\sqrt{|\lambda|}x). \tag{7.18}$$

De facto, com $\lambda < 0$ temos a equação do oscilador (7.7) com $\alpha = 0$ e $-\beta^2 = \lambda$, e (7.18) é a primeira linha da igualdade (7.11).

Da condição $X(0) = 0$ vemos que $a = 0$. A condição $X(\pi) = 0$ implica que $\mathrm{sen}(\sqrt{|\lambda|}\pi) = 0$. Portanto, $\lambda_k = -k^2$, $k = 1, 2, \ldots$, são os valores próprios da aplicação A considerada no espaço de funções $X = X(x)$, duas vezes continuamente diferenciáveis, que verificam as condições $X(0) = X(\pi) = 0$. Os respetivos vetores próprios (ou, podemos dizer, funções próprias) são $X_k(x) = \mathrm{sen}(kx)$.

Por analogia com a definição do produto escalar no espaço de polinómios trigonométricos (veja a Secção 4.6), é natural definir o produto escalar no espaço das funções definidas no intervalo $[0, \pi]$ como

$$\langle f, g\rangle = \int_0^\pi f(x)g(x)\,\mathrm{d}x.$$

Facilmente verificamos que

$$\int_0^\pi \mathrm{sen}(kx)\,\mathrm{sen}(mx)\,\mathrm{d}x = 0, \quad k \neq m, \quad \text{e} \tag{7.19}$$

$$\int_0^\pi \mathrm{sen}^2(kx)\,\mathrm{d}x = \int_0^\pi \frac{1 - \cos(2kx)}{2}\,\mathrm{d}x = \frac{\pi}{2}. \tag{7.20}$$

Portanto, as funções próprias $X_k(x) = \mathrm{sen}(kx)$ são ortogonais. Vamos procurar a solução da equação (7.15) na forma

$$u(x,t) = \sum_{k=1}^{\infty} c_k(t)\,\mathrm{sen}(kx), \tag{7.21}$$

onde $c_k(t)$ são funções do tempo t. Substituindo (formalmente) esta série na equação (7.15), obtemos

$$\sum_{k=1}^{\infty} \dot{c}_k(t)\,\mathrm{sen}(kx) = -\sum_{k=1}^{\infty} k^2 c_k(t)\,\mathrm{sen}(kx).$$

Calculando com a ajuda de (7.19) e (7.20) o produto escalar de ambos os membros desta igualdade com $\mathrm{sen}(kx)$, obtemos

$$\dot{c}_k(t) = -k^2 c_k(t),$$

donde concluímos que

$$c_k(t) = c_k(0)\exp(-k^2 t). \tag{7.22}$$

Agora é necessário encontrarmos os valores iniciais $c_k(0)$ a partir da condição inicial $u(x,0) = x(\pi - x)$. Acontece que as funções próprias da aplicação A, a saber $\operatorname{sen}(kx)$, formam uma base no espaço de funções, ou seja, é possível a representação

$$x(\pi - x) = \sum_{k=1}^{\infty} c_k(0) \operatorname{sen}(kx).$$

(Esta afirmação não é trivial e a sua demonstração está para além do escopo do nosso curso.) Para determinarmos os coeficientes $c_k(0)$, calculamos com a ajuda de (7.19) e (7.20) o produto escalar de ambos os membros desta igualdade com $\operatorname{sen}(kx)$, vindo

$$\int_0^{\pi} x(\pi - x) \operatorname{sen}(kx)\, \mathrm{d}x = c_k(0)\frac{\pi}{2}.$$

Integrando duas vezes por partes, concluímos que

$$\int_0^{\pi} x(\pi - x) \operatorname{sen}(kx)\, \mathrm{d}x = \frac{2}{k^3}(1 - \cos(k\pi)) = \begin{cases} \frac{4}{k^3}, & \text{se } k = 2m+1, \\ 0, & \text{se } k = 2m, \end{cases}$$

onde $m = 0, 1, 2, \ldots$. Portanto, os coeficientes não nulos são $c_k(0) = \frac{8}{\pi(2m+1)^3}$ e de (7.21) juntamente com (7.22) obtemos a solução

$$u(x,t) = \sum_{m=0}^{\infty} \frac{8}{\pi(2m+1)^3} \exp(-(2m+1)^2 t) \operatorname{sen}((2m+1)x).$$

A matéria desta secção foi tratada de uma maneira informal. No entanto, é possível demonstrar que a série obtida converge e que é verdadeiramente a solução do problema.

8. Espaços afins

8.1. Introdução

No primeiro capítulo, introduzimos informalmente vetor como um conjunto de segmentos orientados equivalentes. Vamos agora formalizar este conceito através da definição de espaços afins. A teoria destes espaços está muito ligada à teoria dos espaços lineares. Por exemplo, em vez de combinações lineares nos espaços lineares e aplicações lineares vão aparecer agora combinações afins e aplicações afins, que, como veremos, são conceitos úteis na resolução de vários problemas.

8.2. Definição

Seja $\mathbb{X}$ um espaço linear de dimensão n.

Definição 8.1. Espaço afim

Dizemos que o conjunto $\mathbb{V}$ é um *espaço afim* de dimensão n se a cada par ordenado de elementos (pontos) $P, Q \in \mathbb{V}$ corresponde um único vetor de $\mathbb{X}$ designado por $\overrightarrow{PQ}$ e

1. para cada $P \in \mathbb{V}$ e cada $x \in \mathbb{X}$ existe um único ponto $Q \in \mathbb{V}$ tal que $\overrightarrow{PQ} = x$;
2. para cada três pontos $P, Q, S \in \mathbb{V}$ verifica-se a igualdade $\overrightarrow{PQ} + \overrightarrow{QS} = \overrightarrow{PS}$.

O espaço $\mathbb{X}$ é o espaço de vetores de $\mathbb{V}$ e os elementos de $\mathbb{X}$ são vetores de $\mathbb{V}$.

Da definição do espaço afim imediatamente concluímos que

$$\overrightarrow{AA} + \overrightarrow{AB} = \overrightarrow{AB},$$

ou seja, que o vetor correspondente a um par de pontos coincidentes é o vetor zero. Então, da igualdade

$$\overrightarrow{AB} + \overrightarrow{BA} = \overrightarrow{AA}$$

vemos que $\overrightarrow{AB} = -\overrightarrow{BA}$.

A partir de um espaço linear $\mathbb{X}$ é possível construir um espaço afim. Com efeito, podemos considerar os pontos de $\mathbb{X}$ como elementos de $\mathbb{V}$ e a cada par de pontos $x, y \in \mathbb{X}$ façamos corresponder o vetor $\overrightarrow{xy} = y - x$. A dimensão de um espaço afim $\mathbb{V}$, que é representada por $\dim(\mathbb{V})$, é a dimensão do espaço vetorial $\mathbb{X}$ associado. Neste caso vamos usar a notação $\mathbb{V}^n$. Por exemplo, os espaços $\mathbb{V}^2$ e $\mathbb{V}^3$ considerados nos primeiros dois capítulos são espaços afins de dimensão dois e três, respetivamente.

8.3. Combinações afins e centro de massa

Nos espaços lineares, trabalhamos sempre com combinações lineares de vetores. Nos espaços afins é mais natural considerar outro tipo de combinações que são combinações lineares com uma condição adicional para os coeficientes.

Definição 8.2. Combinação afim

Sejam Q e P_i, $i = \overline{1,N}$, pontos no espaço $\mathbb{V}$, e α_i, $i = \overline{1,N}$, números que verificam a condição

$$\sum_{i=1}^{N} \alpha_i = 1.$$

Chamamos *combinação afim* dos pontos P_i com coeficientes α_i, $i = \overline{1,N}$, ao ponto P tal que

$$\overrightarrow{QP} = \sum_{i=1}^{N} \alpha_i \overrightarrow{QP_i}.$$

Na definição anterior pode parecer que uma combinação afim depende da escolha de ponto Q. O teorema seguinte mostra que isto não é assim.

Teorema 8.1

A definição de uma combinação afim não depende da escolha do ponto Q.

Demonstração. De facto, seja $Q' \in \mathbb{V}$. Então temos

$$\begin{aligned}\overrightarrow{Q'P'} &= \sum_{i=1}^{N} \alpha_i \overrightarrow{Q'P_i} = \sum_{i=1}^{N} \alpha_i (\overrightarrow{QP_i} + \overrightarrow{Q'P_i} + \overrightarrow{P_iQ}) = \sum_{i=1}^{N} \alpha_i \overrightarrow{QP_i} + \overrightarrow{Q'Q} \\ &= \overrightarrow{QP} + \overrightarrow{Q'Q} = \overrightarrow{Q'P}.\end{aligned}$$

Portanto, o ponto P' construído através do ponto Q', coincide com o ponto P. ∎

Para entender melhor o conceito de combinação afim, vamos dar-lhe uma interpretação mecânica. Recordemos que *baricentro* ou o *centro de massa* é o ponto hipotético onde toda a massa de um sistema físico está concentrada e que se move como se todas as forças externas estivessem aí aplicadas.

Definição 8.3. Centro de massa

Sejam P_i, $i = \overline{1,N}$, pontos de um espaço afim $\mathbb{V}$. A cada ponto está atribuída uma "massa" $m_i > 0$, $i = \overline{1,N}$. Dizemos que um ponto P é o *centro de massa ou baricentro* dos pontos P_i se se verifica a igualdade

$$\sum_{i=1}^{N} m_i \overrightarrow{PP_i} = 0.$$

A ligação entre esta definição e a definição anterior fica clara através do seguinte teorema.

Teorema 8.2

O centro de massa tem as seguintes propriedades:

1. Se existe um ponto Q tal que
$$\overrightarrow{QP} = \frac{\sum_{i=1}^{N} m_i \overrightarrow{QP_i}}{\sum_{i=1}^{N} m_i}, \tag{8.1}$$
então P é centro de massa dos pontos P_i, $i = \overline{1,N}$;
2. O centro de massa está definido de maneira única;
3. Se P é o centro de massa dos pontos P_i, $i = \overline{1,N}$, então para qualquer Q temos (8.1);
4. (Lei da alavanca ou lei de Arquimedes) Se P é o centro de massa de dois pontos P_1 e P_2 com massas m_1 e m_2, então
$$m_1|\overrightarrow{PP_1}| = m_2|\overrightarrow{PP_2}|; \tag{8.2}$$
5. Sejam P o centro de massa dos pontos P_i, $i = \overline{1,N}$, e S o centro de massa dos pontos P_i, $i = \overline{1,M}$, $M < N$. Então, o centro de massa dos pontos S, a que está atribuído a massa $\mu = \sum_{i=1}^{M} m_i$, e P_i, $i = M+1, \ldots, N$, com massas m_i, $i = M+1, \ldots, N$, coincide com P.

Antes de demonstrar o teorema, façamos a seguinte observação. Fazendo
$$\alpha_i = \frac{m_i}{\sum_{i=1}^{N} m_i}$$
em (8.1) vemos que o ponto P é uma combinação afim dos pontos P_i com coeficientes positivos. Nos problemas mecânicos e geométricos, os números m_i, $i = \overline{1,N}$, são normalmente positivos, mas em alguns casos podem ser negativos ou, até, complexos. O Teorema 8.2 continua válido sempre que as somas $\sum_{i=1}^{N} m_i$ e $\sum_{i=1}^{M} m_i$ são diferentes de zero.

Vamos agora demonstrar o teorema.

Demonstração.

1. Suponhamos que existe Q tal que se verifica (8.1), então temos
$$\sum_{i=1}^{N} m_i \overrightarrow{PP_i} = \sum_{i=1}^{N} m_i (\overrightarrow{QP_i} - \overrightarrow{QP}) = \sum_{i=1}^{N} m_i \overrightarrow{QP_i} - \left(\sum_{i=1}^{N} m_i\right) \overrightarrow{QP} = 0.$$
2. A segunda afirmação segue imediatamente de (8.1). De facto, escolhemos um ponto Q. Então, o ponto P está completamente definido pela igualdade (8.1).
3. Esta propriedade do centro de massa é uma consequência imediata do Teorema 8.1.

4. Para o caso de dois pontos P_1 e P_2 escolhemos $Q = P$. Então, de (8.1) obtemos

$$0 = m_1\overrightarrow{PP_1} + m_2\overrightarrow{PP_2}.$$

Logo, temos (8.2).

5. Reescrevendo (8.1) como

$$\overrightarrow{QP} = \frac{\sum_{i=1}^{N} m_i\overrightarrow{QP_i}}{\sum_{i=1}^{N} m_i} = \frac{\left(\sum_{i=1}^{M} m_i\right)\frac{\sum_{i=1}^{M} m_i\overrightarrow{QP_i}}{\sum_{i=1}^{M} m_i} + \sum_{M+1}^{N} m_i\overrightarrow{QP_i}}{\sum_{i=1}^{M} m_i + \sum_{i=M+1}^{N} m_i}$$

$$= \frac{\mu\overrightarrow{QS} + \sum_{M+1}^{N} m_i\overrightarrow{QP_i}}{\mu + \sum_{M+1}^{N} m_i},$$

onde $\mu = \sum_{i=1}^{M} m_i$, temos o pretendido.

■

Vamos agora ver como o centro de massa pode ser utilizado na resolução de problemas geométricos.

Exemplo 8.1. O primeiro problema, que já foi resolvido no primeiro capítulo utilizando o cálculo vetorial, é mostrar que as medianas de um triângulo têm um ponto de interseção comum. Vamos agora apreciar a elegância com que se demonstra este facto utilizando o conceito do centro de massa. Esta demonstração é da autoria de Arquimedes. Consideremos os vértices P_1, P_2 e P_3 de um triângulo. A estes vértices associamos massas iguais a um. Então, pela lei de alavanca (parte 4 do Teorema 8.2), os centros de massa dos lados do triângulo são os respetivos pontos médios do triângulo. Pela parte 5 do Teorema 8.2, o centro de massa P do triângulo (que é único) está em cada uma das medianas (e divide as medianas na proporção $1:2$), isto é, as medianas têm um ponto comum.

□

A resolução do problema anterior é clássica e quase óbvia. Mas como utilizar o método do centro de massa para resolver outros problemas? Na utilização do método do centro de massa é necessário distribuir adequadamente as massas entre os pontos para ter o centro de massa no ponto que nos interessa.

Exemplo 8.2. Vamos demonstrar que as alturas de um triângulo têm um ponto comum.

Para simplificar consideremos um triângulo com os vértices P_1, P_2 e P_3 e ângulos agudos α, β e γ (veja a Fig. 8.1). Designemos por H_1, H_2 e H_3 as bases das alturas. (Na Fig. 8.1 só está mostrada a altura P_1, H_1.) Utilizando a lei de alavanca vamos tentar escolher as massas atribuídas aos vértices de modo a que as bases das alturas sejam centros de massa dos lados do triângulo. Obviamente, temos

$$|\overrightarrow{P_3H_1}| = |\overrightarrow{P_1P_3}|\cos\gamma \quad \text{e} \quad |\overrightarrow{P_2H_1}| = |\overrightarrow{P_1P_2}|\cos\beta.$$

Portanto, m_2 e m_3 têm de satisfazer a condição

$$m_3|\overrightarrow{P_1P_3}|\cos\gamma = m_2|\overrightarrow{P_1P_2}|\cos\beta. \tag{8.3}$$

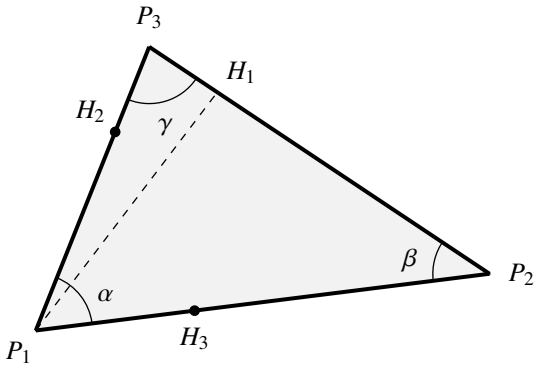


Figura 8.1
Triângulo com vértices P_1, P_2 e P_3 e ângulos agudos.

Analogamente obtemos as condições

$$m_1|\overrightarrow{P_1P_2}|\cos\alpha = m_3|\overrightarrow{P_2P_3}|\cos\gamma \quad \text{e} \quad m_1|\overrightarrow{P_1P_3}|\cos\alpha = m_2|\overrightarrow{P_2P_3}|\cos\beta.$$

Existem massas que verificam todas as três condições? A resposta é sim. Fazendo $m_1 = 1$, das últimas duas condições encontramos

$$m_2 = \frac{|\overrightarrow{P_1P_3}|\cos\alpha}{|\overrightarrow{P_2P_3}|\cos\beta} \quad \text{e} \quad m_3 = \frac{|\overrightarrow{P_1P_2}|\cos\alpha}{|\overrightarrow{P_2P_3}|\cos\gamma}.$$

Substituindo estes valores na igualdade (8.3) obtemos a identidade

$$\frac{|\overrightarrow{P_1P_3}||\overrightarrow{P_1P_2}|\cos\alpha}{|\overrightarrow{P_2P_3}|} = \frac{|\overrightarrow{P_1P_3}||\overrightarrow{P_1P_2}|\cos\alpha}{|\overrightarrow{P_2P_3}|}.$$

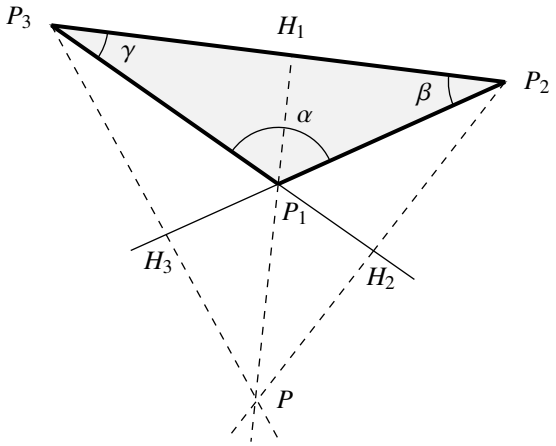


Figura 8.2
Triângulo com vértices P_1, P_2 e P_3 e um ângulo obtuso.

Pela parte 5 do Teorema 8.2 o centro de massa do triângulo, P, pertence a cada uma das alturas. Portanto, as alturas têm um ponto comum. Notemos que as massas neste exemplo podem ser negativas quando um dos ângulos do triângulo é maior do que $\frac{\pi}{2}$. Por exemplo, se $\alpha > \frac{\pi}{2}$ (veja a Fig. 8.2), as massas m_2 e m_3 são negativas e o ponto de interseção das alturas está fora do triângulo. Sugerimos ao leitor analisar a situação quando um dos ângulos é igual a $\frac{\pi}{2}$.

□

8.4. Coordenadas baricêntricas

Consideremos um espaço afim $\mathbb{V}^n$ e os pontos $P_0, P_1, \ldots, P_n$ nesse espaço. Suponhamos que os vetores $\overrightarrow{P_0P_1}, \overrightarrow{P_0P_2}, \ldots, \overrightarrow{P_0P_n}$ são linearmente independentes. Neste caso, vamos dizer que o conjunto das combinações afins dos pontos P_i, $i = \overline{0,n}$, com coeficientes não negativos é um *simplex*, que tem um papel importante na geometria. Aos pontos P_i chamamos *vértices do simplex*.

Seja $Q \in \mathbb{V}^n$. Quando m_i, $i = \overline{0,n}$, verificam somente a condição $\sum_{i=0}^{n} m_i = 1$ (podendo ter sinais arbitrários), qualquer vetor $\overrightarrow{QP}$ do espaço pode ser representado (unicamente) na forma

$$\overrightarrow{QP} = \sum_{i=0}^{n} m_i \overrightarrow{QP_i}. \tag{8.4}$$

De facto, como os vetores $\overrightarrow{P_0P_1}, \overrightarrow{P_0P_2}, \ldots, \overrightarrow{P_0P_n}$ são linearmente independentes, existem números m_i, $i = \overline{1,n}$, tais que

$$\overrightarrow{P_0P} = \sum_{i=1}^{n} m_i \overrightarrow{P_0P_i}.$$

Como

$$\overrightarrow{P_0P_i} = \overrightarrow{PP_i} - \overrightarrow{PP_0},$$

multiplicando estas igualdades por m_i e adicionando, obtemos

$$\overrightarrow{P_0P} = \sum_{i=1}^{n} m_i \left(\overrightarrow{PP_i} - \overrightarrow{PP_0}\right),$$

ou seja

$$0 = \sum_{i=1}^{n} m_i \overrightarrow{PP_i} + \left(1 - \sum_{i=1}^{n} m_i\right) \overrightarrow{PP_0}.$$

Logo, temos

$$0 = \sum_{i=0}^{n} m_i \overrightarrow{PP_i},$$

onde

$$m_0 = 1 - \sum_{i=1}^{n} m_i.$$

Obviamente $\sum_{i=0}^{n} m_i = 1$. Portanto, P é o centro de massa dos pontos P_i, $i = \overline{0,n}$, com massas (não necessariamente positivas) m_i, $i = \overline{0,n}$. Da terceira parte do Teorema 8.2 temos (8.4). Aos números m_i, $i = \overline{0,n}$, chamamos *coordenadas baricêntricas* do ponto P. Estas coordenadas são muito úteis na resolução de vários problemas.

8.5. Subespaços afins

Recordemos que um conjunto $\mathbb{L}$ de um espaço linear $\mathbb{X}$ é subespaço de $\mathbb{X}$ se $\mathbb{L}$ é um espaço linear, isto é, se qualquer combinação linear de vetores de $\mathbb{L}$ pertence a $\mathbb{L}$. Um subespaço de um espaço afim vamos definir assim.

Definição 8.4. Subespaço afim

Ao conjunto $\mathbb{S} \subset \mathbb{V}$ vamos chamar subespaço afim, se qualquer combinação afim dos seus pontos também pertence a $\mathbb{S}$.

Consideremos um exemplo.

Exemplo 8.3. O conjunto

$$\mathbb{S} = \{x \in \mathbb{R}^n : Ax = b\},$$

que contém todas as soluções do sistema possível $Ax = b$, é um subespaço afim. De facto, sejam $x_i \in \mathbb{S}$ e $\alpha_i \in \mathbb{R}$, $i = \overline{1,N}$, tais que $\sum_{i=1}^{N} \alpha_i = 1$. Então temos

$$A\sum_{i=1}^{N} \alpha_i x_i = \sum_{i=1}^{N} \alpha_i b = b.$$

□

Os subespaços afins de $\mathbb{V}$ podem ser caracterizados em termos de subespaços de $\mathbb{X}$, o espaço de vetores de $\mathbb{V}$.

Teorema 8.3

Seja $\mathbb{S} \subset \mathbb{V}$ um subespaço afim. Então, todos os vetores $\overrightarrow{P_1P_2}$ tais que $P_1, P_2 \in \mathbb{S}$ formam um subespaço $\mathbb{L} \subset \mathbb{X}$.

Demonstração. Sejam $Q, P_1, P_2, P_1', P_2' \in \mathbb{S}$ e $\alpha, \beta \in \mathbb{R}$. Então, como

$$-\alpha + \alpha + (1-\beta) + \beta = 1,$$

existe $P \in \mathbb{S}$ tal que

$$\overrightarrow{QP} = -\alpha\overrightarrow{QP_1} + \alpha\overrightarrow{QP_2} + (1-\beta)\overrightarrow{QP_1'} + \beta\overrightarrow{QP_2'} = \alpha\overrightarrow{P_1P_2} + \beta\overrightarrow{P_1'P_2'} + \overrightarrow{QP_1'}.$$

Portanto, temos que

$$\alpha\overrightarrow{P_1P_2} + \beta\overrightarrow{P_1'P_2'} = \overrightarrow{QP} - \overrightarrow{QP_1'} = \overrightarrow{P_1'P},$$

ou seja, a combinação linear dos vetores $\overrightarrow{P_1P_2}$ e $\overrightarrow{P_1'P_2'}$ é um vetor formado por um par de pontos de $\mathbb{S}$.

■

8.6. Aplicações afins

Nos espaços afins, às aplicações lineares dos espaços lineares correspondem aplicações afins dos espaços afins.

Definição 8.5. Aplicação afim

Sejam $\mathbb{V}$ e $\mathbb{W}$ espaços afins. Os respetivos espaços de vetores são $\mathbb{X}$ e $\mathbb{Y}$. A uma aplicação $\mathcal{A}$ que transforma o espaço $\mathbb{V}$ em $\mathbb{W}$ vamos chamar *afim* se ela preserva as combinações afins no seguinte sentido. Sejam Q, P e P_i, $i = \overline{1,N}$, pontos no espaço $\mathbb{V}$, e α_i, $i = \overline{1,N}$, números que verificam as condições

$$\sum_{i=1}^{N} \alpha_i = 1 \quad \text{e} \quad \overrightarrow{QP} = \sum_{i=1}^{N} \alpha_i \overrightarrow{QP_i}.$$

Então temos

$$\overrightarrow{\mathcal{A}(Q)\mathcal{A}(P)} = \sum_{i=1}^{N} \alpha_i \overrightarrow{\mathcal{A}(Q)\mathcal{A}(P_i)}.$$

Consideremos um exemplo.

Exemplo 8.4. A aplicação $\mathcal{A} : \mathbb{V}^2 \to \mathbb{V}^2$ definida por $\mathcal{A}(x^1, x^2) = (x^1 + 2x^2 + 2, x^2 + 2)$ é uma aplicação afim em $\mathbb{V}^2$. Esta aplicação corresponde a uma deformação seguida de uma translação e pode ser reescrita da forma $\mathcal{A}(x^1, x^2) = A(x^1, x^2) + (2, 2)$, onde $A \in \mathcal{L}(\mathbb{R}^2, \mathbb{R}^2)$ definida por

$$A(x^1, x^2) = (x^1 + 2x^2, x^2) \tag{8.5}$$

é uma aplicação linear. □

A aplicação afim faz corresponder a cada par de pontos $P_1, P_2 \in \mathbb{V}$ um par de pontos $\mathcal{A}(P_1), \mathcal{A}(P_2) \in \mathbb{W}$. Portanto, temos uma aplicação A que faz corresponder a um vetor $\overrightarrow{P_1P_2} \in \mathbb{X}$ um vetor $\overrightarrow{\mathcal{A}(P_1)\mathcal{A}(P_2)} \in \mathbb{Y}$.

Teorema 8.4

A aplicação

$$A(\overrightarrow{P_1P_2}) = \overrightarrow{\mathcal{A}(P_1)\mathcal{A}(P_2)}$$

é linear.

Demonstração. Tal como na demonstração do Teorema 8.3, consideremos Q, P_1, P_2, P_1', $P_2' \in \mathbb{V}$ e $\alpha, \beta \in \mathbb{R}$. Seja $P \in \mathbb{V}$ a combinação afim dos pontos P_1, P_2, P_1', P_2' com coeficientes $-\alpha$, α, $(1-\beta)$, β. Então temos

$$\overrightarrow{QP} = -\alpha\overrightarrow{QP_1} + \alpha\overrightarrow{QP_2} + (1-\beta)\overrightarrow{QP_1'} + \beta\overrightarrow{QP_2'} = \alpha\overrightarrow{P_1P_2} + \beta\overrightarrow{P_1'P_2'} + \overrightarrow{QP_1'}.$$

Daqui resulta que

$$\alpha\overrightarrow{P_1P_2} + \beta\overrightarrow{P_1'P_2'} = \overrightarrow{QP} - \overrightarrow{QP_1'} = \overrightarrow{P_1'P}. \tag{8.6}$$

A aplicação afim $\mathcal{A}$ preserva as combinações afins. Logo, temos

$$\overrightarrow{\mathcal{A}(Q)\mathcal{A}(P)} = -\alpha\overrightarrow{\mathcal{A}(Q)\mathcal{A}(P_1)} + \alpha\overrightarrow{\mathcal{A}(Q)\mathcal{A}(P_2)} + (1-\beta)\overrightarrow{\mathcal{A}(Q)\mathcal{A}(P_1')}$$
$$+ \beta\overrightarrow{\mathcal{A}(Q)\mathcal{A}(P_2')} = \alpha\overrightarrow{\mathcal{A}(P_1)\mathcal{A}(P_2)} + \beta\overrightarrow{\mathcal{A}(P_1')\mathcal{A}(P_2')} + \overrightarrow{\mathcal{A}(Q)\mathcal{A}(P_1')}.$$

Daqui obtemos

$$\alpha\overrightarrow{\mathcal{A}(P_1)\mathcal{A}(P_2)} + \beta\overrightarrow{\mathcal{A}(P_1')\mathcal{A}(P_2')} = \overrightarrow{\mathcal{A}(Q)\mathcal{A}(P)} - \overrightarrow{\mathcal{A}(Q)\mathcal{A}(P_1')} = \overrightarrow{\mathcal{A}(P_1')\mathcal{A}(P)}. \quad (8.7)$$

Como

$$A(\overrightarrow{P_1P_2}) = \overrightarrow{\mathcal{A}(P_1)\mathcal{A}(P_2)},$$
$$A(\overrightarrow{P_1'P_2'}) = \overrightarrow{\mathcal{A}(P_1')\mathcal{A}(P_2')} \quad \text{e}$$
$$A(\overrightarrow{P_1'P}) = \overrightarrow{\mathcal{A}(P_1')\mathcal{A}(P)},$$

das igualdades (8.6) e (8.7) temos

$$A(\alpha\overrightarrow{P_1P_2} + \beta\overrightarrow{P_1'P_2'}) = \alpha A(\overrightarrow{P_1P_2}) + \beta A(\overrightarrow{P_1'P_2'}).$$

■

Notemos que, no Exemplo 8.4, a aplicação linear A, induzida pela aplicação afim $\mathcal{A}$, é dada pela fórmula (8.5).

8.7. Cadeias de Márkov*

Os espaços afins são um ambiente natural para resolver vários problemas vindos de diversas áreas de conhecimento. Nesta secção vamos ver um destes problemas — as cadeias de Márkov.

Consideremos um sistema que pode estar num de n estados e em que a probabilidade de transição do estado j para o estado i depende só do estado j e é igual a a_i^j. Como o sistema chegou ao estado j não tem importância. A este processo estocástico damos o nome de *Cadeia de Márkov*. Obviamente $\sum_{i=1}^n a_i^j = 1$, $j = \overline{1,n}$, porque a probabilidade de transição para um dos estados é igual a um.

Dizemos que a matriz

$$A = \begin{pmatrix} a_1^1 & \cdots & a_1^n \\ \vdots & \ddots & \vdots \\ a_n^1 & \cdots & a_n^n \end{pmatrix}$$

é *estocástica* se $a_i^j \geqslant 0$ e $\sum_{i=1}^n a_i^j = 1$ para $j = \overline{1,n}$. Notemos que, se as coordenadas do vetor

$$p = \begin{pmatrix} p^1 \\ \vdots \\ p^n \end{pmatrix} \in \mathbb{R}^n$$

verificam as condições

$$p^i \geqslant 0, i = \overline{1,n}, \quad \text{e} \quad \sum_{i=1}^{n} p^i = 1, \tag{8.8}$$

então as coordenadas do vetor Ap também verificam (8.8). De facto, a primeira condição é óbvia e relativamente à segunda temos

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i^j p^j = \sum_{j=1}^{n}\sum_{i=1}^{n} a_i^j p^j = \sum_{j=1}^{n} p^j = 1.$$

Seja o vetor p, que verifica (8.8), uma distribuição inicial de probabilidades, isto é, no momento inicial de tempo, o sistema está no estado i com probabilidade p^i, $i = \overline{1,n}$. Com o passar do tempo, esta distribuição de probabilidades evolui de acordo com a sucessão de vetores

$$Ap, A^2p, \ldots, A^kp, \ldots.$$

Vamos, agora, estudar algumas propriedades das matrizes estocásticas e das cadeias de Márkov.

Sejam e_i, $i = \overline{1,n}$, os vetores da base canónica em $\mathbb{R}^n$. Consideremos o simplex

$$\Sigma = \left\{ p = \sum_{i=1}^{n} p^i e_i : p^i \geqslant 0, \sum_{i=1}^{n} p^i = 1 \right\}$$

e uma matriz estocástica $A = (a_i^j)$ de dimensões $n \times n$ tal que $a_i^j > a > 0$, $i, j = \overline{1,n}$. Obviamente, $a < \frac{1}{2}$. De facto, caso contrário, a soma dos elementos de uma coluna de A seria superior a 1.

Notemos que, nesta nossa consideração, temos estruturas de um espaço afim que foram introduzidas neste capítulo: o espaço de pontos é $\mathbb{R}^n$, o respetivo espaço de vetores é o mesmo $\mathbb{R}^n$, o conjunto

$$\mathbb{S} = \left\{ p = \sum_{i=1}^{n} p^i e_i : \sum_{i=1}^{n} p^i = 1 \right\}$$

é um subespaço afim e a aplicação A é uma aplicação afim que deixa o subespaço afim $\mathbb{S}$ invariante, isto é, $A\mathbb{S} \subset \mathbb{S}$.

Vamos estabelecer uma propriedade importante da aplicação A. Recordemos que a definição da norma $|\cdot|_1$ é

$$|x|_1 = \sum_{i=1}^{n} |x^i|.$$

Lema 8.1

Para quaisquer vetores $p, q \in \Sigma$ temos a desigualdade

$$|A(p-q)|_1 \leqslant (1-a)|p-q|_1.$$

Demonstração. Sejam $p, q \in \Sigma$. Consideremos os conjuntos de índices

$$J^+ = \{j : p^j - q^j \geqslant 0\} \quad \text{e} \quad J^- = \{j : p^j - q^j < 0\}.$$

Como

$$0 = \sum_{j=1}^{n}(p^j - q^j) = \sum_{j\in J^+}(p^j - q^j) + \sum_{j\in J^-}(p^j - q^j),$$

estes conjuntos não são vazios sempre que $p \neq q$. Como

$$\sum_{j=1}^{n}|p^j - q^j| = \sum_{j\in J^+}(p^j - q^j) + \sum_{j\in J^-}(q^j - p^j)$$

e

$$\sum_{j\in J^+}(p^j - q^j) = \sum_{j\in J^-}(q^j - p^j),$$

temos

$$\sum_{j\in J^+}(p^j - q^j) = \sum_{j\in J^-}(q^j - p^j) = \frac{1}{2}\sum_{j=1}^{n}|p^j - q^j|.$$

Definamos, ainda, os conjuntos de índices

$$I^+ = \left\{i : \sum_{j\in J^+} a_i^j(p^j - q^j) \geqslant \sum_{j\in J^-} a_i^j(q^j - p^j)\right\}$$

e

$$I^- = \left\{i : \sum_{j\in J^+} a_i^j(p^j - q^j) < \sum_{j\in J^-} a_i^j(q^j - p^j)\right\}.$$

O conjunto I^+ nunca é vazio. De facto, a igualdade $I^+ = \emptyset$ implica que para todo $i = \overline{1,n}$ temos

$$\sum_{j\in J^+} a_i^j(p^j - q^j) < \sum_{j\in J^-} a_i^j(q^j - p^j).$$

Adicionando estas desigualdades, obtemos

$$\sum_{i=1}^{n}\sum_{j\in J^+} a_i^j(p^j - q^j) < \sum_{i=1}^{n}\sum_{j\in J^-} a_i^j(q^j - p^j),$$

ou seja,

$$0 = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i^j(p^j - q^j) < 0,$$

o que é uma contradição.

Suponhamos que $I^- = \emptyset$. Então temos

$$|A(p-q)|_1 = \sum_{i=1}^{n} \left| \sum_{j=1}^{n} a_i^j (p^j - q^j) \right| = \sum_{i=1}^{n} \left| \sum_{j \in J^+} a_i^j (p^j - q^j) + \sum_{j \in J^-} a_i^j (p^j - q^j) \right|$$
$$\leqslant \sum_{i=1}^{n} \sum_{j \in J^+} a_i^j (p^j - q^j) = \sum_{j \in J^+} (p^j - q^j) = \frac{1}{2} \sum_{j=1}^{n} |p^j - q^j| < (1-a)|p-q|_1.$$

Agora consideremos o caso quando $I^- \neq \emptyset$. Então temos

$$\sum_{i \in I^+} a_i^j < 1 - a \quad \text{e} \quad \sum_{i \in I^-} a_i^j < 1 - a, \quad j = \overline{1,n},$$

pelo que

$$|A(p-q)|_1 = \sum_{i=1}^{n} \left| \sum_{j=1}^{n} a_i^j (p^j - q^j) \right| = \sum_{i=1}^{n} \left| \sum_{j \in J^+} a_i^j (p^j - q^j) + \sum_{j \in J^-} a_i^j (p^j - q^j) \right|$$
$$\leqslant \sum_{i=1}^{n} \max \left\{ \sum_{j \in J^+} a_i^j (p^j - q^j), \sum_{j \in J^-} a_i^j (q^j - p^j) \right\}$$
$$= \sum_{i \in I^+} \sum_{j \in J^+} a_i^j (p^j - q^j) + \sum_{i \in I^-} \sum_{j \in J^-} a_i^j (q^j - p^j)$$
$$< (1-a) \sum_{j \in J^+} (p^j - q^j) + (1-a) \sum_{j \in J^-} (q^j - p^j) = (1-a)|p-q|_1,$$

o que demonstra o lema. ■

Vamos agora demonstrar o resultado principal desta secção. Na demonstração vamos recorrer a métodos de Análise Matemática (os resultados de Álgebra Linear são muito utilizados em Análise Matemática mas o contrário também é útil).

Teorema 8.5

Seja $p \in \Sigma$. Então, a sucessão $A^k p$, $k = 1, 2, \ldots$, converge para um vetor $\hat{p} \in \Sigma$ que verifica

$$A\hat{p} = \hat{p},$$

isto é, $\hat{p}$ é um vetor próprio da matriz A com o valor próprio um. Este vetor é único e as suas coordenadas são positivas.

Demonstração. Mostremos que a sucessão $A^k p$ verifica a condição de Cauchy. De facto, seja $\epsilon > 0$. Então, do lema anterior, temos

$$\begin{aligned}
|A^{k+l}p - A^k p|_1 &= |A^k(A^l p - p)|_1 < (1-a)^k |A^l p - p|_1 \\
&= (1-a)^k |A^l p - A^{l-1}p + A^{l-1}p - A^{l-2}p + \cdots - Ap + Ap - p|_1 \\
&\leqslant (1-a)^k (|A^l p - A^{l-1}p|_1 + |A^{l-1}p - A^{l-2}p|_1 + \cdots + |Ap - p|_1) \\
&< (1-a)^k ((1-a)^{l-1}|Ap - p|_1 + (1-a)^{l-2}|Ap - p|_1 + \cdots + |Ap - p|_1) \\
&< (1-a)^k \sum_{l=0}^{\infty} (1-a)^l |Ap - p|_1 = \frac{(1-a)^k}{a} |Ap - p|_1 < \epsilon,
\end{aligned}$$

sempre que k é suficientemente grande. (Aqui utilizámos a fórmula para a soma da progressão geométrica infinita com o coeficiente $(1-a) < 1$.) Portanto, existe $\lim_{k\to\infty} A^k p = \hat{p}$. Como $A^k p \in \Sigma$ temos que $\hat{p} \in \Sigma$. Passando ao limite na igualdade

$$A(A^k p) = A^{k+1} p,$$

obtemos $A\hat{p} = \hat{p}$. Agora, suponhamos que em Σ existe $\bar{p} \neq \hat{p}$, um outro vetor próprio da matriz A correspondente ao valor próprio um. Então temos

$$|\hat{p} - \bar{p}|_1 = |A\hat{p} - A\bar{p}|_1 = |A(\hat{p} - \bar{p})|_1 < (1-a)|\hat{p} - \bar{p}|_1,$$

o que é uma contradição. Como todos os elementos da matriz A são positivos, todas as coordenadas do vetor $\hat{p}$ são positivas. ■

O exemplo que vamos ver a seguir faz uso do seguinte corolário.

Corolário 8.1

Suponhamos que A é uma matriz estocástica e existe $k > 1$ tal que todos os elementos da matriz A^k são positivos. Então, existe $\hat{p}$, um vetor próprio da matriz A com o valor próprio um. Este vetor é único e as suas coordenadas são positivas.

Demonstração. Como a matriz A é uma matriz estocástica obvimente a matriz A^k também o é. Pelo Teorema 8.5, existe um único vetor $\hat{p} \in \Sigma$ com coordenadas positivas, tal que

$$A^k \hat{p} = \hat{p}.$$

Logo, temos

$$A\hat{p} = A(A^k \hat{p}) = A^k(A\hat{p}).$$

Então, $A\hat{p} \in \Sigma$ é um vetor próprio de A^k com o valor próprio um. Sabemos que este vetor é único. Portanto, temos $A\hat{p} = \hat{p}$. ■

A Teoria das Cadeias de Márkov tem muitas aplicações. Consideremos um exemplo vindo da genética.

Exemplo 8.5. O tipo mais simples de herança de características em animais ocorre quando uma característica é governada por um par de genes, cada um dos quais pode ser de dois tipos, por exemplo, G e g. Um indivíduo pode ter uma combinação GG ou Gg (que é geneticamente o mesmo que gG) ou gg. Muitas vezes os tipos GG e Gg são indistinguíveis em aparência, e então dizemos que o gene G domina o gene g. Dizemos que um indivíduo é *dominante* se tiver genes GG, *recessivo* se tiver gg e *híbrido* se tiver genes Gg.

No acasalamento de dois animais, os descendentes herdam um gene do par de cada progenitor, e a suposição básica da genética é que esses genes são selecionados aleatoriamente. Esta suposição determina a probabilidade de ocorrência de cada tipo de descendente. O descendente de dois progenitores dominantes deve ser dominante, de dois progenitores recessivos deve ser recessivo e de um dominante e um recessivo deve ser híbrido. No acasalamento de um animal dominante e um híbrido, cada descendente deve obter um G do primeiro e tem chances iguais de obter G ou g do último. Portanto, há uma probabilidade igual de obter um descendente dominante ou híbrido. Novamente, no cruzamento de um recessivo com um híbrido, há uma chance igual de obter descendente recessivo ou híbrido. No acasalamento de dois híbridos, o descendente tem chances iguais de obter G ou g de cada progenitor. Portanto, as probabilidades são $\frac{1}{4}$ para GG, $\frac{1}{2}$ para Gg e $\frac{1}{4}$ para gg.

Consideremos uma sucessão de acasalamentos. Começamos com um animal cujo caráter genético é conhecido e acasalamos com um híbrido. Assumimos que há pelo menos um descendente. Um descendente é escolhido aleatoriamente e é cruzado com um híbrido e este processo é repetido por várias gerações. O tipo genético do descendente escolhido em gerações sucessivas pode ser representado por uma cadeia de Márkov. Os estados são dominante, híbrido e recessivo e indicados por GG, Gg e gg, respetivamente. A respetiva matriz de transição é

$$A = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & 0 \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{4} & \frac{1}{2} \end{pmatrix}.$$

Esta matriz contém zeros, mas o seu quadrado, A^2, tem todos os elementos positivos. É fácil verificar que o único vetor próprio desta matriz A, correspondente ao valor próprio um e que representa a distribuição final de probabilidades de obter as combinações GG, Gg e gg, é

$$\hat{p} = \begin{pmatrix} \frac{1}{4} \\ \frac{1}{2} \\ \frac{1}{4} \end{pmatrix}.$$

A análise deste tipo é muito útil para o trabalho dos criadores de gado ou de novas variedades de cereais e outras culturas.

□

9. Formas quadráticas

9.1. Introdução

Até agora, nos espaços Euclidianos, trabalhamos só com funções lineares e aplicações lineares. Neste capítulo vamos estudar uma classe de funções não lineares. Depois das funções lineares ax, as funções mais simples são as funções quadráticas ax^2. No caso de espaços de dimensão maior do que um, a função mais simples a seguir à função linear $\langle a, x\rangle$ é a função

$$f(x) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i^j x^i x^j.$$

Estas funções, chamadas *formas quadráticas*, desempenham um papel muito importante em muitos ramos da Matemática.

9.2. Formas quadráticas e matrizes

Sejam $\mathbb{X}$ um espaço Euclidiano real de dimensão n e $A \in \mathcal{L}(\mathbb{X}, \mathbb{X})$ uma aplicação linear simétrica. Consideremos uma base ortonormal $\mathcal{E} = \{e_1, \ldots, e_n\}$. Nesta base, à aplicação A corresponde a matriz simétrica

$$A = \begin{pmatrix} a_1^1 & a_1^2 & \cdots & a_1^n \\ a_2^1 & a_2^2 & \cdots & a_2^n \\ \vdots & \vdots & \ddots & \vdots \\ a_n^1 & a_n^2 & \cdots & a_n^n \end{pmatrix}.$$

A forma quadrática

$$f(x) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i^j x^i x^j$$

pode ser escrita utilizando o produto escalar como

$$f(x) = \langle x, Ax\rangle,$$

ou ainda, considerando x como vetor-coluna, como

$$f(x) = x^{\mathsf{T}} Ax.$$

Temos então três maneiras de representar uma forma quadrática. No que se segue vamos utilizar todas estas representações, escolhendo aquela que é mais adequada à questão que está a ser considerada.

Consideremos uma aplicação linear Q invertível. Recordemos a fórmula da matriz transposta de um produto de matrizes: $(AB)^{\mathsf{T}} = B^{\mathsf{T}} A^{\mathsf{T}}$. Fazendo $x = Qy$, obtemos

$$f(x) = f(Qy) = (Qy)^{\mathsf{T}} AQy = y^{\mathsf{T}}(Q^{\mathsf{T}} AQ)y = y^{\mathsf{T}} By,$$

onde $B = Q^{\mathsf{T}} AQ$. Portanto, após a mudança de variável $x = Qy$, a matriz da forma quadrática transforma-se segundo a regra

$$B = Q^{\mathsf{T}} AQ. \tag{9.1}$$

Vamos agora ver como se transforma a matriz de uma forma quadrática quando fazemos uma mudança de base.

Na base ortonormal $\mathcal{E}$, à forma quadrática f corresponde a matriz simétrica $A_{\mathcal{E}}$. Passando à base $\mathcal{F} = \{f_1, \ldots, f_n\}$, definida como $f_k = Pe_k$, $k = \overline{1,n}$, onde P é uma aplicação linear invertível, temos que a relação entre os vetores x na base $\mathcal{E}$ e y na base $\mathcal{F}$ é

$$x = Py.$$

A matriz da aplicação A na base $\mathcal{F}$ tem a forma

$$A_{\mathcal{F}} = P^{-1} A_{\mathcal{E}} P.$$

Logo,

$$A_{\mathcal{E}} = P A_{\mathcal{F}} P^{-1},$$

pelo que

$$\langle x, A_{\mathcal{E}} x\rangle = \langle x, PA_{\mathcal{F}} P^{-1} x\rangle = \langle Py, PA_{\mathcal{F}} P^{-1} Py\rangle = \langle y, P^{\mathsf{T}} PA_{\mathcal{F}} y\rangle. \tag{9.2}$$

Assim temos a relação entre as matrizes da forma quadrática nas bases $\mathcal{E}$ e $\mathcal{F}$. A matriz $P^{\mathsf{T}} P$ é a matriz de Gram construida de produtos escalares das colunas (ou linhas) da matriz P.

Suponhamos que a base $\mathcal{F}$ é ortonormal. Então, o produto escalar dos vetores

$$x = \sum_{i=1}^{n} \xi^i e_i \quad \text{e} \quad y = \sum_{j=1}^{n} \eta^j e_j$$

é igual a

$$\begin{aligned}\langle x, y\rangle &= \left\langle \sum_{i=1}^{n} \xi^i e_i, \sum_{j=1}^{n} \eta^j e_j \right\rangle = \sum_{i=1}^{n} \xi^i \eta^i = \left\langle \sum_{i=1}^{n} \xi^i f_i, \sum_{j=1}^{n} \eta^j f_j \right\rangle \\ &= \left\langle \sum_{i=1}^{n} \xi^i Pe_i, \sum_{j=1}^{n} \eta^j Pe_j \right\rangle = \langle Px, Py\rangle.\end{aligned}$$

Portanto, neste caso, a aplicação P preserva o produto escalar e, logo, é uma aplicação ortogonal. A matriz inversa de P é P^{T}, e o último termo da igualdade (9.2) transforma-se em

$$\langle y, P^{\mathsf{T}} PA_{\mathcal{F}} y\rangle = \langle y, A_{\mathcal{F}} y\rangle = \langle y, P^{\mathsf{T}} A_{\mathcal{E}} Py\rangle.$$

Portanto, a mudança de variáveis $x = Qy$ numa forma quadrática $\langle x, Ax\rangle$ significa passar de uma base ortonormal para uma outra base ortonormal.

9.3. Diagonalização

Fazendo uma mudança de variável, alteramos a matriz da forma quadrática. É sempre interessante encontrar variáveis com as quais a matriz tem a forma mais simples possível, uma forma diagonal.

Teorema 9.1

Seja uma forma quadrática

$$f(x) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i^j x^i x^j.$$

Então, existe uma matriz Q invertível tal que a matriz $Q^{\mathsf{T}}AQ$ é diagonal.

Demonstração. O teorema é óbvio no caso de espaços de dimensão $n = 1$, pois qualquer forma quadrática tem a forma ax^2. Suponhamos que o teorema já está demonstrado para espaços de dimensão menor do que n. Consideremos uma forma quadrática

$$f(x) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i^j x^i x^j.$$

Suponhamos que $a_1^1 \neq 0$. Então, a forma quadrática

$$g(x) = f(x) - \frac{1}{a_1^1}(a_1^1x^1 + a_1^2x^2 + \cdots + a_1^nx^n)^2, \tag{9.3}$$

como é fácil de ver, não depende de x^1. De facto, os termos do quadrado

$$\frac{1}{a_1^1}(a_1^1x^1 + a_1^2x^2 + \cdots + a_1^nx^n)^2$$

que têm a variável x^1 formam a soma

$$a_1^1x^1x^1 + 2a_1^2x^1x^2 + \cdots + 2a_1^nx^1x^n.$$

Portanto, fazendo

$$y^1 = a_1^1x^1 + a_1^2x^2 + \cdots + a_1^nx^n \quad \text{e} \quad y^k = x^k, k = \overline{2,n}, \tag{9.4}$$

obtemos

$$f(x) = \frac{1}{a_1^1}(y^1)^2 + g(x) = \frac{1}{a_1^1}(y^1)^2 + g(y^2, y^3, \ldots, y^n).$$

Pela hipótese da indução, sem perda de generalidade, a forma $g(y^2, y^3, \ldots, y^n)$ já tem matriz diagonal. Como $a_1^1 \neq 0$, sabendo y^k, $k = \overline{1,n}$, de (9.4) podemos encontrar x^k. Portanto, a matriz de transição das variáveis x^k para as variáveis y^k é invertível.

Consideremos agora o caso quando $a_k^k = 0$, $k = \overline{1,n}$. Existe pelo menos um coeficiente $a_k^m \neq 0$. Suponhamos que $a_1^2 \neq 0$. Introduzindo as coordenadas z^k

$$x^1 = z^1 - z^2, \quad x^2 = z^1 + z^2 \quad \text{e} \quad z^k = x^k, k = \overline{3,n}, \tag{9.5}$$

obtemos uma forma quadrática onde o termo $a_1^2 x^1 x^2$ toma a forma

$$2a_1^2 x^1 x^2 = 2a_1^2 (z^1)^2 - 2a_1^2 (z^2)^2,$$

pelo que a matriz de transição (9.5) das variáveis x^k para as variáveis z^k é invertível. Em variáveis z^k a forma já tem termos com quadrados das variáveis e podemos aplicar o raciocínio anterior.

■

O método utilizado na demonstração deste teorema é conhecido pelo *método de diagonalização de formas quadráticas de Lagrange*.

Há várias possibilidades para diagonalizar uma forma quadrática. Uma dessas possibilidades, vantajosa em muitos aspetos, é através de matrizes ortogonais.

Teorema 9.2

Seja A uma matriz simétrica. Então, existe uma matriz ortogonal Q tal que $Q^{\mathsf{T}} A Q$ é uma matriz diagonal.

Demonstração. Seja $A \in \mathcal{L}(\mathbb{X}, \mathbb{X})$ uma aplicação simétrica cuja matriz é A. Consideremos a matriz Q cujas colunas e_k, $k = \overline{1,n}$, são os vetores próprios da aplicação A e que formam uma base ortonormal em $\mathbb{X}$ (veja o Teorema 7.3). É fácil ver que $Q^{\mathsf{T}} Q = I$, ou seja, a matriz Q é ortogonal. A matriz A nesta base é diagonal:

$$Q^{-1} A Q = Q^{\mathsf{T}} A Q = Q^{\mathsf{T}} \begin{pmatrix} \lambda_1 e_1 & \cdots & \lambda_n e_n \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix},$$

onde λ_k, $k = \overline{1,n}$, são os valores próprios que correspondem aos vetores próprios e_k, $k = \overline{1,n}$.

■

Vamos aplicar a teoria desenvolvida para diagonalizar a matriz de uma forma quadrática.

Exemplo 9.1. Consideremos a forma quadrática

$$f(x^1, x^2, x^3) = (x^1)^2 + 4x^1 x^3 + (x^2)^2 + 2x^2 x^3 + 4(x^3)^2.$$

A matriz desta forma quadrática é

$$A = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 2 & 1 & 4 \end{pmatrix}.$$

Existe um número infinito de aplicações lineares Q que diagonalizam a forma quadrática e o número de resultados da diagonalização também é infinito. De facto, se em coordenadas $y = Qx$ a matriz da forma quadrática é diagonal, ela é também diagonal

em coordenadas $y' = DQx$, onde D é uma matriz diagonal. Vamos diagonalizar a matriz utilizando três processos.

Primeiro, vamos diagonalizar a forma quadrática utilizando o método de Lagrange. Juntamos os termos com x^1 e adicionamos e subtraímos os termos necessários para obter um quadrado completo:

$$\begin{aligned} f &= (x^1)^2 + 4x^1x^3 + 4(x^3)^2 - 4(x^3)^2 + (x^2)^2 + 2x^2x^3 + 4(x^3)^2 \\ &= (x^1 + 2x^3)^2 + (x^2)^2 + 2x^2x^3. \end{aligned}$$

Fazendo $y^1 = x^1 + 2x^3$ temos

$$f = (y^1)^2 + (x^2)^2 + 2(x^2)x^3.$$

Agora juntamos os termos com x^2 e adicionamos e subtraímos os termos necessários para obter um quadrado completo:

$$f = (y^1)^2 + (x^2)^2 + 2x^2x^3 + (x^3)^2 - (x^3)^2 = (y^1)^2 + (x^2 + x^3)^2 - (x^3)^2.$$

Fazendo $y^2 = x^2 + x^3$ e $y^3 = x^3$ chegamos à forma quadrática diagonalizada:

$$f(y^1, y^2, y^3) = (y^1)^2 + (y^2)^2 - (y^3)^2.$$

A matriz da respetiva mudança de coordenadas $y = Qx$ é

$$Q = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

A sua inversa, $P = Q^{-1}$ que corresponde à mudança de coordenadas $x = Py$ é

$$P = \begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Outro método é baseado nas operações elementares sobre matrizes. Recordemos que as operações elementares (multiplicação de uma linha ou coluna por um número, adição de uma linha ou coluna a uma outra e troca de linhas ou colunas) são equivalentes à multiplicação à esquerda ou à direita da matriz por matrizes $M(i, \lambda)$, $S(i, j)$ e $T(i, j)$ (veja a Secção 5.5). Os determinantes destas matrizes são λ, 1 e -1, respetivamente. Portanto, é possível procurar a mudança de coordenadas $x = Py$ construindo a matriz P como um produto das matrizes de operações elementares: $P = E_1 \dots E_m$. O objetivo é obter uma matriz diagonal D tal que

$$f = \langle x, Ax \rangle = \langle Py, APy \rangle = \langle y, P^{\mathsf{T}} APy \rangle = \langle y, Dy \rangle.$$

Em termos práticos isto significa que é necessário aplicar as mesmas operações elementares às linhas e às colunas por forma a diagonalizar a matriz A. Fazendo, em paralelo, as mesmas operações com as colunas da matriz I obtemos a matriz $P = E_1 \dots E_m$. Vamos aplicar este método à diagonalização da forma quadrática f.

Para o fazer, tal como na Secção 3.3, onde foram introduzidas as notações para as operações elementares com linhas, é útil introduzir as respetivas notações para as operações com colunas. A troca da i-ésima coluna com a j-ésima coluna vamos denotar por $c_i \leftrightarrow c_j$ e a substituição da i-ésima coluna pela soma da j-ésima coluna multiplicada por a e da k-ésima coluna vamos denotar por $c_i \leftarrow ac_j + c_k$. Então, o algoritmo de diagonalização de formas quadráticas é o seguinte. Primeiro construimos a matriz $(A|I)$. A seguir aplicando as operações elementares vamos obter uma sequência de matrizes $(A_k|P_k)$, $k = \overline{0, K}$, onde $(A_0|P_0) = (A|I)$ e $(A_K|P_K) = (D|P)$. Nesta construção aplicamos as mesmas operações elementares às linhas e colunas da matriz A_k e só às colunas da matriz P_k.

Vamos agora ver como isto funciona no caso do nosso exemplo. Começamos por construir a matriz

$$(A|I) = \begin{pmatrix} 1 & 0 & 2 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 2 & 1 & 4 & 0 & 0 & 1 \end{pmatrix}.$$

Efetuamos as operações elementares $l_3 \leftarrow l_3 - 2l_1$ e $c_3 \leftarrow c_3 - 2c_1$. Como resultado temos a matriz

$$\begin{pmatrix} 1 & 0 & 0 & 1 & 0 & -2 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

(transformamos as linhas e colunas da matriz A e só as colunas da matriz I). Depois das operações $l_3 \leftarrow l_3 - l_2$ e $c_3 \leftarrow c_3 - c_2$, chegamos ao resultado

$$(D|P) = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & -2 \\ 0 & 1 & 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & 0 & 0 & 1 \end{pmatrix}.$$

Pode acontecer que durante o processo de diagonalização pelo método de Lagrange chegamos à situação quando só restam os termos $x^i x^j$, $i \neq j$. Neste caso é necessário utilizar a mudança de coordenadas (9.5)

$$x^i = z^i - z^j \quad \text{e} \quad x^j = z^i + z^j.$$

Por exemplo, em caso da forma $\phi = 2x^1x^2$, fazendo $x^1 = z^1 - z^2$ e $x^2 = z^1 + z^2$ obtemos $\phi = 2(z^1)^2 - 2(z^2)^2$. Na diagonalização pelo método de operações elementares a esta situação corresponde uma matriz que tem forma

$$\begin{pmatrix} D' & 0 \\ 0 & A' \end{pmatrix},$$

onde D' é uma matriz diagonal e a matriz A' tem na diagonal somente zeros. Neste caso é necessário fazer troca de linhas e colunas (na matriz correspondente à I só de colunas) ou, usando outras operações elementares, transformar a matriz A' a uma

matriz com pelo menos um elemento não nulo na diagonal. Por exemplo, no caso da forma quadrática ϕ, as operações $l_1 \leftarrow l_1 + l_2$ e $c_1 \leftarrow c_1 + c_2$ aplicadas à matriz

$$\begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

resultam na matriz

$$\begin{pmatrix} 2 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \end{pmatrix},$$

e depois das operações elementares $l_2 \leftarrow l_2 - \frac{1}{2}l_1$ e $c_2 \leftarrow c_2 - \frac{1}{2}c_1$ obtemos

$$\begin{pmatrix} 2 & 0 & 1 & -\frac{1}{2} \\ 0 & -\frac{1}{2} & 1 & \frac{1}{2} \end{pmatrix}.$$

O terceiro método que vamos utilizar baseia-se na construção da base ortonormal formada pelos vetores próprios da matriz A. A equação característica da matriz A é

$$0 = \det \begin{pmatrix} 1-\lambda & 0 & 2 \\ 0 & 1-\lambda & 1 \\ 2 & 1 & 4-\lambda \end{pmatrix} = (1-\lambda)(\lambda^2 - 5\lambda - 1).$$

Portanto, os valores próprios são $\lambda_1 = 1$, $\lambda_2 = \frac{5+\sqrt{29}}{2}$ e $\lambda_3 = \frac{5-\sqrt{29}}{2}$. Os respetivos vetores próprios normalizados são

$$v_1 = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ -2 \\ 0 \end{pmatrix}, \; v_2 = \frac{\sqrt{2}}{\sqrt{29+3\sqrt{29}}} \begin{pmatrix} 2 \\ 1 \\ \frac{3+\sqrt{29}}{2} \end{pmatrix} \quad \text{e} \quad v_3 = \frac{\sqrt{2}}{\sqrt{29-3\sqrt{29}}} \begin{pmatrix} 2 \\ 1 \\ \frac{3-\sqrt{29}}{2} \end{pmatrix}.$$

Na base formada pelos vetores próprios a matriz da forma quadrática é

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{5+\sqrt{29}}{2} & 0 \\ 0 & 0 & \frac{5-\sqrt{29}}{2} \end{pmatrix}.$$

As colunas da matriz P são os vetores v_1, v_2 e v_3. □

9.4. Lei de inércia

Como já sabemos, uma forma quadrática real

$$f(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i^j x^i x^j$$

pode ser diagonalizada aplicando uma mudança de variável $x = Cy$ com $\det(C) \neq 0$, vindo

$$f(y) = \sum_{k=1}^{n} b_k^k (y^k)^2.$$

Fazendo

$$y^k = \begin{cases} \frac{1}{\sqrt{|b_k^k|}} z^k, & \text{se } b_k^k \neq 0, \\ z^k, & \text{se } b_k^k = 0, \end{cases}$$

obtemos a forma

$$f(z) = \sum_{k=1}^{n} s_k^k (z^k)^2,$$

onde $s_k^k = \pm 1$ ou $s_k^k = 0$. Esta observação leva-nos à seguinte definição.

Definição 9.1. Forma canónica, característica e assinatura

Dizemos que uma forma quadrática está na sua forma *canónica* se está escrita como

$$f(z) = \sum_{k=1}^{n} s_k^k (z^k)^2,$$

onde $s_k^k = \pm 1$ ou $s_k^k = 0$.
Ao número de coeficientes não nulos damos o nome de *característica da forma quadrática* e chamamos *assinatura da forma quadrática* à diferença entre o número de coeficientes s_k^k positivos e negativos.

Para ilustrar a utilidade e a importância desta definição, consideremos um exemplo. Seja f a forma quadrática

$$f(x^1, x^2, x^3) = 2x^1x^2 - 6x^2x^3 + 2x^1x^3.$$

A aplicação linear

$$x^1 = \frac{1}{2}y^1 + \frac{1}{2}y^2 + 3y^3,$$
$$x^2 = \frac{1}{2}y^1 - \frac{1}{2}y^2 - y^3,$$
$$x^3 = y^3,$$

transforma $f(x^1, x^2, x^3)$ em

$$f(y^1, y^2, y^3) = \frac{1}{2}(y^1)^2 - \frac{1}{2}(y^2)^2 + 6(y^3)^2.$$

Então, a forma canónica de f é

$$f(z^1, z^2, z^3) = (z^1)^2 - (z^2)^2 + (z^3)^2.$$

Por outro lado, a aplicação linear

$$x^1 = y^1 + 3y^2 + 2y^3,$$
$$x^2 = y^1 - y^2 - 2y^3,$$
$$x^3 = y^2,$$

transforma $f(x^1, x^2, x^3)$ em

$$f(y^1, y^2, y^3) = 2(y^1)^2 + 6(y^2)^2 - 8(y^3)^2.$$

Neste caso, temos uma outra forma canónica

$$f(z^1, z^2, z^3) = (z^1)^2 + (z^2)^2 - (z^3)^2.$$

O que têm estas formas em comum? A resposta é dada pelo teorema conhecido como *lei de inércia*.

Teorema 9.3. (Lei de inércia)

O número de quadrados positivos e negativos numa forma canónica de uma forma quadrática real não depende da aplicação linear real invertível que é utilizada para transformar a forma quadrática nessa sua forma canónica.

Demonstração. Consideremos a matriz real simétrica A de ordem n e a matriz $C^{\mathrm{T}}AC$, tal que existe a matriz inversa C^{-1}. A dimensão do núcleo da matriz A coincide com a dimensão do núcleo da matriz $C^{\mathrm{T}}AC$. Portanto, as características das formas quadráticas correspondentes a estas matrizes são iguais.

Suponhamos que uma forma f é transformada em duas formas canónicas através de duas mudanças de variáveis distintas $y(x) = Px$ e $z(x) = Qx$ com matrizes P e Q invertíveis. O resultado destas transformações é

$$(y^1)^2 + \cdots + (y^k)^2 - (y^{k+1})^2 - \cdots - (y^r)^2 \tag{9.6}$$

e

$$(z^1)^2 + \cdots + (z^l)^2 - (z^{l+1})^2 - \cdots - (z^r)^2. \tag{9.7}$$

As respetivas assinaturas destas formas canónicas são $k - (r-k)$ e $l - (r-l)$. Suponhamos que $k < l$. Consideremos o sistema

$$y^1(x) = 0, \ldots, y^k(x) = 0, z^{l+1}(x) = 0, \ldots, z^r(x) = 0, \ldots, z^n(x) = 0. \tag{9.8}$$

Este sistema homogéneo tem $n - l + k$ equações e n variáveis $x^1, \ldots, x^n$. Portanto, o sistema tem uma solução não trivial $\alpha = (\alpha^1, \ldots, \alpha^n)$. Fazendo $x = \alpha$ obtemos $y = y(\alpha)$ e $z = z(\alpha)$. Então, de (9.6), (9.7) e (9.8) segue

$$-(y^{k+1})^2(\alpha) - \cdots - (y^r)^2(\alpha) = (z^1)^2(\alpha) + \cdots + (z^l)^2(\alpha).$$

Logo

$$(z^1)^2(\alpha) + \cdots + (z^l)^2(\alpha) = 0.$$

Juntamente com (9.8) isto implica que $z(\alpha) = 0$. Portanto, o sistema de n equações lineares

$$z(x) = Qx = 0,$$

com matriz invertível, tem uma solução não trivial $x = \alpha$, o que é uma contradição. Portanto, $k \geqslant l$. Analogamente mostramos que $l \geqslant k$. ■

A lei de inércia permite responder à pergunta quando uma forma quadrática pode, através de uma mudança de variáveis invertível, ser transformada numa outra.

Corolário 9.1

Sejam A e B duas matrizes simétricas. As características e assinaturas das formas quadráticas $f(x) = \langle x, Ax\rangle$ e $g(x) = \langle x, Bx\rangle$ coincidem, se e só se existe uma mudança de variáveis invertível definida pela matriz Q tal que $B = Q^{\mathsf{T}}AQ$.

Demonstração. De facto, existem matrizes invertíveis Q_A e Q_B tais que as matrizes

$$C_A = Q_A^{\mathsf{T}}AQ_A \quad \text{e} \quad C_B = Q_B^{\mathsf{T}}BQ_B$$

têm forma canónica. Como as características e assinaturas das formas quadráticas f e g coincidem, a diferença entre C_A e C_B é só na ordem em que aparecem os elementos diagonais. Uma delas pode ser obtida da outra através da reordenação das variáveis. Esta reordenação é uma mudança de variáveis invertível. ■

9.5. Formas quadráticas definidas positivas

Existe uma classe de formas quadráticas muito importante — as formas que tomam valores positivos quando $x \neq 0$. Estas formas aparecem em muitos problemas práticos e o estudo das suas propriedades é necessário para a resolução desses problemas.

Definição 9.2. Forma quadrática definida positiva

Dizemos que a forma quadrática

$$f(x) = \langle x, Ax\rangle$$

é *definida positiva* se

$$\langle x, Ax\rangle > 0$$

para todos os vetores $x \neq 0$.

Quando a forma quadrática é definida positiva, vamos dizer que a respetiva matriz A também é *definida positiva*.

Apresentamos agora dois critérios para verificar se uma matriz é definida positiva.

Teorema 9.4

Uma forma quadrática é definida positiva se e só se a sua diagonalização tem todos os coeficientes b_k^k positivos.

Demonstração. Seja a forma quadrática $f(x) = \langle x, Ax\rangle$. Pelo Teorema 9.1 existe uma matriz Q invertível tal que passando às variáveis $y = Q^{-1}x$ a matriz $B = Q^{\mathsf{T}}AQ$ é diagonal. Como

$$0 < \langle x, Ax\rangle = \langle Qy, AQy\rangle = \langle y, Q^{\mathsf{T}}AQy\rangle = \sum_{k=1}^{n} b_k^k (y^k)^2,$$

concluímos que f é definida positiva se e só se a sua diagonalização tem todos os coeficientes b_k^k positivos. ∎

O seguinte teorema dá-nos uma condição necessária e suficiente para uma forma quadrática ser definida positiva baseada no cálculo de determinantes.

Teorema 9.5. (Critério de Sylvester)

A forma quadrática

$$f(x) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i^j x^i x^j$$

é definida positiva se e só se os determinantes das matrizes com elementos (a_i^j), $i, j = \overline{1,k}$, $k = \overline{1,n}$, ou seja, os determinantes

$$a_1^1, \quad \det\begin{pmatrix} a_1^1 & a_1^2 \\ a_2^1 & a_2^2 \end{pmatrix}, \ldots, \det\begin{pmatrix} a_1^1 & a_1^2 & \cdots & a_1^n \\ a_2^1 & a_2^2 & \cdots & a_2^n \\ \vdots & \vdots & \ddots & \vdots \\ a_n^1 & a_n^2 & \cdots & a_n^n \end{pmatrix}, \tag{9.9}$$

são todos positivos.

Demonstração. O teorema é óbvio no caso de espaços de dimensão $n = 1$, pois qualquer forma quadrática tem a forma ax^2. Suponhamos que o teorema já está demonstrado para espaços de dimensão menor do que n. A forma quadrática

$$f(x) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i^j x^i x^j$$

pode ser representada como

$$f(x) = g(x^1, \ldots, x^{n-1}) + 2\sum_{i=1}^{n-1} a_i^n x^i x^n + a_n^n (x^n)^2, \tag{9.10}$$

onde

$$g(x^1, \ldots, x^{n-1}) = \sum_{i=1}^{n-1}\sum_{j=1}^{n-1} a_i^j x^i x^j.$$

Obviamente os determinantes

$$a_1^1, \quad \det\begin{pmatrix} a_1^1 & a_1^2 \\ a_2^1 & a_2^2 \end{pmatrix}, \ldots, \det\begin{pmatrix} a_1^1 & a_1^2 & \cdots & a_1^{n-1} \\ a_2^1 & a_2^2 & \cdots & a_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n-1}^1 & a_{n-1}^2 & \cdots & a_{n-1}^{n-1} \end{pmatrix} \tag{9.11}$$

da matriz da forma g coincidem com os respetivos primeiros $n-1$ determinantes da matriz A. Notemos ainda que

$$\det(Q^{\mathsf{T}} A Q) = \det(Q^{\mathsf{T}}) \det(A) \det(Q) = (\det Q)^2 \det(A). \tag{9.12}$$

Portanto, o sinal do determinante da matriz de uma forma quadrática mantém-se quando se faz uma mudança de variáveis.

Suponhamos que a forma quadrática f é definida positiva. Considerando só vetores com $x^n = 0$ chegamos à conclusão que a forma g também é definida positiva. Logo, todos os determinantes (9.11) da matriz da forma g e, portanto, os respetivos primeiros $n-1$ determinantes da matriz A são positivos. O determinante de A é positivo porque graças ao Teorema 9.1 existem variáveis com as quais a forma f tem uma matriz diagonal B com todos os elementos diagonais positivos. De (9.12) vemos que o sinal do determinante de A coincide com o sinal do determinante de B.

Suponhamos agora que todos os determinantes (9.9) da matriz A são positivos. Logo, são positivos todos os determinantes (9.11) da matriz da forma g e pela hipótese da indução esta forma é definida positiva. Pelo Teorema 9.1 existem variáveis y^k, $k = \overline{1, n-1}$, tais que a forma g tem uma matriz diagonal com todos os elementos diagonais b_k^k positivos. Fazendo $y^n = x^n$ e utilizando (9.10) vemos que nas variáveis y^k a forma f se escreve como

$$f(x) = \sum_{k=1}^{n-1} b_k^k (y^k)^2 + 2 \sum_{k=1}^{n-1} c_k^n y^k y^n + b_n^n (y^n)^2$$

(os valores exatos dos coeficientes c_k^n não têm importância). Notemos que

$$\sum_{k=1}^{n-1} b_k^k (y^k)^2 + 2 \sum_{k=1}^{n-1} c_k^n y^k y^n + b_n^n (y^n)^2 = \sum_{k=1}^{n-1} b_k^k \left(y^k + \frac{c_k^n}{b_k^k} y^n \right)^2 + c (y^n)^2,$$

onde

$$c = b_n^n - \sum_{k=1}^{n-1} \frac{(c_k^n)^2}{b_k^k}.$$

Portanto, em variáveis

$$z^k = y^k + \frac{c_k^n}{b_k^k} y^n, k = \overline{1, n-1}, \quad \text{e} \quad z^n = y^n,$$

representamos a forma f como

$$f(x) = \sum_{k}^{n-1} b_k^k (z^k)^2 + c (z^n)^2. \tag{9.13}$$

Mostremos que $c > 0$. O determinante da matriz da forma (9.13) é $b_1^1 b_2^2 \cdots c$. Este determinante é positivo porque a forma (9.13) é obtida como resultado de duas mudanças de variáveis e, segundo (9.12), tem o sinal do determinante da matriz A que é positivo. Como $b_k^k > 0$, $k = \overline{1, n-1}$, vemos que $c > 0$. De (9.13) obtemos que a forma f é definida positiva. ■

Na resolução de um problema muitas vezes é necessário definir uma estrutura Euclidiana no espaço, ou seja, definir um produto escalar. A escolha deste produto escalar depende de nós e deve ser adequada ao problema que estamos a resolver. Uma maneira de fazer isto é através de formas quadráticas definidas positivas. Mais precisamente, seja $f(x) = \langle x, Ax \rangle$ uma forma quadrática definida positiva. Então, usando a matriz A podemos definir um outro produto escalar no espaço segundo a fórmula

$$\langle x, y \rangle_A = \langle x, Ay \rangle.$$

É fácil verificar todas as propriedades do produto escalar.

Também é utilizada a seguinte terminologia.

Definição 9.3. Forma quadrática definida negativa

Dizemos que a forma quadrática

$$f(x) = \langle x, Ax \rangle$$

é *definida negativa* se

$$\langle x, Ax \rangle < 0$$

para todos os vetores $x \neq 0$.

Quando a forma quadrática é definida negativa, vamos dizer que a respetiva matriz A também é *definida negativa*.

Definição 9.4. Formas quadráticas semidefinidas

Dizemos que a forma quadrática

$$f(x) = \langle x, Ax \rangle$$

é *semidefinida positiva (negativa)* se

$$\langle x, Ax \rangle \geqslant 0 \ (\leqslant 0).$$

Quando a forma quadrática é semidefinida positiva (negativa), vamos dizer que a respetiva matriz A também é *semidefinida positiva (negativa)*.

9.6. Diagonalização simultânea de um par de formas quadráticas

Consideremos um par de formas quadráticas $f(x) = \langle x, Ax\rangle$ e $g(x) = \langle x, Bx\rangle$. Suponhamos que a forma f é definida positiva. Mostremos que existe uma mudança de variáveis que diagonaliza simultaneamente ambas as formas. Este facto é muito importante no estudo de pequenas oscilações de sistemas mecânicos e na sua demonstração as aplicações ortogonais têm um papel importante.

Teorema 9.6

Sejam $f(x) = \langle x, Ax\rangle$ e $g(x) = \langle x, Bx\rangle$ formas quadráticas das quais f é definida positiva. Então, existe uma mudança de variáveis que diagonaliza simultaneamente ambas as formas.

Demonstração. Seja P uma matriz que diagonaliza a matriz A, ou seja, tal que

$$P^{\mathrm{T}}AP = \begin{pmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & a_n \end{pmatrix},$$

$a_k > 0$, $k = \overline{1,n}$. A matriz P pode ser obtida, por exemplo, através do método de Lagrange (veja o Teorema 9.1). A composição da aplicação com matriz P e da aplicação com a matriz

$$S = \begin{pmatrix} \frac{1}{\sqrt{a_1}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{a_2}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{1}{\sqrt{a_n}} \end{pmatrix}$$

transforma f numa soma de quadrados de variáveis

$$f(y) = (y^1)^2 + \cdots + (y^n)^2,$$

isto é, em variáveis y introduzidas como $x = SPy$ temos a matriz da forma f igual à matriz identidade $P^{\mathrm{T}}S^{\mathrm{T}}ASP = I$. Nestas variáveis, g tem forma

$$g(y) = \langle y, Gy\rangle.$$

Agora consideremos uma aplicação ortogonal $y = Qz$ que diagonaliza $g(y)$

$$h(z) = \mu_1(z^1)^2 + \cdots + \mu_n(z^n)^2.$$

Aqui μ_k, $k = \overline{1,n}$, são os valores próprios da matriz G. Nas variáveis z^k, $k = \overline{1,n}$, a forma f continua a ser uma soma dos quadrados

$$f(z) = (z^1)^2 + \cdots + (z^n)^2.$$

■

A condição que uma das formas é definida positiva é importante. Por exemplo, não é possível diagonalizar simultaneamente as formas

$$(x^1)^2 - (x^2)^2 \quad \text{e} \quad 2x^1x^2.$$

De facto, suponhamos que existe uma matriz

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad \det\begin{pmatrix} a & b \\ c & d \end{pmatrix} \neq 0$$

tal que

$$\begin{pmatrix} a & c \\ b & d \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a^2 - c^2 & ab - cd \\ ab - cd & b^2 - d^2 \end{pmatrix} = \begin{pmatrix} p & 0 \\ 0 & q \end{pmatrix}$$

e

$$\begin{pmatrix} a & c \\ b & d \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 2ac & ad + bc \\ ad + bc & 2bd \end{pmatrix} = \begin{pmatrix} r & 0 \\ 0 & s \end{pmatrix}.$$

Então temos

$$ad - bc \neq 0, \tag{9.14}$$

$$ab - cd = 0 \quad \text{e} \tag{9.15}$$

$$ad + bc = 0. \tag{9.16}$$

Adicionando (9.14) e (9.16) vemos que $ad \neq 0$. Logo (9.16) implica que $bc \neq 0$. Multiplicando (9.15) por c e (9.16) por a e subtraindo obtemos $-a^2d = c^2d$. Como $d \neq 0$ temos $-a^2 = c^2$ e, portanto, $a = c = 0$, o que é uma contradição.

Exemplo 9.2. Vamos determinar a mudança de coordenadas que permite diagonalizar simultaneamente as formas quadráticas $f(x)$ e $g(x)$ com as matrizes

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 5 \end{pmatrix} \quad \text{e} \quad B = \begin{pmatrix} 4 & 8 \\ 8 & -4 \end{pmatrix},$$

respetivamente. A matriz A é definida positiva. Como

$$f(x) = \langle x, Ax\rangle = (x^1)^2 + 2x^1x^2 + 5(x^2)^2 = (x^1 + x^2)^2 + 4(x^2)^2,$$

fazendo

$$y^1 = x^1 + x^2 \quad \text{e} \quad y^2 = 2x^2,$$

obtemos

$$f(y) = (y^1)^2 + (y^2)^2.$$

Como

$$x^1 = y^1 - \frac{1}{2}y^2 \quad \text{e} \quad x^2 = \frac{1}{2}y^2,$$

temos

$$\begin{aligned} g(y) &= 4\left(y^1 - \frac{1}{2}y^2\right)^2 + 16\left(y^1 - \frac{1}{2}y^2\right)\frac{1}{2}y^2 - 4\frac{1}{4}(y^2)^2 \\ &= 4(y^1)^2 + 4y^1y^2 - 4(y^2)^2. \end{aligned}$$

A matriz desta forma quadrática é

$$\begin{pmatrix} 4 & 2 \\ 2 & -4 \end{pmatrix}.$$

A sua equação característica

$$\det\begin{pmatrix} 4-\lambda & 2 \\ 2 & -4-\lambda \end{pmatrix} = \lambda^2 - 20 = 0$$

tem raízes $\pm 2\sqrt{5}$. Os respetivos vetores próprios normalizados são

$$\frac{1}{\sqrt{10-4\sqrt{5}}}\begin{pmatrix} -1 \\ 2-\sqrt{5} \end{pmatrix} \quad \text{e} \quad \frac{1}{\sqrt{10+4\sqrt{5}}}\begin{pmatrix} -1 \\ 2+\sqrt{5} \end{pmatrix}.$$

Fazendo

$$\begin{aligned} y^1 &= -\frac{1}{\sqrt{10-4\sqrt{5}}}z^1 - \frac{1}{\sqrt{10+4\sqrt{5}}}z^2 \quad \text{e} \\ y^2 &= \frac{2-\sqrt{5}}{\sqrt{10-4\sqrt{5}}}z^1 + \frac{2+\sqrt{5}}{\sqrt{10+4\sqrt{5}}}z^2, \end{aligned}$$

obtemos

$$f(z) = (z^1)^2 + (z^2)^2 \quad \text{e} \quad g(z) = 2\sqrt{5}(z^1)^2 - 2\sqrt{5}(z^2)^2.$$

A respetiva mudança de variáveis é

$$\begin{aligned} z^1 &= -\frac{1}{\sqrt{10-4\sqrt{5}}}y^1 + \frac{2-\sqrt{5}}{\sqrt{10-4\sqrt{5}}}y^2 \\ &= -\frac{1}{\sqrt{10-4\sqrt{5}}}(x^1 + x^2) + 2\frac{2-\sqrt{5}}{\sqrt{10-4\sqrt{5}}}x^2 \quad \text{e} \\ z^2 &= -\frac{1}{\sqrt{10+4\sqrt{5}}}y^1 + \frac{2+\sqrt{5}}{\sqrt{10+4\sqrt{5}}}y^2 \\ &= -\frac{1}{\sqrt{10+4\sqrt{5}}}(x^1 + x^2) + 2\frac{2+\sqrt{5}}{\sqrt{10+4\sqrt{5}}}x^2. \end{aligned}$$

□

9.7. Formas quadráticas hermitianas

Nos espaços Euclidianos complexos também é relevante considerar formas quadráticas. Seja $\mathbb{X}$ um espaço Euclidiano complexo de dimensão n. Consideremos uma matriz complexa $A = (a_j^k)$. À função $f : \mathbb{X} \to \mathbb{C}$ definida por

$$f(x) = \sum_{j=1}^{n} \sum_{k=1}^{n} a_j^k x^j \bar{x}^k \tag{9.17}$$

damos o nome de *forma quadrática hermitiana*. De entre as formas hermitianas existe uma classe muito importante que encontramos em várias áreas da Ciência.

Definição 9.5. Forma quadrática hermitiana simétrica

Se a matriz A da forma (9.17) é simétrica no sentido $a_i^j = \bar{a}_j^i$, à respetiva forma damos o nome de *forma quadrática hermitiana simétrica*.

No que segue, vamos considerar só formas simétricas, por ser o caso mais interessante do ponto de vista das aplicações. Estas formas só tomam valores reais, pois

$$\overline{f(x)} = f(x).$$

Vamos agora demonstrar um teorema importante sobre diagonalização de formas quadráticas hermitianas simétricas.

Teorema 9.7

No espaço $\mathbb{C}^n$ existe uma base tal que a forma quadrática hermitiana simétrica escreve-se como

$$f(x) = \sum_{k=1}^{n} \lambda_k y^k \bar{y}^k = \sum_{k=1}^{n} \lambda_k |y^k|^2,$$

onde λ_k, $k = \overline{1,n}$, são coeficientes reais.

Demonstração. A demonstração é praticamente igual à demonstração do Teorema 9.1. Só são necessárias pequenas alterações, nomeadamente, em vez da igualdade (9.3) utilizamos a igualdade

$$g(x) = f(x) - \frac{1}{a_1^1}(a_1^1 x^1 + a_1^2 x^2 + \cdots + a_1^n x^n)\overline{(a_1^1 x^1 + a_1^2 x^2 + \cdots + a_1^n x^n)}.$$

Obviamente a forma g não depende da variável x^1. Em vez da transformação (9.5) utilizamos a transformação

$$x^1 = z^1 + z^2, \quad x^2 = z^1 + \mathrm{i} z^2,$$

que transforma a soma

$$a_1^2 x^1 \bar{x}^2 + \bar{a}_1^2 \bar{x}^1 x^2$$

em

$$(a_1^2 + \bar{a}_1^2)z^1\bar{z}^1 - \mathtt{i}(a_1^2 - \bar{a}_1^2)z^2\bar{z}^2 + \cdots,$$

onde pelo menos um dos coeficientes $(a_1^2 + \bar{a}_1^2)$ e $(a_1^2 - \bar{a}_1^2)$ é diferente de zero. ∎

9.8. Equações diferenciais: estabilidade*

Um dos tópicos mais importantes da teoria das equações diferenciais é a teoria de estabilidade. Nesta secção vamos ver como a teoria das formas quadráticas pode ser útil na análise dos problemas dessa teoria.

Consideremos a equação diferencial linear

$$\dot{x} = Ax, \tag{9.18}$$

onde A é uma matriz de dimensões $n \times n$. Esta equação admite a solução trivial $x(t) \equiv 0$. Ao ponto $x = 0$ chamamos um *ponto de equilíbrio*. Isto significa que o sistema físico descrito pela equação (9.18) pode permanecer neste ponto durante tempo infinito. O que acontece ao sistema quando o ponto inicial $x(0)$ é diferente de zero? O sistema permanece numa vizinhança do ponto de equilíbrio $x = 0$? Tende para este ponto ou afasta-se dele? Estas questões são muito importantes do ponto de vista prático e são objetos de estudo da teoria de estabilidade.

Em primeiro lugar, vamos ver como evolui uma forma quadrática ao longo de uma trajetória do sistema (9.18). Sejam $V(x) = \langle x, Vx\rangle$ uma forma quadrática e $x(t)$ uma solução da equação (9.18). Então,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}V(x(t)) &= \frac{\mathrm{d}}{\mathrm{d}t}\langle x(t), Vx(t)\rangle = \langle \dot{x}(t), Vx(t)\rangle + \langle x(t), V\dot{x}(t)\rangle \\ &= \langle Ax(t), Vx(t)\rangle + \langle x(t), VAx(t)\rangle = \langle x(t), (A^{\mathsf{T}}V + VA)x(t)\rangle.\end{aligned}$$

Para estudar a estabilidade do ponto de equilíbrio $x = 0$, vamos construir uma forma quadrática $V(x) = \langle x, Vx\rangle$ que verifica a igualdade

$$\langle x(t), (A^{\mathsf{T}}V + VA)x(t)\rangle = \lambda\langle x, Vx\rangle + \langle x, Wx\rangle \tag{9.19}$$

para todos os $x \in \mathbb{R}^n$. Aqui, W é uma matriz simétrica dada de dimensões $n \times n$. A equação (9.19) pode ser escrita como a equação matricial

$$A^{\mathsf{T}}V + VA = \lambda V + W.$$

Esta equação também pode ser considerada como um sistema linear com $\frac{n(n+1)}{2}$ incógnitas v_i^j, $i = \overline{1,n}$, $j = \overline{1,n}$, $i \geqslant j$ (os elementos da matriz simétrica V). Este sistema de equações lineares é possível e determinado para qualquer matriz simétrica W, sempre que o sistema linear

$$A^{\mathsf{T}}V + VA - \lambda V = 0 \tag{9.20}$$

só admite a solução trivial, ou seja, sempre que λ não é raiz da equação característica da aplicação linear

$$V \to A^{\mathrm{T}}V + VA. \tag{9.21}$$

O teorema seguinte, demonstrado por Lyapunov, descreve todas as raízes da equação característica desta aplicação.

Teorema 9.8

Todas as raízes da equação característica da aplicação (9.21) têm forma

$$\lambda = m_1\lambda_1 + \cdots + m_n\lambda_n, \tag{9.22}$$

onde λ_i, $i = \overline{1,n}$, são os valores próprios da aplicação A e $m_i \in \{0,1,2\}$, $i = \overline{1,n}$, são números que verificam a condição

$$m_1 + \cdots + m_n = 2. \tag{9.23}$$

Demonstração. Sejam v_i, $i = \overline{1,n}$, os vetores próprios da matriz A^{T} que correspondem aos valores próprios λ_i, $i = \overline{1,n}$. Consideremos a forma quadrática

$$V(x) = \langle v_1, x\rangle^{m_1} \cdots \langle v_n, x\rangle^{m_n},$$

onde $m_i \in \{0,1,2\}$, $i = \overline{1,n}$, são números que verificam a condição (9.23). Como se verifica a condição (9.23) temos

$$V(x) = \langle v_i, x\rangle\langle v_j, x\rangle. \tag{9.24}$$

Os índices i e j, que podem ser iguais, correspondem aos números m_i e m_j diferentes de zero. Os outros números m_k, $k \neq i$, $k \neq j$ são iguais a zero. Logo, atendendo a (9.24), temos

$$\begin{aligned}\langle x, (A^{\mathrm{T}}V + VA)x\rangle &= \langle Ax, Vx\rangle + \langle x, VAx\rangle = \langle v_i, Ax\rangle\langle v_j, x\rangle + \langle v_i, x\rangle\langle v_j, Ax\rangle \\ &= \langle A^{\mathrm{T}}v_i, x\rangle\langle v_j, x\rangle + \langle v_i, x\rangle\langle A^{\mathrm{T}}v_j, x\rangle = (\lambda_i + \lambda_j)\langle v_i, x\rangle\langle v_j, x\rangle \\ &= (\lambda_i + \lambda_j)V(x) = (m_1\lambda_1 + \cdots + m_n\lambda_n)V(x).\end{aligned}$$

Portanto, a matriz V correspondente à forma quadrática $V(x)$ é uma solução não trivial da equação (9.20) em que λ verifica (9.22). Assim, cada $\lambda = m_1\lambda_1 + \cdots + m_n\lambda_n$, onde $m_i \in \{0,1,2\}$, $i = \overline{1,n}$, são números que verificam a condição (9.23), é um valor próprio da aplicação (9.21).

Resta demonstrar que cada valor próprio da aplicação (9.21) pode ser representado nesta forma. Se todos os números que têm forma (9.22) com $m_i \in \{0,1,2\}$, $i = \overline{1,n}$, que verificam a condição (9.23), são diferentes, temos $\frac{n(n+1)}{2}$ valores próprios diferentes da aplicação (9.21) e o teorema está demonstrado. No caso geral, aproximamos a matriz A por uma sucessão de matrizes que têm valores próprios (9.22) com $m_i \in \{0,1,2\}$, $i = \overline{1,n}$, que verificam a condição (9.23), diferentes, e depois passamos ao limite. ■

Uma das consequências mais importantes do teorema anterior é o seguinte resultado.

Teorema 9.9

Seja A uma matriz de dimensões $n \times n$ tal que todos os seus valores próprios $\lambda_i = \alpha_i + \mathtt{i}\beta_i$, $i = \overline{1,n}$, têm parte real α_i negativa. Então, existe uma forma quadrática definida positiva $V(x) = \langle x, Vx\rangle$ que verifica

$$\langle x, (A^{\mathsf{T}}V + VA)x\rangle \leqslant -\frac{1}{2}\langle x, x\rangle, \tag{9.25}$$

para todos $x \in \mathbb{R}^n$.

Demonstração. Se os valores próprios $\lambda_i = \alpha_i + \mathtt{i}\beta_i$, $i = \overline{1,n}$, com $\alpha_i < 0$, da matriz A são todos diferentes, a solução geral da equação (9.18) tem a forma

$$x(t) = \sum_{i=1}^{n} c_i v_i \exp((\alpha_i + \mathtt{i}\beta_i)t) = \sum_{i=1}^{n} c_i v_i \exp(\alpha_i t)(\cos(\beta_i t) + \mathtt{i}\,\mathrm{sen}(\beta_i t)),$$

onde v_i são os vetores próprios correspondentes aos valores próprios λ_i e c_i são constantes arbitrárias. Portanto, $x(t)$ tende para zero quando t tende para infinito. Do Teorema 9.8 vemos que zero não é um valor próprio da aplicação (9.21). De facto, a parte real da soma (9.22) é um número negativo. Logo, existe uma forma quadrática $V_0(x) = \langle x, V_0 x\rangle$ que verifica a equação

$$A^{\mathsf{T}}V_0 + V_0 A = -I.$$

Suponhamos que existe $x_0 \in \mathbb{R}^n$ tal que $V_0(x_0) < 0$. Obviamente $x_0 \neq 0$. Consideremos a solução $x(t)$ da equação diferencial (9.18) que verifica a condição inicial $x(0) = x_0$. Então temos

$$\frac{\mathrm{d}}{\mathrm{d}t}V_0(x(t)) = -|x(t)|^2 \leqslant 0,$$

pelo que $V(x(t)) \leqslant V(x_0) < 0$, isto é, $x(t)$ não tende para zero, o que é uma contradição. Portanto, $V_0(x) \geqslant 0$ qualquer que seja $x \in \mathbb{R}^n$.

Agora consideremos o caso quando os valores próprios $\lambda_i = \alpha_i + \mathtt{i}\beta_i$, $\alpha_i < 0$, da matriz A não são todos diferentes. Existe uma sucessão de matrizes A_k com valores próprios todos diferentes que converge para a matriz A. Para k suficientemente grandes, as partes reais dos valores próprios das matrizes A_k são todas negativas. Logo, para cada uma destas matrizes existe uma matriz semidefinida positiva V_k, que é a solução da equação

$$A_k^{\mathsf{T}}V_k + V_k A_k = -I.$$

Pelo Corolário 6.6, a sucessão das matrizes V_k converge para uma matriz $\hat{V}$ que é semidefinida positiva e que verifica a equação

$$A^{\mathsf{T}}\hat{V} + \hat{V}A = -I.$$

Seja $\epsilon > 0$. Consideremos a matriz $V = \hat{V} + \epsilon I$. Obviamente esta matriz é definida positiva. Para esta matriz temos

$$A^{\mathsf{T}}V + VA = A^{\mathsf{T}}(\hat{V} + \epsilon I) + (\hat{V} + \epsilon I)A = A^{\mathsf{T}}\hat{V} + \hat{V}A + \epsilon(A^{\mathsf{T}} + A) = -I + \epsilon(A^{\mathsf{T}} + A).$$

Logo, temos

$$\langle x, (A^{\mathsf{T}}V + VA)x\rangle = -|x|^2 + \epsilon\langle x, (A^{\mathsf{T}} + A)x\rangle \leqslant -\frac{1}{2}\langle x, x\rangle,$$

sempre que ϵ é suficientemente pequeno. ■

Vamos agora estudar a estabilidade de pontos de equilíbrio de sistemas de equações diferenciais ordinárias não lineares. Para tal, seja $f : \mathbb{R}^n \to \mathbb{R}^n$ uma função e a equação diferencial

$$\dot{x} = f(x), \tag{9.26}$$

que tem ponto de equilíbrio $x = 0$, isto é, $f(0) = 0$. Suponhamos que $f(x) = Ax + g(x)$, onde A é uma aplicação linear cujos valores próprios têm partes reais negativas e $g(x)$ é uma função que verifica numa vizinhança de zero, $U = \{x \in \mathbb{R}^n : |x| < \rho\}$, a condição $|g(x)| \leqslant \delta(x)|x|$, onde $\lim_{x\to 0} \delta(x) = 0$.

O resultado da teoria de estabilidade mais importante do ponto de vista prático é o seguinte teorema.

Teorema 9.10

As soluções da equação (9.26) tendem para zero, sempre que o ponto inicial está suficientemente próximo de zero.

Demonstração. Pelo Teorema 9.9 existe uma forma quadrática definida positiva $V(x) = \langle x, Vx\rangle$ que verifica a condição (9.25). Existe $\bar{\rho} > 0$ tal que $\delta(x) < \frac{1}{8|V|}$, sempre que $|x| < \bar{\rho}$. (Aqui, $|V|$ é a norma da aplicação V definida na Secção 5.10.) Consideremos uma solução $x(t)$ da equção (9.26) com ponto inicial x_0 próximo de zero. Então temos

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}V(x(t)) &= \langle \dot{x}(t), Vx(t)\rangle + \langle x(t), V\dot{x}(t)\rangle \\
&= \langle Ax(t) + g(x(t)), Vx(t)\rangle + \langle x(t), V(Ax(t) + g(x(t)))\rangle \\
&= \langle Ax(t), Vx(t)\rangle + \langle x(t), VAx(t)\rangle + \langle g(x(t)), Vx(t)\rangle + \langle x(t), Vg(x(t))\rangle \\
&\leqslant -\frac{1}{2}|x(t)|^2 + 2|V|\delta(x(t))|x(t)|^2 = -\frac{1}{2}(1 - 4|V|\delta(x(t)))|x(t)|^2 \\
&\leqslant -\frac{1}{4}|x(t)|^2,
\end{aligned}$$

sempre que $|x(t)| < \mu = \min\{\rho, \bar{\rho}\}$. Seja $\nu = \min\{V(x) : |x| = \mu\}$. Portanto, a função $V(x(t))$ é monótona decrescente sempre que o ponto inicial x_0 pertence à vizinhança de zero $\{x : V(x) < \nu\}$. A solução $x(t)$ da equação (9.26) nunca sai desta vizinhança. Suponhamos que $V(x(t))$ tende para um valor $\bar{V} > 0$ quando t tende para infinito. Isto significa que a derivada da função $V(x(t))$ é sempre negativa pois $\frac{\mathrm{d}}{\mathrm{d}t}V(x(t)) \leqslant -\frac{1}{4}\mu^2$, o que é uma contradição. ■

Vamos ilustrar o teorema anterior através de um exemplo.

Exemplo 9.3. Consideremos o pêndulo (veja a Fig. 9.1), isto é, um ponto de massa m preso a um fio inextensível de comprimento L que oscila em torno de um ponto fixo num meio denso.

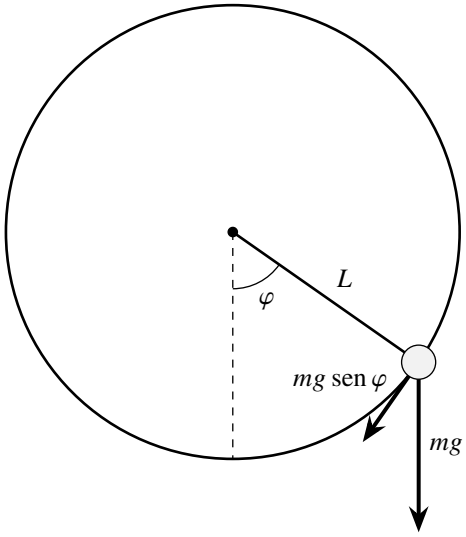


Figura 9.1
Pêndulo.

A velocidade com que o ponto de massa m se move ao longo da circunferência de raio L é $L\dot{\varphi}$. A força aplicada ao ponto ao longo da tangente à circunferência é $mg \operatorname{sen} \varphi$. Portanto, pela segunda lei de Newton o movimento do pêndulo descreve-se pela equação diferencial

$$mL\ddot{\varphi} = -\sigma L\dot{\varphi} - mg \operatorname{sen} \varphi.$$

Aqui, $-\sigma L\dot{\varphi}$, onde $\sigma > 0$, é a força de resistência do meio denso. Podemos escrever esta equação na forma de um sistema de equações diferenciais equivalente

$$\begin{cases} \dot{\varphi} = \omega, \\ \dot{\omega} = -2\alpha\dot{\varphi} - \beta^2 \operatorname{sen} \varphi, \end{cases}$$

onde $\alpha = \frac{\sigma}{2m}$ e $\beta^2 = \frac{g}{L}$. Pelo Teorema do valor médio de Lagrange existe $\psi \in [-\varphi, \varphi]$ tal que

$$|\varphi - \operatorname{sen} \varphi| = |(1 - \cos \psi)\varphi| \leqslant (1 - \cos \varphi)|\varphi|.$$

Como $\lim_{\varphi \to 0}(1 - \cos \varphi) = 0$, vemos que o sistema

$$\begin{pmatrix} \dot{\varphi} \\ \dot{\omega} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\beta^2 & -2\alpha \end{pmatrix} \begin{pmatrix} \varphi \\ \omega \end{pmatrix} + \begin{pmatrix} 0 \\ \beta^2(\varphi - \operatorname{sen} \varphi) \end{pmatrix}$$

verifica as condições do Teorema 9.10. Portanto, as soluções do sistema com valores iniciais $\varphi(0)$ e $\omega(0)$ suficientemente pequenos tendem para o ponto de equilíbrio $\varphi = 0$ e $\omega = 0$.

□

10. Cónicas

10.1. Introdução

As cónicas podem ser definidas como conjuntos de pontos em $\mathbb{R}^2$ cujas coordenadas satisfazem uma equação de segundo grau. Dentro da classe das cónicas existem três tipos de cónicas relevantes: elipses, hipérboles e parábolas. Estas cónicas têm aplicação, por exemplo, na Física, onde correspondem a órbitas e a trajetórias de corpos celestes e de partículas, e na Arquitetura, onde servem de modelos para estruturas (pontes, por exemplo). No capítulo seguinte vamos justificar a utilização do termo cónicas para designar estas curvas.

Um objetivo deste capítulo é a dedução das propriedades geométricas das cónicas através da análise das equações de segundo grau. Um outro objetivo é a classificação das cónicas através da transformação da equação de segundo grau numa expressão mais simples através de mudanças de variáveis. Para isso recorremos à teoria das formas quadráticas do capítulo anterior.

10.2. Cónicas relevantes

Vamos agora definir as cónicas relevantes.

Definição 10.1. Elipse, circunferência

Sejam $a, b, \rho > 0$. Chamamos *elipse* à cónica cuja equação é

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,\ a \neq b,$$

e *circunferência* à cónica cuja equação é

$$x^2 + y^2 = \rho^2.$$

Exemplo 10.1. A cónica $\frac{x^2}{9} + \frac{y^2}{4} = 1$ é uma elipse (veja a Fig. 10.1).

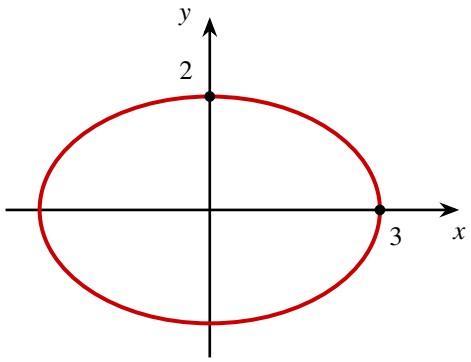


Figura 10.1
Elipse $\frac{x^2}{9} + \frac{y^2}{4} = 1$.

□

Exemplo 10.2. A cónica $x^2 + y^2 = 2$ é uma circunferência (veja a Fig. 10.2).

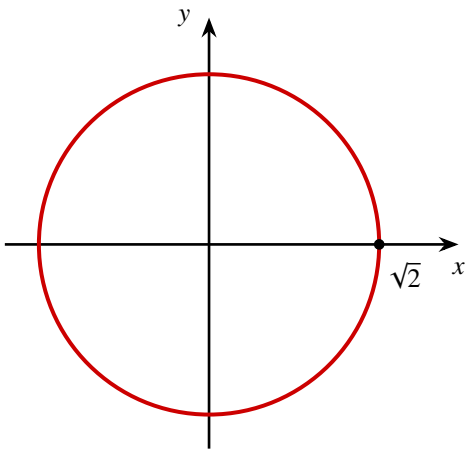


Figura 10.2
Circunferência $x^2 + y^2 = 2$.

□

Definição 10.2. Hipérbole

Sejam $a, b > 0$. Chamamos *hipérbole* à cónica cuja equação é

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1.$$

Exemplo 10.3. A cónica $x^2 - y^2 = 1$ é uma hipérbole (veja a Fig. 10.3).

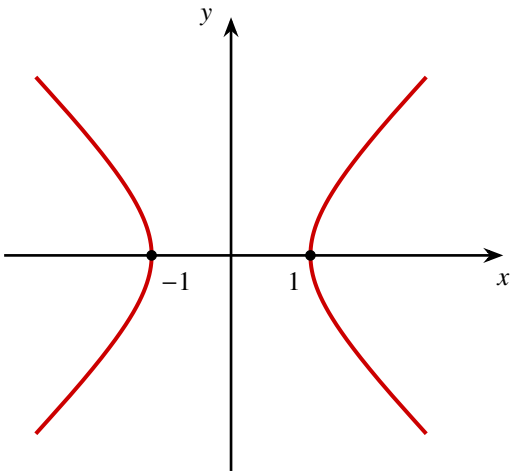


Figura 10.3
Hipérbole $x^2 - y^2 = 1$.

□

Definição 10.3. Parábola

Seja $p > 0$. Chamamos *parábola* à cónica cuja equação é

$$y^2 = 2px.$$

Exemplo 10.4. A cónica $y^2 = 2x$, que já está na forma reduzida (com $p = 1$ na equação $y^2 = 2px$), é uma parábola (veja a Fig. 10.4).

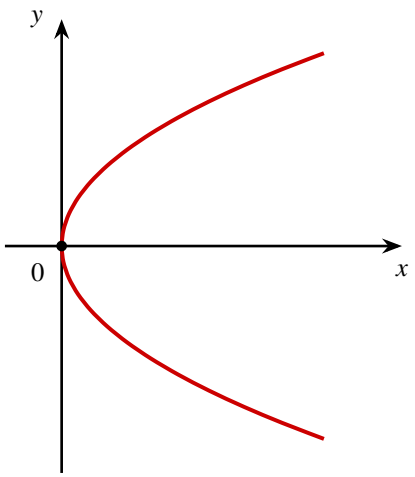


Figura 10.4
Parábola $y^2 = 2x$.

□

O facto de exigirmos $a, b, \rho > 0$ deve-se a um certo significado geométrico destas constantes, que será esclarecido posteriormente.

Nas três secções seguintes vamos estudar algumas das propriedades mais importantes das cónicas não degeneradas na forma reduzida.

10.3. Elipses

Consideremos uma elipse

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \tag{10.1}$$

onde $a \geqslant b$. Desta equação vemos que uma elipse está contida no retângulo gerado pelas retas paralelas aos eixos de coordenadas e que passam pelos pontos $(a, 0)$, $(-a, 0)$, $(0, b)$ e $(0, -b)$. A a e b damos os nomes *semieixo maior* e *semieixo menor*, respetivamente. As elipses são simétricas em relação aos eixos de coordenadas e ao ponto $(0, 0)$, sendo possível considerá-las como resultado de achatamento da circunferência de centro $(0, 0)$ e raio a, em que todas as ordenadas se multiplicam por $\frac{b}{a}$.

Seja $c^2 = a^2 - b^2$. Os *focos* de uma elipse são os pontos F_1 e F_2 com coordenadas $(c, 0)$ e $(-c, 0)$ (veja a Fig. 10.5). No caso de uma circunferência, $a = b$, os focos coincidem com o ponto zero. Ao valor $e = \frac{c}{a}$ damos o nome de *excentricidade*. Temos sempre $e < 1$, vindo $e = 0$ no caso de circunferências.

A elipse é uma curva com propriedades muito interessantes e importantes do ponto de vista de aplicações. Vamos estabelecer estas propriedades nos seguintes teoremas.

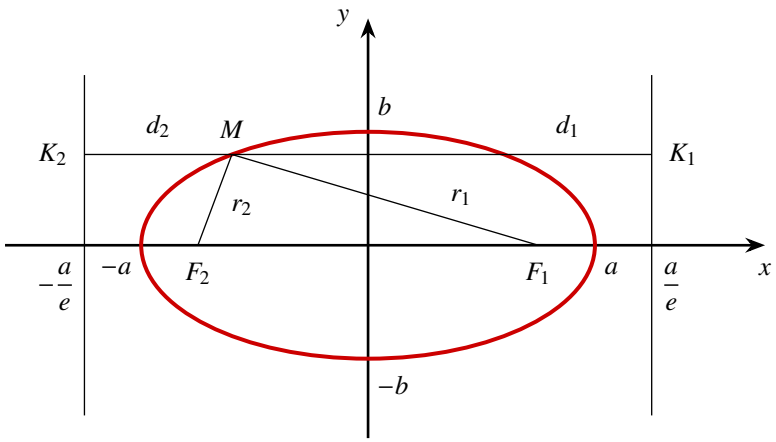


Figura 10.5
Elipse.

Teorema 10.1

As distâncias entre um ponto M da elipse com coordenadas (x, y) e os seus focos são

$$r_1 = |F_1M| = a - ex \quad \text{e} \quad r_2 = |F_2M| = a + ex$$

(veja a Fig. 10.5).

Demonstração. Como $c^2 = a^2 - b^2$, então

$$r_1 = \sqrt{(x-c)^2 + y^2} = \sqrt{x^2 - 2cx + c^2 + b^2 - \frac{b^2x^2}{a^2}} = \sqrt{a^2 - 2cx + \frac{c^2x^2}{a^2}}$$

$$= \sqrt{\left(a - \frac{cx}{a}\right)^2} = \sqrt{(a-ex)^2}.$$

Como $e < 1$ e $x \leqslant a$, temos $a - ex > 0$. Isto demonstra a primeira igualdade. Analogamente provamos a segunda igualdade. ■

Teorema 10.2

Um ponto pertence a uma elipse se e só se a soma das distâncias entre o ponto e os focos é igual a $2a$.

Demonstração. Se o ponto pertence à elipse, do teorema anterior temos $r_1 + r_2 = 2a$.

Agora suponhamos que $r_1 + r_2 = 2a$. Então temos

$$\sqrt{(x-c)^2 + y^2} = 2a - \sqrt{(x+c)^2 + y^2}.$$

Esta igualdade elevada ao quadrado dá

$$xc + a^2 = a\sqrt{(x+c)^2 + y^2}.$$

O quadrado desta igualdade é a equação da elipse (10.1). ■

Este teorema explica como efetivamente como desenhar uma elipse. Fixando as pontas de uma corda de comprimento $2a$ nos dois focos, esticando a corda com um lápis e mantendo a corda esticada, desenhamos uma curva que vai dar metade da elipse.

As *diretrizes* da elipse são as retas definidas por

$$x = \frac{a}{e} \quad \text{e} \quad x = -\frac{a}{e}.$$

As diretrizes permitem ligar a posição de um ponto numa elipse à excentricidade.

Designemos por $d_1 = |K_1M|$ e $d_2 = |K_2M|$ as distâncias entre o ponto M da elipse e as diretrizes correspondentes.

Teorema 10.3

Um ponto pertence a uma elipse se e só se $\frac{r_1}{d_1} = e$ ou $\frac{r_2}{d_2} = e$ (veja a Fig. 10.5).

Demonstração. Demonstremos o teorema para o foco F_2 que tem coordenadas $(-c, 0)$. Seja M um ponto da elipse com coordenadas (x, y), então (veja o Teorema 10.1)

$$d_2 = x + \frac{a}{e} = \frac{1}{e}(ex + a) = \frac{|F_2M|}{e} = \frac{r_2}{e}.$$

Agora suponhamos que o ponto verifica a condição $\frac{r_2}{d_2} = e$, ou seja,

$$\frac{\sqrt{(x+c)^2 + y^2}}{x + \frac{a}{e}} = e.$$

Como $e = \frac{c}{a}$, o quadrado desta igualdade é a equação da elipse. Para o foco F_1 o teorema demonstra-se analogamente. ■

Consideremos a equação da elipse. Esta equação define y como função de x:

$$\frac{x^2}{a^2} + \frac{y^2(x)}{b^2} = 1.$$

Derivando esta igualdade obtemos

$$\frac{2x}{a^2} + \frac{2yy'}{b^2} = 0.$$

Seja (x_0, y_0) um ponto da elipse com $y_0 \neq 0$. Então temos

$$y'(x_0) = -\frac{b^2 x_0}{a^2 y_0}.$$

Portanto, a equação da reta tangente à elipse no ponto (x_0, y_0) é

$$y - y_0 = -\frac{b^2 x_0}{a^2 y_0}(x - x_0).$$

Lema 10.1

Consideremos a Fig. 10.6. Suponhamos que

$$\frac{|AB|}{|BC|} = \frac{|AD|}{|CD|}. \tag{10.2}$$

Então temos

$$\widehat{CBD} = \widehat{DBE}.$$

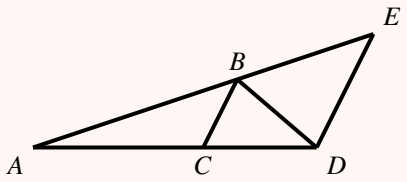


Figura 10.6
Ilustração do Lema 10.1.

Demonstração. Na reta que atravessa os pontos A e B, é possível escolhermos um ponto E por forma a que os segmentos BC e ED sejam paralelos. Então, $\widehat{CBD} = \widehat{BDE}$ e basta demonstrar que $|BE| = |ED|$, pois isto implica que o triângulo BDE é equilátero e portanto $\widehat{CBD} = \widehat{BDE} = \widehat{DBE}$. Os triângulos ABC e AED são semelhantes. Portanto, temos

$$\frac{|AB| + |BE|}{|AB|} = \frac{|AD|}{|AD| - |CD|} = \frac{|ED|}{|BC|}. \tag{10.3}$$

De (10.2) temos

$$\frac{|AD|}{|AD| - |CD|} = \frac{|AB|}{|AB| - |BC|}.$$

Então, de (10.3) obtemos

$$1 + \frac{|BE|}{|AB|} = \frac{|AB|}{|AB| - |BC|},$$

donde

$$|BE| = \frac{|AB||BC|}{|AB| - |BC|}. \tag{10.4}$$

Novamente, utilizando (10.3), temos

$$\frac{|ED|}{|BC|} = \frac{|AD|}{|AD| - |CD|} = \frac{|AB|}{|AB| - |BC|}.$$

Daqui obtemos

$$|ED| = \frac{|AB||BC|}{|AB| - |BC|}.$$

Comparando isto com (10.4) temos o resultado pretendido. ∎

Teorema 10.4

Sejam F_1 e F_2 os focos de uma elipse e M um ponto da elipse com coordenadas (x_0, y_0), que é diferente dos pontos $(\pm a, 0)$. Então a tangente ML é bissetriz do ângulo adjacente suplementar ao ângulo $\widehat{F_1MF_2}$ cuja amplitude é igual a $\pi - \widehat{F_1MF_2}$ (veja a Fig. 10.7).

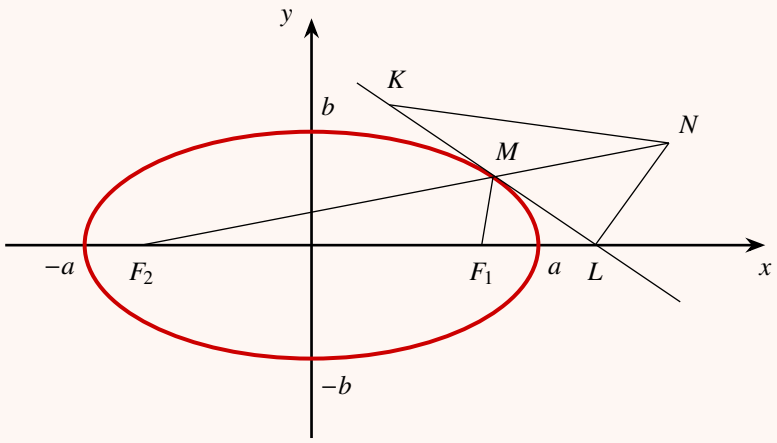


Figura 10.7 Ilustração do Teorema 10.4.

Demonstração. As coordenadas do ponto L são $(\frac{a^2}{x_0}, 0)$. A desigualdade $|\frac{a^2}{x_0}| > a$ implica que L não pertence ao segmento F_1F_2. Como

$$|F_1L| = |\frac{a^2}{x_0} - c| \quad \text{e} \quad |F_2L| = |\frac{a^2}{x_0} + c|,$$

utilizando o Teorema 10.1, temos

$$\frac{|F_1L|}{|F_2L|} = \frac{|\frac{a^2}{x_0} - c|}{|\frac{a^2}{x_0} + c|} = \frac{|a - ex_0|}{|a + ex_0|} = \frac{|MF_1|}{|MF_2|}.$$

Aplicando o Lema 10.1 temos o que a tangente ML é bissetriz do ângulo adjacente suplementar ao ângulo $\widehat{F_1MF_2}$, cuja amplitude é igual a $\pi - \widehat{F_1MF_2} = \pi - 2\widehat{F_1ML}$. ■

Deste teorema vemos que os ângulos $\widehat{F_2MK} = \widehat{LMN}$ e $\widehat{F_1ML}$ são iguais. Por exemplo, isto significa o seguinte. Consideremos uma fonte de luz colocada no foco F_2. Então, todos os raios emitidos por esta fonte de luz, depois da reflexão na elipse, atravessam o foco F_1.

10.4. Hipérboles

Consideremos uma hipérbole

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1. \tag{10.5}$$

Desta equação concluímos que $|x| \geqslant a$ para todos os pontos da hipérbole. Os números a e b são, respetivamente, *semieixo real* e *semieixo imaginário* da hipérbole. Uma hipérbole é simétrica em relação aos eixos de coordenadas e ao ponto $(0, 0)$.

Para estudar a forma da hipérbole, consideremos a sua interseção com a reta $y = kx$, onde $k \in \mathbb{R}$. As abcissas dos pontos de interseção encontramos da equação

$$\frac{x^2}{a^2} - \frac{k^2x^2}{b^2} = 1.$$

Se $D = b^2 - a^2k^2 > 0$, temos

$$x = \pm\frac{ab}{\sqrt{D}}.$$

Portanto, há dois pontos de interseção:

$$\left(\frac{ab}{\sqrt{D}}, \frac{abk}{\sqrt{D}}\right) \quad \text{e} \quad \left(-\frac{ab}{\sqrt{D}}, -\frac{abk}{\sqrt{D}}\right).$$

Estes pontos são simétricos em relação ao ponto $(0,0)$. Basta estudar o comportamento de um deles, por exemplo, do primeiro. Quando $k = 0$ a abcissa do ponto de interseção é a. Quando k cresce e se aproxima do valor $\frac{b}{a}$, a abcissa do ponto de interseção tende para infinito. A reta $y = \frac{b}{a}x$ não tem pontos de interseção com a hipérbole. As retas com $k > \frac{b}{a}$ também não intersetam a hipérbole. O mesmo podemos dizer sobre a reta $y = -\frac{b}{a}x$. Ela não interseta a hipérbole e separa as retas que intersectam a hipérbole das que não a intersectam. Às retas $y = \pm\frac{b}{a}x$ damos o nome de *assíntotas* da hipérbole. Destes raciocínios podemos concluir que a hipérbole tem a forma apresentada na Fig. 10.8. Ela consiste de duas partes conhecidas como *ramos* da hipérbole.

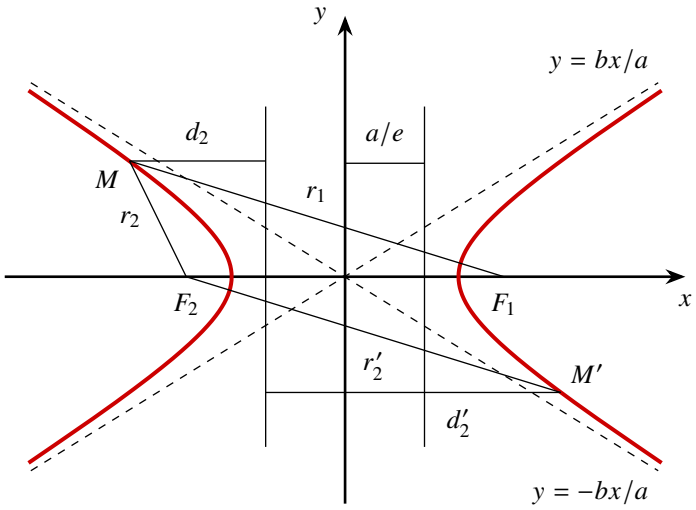


Figura 10.8
Hipérbole.

Podemos escrever as equações das assíntotas como

$$\frac{bx - ay}{\sqrt{a^2 + b^2}} = 0 \quad \text{e} \quad \frac{bx + ay}{\sqrt{a^2 + b^2}} = 0,$$

ou, usando o produto vetorial, como

$$\left[\frac{(\pm a, b)}{\sqrt{a^2 + b^2}}, (x, y)\right] = 0.$$

Esta observação é importante para a demonstração do seguinte lema.

Teorema 10.5

O produto das distâncias entre um ponto da hipérbole e as assíntotas é a constante $\frac{a^2b^2}{a^2+b^2}$.

Demonstração. A distância entre a reta gerada pelo vetor $\frac{(b,\pm a)}{\sqrt{a^2+b^2}}$, que tem norma um, e o ponto com coordenadas (x,y) é igual ao produto da norma do vetor (x,y) e do seno do ângulo entre os vetores $\frac{(b,\pm a)}{\sqrt{a^2+b^2}}$ e (x,y), ou seja, é igual à norma do produto vetorial

$$\left[\frac{(b,\pm a)}{\sqrt{a^2+b^2}},(x,y)\right].$$

Portanto, as distâncias entre o ponto M com coordenadas (x,y) que pertence à hipérbole e as assíntotas são

$$h_1=\frac{|bx-ay|}{\sqrt{a^2+b^2}} \quad \text{e} \quad h_2=\frac{|bx+ay|}{\sqrt{a^2+b^2}}.$$

Daqui temos

$$h_1h_2=\frac{|b^2x^2-a^2y^2|}{a^2+b^2}=\frac{a^2b^2}{a^2+b^2}.$$

■

Deste lema deduzimos o seguinte resultado muito importante sobre as assíntotas de uma hipérbole.

Lema 10.2

Se um ponto se move ao longo da hipérbole e o módulo da sua abcissa tende para infinito, então a distância entre o ponto e uma das assíntotas tende para zero.

Demonstração. De facto, uma das distâncias tende para o infinito. Portanto, a outra deve ter como limite zero.

■

Seja $c^2=a^2+b^2$. Os *focos* da hipérbole são os pontos F_1 e F_2 com coordenadas $(c,0)$ e $(-c,0)$, respetivamente. Ao valor $e=\frac{c}{a}$ damos o nome de *excentricidade*. Obviamente $e>1$.

Teorema 10.6

As distâncias entre um ponto M da hipérbole com coordenadas (x,y) e os seus focos são

$$r_1=|F_1M|=|a-ex| \quad \text{e} \quad r_2=|F_2M|=|a+ex|$$

(veja a Fig. 10.8).

Demonstração. Como $c^2 = a^2 + b^2$, é fácil ver que

$$r_1 = \sqrt{(x-c)^2 + y^2} = \sqrt{x^2 - 2cx + c^2 - b^2 + \frac{b^2x^2}{a^2}} = \sqrt{a^2 - 2cx + \frac{c^2x^2}{a^2}}$$
$$= \sqrt{\left(a - \frac{cx}{a}\right)^2} = |a - ex|.$$

Isto demonstra a primeira igualdade. Demonstramos a segunda analogamente. ■

Tal como no caso das elipses, podemos caracterizar os pontos de uma hipérbole através das suas distâncias aos focos.

Teorema 10.7

Um ponto pertence à hipérbole se e só se

$$|r_1 - r_2| = 2a. \tag{10.6}$$

Demonstração. Para $x \geqslant a$ (ramo direito da hipérbole) temos

$$r_1 = ex - a \quad \text{e} \quad r_2 = ex + a,$$

e para $x \leqslant -a$ (ramo esquerdo da hipérbole) temos

$$r_1 = a - ex \quad \text{e} \quad r_2 = -ex - a.$$

Portanto, temos $r_2 - r_1 = 2a$ para o ramo direito e $r_1 - r_2 = 2a$ para o ramo esquerdo. Em ambos os casos verifica-se a igualdade (10.6).

Agora suponhamos que se verifica a igualdade (10.6), ou, na forma equivalente

$$\pm\sqrt{(x-c)^2 + y^2} = 2a \pm \sqrt{(x+c)^2 + y^2}.$$

O quadrado desta igualdade é

$$xc + a^2 = \pm a\sqrt{(x+c)^2 + y^2}.$$

O quadrado desta igualdade é a equação da hipérbole. ■

As *diretrizes* da hipérbole são as retas

$$x = \pm\frac{a}{e}.$$

A diretriz e o foco que estão de um lado do eixo das ordenadas são correspondentes.

Designemos por d_1 e d_2 as distâncias entre o ponto M da hipérbole e as diretrizes correspondentes.

Teorema 10.8

Um ponto pertence a uma hipérbole se e só se $\frac{r_1}{d_1} = e$ ou $\frac{r_2}{d_2} = e$.

Demonstração. Demonstremos este teorema para o foco F_2. Obviamente

$$d_2 = \left|x + \frac{a}{e}\right| = \frac{1}{e}|ex + a| = \frac{r_2}{e}.$$

Suponhamos que $\frac{r_2}{d_2} = e$, ou seja,

$$\frac{\sqrt{(x+c)^2 + y^2}}{|x + \frac{a}{e}|} = e.$$

Como $e = \frac{c}{a}$, esta última igualdade é equivalente à equação da hipérbole. ■

Considerando y como função de x e derivando a equação da hipérbole, obtemos $y' = \frac{b^2 x}{a^2 y}$. Portanto, a equação da tangente à hipérbole no ponto M com coordenadas (x_0, y_0) é

$$y - y_0 = \frac{b^2 x_0}{a^2 y_0}(x - x_0).$$

A interseção da tangente com o eixo das abcissas é o ponto L com coordenadas $(\frac{a^2}{x_0}, 0)$ (veja a Fig. 10.9).

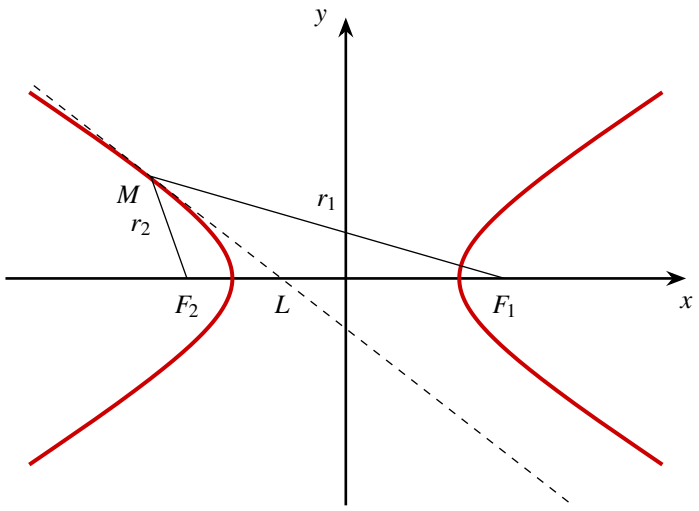


Figura 10.9
Tangente à hipérbole.

O teorema seguinte dá-nos um resultado semelhante ao Teorema 10.4 demonstrado para as elipses.

Teorema 10.9

A tangente à hipérbole num ponto M é a bissetriz do ângulo formado pelos segmentos que ligam M aos focos (veja a Fig. 10.9).

Demonstração. O quociente das áreas dos triângulos F_2ML e LMF_1 é

$$\frac{y\frac{|F_2L|}{2}}{y\frac{|F_1L|}{2}} = \frac{|F_2L|}{|F_1L|} = \frac{c+\frac{a^2}{x_0}}{c-\frac{a^2}{x_0}} = -\frac{a^2+cx_0}{a^2-cx_0}. \tag{10.7}$$

Também é possível escrever este quociente utilizando a definição da área em termos do produto vetorial, ou seja,

$$\frac{|F_2M||ML|\frac{\operatorname{sen}(\widehat{F_2ML})}{2}}{|F_1M||ML|\frac{\operatorname{sen}(\widehat{LMF_1})}{2}} = \frac{r_2\operatorname{sen}(\widehat{F_2ML})}{r_1\operatorname{sen}(\widehat{LMF_1})} = \frac{(-a-ex_0)\operatorname{sen}(\widehat{F_2ML})}{(a-ex_0)\operatorname{sen}(\widehat{LMF_1})}$$

$$= -\frac{(a^2+cx_0)\operatorname{sen}(\widehat{F_2ML})}{(a^2-cx_0)\operatorname{sen}(\widehat{LMF_1})}.$$

Como o quociente (10.7) e esta expressão coincidem, temos $\operatorname{sen}(\widehat{F_2ML}) = \operatorname{sen}(\widehat{LMF_1})$.

■

10.5. Parábolas

Consideremos uma parábola

$$y^2 = 2px, \tag{10.8}$$

onde $p > 0$. Obviamente esta curva é simétrica em relação ao eixo de abcissas. O *foco* da parábola é o ponto F com coordenadas $(\frac{p}{2}, 0)$. A *diretriz* da parábola é a reta $x = -\frac{p}{2}$ (veja a Fig. 10.10).

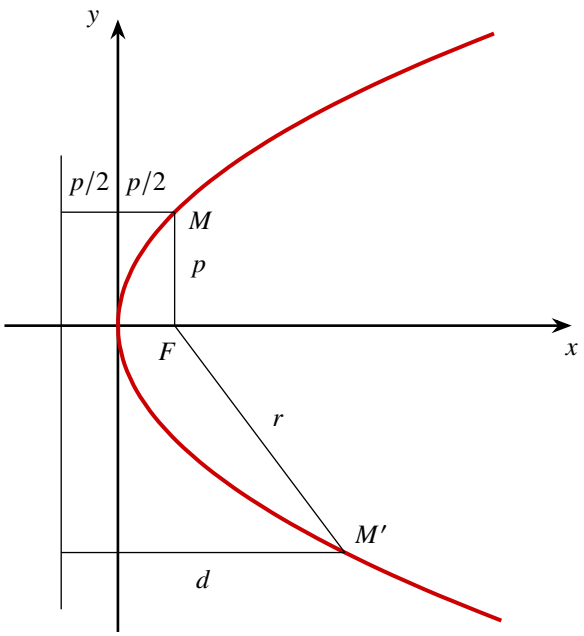


Figura 10.10
Parábola.

O teorema seguinte dá-nos um processo para construir uma parábola dados o foco e a diretriz.

Teorema 10.10

A distância entre um ponto M da parábola com coordenadas (x, y) e o foco é

$$r = x + \frac{p}{2}. \tag{10.9}$$

Demonstração. De facto, como $x \geqslant 0$ temos

$$r = \sqrt{\left(x - \frac{p}{2}\right)^2 + y^2} = \sqrt{\left(x - \frac{p}{2}\right)^2 + 2px} = \sqrt{\left(x + \frac{p}{2}\right)^2} = x + \frac{p}{2}.$$

■

Outra caracterização geométrica de parábolas é dada no seguinte teorema.

Teorema 10.11

Um ponto M pertence à parábola se e só se as distâncias entre o ponto e o foco e entre o ponto e a diretriz são iguais.

Demonstração. Notemos que a distância entre um ponto da parábola e a diretriz também é igual a

$$d = x + \frac{p}{2}.$$

Suponhamos agora que as distâncias entre o ponto M com coordenadas (x, y) e o foco e entre o ponto e a diretriz são iguais, isto é, que

$$\sqrt{\left(x - \frac{p}{2}\right)^2 + y^2} = x + \frac{p}{2}.$$

O quadrado desta igualdade dá a equação da parábola.

■

O valor da *excentricidade* da parábola consideramos igual a 1. Assim, a fórmula $e = \frac{r}{d}$ que relaciona a excentricidade e as distâncias entre o ponto e o foco e entre o ponto e a diretriz, e que foi demonstrada para as elipses e hipérboles, também é válida para as parábolas.

Considerando y como função de x e derivando a equação da parábola, encontramos $y' = \frac{p}{y}$. Portanto, a equação da tangente à parábola no ponto (x_0, y_0) é

$$y - y_0 = \frac{p}{y_0}(x - x_0).$$

Como $y_0^2 = 2px_0$, podemos escrever a equação da tangente como

$$yy_0 = p(x + x_0). \tag{10.10}$$

O seguinte teorema para parábolas corresponde aos Teoremas 10.4 e 10.9 que demonstramos para o caso das elipses e hipérboles, respetivamente.

Teorema 10.12

Sejam M um ponto da parábola e ML um segmento paralelo ao eixo de abcissas. Então, a tangente à parábola no ponto M é bissetriz do ângulo adjacente suplementar ao ângulo $\widehat{FML}$ (veja a Fig. 10.11).

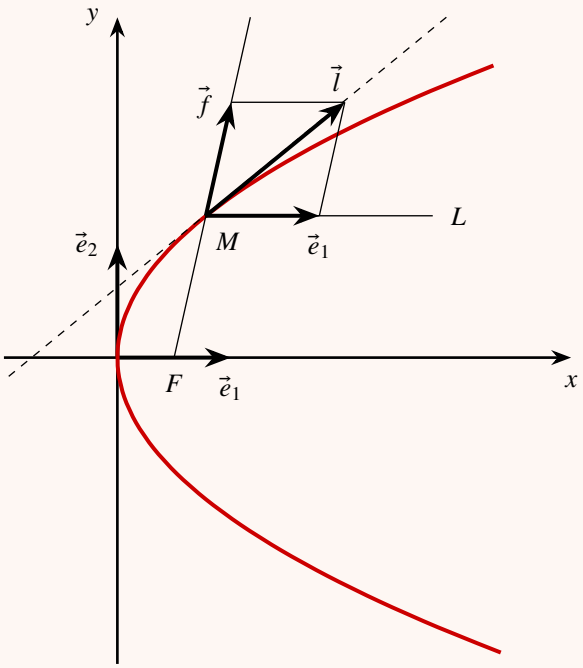


Figura 10.11
Ilustração do Teorema 10.12.

Demonstração. Sejam M um ponto da parábola com coordenadas (x, y) e $\vec{f} = \frac{\overrightarrow{FM}}{|\overrightarrow{FM}|}$. Como

$$\overrightarrow{FM} = \left(x - \frac{p}{2}\right)\vec{e}_1 + y\vec{e}_2,$$

obtemos

$$|\overrightarrow{FM}|^2 = \left(x - \frac{p}{2}\right)^2 + y^2 = x^2 - xp + \frac{p^2}{4} + y^2 = x^2 + xp + \frac{p^2}{4} = \left(x + \frac{p}{2}\right)^2.$$

Logo, temos

$$\vec{l} = \vec{f} + \vec{e}_1 = \left(\frac{x - \frac{p}{2}}{x + \frac{p}{2}} + 1\right)\vec{e}_1 + \frac{y}{x + \frac{p}{2}}\vec{e}_2 = \frac{2x\vec{e}_1 + y\vec{e}_2}{x + \frac{p}{2}}.$$

Portanto, a tangente do ângulo entre o vetor $\vec{l}$ e o eixo de abcissas é igual a

$$\frac{y}{2x} = \frac{yp}{2px} = \frac{yp}{y^2} = \frac{p}{y},$$

ou seja, é igual ao coeficiente angular da tangente à parábola no ponto M com coordenadas (x, y). Como $\vec{l}$ gera a bissetriz do ângulo entre os vetores $\vec{e}_1$ e $\vec{f}$, temos que a tangente à parábola no ponto M também é a bissetriz do mesmo ângulo. ■

Porque é que este teorema é interessante? A fonte de luz que está no foco da parábola projeta toda a luz para o infinito, como, por exemplo, acontece nos faróis máximos de um carro.

10.6. Classificação de cónicas

Vamos definir a forma geral de uma cónica.

Definição 10.4. Cónica, cónica degenerada, cónica não degenerada

Uma *cónica* é o conjunto de pontos de $\mathbb{R}^2$ que verificam uma equação quadrática de duas variáveis, ou seja,

$$X = \{(x^1, x^2) \in \mathbb{R}^2 : f(x^1, x^2) = 0\},$$

em que

$$f(x^1, x^2) = a_1^1(x^1)^2 + 2a_1^2 x^1 x^2 + a_2^2(x^2)^2 + b^1 x^1 + b^2 x^2 + c, \qquad (10.11)$$

com $a_i^j, b^j, c \in \mathbb{R}$ para $i, j = 1, 2$.
Dizemos que a cónica é *degenerada* se $X = \emptyset$ ou se X representa um ponto ou uma reta ou duas retas. Caso contrário, dizemos que a cónica é *não degenerada*.

Quando as formas quadráticas das cónicas não degeneradas estão numa forma especial, a cónica pode ser facilmente identificada a partir da sua expressão.

Notemos que a expressão (10.11) pode ser escrita de forma matricial como

$$f(x) = x^\intercal A x + b^\intercal x + c, \qquad (10.12)$$

em que $x = \begin{pmatrix} x^1 \\ x^2 \end{pmatrix}$, $A = \begin{pmatrix} a_1^1 & a_1^2 \\ a_2^1 & a_2^2 \end{pmatrix}$, onde $a_2^1 = a_1^2$, e $b = \begin{pmatrix} b^1 \\ b^2 \end{pmatrix}$.

O seguinte teorema classifica cónicas não degeneradas.

Teorema 10.13

Toda a cónica não degenerada pode ser transformada numa das seguintes *formas reduzidas (ou canónicas)*, onde $d_1, d_2 > 0$:

1. Elipse
$$d_1(y^1)^2 + d_2(y^2)^2 - 1 = 0;$$
2. Hipérbole
$$d_1(y^1)^2 - d_2(y^2)^2 - 1 = 0;$$
3. Parábola
$$d_1(y^1)^2 - y^2 = 0.$$

Demonstração. Como a matriz A é simétrica, existe uma matriz ortogonal Q, tal que $Q^\intercal A Q = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$ é uma matriz diagonal, sendo λ_1 e λ_2 os valores próprios da matriz A (veja o Teorema 9.2). Fazendo a substituição $x = Qy$, em que $x = \begin{pmatrix} x^1 \\ x^2 \end{pmatrix}$ e $y = \begin{pmatrix} y^1 \\ y^2 \end{pmatrix}$,

na expressão (10.12), obtemos a equação

$$y^{\mathsf{T}}Q^{\mathsf{T}}AQy + b^{\mathsf{T}}Qy + c = 0.$$

Assim, concluímos que, aplicando uma transformação ortogonal de coordenadas, a expressão (10.11) dá origem à equação

$$\lambda_1(y^1)^2 + \lambda_2(y^2)^2 + c_1y^1 + c_2y^2 + c = 0. \tag{10.13}$$

Analisemos os seguintes casos.

1. $\lambda_1 \neq 0$ e $\lambda_2 \neq 0$:

Efetuando a transformação de coordenadas dada por $y^i = z^i - \frac{c_i}{2\lambda_i}$, $i = 1, 2$, obtemos a partir de (10.13) a equação

$$\lambda_1(z^1)^2 + \lambda_2(z^2)^2 + d = 0, \tag{10.14}$$

onde $d = c - \frac{c_1^2}{4\lambda_1} - \frac{c_2^2}{4\lambda_2}$.

Se $d = 0$, então (10.14) define um ponto ou duas retas, sendo a cónica degenerada.

Se $d \neq 0$, então dividindo (10.14) por $-d$, obtemos

$$d_1(z^1)^2 + d_2(z^2)^2 - 1 = 0, \tag{10.15}$$

em que $d_i = -\frac{\lambda_i}{d}$, $i = 1, 2$. Neste caso, se d_1 e d_2 forem ambos negativos, então $X = \emptyset$, logo a cónica é degenerada. Se d_1 e d_2 forem ambos positivos, a equação (10.15) representa uma elipse. Se d_1 e d_2 têm sinais diferentes, a equação representa uma hipérbole.

2. $\lambda_1 \neq 0$, $\lambda_2 = 0$ e $c_2 \neq 0$:

A equação (10.13) expressa-se da forma

$$\lambda_1(y^1)^2 + c_1y^1 + c_2y^2 + c = 0.$$

Usando a substituição $y^1 = z^1 - \frac{c_1}{2\lambda_1}$ resulta que

$$\lambda_1(z^1)^2 + c_2y^2 + e = 0, \tag{10.16}$$

com $e = -\frac{c_1^2}{4\lambda_1} + c$. Fazendo a substituição $y^2 = z^2 - \frac{e}{c_2}$ em (10.16), concluímos que

$$\lambda_1(z^1)^2 + c_2z^2 = 0$$

e dividindo por $-c_2$ obtemos

$$d_1(z^1)^2 - z^2 = 0, \tag{10.17}$$

onde $d_1 = -\frac{\lambda_1}{c_2}$. A equação (10.17) representa uma parábola.

Todos os restantes casos correspondem a cónicas degeneradas. ■

Consideremos alguns exemplos onde em vez de x^1 vamos usar x e em vez de x^2 vamos usar y para simplificar a notação como no início deste capítulo.

Exemplo 10.5. Consideremos a cónica

$$x^2 + 8x + 4y^2 - 24y = -36.$$

Para classificarmos a cónica vamos determinar a mudança de coordenadas de modo a transformar a cónica na forma reduzida. Atendendo a que $x^2 + 8x = (x+4)^2 - 16$ e $4y^2 - 24y = 4((y-3)^2 - 9)$, obtemos

$$\frac{(x+4)^2}{4^2} + \frac{(y-3)^2}{2^2} = 1.$$

A mudança de coordenadas $x' = x + 4$, $y' = y - 3$ transforma a cónica na forma

$$\frac{(x')^2}{4^2} + \frac{(y')^2}{2^2} = 1,$$

que é a equação reduzida de uma elipse. A origem $x' = 0$, $y' = 0$ do novo sistema de coordenadas x' e y' está em $x = -4$, $y = 3$. Podemos ver na Fig. 10.12 o esboço da elipse com semieixos $a = 4$ e $b = 2$. (Nesta figura, as coordenadas no sistema original (x, y) estão representadas a preto e as coordenadas no sistema (x', y') estão representadas a azul.) Os focos da elipse, em coordenadas x' e y', são os pontos $F_1 = (\sqrt{12}, 0)$ e $F_2 = (-\sqrt{12}, 0)$. A excentricidade da elipse é $e = \sqrt{\frac{3}{4}} < 1$.

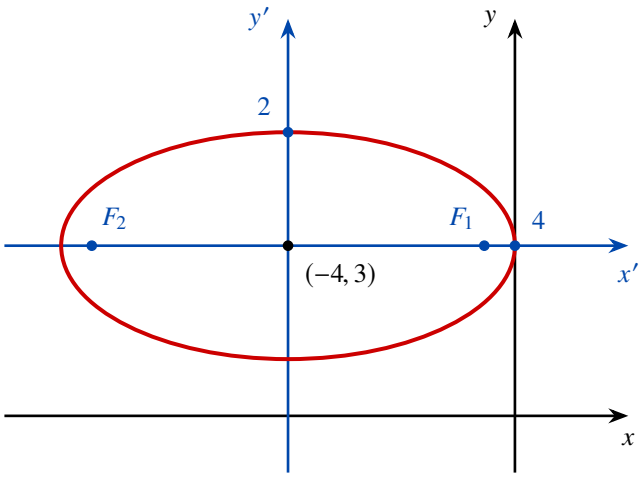


Figura 10.12
Cónica $x^2 + 8x + 4y^2 - 24y = -36$.

□

Exemplo 10.6. Consideremos a cónica

$$2x^2 + 4x - 4y = -6.$$

Para transformarmos a cónica na forma reduzida, completamos primeiro o quadrado da expressão que depende da coordenada x, vindo

$$(x+1)^2 = 2(y-1).$$

Usando a mudança de coordenadas

$$x' = x + 1 \quad \text{e} \quad y' = y - 1,$$

obtemos a equação na forma reduzida

$$(x')^2 = 2y',$$

que representa uma parábola (veja a Fig. 10.13). A origem $x' = 0$, $y' = 0$ do novo sistema de coordenadas x' e y' está em $x = -1$, $y = 1$. O foco da parábola em coordenadas x' e y' é o ponto $F = \left(0, \frac{1}{2}\right)$ e a diretriz da parábola é a reta $y' = -\frac{1}{2}$. (Na figura, as coordenadas no sistema original (x, y) estão representadas a preto e as coordenadas no sistema (x', y') estão representadas a azul.)

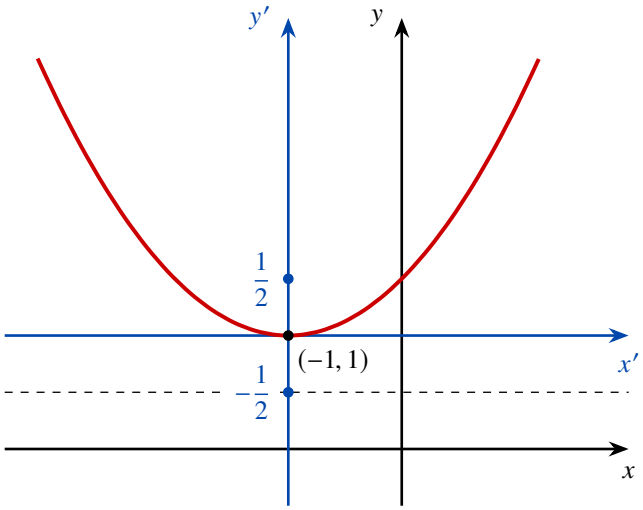


Figura 10.13
Cónica
$2x^2 + 4x - 4y = -6$.

□

Exemplo 10.7. Consideremos a cónica

$$10x^2 - 8xy + 4y^2 = 12.$$

Neste caso vamos aplicar uma transformação ortogonal de coordenadas para transformar a cónica na forma reduzida. A matriz da forma quadrática é $A = \left(\begin{smallmatrix} 10 & -4 \\ -4 & 4 \end{smallmatrix}\right)$. A equação característica da matriz A é

$$\lambda^2 - 14\lambda + 24 = 0.$$

As suas raízes são $\lambda = 2$ e $\lambda = 12$. Os respetivos vetores próprios normalizados são

$$\vec{q}_1 = \begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix} \quad \text{e} \quad \vec{q}_2 = \begin{pmatrix} -\frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix}.$$

A matriz Q que transforma A numa matriz diagonal (veja o Teorema 9.2), é

$$Q = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix}.$$

O determinante desta matriz é um. Assim, a mudança de coordenadas dada por

$$x = \frac{1}{\sqrt{5}}x' - \frac{2}{\sqrt{5}}y',$$
$$y = \frac{2}{\sqrt{5}}x' + \frac{1}{\sqrt{5}}y'$$

transforma a cónica na forma reduzida

$$\frac{(x')^2}{6} + \frac{(y')^2}{1} = 1.$$

Assim, concluímos que a cónica é uma elipse com semieixos $a = \sqrt{6}$ e $b = 1$ (veja a Fig. 10.14). Os focos da elipse, em coordenadas x' e y', são os pontos $F_1 = (\sqrt{5}, 0)$ e $F_2 = (-\sqrt{5}, 0)$. A excentricidade da elipse é $e = \sqrt{\frac{5}{6}} < 1$. A origem no sistema de coordenadas (x', y') coincide com a origem no sistema de coordenadas inicial. (Na figura, as coordenadas no sistema original (x, y) estão representadas a preto e as coordenadas no sistema (x', y') estão representadas a azul. Os eixos do sistema de coordenadas (x', y') têm direção dos vetores próprios, que estão representados a preto.)

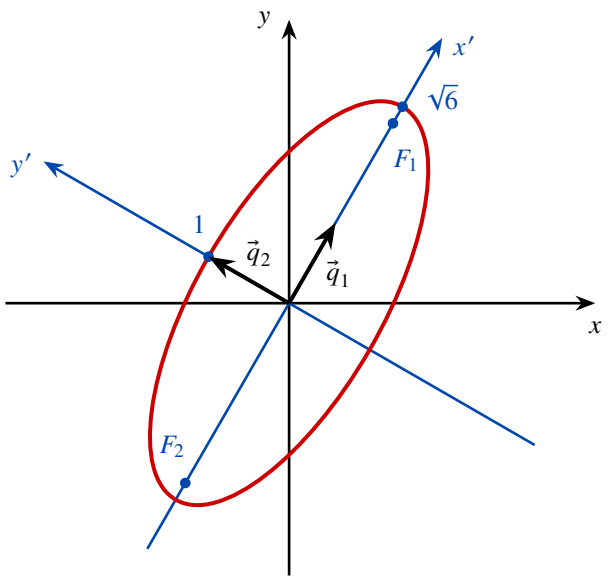


Figura 10.14
Cónica
$10x^2 - 8xy + 4y^2 = 12.$

□

Exemplo 10.8. Consideremos a cónica

$$4xy - 3y^2 - 4x + 10y - 6 = 0.$$

A matriz da forma quadrática é $A = \left(\begin{smallmatrix} 0 & 2 \\ 2 & -3 \end{smallmatrix}\right)$. Os valores próprios desta matriz são $\lambda_1 = 1$ e $\lambda_2 = -4$. A matriz dos respetivos vetores próprios normalizados é

$$Q = \frac{1}{\sqrt{5}}\begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix}.$$

A mudança de coordenadas

$$x = \frac{1}{\sqrt{5}}(2x' - y') \quad \text{e} \quad y = \frac{1}{\sqrt{5}}(x' + 2y'),$$

transforma a equação da cónica em

$$(x')^2 - 4(y')^2 + \frac{2}{\sqrt{5}}x' + \frac{24}{\sqrt{5}}y' - 6 = 0,$$

ou, na forma equivalente,

$$4\left(y' - \frac{3}{\sqrt{5}}\right)^2 - \left(x' + \frac{1}{\sqrt{5}}\right)^2 = 1.$$

Fazendo

$$x'' = x' + \frac{1}{\sqrt{5}} \quad \text{e} \quad y'' = y' - \frac{3}{\sqrt{5}},$$

obtemos

$$-(x'')^2 + 4(y'')^2 = 1.$$

Portanto, a cónica é uma hipérbole. A mudança de variáveis que transforma a cónica na sua forma canónica é

$$x'' = \frac{2x + y + 1}{\sqrt{5}} \quad \text{e} \quad y'' = \frac{-x + 2y - 3}{\sqrt{5}}.$$

As assíntotas são $y'' = \pm\frac{1}{2}x''$ ou, em coordenadas x e y, $y = 1$ e $y = \frac{4x+7}{3}$. Os focos da hipérbole, em coordenadas x'' e y'', são os pontos $F_1 = (0, \frac{5}{4})$ e $F_2 = (0, -\frac{5}{4})$. A excentricidade da elipse é $e = \sqrt{\frac{5}{4}} > 1$. A origem do sistema (x'', y''), em coordenadas x e y, é o ponto $(-1, 1)$. Agora é fácil construir a hipérbole (veja a Fig. 10.15). (Na figura, as coordenadas no sistema original (x, y) estão representadas a preto e as coordenadas no sistema (x'', y'') estão representadas a azul.)

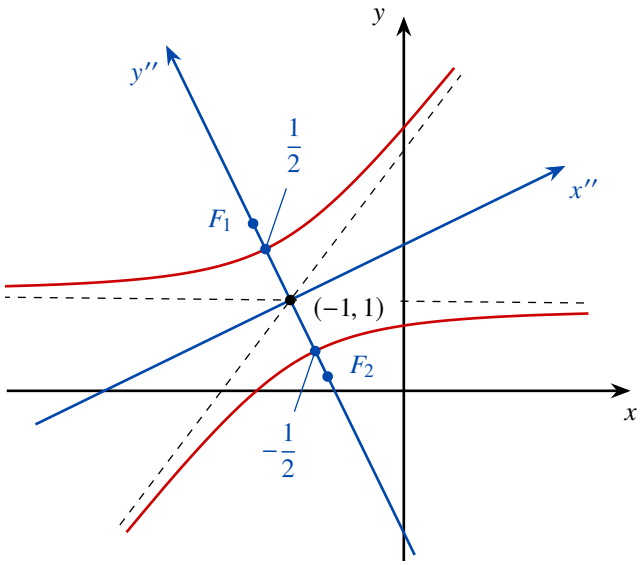


Figura 10.15
Cónica $4xy - 3y^2 - 4x + 10y - 6 = 0$.

□

Exemplo 10.9. Consideremos a cónica

$$2x^2 - 2xy + 2y^2 - 2\sqrt{2}x + 4\sqrt{2}y = 8.$$

A matriz da forma quadrática é $A = \left(\begin{smallmatrix} 2 & -1 \\ -1 & 2 \end{smallmatrix}\right)$. Os valores próprios desta matriz são $\lambda_1 = 1$ e $\lambda_2 = 3$. A matriz dos respetivos vetores próprios normalizados é

$$Q = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

A mudança das coordenadas

$$x = \frac{1}{\sqrt{2}}(x' - y') \quad \text{e} \quad y = \frac{1}{\sqrt{2}}(x' + y'),$$

transforma a equação da cónica em

$$(x')^2 + 3(y')^2 + 2x' + 6y' - 8 = 0.$$

Completando os quadrados obtemos

$$\frac{(x'+1)^2}{12} + \frac{(y'+1)^2}{4} = 1.$$

Aplicando agora a mudança de coordenadas

$$x'' = x' + 1 \quad \text{e} \quad y'' = y' + 1,$$

a equação é transformada na forma reduzida

$$\frac{(x'')^2}{12} + \frac{(y'')^2}{4} = 1,$$

que representa uma elipse.

Logo, para transformarmos a cónica na forma reduzida aplicamos primeiro uma rotação de eixos e de seguida uma translação de eixos, obtendo

$$x'' = \frac{1}{\sqrt{2}}x + \frac{1}{\sqrt{2}}y + 1 \quad \text{e} \quad y'' = \frac{1}{\sqrt{2}}y - \frac{1}{\sqrt{2}}x + 1.$$

Na Fig. 10.16 vemos o esboço da elipse com semieixos $a = \sqrt{12}$ e $b = 2$. Os focos da elipse, em coordenadas x'' e y'', são os pontos $F_1 = (\sqrt{8}, 0)$ e $F_2 = (-\sqrt{8}, 0)$. A excentricidade da elipse é $e = \sqrt{\frac{2}{3}} < 1$. A origem do sistema (x'', y''), em coordenadas x' e y', é o ponto $(-1, -1)$. (Na figura, as coordenadas no sistema original (x, y) estão representadas a preto, as coordenadas no sistema (x', y') estão representadas a azul e as coordenadas no sistema (x'', y'') a verde. Os eixos do sistema de coordenadas (x', y') têm direção dos vetores próprios, que estão representados a preto.)

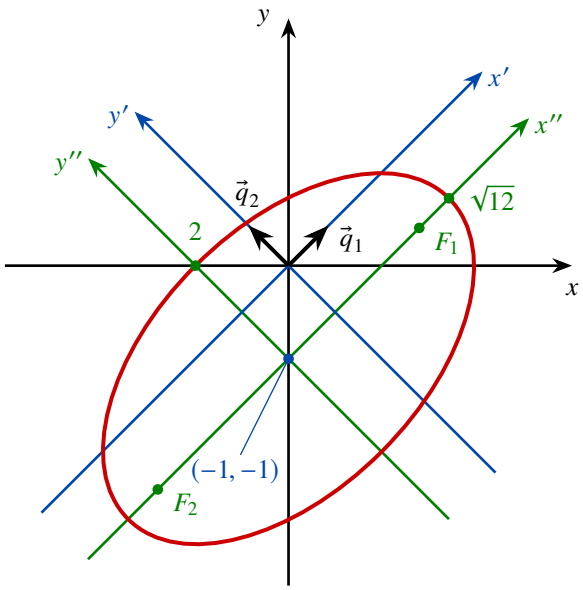


Figura 10.16
Cónica $2x^2 - 2xy + 2y^2 - 2\sqrt{2}x + 4\sqrt{2}y = 8$.

□

11. Quádricas

11.1. Introdução

As quádricas podem ser definidas como conjuntos de pontos em $\mathbb{R}^3$ cujas coordenadas satisfazem uma equação de segundo grau. Dentro do conjunto das quádricas existem as seguintes classes principais de quádricas: elipsoides, hiperboloides, paraboloides, cones e cilindros. As formas geométricas e algébricas das quádricas estão relacionadas com as das cónicas da seguinte maneira. A projeção de uma quádrica num plano qualquer definido em $\mathbb{R}^3$, depois de identificadas duas coordenadas nesse plano, vai representar uma cónica (degenerada ou não degenerada). Como veremos, o nome "cónica" está relacionado com o cone, pelo facto de ser possível obter as três cónicas principais (elipses, hipérboles e parábolas) através da interseção de um cone com um plano.

Quanto às aplicações, algumas das quádricas têm servido de modelos na arquitetura. Por exemplo, o hiperboloide de uma folha é um elemento da construção de torres radiotelevisivas, alguns faróis e de torres de arrefecimento de centrais elétricas. A justificação desta utilização consiste na simplicidade da sua construção. Como veremos, através de cada ponto desta quádrica passam duas retas que lhe pertencem. Portanto, a quádrica pode ser constituída por barras retas. A quádrica paraboloide hiperbólico, que tem a forma de uma sela, também possui a mesma propriedade, o que novamente justifica que algumas coberturas de edifícios como estádios de futebol e igrejas têm esta forma.

Outros exemplos são a quádrica paraboloide, que é usada na construção de antenas parabólicas ou dos espelhos dos faróis máximos, e a quádrica elipsoide, que é usada na área das geociências como modelo da forma do planeta Terra para cálculos geodésicos.

O objetivo deste capítulo é a identificação e a classificação das quádricas através da transformação da equação de segundo grau numa expressão mais simples através de mudanças de variáveis. Para isso vamos usar o procedimento do capítulo anterior, recorrendo à teoria das formas quadráticas.

11.2. Quádricas relevantes

Iniciamos o estudo com as definições das quádricas relevantes.

Definição 11.1. Elipsoide, esfera

Sejam $a, b, c, \rho > 0$. Chamamos *elipsoide* à quádrica cuja equação é

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1, \ a, b \text{ e } c \text{ não todos iguais,}$$

e *esfera* à quádrica cuja equação é

$$x^2 + y^2 + z^2 = \rho^2.$$

Exemplo 11.1. A quádrica $x^2 + 5y^2 + 3z^2 = 1$ é um elipsoide (veja a Fig. 11.1).

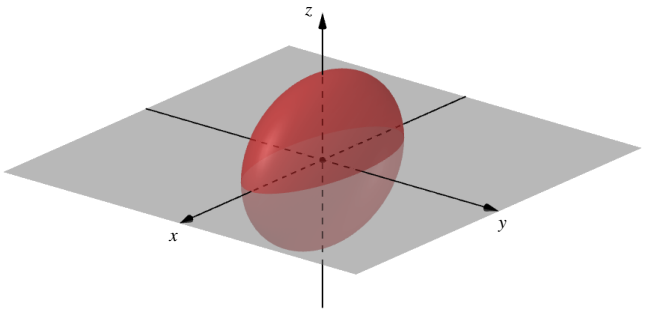


Figura 11.1
Elipsoide
$x^2 + 5y^2 + 3z^2 = 1$.

□

Definição 11.2. Hiperboloide de uma folha

Sejam $a, b, c > 0$. Chamamos *hiperboloide de uma folha* à quádrica cuja equação é

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1.$$

Exemplo 11.2. A quádrica $x^2 + y^2 - z^2 = 1$ é um hiperboloide de uma folha (veja a Fig. 11.2).

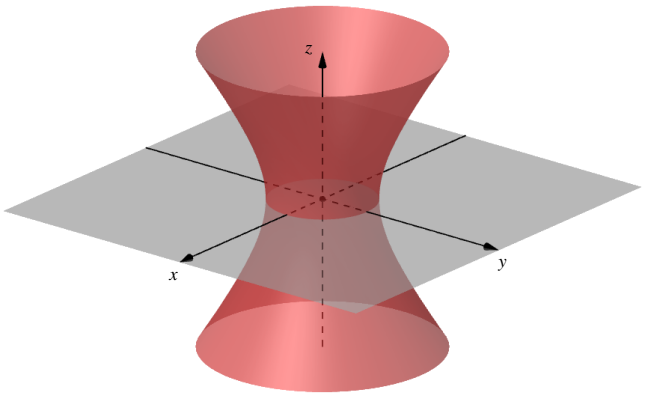


Figura 11.2
Hiperboloide de uma folha $x^2 + y^2 - z^2 = 1$.

□

Definição 11.3. Hiperboloide de duas folhas

Sejam $a, b, c > 0$. Chamamos *hiperboloide de duas folhas* à quádrica cuja equação é

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1.$$

Exemplo 11.3. A quádrica de equação $\frac{x^2}{3} - 2y^2 - 2z^2 = 1$ é um hiperboloide de duas folhas (veja a Fig. 11.3).

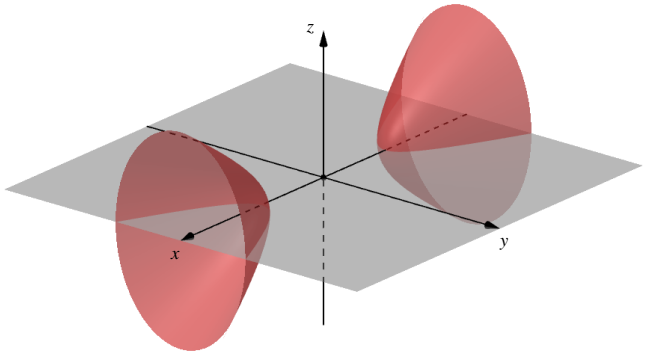


Figura 11.3
Hiperboloide de duas folhas
$\frac{x^2}{3} - 2y^2 - 2z^2 = 1$.

□

Definição 11.4. Cone elíptico, cone circular

Sejam $a, b, c, \rho > 0$. Chamamos *cone elíptico* à quádrica cuja equação é

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 0$$

e *cone circular* à quádrica cuja equação é

$$x^2 + y^2 - \rho z^2 = 0$$

Exemplo 11.4. A quádrica $x^2 + y^2 - z^2 = 0$ é um cone circular (veja a Fig. 11.4).

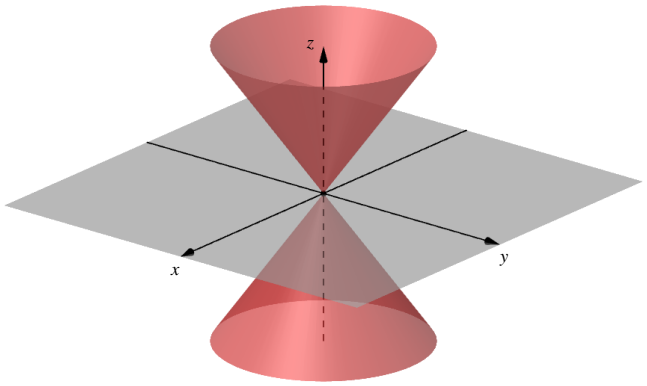


Figura 11.4
Cone circular
$x^2 + y^2 - z^2 = 0$.

□

Definição 11.5. Cilindro elíptico, cilindro circular

Sejam $a, b, \rho > 0$. Chamamos *cilindro elíptico* à quádrica cuja equação é

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, a \neq b,$$

e *cilindro circular* à quádrica cuja equação é

$$x^2 + y^2 = \rho^2.$$

Exemplo 11.5. A quádrica $x^2 + 2y^2 = 1$ é um cilindro elíptico (veja a Fig. 11.5).

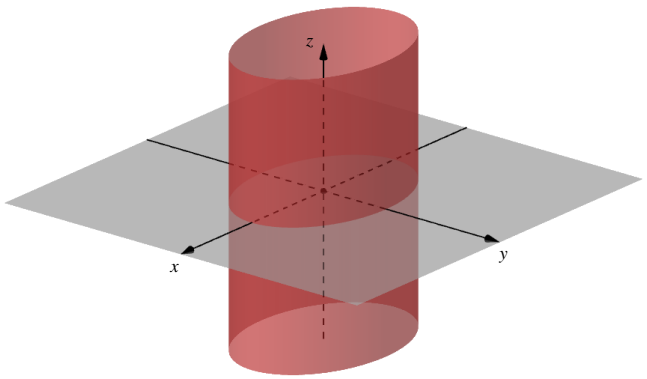


Figura 11.5
Cilindro elíptico
$x^2 + 2y^2 = 1$.

□

Definição 11.6. Cilindro hiperbólico

Sejam $a, b > 0$. Chamamos *cilindro hiperbólico* à quádrica cuja equação é

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1.$$

Exemplo 11.6. A quádrica $x^2 - 4y^2 = 1$ é um cilindro hiperbólico (veja a Fig. 11.6).

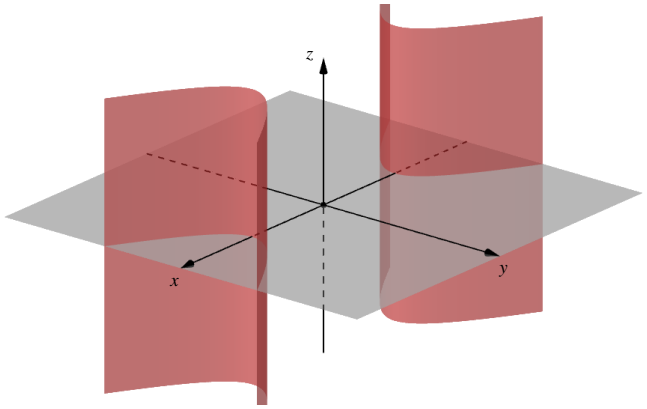


Figura 11.6
Cilindro hiperbólico
$x^2 - 4y^2 = 1$.

□

Definição 11.7. Paraboloide elíptico, paraboloide circular

Sejam $a, b, \rho > 0$ e $p \neq 0$. Chamamos *paraboloide elíptico* à quádrica cuja equação é

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 2pz, a \neq b,$$

e *paraboloide circular* à quádrica cuja equação é

$$x^2 + y^2 = 2p\rho^2 z.$$

Exemplo 11.7. A quádrica $x^2 + y^2 = z$ é um paraboloide circular (veja a Fig. 11.7).

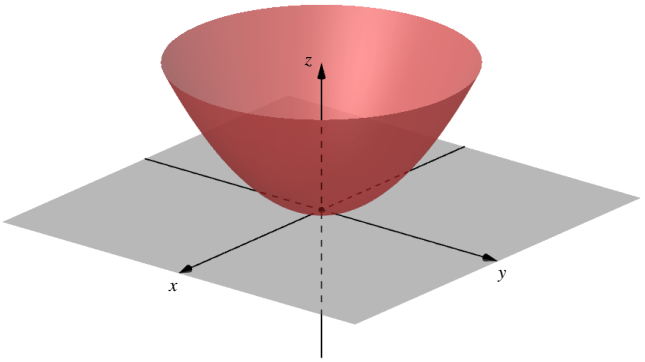


Figura 11.7
Paraboloide circular
$x^2 + y^2 = z$.

□

Definição 11.8. Paraboloide hiperbólico

Sejam $a, b > 0$ e $p \neq 0$. Chamamos *paraboloide hiperbólico* à quádrica cuja equação é

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 2pz.$$

Exemplo 11.8. A quádrica $x^2 - y^2 = z$ é um paraboloide hiperbólico (veja a Fig. 11.8).

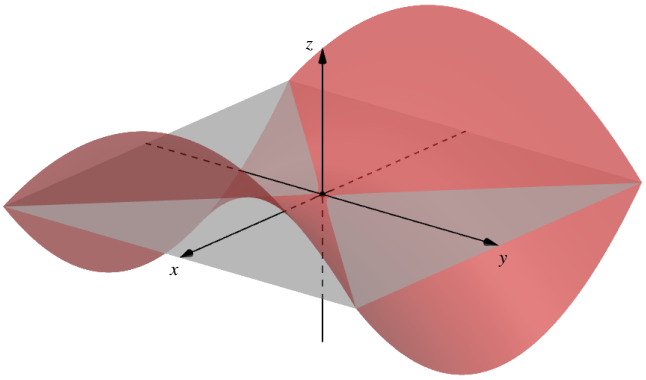


Figura 11.8
Paraboloide hiperbólico
$x^2 - y^2 = z$.

□

Definição 11.9. Cilindro parabólico

Seja $p > 0$. Chamamos *cilindro parabólico* à quádrica cuja equação é

$$x^2 = 2py.$$

Exemplo 11.9. A quádrica $4x^2 = y$ é um cilindro parabólico (veja a Fig. 11.9).

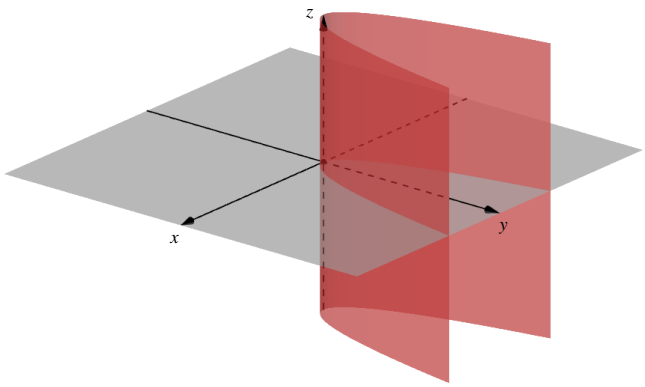


Figura 11.9
Cilindro parabólico $4x^2 = y$.

□

11.3. Porque chamamos "cónicas" às cónicas?

Como já foi mencionado na introdução a este capítulo, é possível obter as três cónicas principais (elipses, hipérboles e parábolas) através da interseção de um cone com um plano. Vamos agora ver isto em detalhe. A equação geral de um cone circular é

$$x^2 + y^2 - \rho^2 z^2 = 0, \quad \rho > 0.$$

A interseção deste cone com o plano $z = 1$ é a circunferência $x^2 + y^2 = \rho^2$. A interseção do cone com o plano $x = b$ é a hipérbole definida pela equação

$$\rho^2 z^2 - y^2 = b^2.$$

Consideremos o plano Π definido pela equação

$$z = \alpha x + 1, \tag{11.1}$$

$\alpha > 0$. Introduzindo novas coordenadas (x', y', z') definidas como

$$x = \frac{x' - \alpha z'}{\sqrt{1+\alpha^2}} - \frac{1}{\alpha}, \quad y = y' \quad \text{e} \quad z = \frac{\alpha x' + z'}{\sqrt{1+\alpha^2}},$$

temos a equação do plano Π na forma $z' = 0$. Substituindo (11.1) na equação do cone obtemos

$$x^2 + y^2 - \rho^2\alpha^2 x^2 - 2\rho^2\alpha x - \rho^2 = 0. \tag{11.2}$$

Suponhamos que $\rho^2\alpha^2 \neq 1$. Então, a última igualdade pode ser reescrita na forma

$$\frac{\left(x - \frac{\rho^2\alpha}{1-\rho^2\alpha^2}\right)^2}{\frac{\rho^2}{(1-\rho^2\alpha^2)^2}} + \frac{y^2}{\frac{\rho^2}{1-\rho^2\alpha^2}} = 1.$$

No plano Π, isto é, quando $z' = 0$, obtemos a equação

$$\frac{\left(\frac{x'}{\sqrt{1+\alpha^2}} - \frac{1}{\alpha} - \frac{\rho^2\alpha}{1-\rho^2\alpha^2}\right)^2}{\frac{\rho^2}{(1-\rho^2\alpha^2)^2}} + \frac{y^2}{\frac{\rho^2}{1-\rho^2\alpha^2}} = 1.$$

ou, na forma equivalente

$$\frac{\left(x' - \left(\frac{\sqrt{1+\alpha^2}}{\alpha} + \frac{\rho^2\alpha\sqrt{1+\alpha^2}}{1-\rho^2\alpha^2}\right)\right)^2}{\frac{\rho^2(1+\alpha^2)}{(1-\rho^2\alpha^2)^2}} + \frac{y'^2}{\frac{\rho^2}{1-\rho^2\alpha^2}} = 1.$$

Introduzindo as coordenadas (x'', y'') definidas como

$$x'' = x' - \frac{\sqrt{1+\alpha^2}}{\alpha} - \frac{\rho^2\alpha\sqrt{1+\alpha^2}}{1-\rho^2\alpha^2} \quad \text{e} \quad y'' = y',$$

obtemos a equação

$$\frac{(x'')^2}{a^2} \pm \frac{(y'')^2}{b^2} = 1,$$

onde

$$a^2 = \frac{\rho^2(1+\alpha^2)}{(1-\rho^2\alpha^2)^2} \tag{11.3}$$

e

$$b^2 = \frac{\rho^2}{|1-\rho^2\alpha^2|}. \tag{11.4}$$

O sinal "$+$" corresponde à desigualdade $\rho^2\alpha^2 < 1$. Neste caso temos a equação de uma elipse. O sinal "$-$" corresponde à desigualdade $\rho^2\alpha^2 > 1$. Neste caso temos a equação de uma hipérbole. É fácil de ver que, dados $a > 0$ e $b > 0$, é sempre possível escolher ρ e α que satisfaçam as igualdades (11.3) e (11.4). De facto, das igualdades (11.3) e (11.4) temos

$$\rho^4\alpha^4 - 2\rho^2\alpha^2 + 1 = \frac{\rho^2(1+\alpha^2)}{a^2} \quad \text{e} \quad \rho^4\alpha^4 - 2\rho^2\alpha^2 + 1 = \frac{\rho^4}{b^4}.$$

Portanto, podemos sempre escolher ρ e α que satisfaçam a condição

$$\frac{a^2}{b^4} = \frac{1+\alpha^2}{\rho^2}$$

e tais que $\rho^2\alpha^2 \neq 1$.

No caso $\rho^2\alpha^2 = 1$, a equação (11.2) toma a forma

$$y^2 - 2\rho^2\alpha x - \rho^2 = 0,$$

que nas coordenadas

$$x'' = x' - \frac{\sqrt{1+\alpha^2}}{2\alpha} \quad \text{e} \quad y'' = y',$$

é a equação de uma parábola

$$(y'')^2 = \frac{2\rho^2\alpha}{\sqrt{1+\alpha^2}}x'' = \frac{2}{\alpha\sqrt{1+\alpha^2}}x''.$$

Esta equação permite obter qualquer parábola. De facto, a função $\phi(\alpha) = \alpha\sqrt{1+\alpha^2}$ toma todos os valores entre zero e infinito quando $\alpha \geqslant 0$.

11.4. Classificação de quádricas

Vamos definir a forma geral de uma quádrica.

Definição 11.10. Quádrica, quádrica degenerada, quádrica não degenerada

Uma *quádrica* é o conjunto de pontos de $\mathbb{R}^3$ que verificam uma equação quadrática de três variáveis, ou seja,

$$X = \{(x^1, x^2, x^3) \in \mathbb{R}^3 : f(x^1, x^2, x^3) = 0\},$$

em que

$$f(x^1, x^2, x^3) = a_1^1(x^1)^2 + a_2^2(x^2)^2 + a_3^3(x^3)^2 + 2a_1^2x^1x^2 + 2a_1^3x^1x^3 + 2a_2^3x^2x^3 + b^1x^1 + b^2x^2 + b^3x^3 + c, \quad (11.5)$$

com $a_i^j, b^j, c \in \mathbb{R}$ para $i, j = 1, 2, 3$.
Dizemos que a quádrica é *degenerada* se $X = \emptyset$ ou se X representa um ponto, uma reta, um plano, dois planos paralelos, ou dois planos oblíquos. Caso contrário, dizemos que a quádrica é *não degenerada*.

Quando as formas quadráticas das quádricas não degeneradas estão numa forma especial, a quádrica pode ser facilmente identificada a partir da sua expressão.

Notemos que a expressão (11.5) pode ser escrita de forma matricial como

$$f(x) = x^\intercal Ax + b^\intercal x + c, \quad (11.6)$$

em que $x = \begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix}$, $A = \begin{pmatrix} a_1^1 & a_1^2 & a_1^3 \\ a_2^1 & a_2^2 & a_2^3 \\ a_3^1 & a_3^2 & a_3^3 \end{pmatrix}$, onde $a_2^1 = a_1^2$, $a_3^1 = a_1^3$ e $a_3^2 = a_2^3$, e $b = \begin{pmatrix} b^1 \\ b^2 \\ b^3 \end{pmatrix}$.

O seguinte teorema classifica as cónicas não degeneradas.

Teorema 11.1

Toda a quádrica não degenerada pode ser transformada numa das seguintes *formas reduzidas (ou canónicas)*, onde $d_1, d_2, d_3 > 0$:

1. Elipsoide

$$d_1(y^1)^2 + d_2(y^2)^2 + d_3(y^3)^2 - 1 = 0;$$

2. Hiperboloide de uma folha

$$d_1(y^1)^2 + d_2(y^2)^2 - d_3(y^3)^2 - 1 = 0;$$

3. Hiperboloide de duas folhas

$$d_1(y^1)^2 - d_2(y^2)^2 - d_3(y^3)^2 - 1 = 0;$$

continua na página seguinte

continuação da página anterior

4. Paraboloide elíptico

$$d_1(y^1)^2 + d_2(y^2)^2 - y^3 = 0;$$

5. Paraboloide hiperbólico

$$d_1(y^1)^2 - d_2(y^2)^2 - y^3 = 0;$$

6. Cone elíptico

$$d_1(y^1)^2 + d_2(y^2)^2 - (y^3)^2 = 0;$$

7. Cilindro elíptico

$$d_1(y^1)^2 + d_2(y^2)^2 - 1 = 0;$$

8. Cilindro hiperbólico

$$d_1(y^1)^2 - d_2(y^2)^2 - 1 = 0;$$

9. Cilindro parabólico

$$d_1(y^1)^2 - y^2 = 0.$$

Demonstração. A demonstração é idêntica à feita para o Teorema 10.13. ■

Consideremos alguns exemplos onde em vez de x^1 vamos usar x, em vez de x^2 vamos usar y e em vez de x^3 vamos usar z para simplificar a notação como no início deste capítulo.

Exemplo 11.10. Consideremos a quádrica

$$4x^2 + 36y^2 - 9z^2 - 16x - 216y + 304 = 0.$$

Vamos determinar a mudança de coordenadas de modo que a quádrica resultante esteja na forma reduzida para a poder classificar. Observamos que a forma quadrática já está na forma diagonal. Notemos que

$$4x^2 - 16x = 4(x^2 - 4x + 4) - 16 = 4(x-2)^2 - 16$$

e

$$36y^2 - 216y = 36(y^2 - 6y + 9) - 324 = 36(y-3)^2 - 324.$$

Substituindo estas expressões na equação da quádrica e efetuando cálculos, obtemos

$$\frac{(x-2)^2}{9} + (y-3)^2 - \frac{z^2}{4} = 1.$$

A mudança de variáveis

$$x' = x - 2, \quad y' = y - 3 \quad \text{e} \quad z' = z$$

transforma a quádrica na forma reduzida

$$\frac{(x')^2}{9} + (y')^2 - \frac{(z')^2}{4} = 1,$$

que é a equação de um hiperboloide de uma folha. □

Exemplo 11.11. Consideremos a quádrica

$$2xy + z = 0.$$

A quádrica pode ser escrita na forma matricial

$$\begin{pmatrix} x & y & z \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0.$$

A equação característica da matriz da forma quadrática é

$$\lambda(\lambda^2 - 1) = 0.$$

Portanto, os valores próprios da matriz são $\lambda = 1$, $\lambda = -1$ e $\lambda = 0$ e os respetivos vetores próprios normalizados são

$$q_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad q_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \quad q_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

A matriz ortogonal Q com $\det(Q) = 1$ que diagonaliza a matriz é

$$Q = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Assim, a mudança de coordenadas

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = Q \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix}$$

dada por

$$x = \frac{1}{\sqrt{2}}x' - \frac{1}{\sqrt{2}}y', \quad y = \frac{1}{\sqrt{2}}x' + \frac{1}{\sqrt{2}}y' \quad \text{e} \quad z = z',$$

transforma a quádrica na forma reduzida

$$(x')^2 - (y')^2 + z' = 0,$$

que é a equação de um paraboloide hiperbólico. □

Exemplo 11.12. Consideremos a quádrica

$$x^2 - y^2 - z^2 + 2\sqrt{3}xy - 6y - 9 = 0.$$

Vamos determinar a mudança de coordenadas que transforma esta quádrica na forma reduzida para a podermos classificar. A quádrica pode ser escrita na forma matricial

$$\begin{pmatrix} x & y & z \end{pmatrix} \begin{pmatrix} 1 & \sqrt{3} & 0 \\ \sqrt{3} & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} 0 & -6 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 9.$$

A equação característica da matriz da forma quadrática é

$$(1+\lambda)(\lambda^2 - 4) = 0.$$

Assim, os valores próprios da matriz são $\lambda = 2$, $\lambda = -2$ e $\lambda = -1$. Os respetivos vetores próprios normalizados são

$$\vec{q}_1 = \begin{pmatrix} \frac{\sqrt{3}}{2} \\ \frac{1}{2} \\ 0 \end{pmatrix}, \quad \vec{q}_2 = \begin{pmatrix} -\frac{1}{2} \\ \frac{\sqrt{3}}{2} \\ 0 \end{pmatrix} \quad \text{e} \quad \vec{q}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

e assim podemos construir a matriz ortogonal

$$Q = \begin{pmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} & 0 \\ \frac{1}{2} & \frac{\sqrt{3}}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

com $\det(Q) = 1$. Depois da diagonalização, obtemos a forma quadrática com a matriz

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Assim, a mudança de coordenadas

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = Q \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix}$$

dada por

$$x = \frac{\sqrt{3}}{2}x' - \frac{1}{2}y', \quad y = \frac{1}{2}x' + \frac{\sqrt{3}}{2}y' \quad \text{e} \quad z = z',$$

transforma a quádrica na equação

$$2(x')^2 - 2(y')^2 - (z')^2 - 3x' - 3\sqrt{3}y' - 9 = 0.$$

Completando os quadrados na expressão anterior, vem

$$2\left(x' - \frac{3}{4}\right)^2 - 2\left(y' + \frac{3\sqrt{3}}{4}\right)^2 - (z')^2 + \frac{27}{8} - \frac{9}{8} - 9 = 0.$$

Usando, agora, a mudança de coordenadas

$$x'' = x' - \frac{3}{4}, \quad y'' = y' + \frac{3\sqrt{3}}{4} \quad \text{e} \quad z'' = z',$$

obtemos forma reduzida

$$2(x'')^2 - 2(y'')^2 - (z'')^2 = \frac{27}{4},$$

que é a equação de um hiperboloide de duas folhas. □

Exemplo 11.13. Consideremos a quádrica

$$x^2 + y^2 + z^2 + 2xy + 2yz + 2xz - x + y = 0.$$

A quádrica pode ser escrita na forma matricial

$$\begin{pmatrix} x & y & z \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} -1 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0.$$

A equação característica da matriz da forma quadrática é

$$\lambda^2(\lambda - 3) = 0$$

e, assim, os valores próprios da matriz são $\lambda = 0$, que é um valor próprio de multiplicidade 2, e $\lambda = 3$. Os respetivos vetores próprios normalizados são

$$q_1 = \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}, \quad q_2 = \begin{pmatrix} -\frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{pmatrix} \quad \text{e} \quad q_3 = \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix},$$

e assim podemos construir a matriz ortogonal

$$Q = \begin{pmatrix} -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{pmatrix},$$

com $\det(Q) = 1$. Depois da diagonalização, obtemos a forma quadrática com a matriz

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

Assim, a mudança de coordenadas

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = Q \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix}$$

dada por

$$x = -\frac{1}{\sqrt{2}}x' - \frac{1}{\sqrt{6}}y' + \frac{1}{\sqrt{3}}z', \quad y = \frac{1}{\sqrt{2}}x' - \frac{1}{\sqrt{6}}y' + \frac{1}{\sqrt{3}}z' \quad \text{e} \quad z = \frac{2}{\sqrt{6}}y' + \frac{1}{\sqrt{3}}z',$$

transforma a quádrica na forma reduzida

$$3(z')^2 + \frac{2}{\sqrt{2}}x' = 0,$$

que é a equação de um cilindro parabólico. □

11.5. Retas na superfície

Vamos ver uma propriedade interessante de duas classes de quádricas, que explica as suas utilizações na construção de torres e coberturas de edifícios.

Através de cada ponto de um hiperboloide de uma folha passa um par de retas (veja a Fig. 11.10). De facto, a equação desta superfície é

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1,$$

ou, de forma equivalente,

$$\left(\frac{x}{a} + \frac{z}{c}\right)\left(\frac{x}{a} - \frac{z}{c}\right) = \left(1 + \frac{y}{b}\right)\left(1 - \frac{y}{b}\right).$$

Consideremos a reta definida pelo sistema de equações lineares

$$\begin{cases} \mu\left(\frac{x}{a} + \frac{z}{c}\right) = \lambda\left(1 + \frac{y}{b}\right), \\ \lambda\left(\frac{x}{a} - \frac{z}{c}\right) = \mu\left(1 - \frac{y}{b}\right), \end{cases}$$

onde λ e μ são constantes. As coordenadas de cada ponto da reta satisfazem ambas as equações e, logo, satisfazem a equação do hiperboloide de uma folha. Escolhendo (x, y), coordenadas de um ponto no hiperboloide de uma folha, podemos encontrar os parâmetros λ e μ (a menos de um multiplicador comum). Podemos aplicar o mesmo raciocínio ao conjunto de retas definidas pelo sistema de equações lineares

$$\begin{cases} \mu'\left(\frac{x}{a} + \frac{z}{c}\right) = \lambda'\left(1 - \frac{y}{b}\right), \\ \mu'\left(\frac{x}{a} - \frac{z}{c}\right) = \lambda'\left(1 + \frac{y}{b}\right). \end{cases}$$

O paraboloide hiperbólico tem a mesma propriedade (veja a Fig. 11.11). Como a sua equação é

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 2z,$$

através de cada seu ponto passam as retas definidas pelos sistemas de equações lineares

$$\begin{cases} \lambda\left(\frac{x}{a} - \frac{y}{b}\right) = \mu, \\ \mu\left(\frac{x}{a} + \frac{y}{b}\right) = 2\lambda z, \end{cases}$$

e

$$\begin{cases} \lambda'\left(\frac{x}{a}+\frac{y}{b}\right)=\mu', \\ \mu'\left(\frac{x}{a}-\frac{y}{b}\right)=2\lambda' z. \end{cases}$$

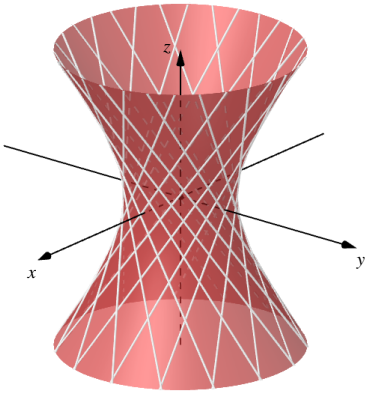


Figura 11.10
Retas no hiperboloide de uma folha.

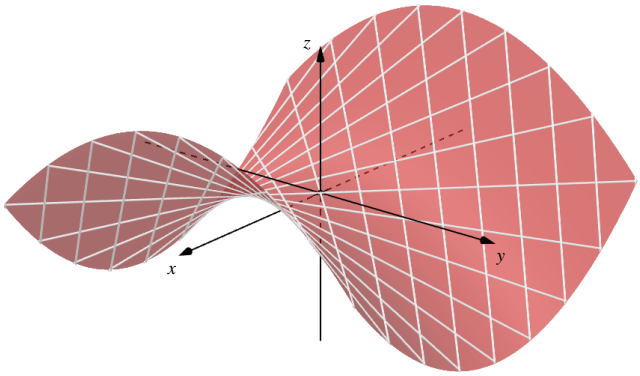


Figura 11.11
Retas no paraboloide hiperbólico.

Devido a esta propriedade, estas superfícies são fáceis de empregar na construção de estruturas pois podem ser concebidas utilizando vigas retas.

12. Forma canónica de Jordan*

12.1. Introdução

Quando no espaço linear $\mathbb{X}$ existe uma base de vetores próprios de uma aplicação linear $A \in \mathcal{L}(\mathbb{X}, \mathbb{X})$, a matriz da aplicação linear nessa base tem a forma diagonal, que é a forma mais simples possível. Mas, como sabemos, a base de vetores próprios pode não existir. Como, então, construir neste caso uma base onde a forma da matriz da aplicação linear é razoavelmente simples? Neste capítulo vamos construir essa base, conhecida como *base de Jordan*, que além dos vetores próprios também contém os chamados *vetores associados*. Nesta base a matriz da aplicação linear tem uma forma diagonal por blocos em que os blocos têm uma estrutura muito simples. Esta forma da matriz é a famosa *forma canónica de Jordan*, que pode ser muito útil na resolução de vários problemas. Por exemplo, com a ajuda da forma canónica de Jordan podemos escrever a solução geral de um sistema de equações diferenciais lineares de coeficientes constantes sem recorrer aos métodos de Análise Matemática como a transformada de Laplace, mas utilizando ideias puramente geométricas, ou seja, através da construção de uma base adequada.

Há muitas demonstrações do Teorema de Jordan sobre a forma canónica. Algumas são muito curtas mas não permitem perceber a natureza deste importante resultado. Nós vamos seguir o caminho mais longo de todos que consiste no estudo detalhado da estrutura dos espaços invariantes de uma aplicação linear. Após este estudo o Teorema de Jordan torna-se praticamente óbvio.

Neste capítulo $\mathbb{X} = \mathbb{C}^n$.

12.2. Teorema Fundamental da Álgebra

Vamos agora demonstrar o Teorema Fundamental da Álgebra sobre a existência de raízes de um polinómio considerado no conjunto dos números complexos. Este teorema já foi utilizado em várias considerações anteriores e é fundamental na demonstração do Teorema de Jordan. Embora conhecido como Teorema Fundamental da Álgebra este teorema é um teorema da Análise Matemática, pois não existem demonstrações deste teorema puramente algébricas. Na verdade, é sempre necessário recorrer, de uma maneira ou outra, ao conceito de continuidade.

Seja

$$P(z) = z^n + a_1 z^{n-1} + \cdots + a_n \tag{12.1}$$

um polinómio de coeficientes complexos com $a_n \neq 0$.

Para demonstrar este teorema, vamos precisar dos seguintes resultados auxiliares.

Lema 12.1

Para todo $A > 0$ existe $r > 0$ tal que $|P(z)| \geqslant A$, sempre que $|z| > r$.

Demonstração. Seja

$$r = \max\left\{1, 2n \max_{k=\overline{1,n}} |a_k|, \sqrt[n]{2A}\right\}.$$

Então, utilizando a desigualdade triangular e a sua consequência imediata, a desigualdade $|a| - |b| \leqslant |a - b|$, para z que verifica a condição $|z| > r$, temos

$$|P(z)| = |z^n + a_1 z^{n-1} + \cdots + a_n| = |z|^n \left|1 + \frac{a_1}{z} + \frac{a_2}{z^2} + \cdots + \frac{a_n}{z^n}\right|$$

$$\geqslant |z|^n \left(1 - \left|\frac{a_1}{z} + \frac{a_2}{z^2} + \cdots + \frac{a_n}{z^n}\right|\right) \geqslant |z|^n \left(1 - n\frac{\max_{k=\overline{1,n}} |a_k|}{|z|}\right) \geqslant \frac{|z|^n}{2} > A.$$

■

Lema 12.2

Seja z_0 um ponto que verifica $P(z_0) \neq 0$. Então, existe z_1 tal que $|P(z_1)| < |P(z_0)|$.

Demonstração. Consideremos primeiro o caso especial quando $P(0) = 1$. Neste caso o polinómio $P(z)$ tem a forma

$$P(z) = 1 + a_{n-k} z^k (1 + H(z)),$$

onde k é um número entre 1 e $n - 1$, ou seja, $a_{n-1} = a_{n-2} = \cdots = a_{n-j+1} = 0$, onde j pertence ao conjunto de $2, \ldots, n$, e o polinómio $H(z)$ verifica a condição $H(0) = 0$. Seja z_1 uma solução da equação

$$z^k = -\delta^k \frac{|a_{n-k}|}{a_{n-k}},$$

onde $\delta > 0$. Notemos que, no conjunto dos números complexos, esta equação tem sempre pelo menos uma solução. É aqui que os números complexos assumem um papel imprescindível. Então, como $|H(z_1)| \leqslant c\delta$, onde $c > 0$ é uma constante, temos

$$|P(z_1)| = |(1 - \delta^k |a_{n-k}|(1 + H(z_1)))| \leqslant |1 - \delta^k |a_{n-k}|| + \delta^{k+1} |a_{n-k}| c$$

$$= 1 - \delta^k |a_{n-k}| + \delta^{k+1} |a_{n-k}| c < 1,$$

sempre que $\delta > 0$ é suficientemente pequeno.

No caso geral, representamos o polinómio $P(z)$ na forma

$$P(z) = ((z - z_0) + z_0)^n + \sum_{k=1}^{n} a_k ((z - z_0) + z_0)^{n-k}$$

$$= (z - z_0)^n + \sum_{k=1}^{n} b^k (z - z_0)^{n-k} = P_0(z - z_0).$$

Obviamente temos

$$b^n = P(z_0) \neq 0.$$

Portanto, introduzindo uma nova variável $\zeta = z - z_0$, obtemos polinómio $Q(\zeta) = \frac{1}{b^n} P_0(\zeta)$ que verifica a condição $Q(0) = 1$. Logo, como acabamos de ver, existe ζ_1 tal que $|Q(\zeta_1)| < |Q(0)| = 1$. Seja $z_1 = \zeta_1 + z_0$. Então temos

$$|P(z_1)| = |P_0(z_1 - z_0)| = |b^n||Q(\zeta_1)| < |b^n||Q(0)| = |P_0(0)| = |P(z_0)|.$$

■

Estamos agora em condições de demonstrar o Teorema Fundamental da Álgebra. O ponto chave desta demonstração é a utilização do Teorema de Weierstrass sobre a existência de mínimo de uma função contínua definida num conjunto fechado e limitado.

Teorema 12.1. (Fundamental da Álgebra)

O polinómio (12.1) tem pelo menos uma raiz complexa.

Demonstração. Fazendo $A = 2|P(0)|$ no Lema 12.1, temos $r > 0$ tal que $|P(z)| > 2|P(0)|$ sempre que $|z| > r$. Pelo Teorema de Weierstrass, a função contínua $|P(z)|$, considerada no disco fechado $D = \{z : |z| \leqslant r\}$, atinge o seu valor mínimo num ponto $z_0 \in D$. Obviamente temos $|P(z_0)| \leqslant |P(0)| < 2|P(0)|$, o que implica que $|P(z_0)| \leqslant |P(z)|$ sempre que $z \notin D$. Portanto, $|P(z_0)| \leqslant |P(z)|$ para qualquer z. Então, aplicando o Lema 12.2 vemos que $P(z_0) = 0$.

■

12.3. Soma direta de subespaços

Seja $\mathbb{X}$ um espaço linear de dimensão n. É fácil verificar que a intersecção de subespaços é um subespaço. Obviamente a dimensão de um subespaço não supera a dimensão do espaço.

Vamos introduzir uma operação importante com conjuntos.

Definição 12.1. Soma geométrica de conjuntos

Consideremos os conjuntos $A_j \subset \mathbb{X}$, $j = \overline{1,m}$. Ao conjunto

$$\sum_{j=1}^{m} A_j = \left\{ x \in \mathbb{X} : x = \sum_{j=1}^{m} a_j, a_j \in A_j \right\}$$

damos o nome de *soma geométrica* dos conjuntos $A_j \subset \mathbb{X}$, $j = \overline{1,m}$.

Obviamente uma soma geométrica de subespaços é um subespaço. Vamos agora definir uma classe especial de somas geométricas de subespaços.

Definição 12.2. Soma direta de subespaços

Vamos dizer que o subespaço $\mathbb{S}$ é uma *soma direta* dos subespaços $\mathbb{X}_1, \mathbb{X}_2, \ldots, \mathbb{X}_m$ se

$$\mathbb{S} = \sum_{j=1}^{m} \mathbb{X}_j \quad \text{e} \quad \mathbb{X}_k \bigcap \sum_{j \neq k} \mathbb{X}_j = \{0\}, \quad k = \overline{1, m}.$$

Neste caso vamos escrever

$$\mathbb{S} = \mathbb{X}_1 \oplus \mathbb{X}_2 \oplus \cdots \oplus \mathbb{X}_m \quad \text{ou} \quad \mathbb{S} = \bigoplus_{j=1}^{m} \mathbb{X}_j.$$

Os seguintes teoremas esclarecem a importância do conceito de soma direta.

Teorema 12.2

Se $\mathbb{X} = \mathbb{X}_1 \oplus \mathbb{X}_2 \oplus \cdots \oplus \mathbb{X}_m$, então para todo $x \in \mathbb{X}$ a representação

$$x = x_1 + x_2 + \cdots + x_m,$$

onde $x_j \in \mathbb{X}_j$, $j = \overline{1, m}$, é única. Em particular, se

$$0 = x_1 + x_2 + \cdots + x_m,$$

onde $x_j \in \mathbb{X}_j$, $j = \overline{1, m}$, então $x_j = 0$, $j = \overline{1, m}$.

Demonstração. Suponhamos que há duas representações

$$x = x_1 + x_2 + \cdots + x_m \quad \text{e} \quad x = y_1 + y_2 + \cdots + y_m,$$

onde $x_j \in \mathbb{X}_j$, $y_j \in \mathbb{X}_j$, $j = \overline{1, m}$. Então temos

$$0 = (x_1 - y_1) + (x_2 - y_2) + \cdots + (x_m - y_m). \tag{12.2}$$

Obviamente

$$x_k - y_k \in \mathbb{X}_k.$$

Por outro lado, de (12.2) vemos que

$$x_k - y_k = \sum_{j \neq k} (y_j - x_j) \in \sum_{j \neq k} \mathbb{X}_j.$$

Logo, temos

$$x_k - y_k \in \mathbb{X}_k \bigcap \sum_{j \neq k} \mathbb{X}_j = \{0\},$$

para todo o k.

■

Demonstremos agora o resultado inverso.

Teorema 12.3

Sejam $\mathbb{X}_j \subset \mathbb{X}$, $j = \overline{1,m}$, subespaços cuja soma geométrica é igual a $\mathbb{X}$. Suponhamos que a igualdade

$$0 = x_1 + x_2 + \cdots + x_m,$$

onde $x_j \in \mathbb{X}_j$ implica $x_j = 0$. Então, $\mathbb{X} = \mathbb{X}_1 \oplus \mathbb{X}_2 \oplus \cdots \oplus \mathbb{X}_m$.

Demonstração. Consideremos um vetor $x_k \in \mathbb{X}_k \bigcap \sum_{j\neq k} \mathbb{X}_j$. Existem vetores $x_j \in \mathbb{X}_j$, $j \neq k$, tais que

$$x_k = x_1 + \cdots + x_{k-1} + x_{k+1} + \cdots + x_m,$$

ou, na forma equivalente,

$$0 = x_1 + \cdots + x_{k-1} - x_k + x_{k+1} + \cdots + x_m.$$

Esta igualdade implica $x_j = 0$, $j = \overline{1,m}$. Portanto, $x_k = 0$. Logo, temos

$$\mathbb{X}_k \bigcap \sum_{j\neq k} \mathbb{X}_j = \{0\}.$$

■

Teorema 12.4

Sejam $\mathbb{X}$ um espaço linear e $\mathbb{M}$ um seu subespaço. Então, existe um subespaço $\mathbb{N}$ tal que $\mathbb{X} = \mathbb{M} \oplus \mathbb{N}$.

Demonstração. Seja $\{e_1, e_2, \ldots, e_m\}$ uma base em $\mathbb{M}$. Pelo Teorema 3.5 existe uma base $\{e_1, \ldots, e_m, e_{m+1} \ldots, e_n\}$ em $\mathbb{X}$. Definamos o subespaço $\mathbb{N}$ como

$$\mathbb{N} = \left\{ \sum_{k=m+1}^{n} \alpha^k e_k : \alpha^k \in \mathbb{C}, k = \overline{m+1,n} \right\}.$$

Como os vetores $\{e_1, \ldots, e_m, e_{m+1} \ldots, e_n\}$ formam uma base em $\mathbb{X}$, para todo o $x \in \mathbb{X}$ existem $\alpha^k \in \mathbb{C}$, $k = \overline{1,n}$, tais que

$$x = \sum_{k=1}^{m} \alpha^k e_k + \sum_{k=m+1}^{n} \alpha^k e_k.$$

Portanto, $x = x_{\mathbb{M}} + x_{\mathbb{N}}$, onde $x_{\mathbb{M}} = \sum_{k=1}^{m} \alpha^k e_k$ e $x_{\mathbb{N}} = \sum_{k=m+1}^{n} \alpha^k e_k$. Consideremos um vetor $y \in \mathbb{M} \cap \mathbb{N}$. Temos

$$y = \sum_{k=1}^{m} \alpha^k e_k = \sum_{k=m+1}^{n} \alpha^k e_k.$$

Portanto

$$0 = \sum_{k=1}^{m} \alpha^k e_k - \sum_{k=m+1}^{n} \alpha^k e_k.$$

Esta igualdade implica $\alpha^k = 0$, $k = \overline{1,n}$. Logo, $y = 0$. O que significa $\mathbb{X} = \mathbb{M} \oplus \mathbb{N}$.

■

12.4. Subespaços invariantes

Sejam $A \in \mathcal{L}(\mathbb{X}, \mathbb{X})$ e $\mathbb{L} \subset \mathbb{X}$ um subespaço invariante de A. Suponhamos que $\dim(\mathbb{L}) = m < n$. Escolhemos uma base $\{e_1, e_2, \ldots, e_m\}$ no subespaço $\mathbb{L}$. Pelo Teorema 3.5 existem vetores $\{e_{m+1}, \ldots, e_n\}$ tais que o sistema $\{e_1, \ldots, e_m, e_{m+1}, \ldots, e_n\}$ é uma base em $\mathbb{X}$. Como $\mathbb{L}$ é um subespaço invariante de A temos

$$Ae_k = \sum_{j=1}^{m} a_j^k e_j, k = \overline{1, m},$$

isto é, a matriz A tem a forma

$$A = \begin{pmatrix} a_1^1 & \cdots & a_1^m & a_1^{m+1} & \cdots & a_1^n \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_m^1 & \cdots & a_m^m & a_m^{m+1} & \cdots & a_m^n \\ 0 & \cdots & 0 & a_{m+1}^{m+1} & \cdots & a_{m+1}^n \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & a_n^{m+1} & \cdots & a_n^n \end{pmatrix}.$$

Inversamente, se a matriz tem esta forma, então o subespaço gerado pelos vetores e_k, $k = \overline{1, m}$, é invariante relativamente a A.

Agora suponhamos que o espaço $\mathbb{X}$ representa-se na forma de uma soma direta de subespaços invariantes da aplicação A, ou seja,

$$\mathbb{X} = \mathbb{X}_1 \oplus \mathbb{X}_2 \oplus \cdots \oplus \mathbb{X}_S.$$

Vamos ver que forma tem a matriz de uma aplicação linear quando o espaço é uma soma direta de subespaços invariantes desta aplicação.

Sejam $\dim(\mathbb{X}_s) = m_s$, $s = \overline{1, S}$. Então, $n = m_1 + m_2 + \cdots + m_S$. Escolhemos uma base $\{e_1, e_2, \ldots, e_n\}$ em $\mathbb{X}$ tal que

$$\{e_1, e_2, \ldots, e_{m_1}\} \subset \mathbb{X}_1,$$

$$\{e_{m_1+1}, e_{m_1+2}, \ldots, e_{m_1+m_2}\} \subset \mathbb{X}_2,$$

$$\vdots$$

$$\{e_{m_1+\cdots+m_{S-1}+1}, e_{m_1+\cdots+m_{S-1}+2}, \ldots, e_{m_1+\cdots+m_{S-1}+m_S}\} \subset \mathbb{X}_S.$$

Nesta base a matriz de A tem uma forma diagonal por blocos, isto é,

$$A = \begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & A_S \end{pmatrix},$$

onde os blocos A_s contêm os elementos a_j^k definidos pelas fórmulas

$$Ae_k = \sum_{j=m_1+\cdots+m_{s-1}+1}^{m_1+\cdots+m_{s-1}+m_s} a_j^k e_j, \quad k = \overline{m_1 + \cdots + m_{s-1} + 1, m_1 + \cdots + m_{s-1} + m_s}.$$

12.5. Forma canónica de uma aplicação nilpotente

A matriz que corresponde à aplicação $A \in \mathcal{L}(\mathbb{X}, \mathbb{X})$ depende da base. Consideremos a questão de escolha da base, tal que a matriz tenha a forma mais simples possível. Nesta secção vamos construir essa base para uma classe especial de aplicações lineares.

Definição 12.3. Aplicação nilpotente

Dizemos que a aplicação linear $A \in \mathcal{L}(\mathbb{X}, \mathbb{X})$ é *nilpotente* se existe um número natural m tal que $A^m = 0$, isto é, $A^m x = 0$ para todo o $x \in \mathbb{X}$.

Seja A uma aplicação nilpotente. Suponhamos que $A^m = 0$ e $A^{m-1} \neq 0$. Consideremos os subespaços

$$\mathbb{M}_k = \{x : A^k x = 0\}, \quad k = \overline{0, m}.$$

Obviamente

$$\{0\} = \mathbb{M}_0 \subset \mathbb{M}_1 \subset \cdots \subset \mathbb{M}_{m-1} \subset \mathbb{M}_m = \mathbb{X}$$

e $\mathbb{M}_{m-1} \neq \mathbb{M}_m$.

Estes subespaços têm representação na forma de somas diretas de subespaços descrita no seguinte teorema.

Teorema 12.5

Existem subespaços $\mathbb{N}_k$, $k = \overline{0, m-1}$, tais que

$$\begin{aligned}
&\mathbb{M}_m = \mathbb{M}_{m-1} \oplus \mathbb{N}_{m-1}, \\
&\mathbb{M}_{m-1} = \mathbb{M}_{m-2} \oplus A\mathbb{N}_{m-1} \oplus \mathbb{N}_{m-2}, \\
&\qquad \vdots \\
&\mathbb{M}_{m-k+1} = \mathbb{M}_{m-k} \oplus A^{k-1}\mathbb{N}_{m-1} \oplus A^{k-2}\mathbb{N}_{m-2} \oplus \cdots \oplus \mathbb{N}_{m-k}, \\
&\qquad \vdots \\
&\mathbb{M}_1 = \mathbb{M}_0 \oplus A^{m-1}\mathbb{N}_{m-1} \oplus A^{m-2}\mathbb{N}_{m-2} \oplus \cdots \oplus \mathbb{N}_0.
\end{aligned} \tag{12.3}$$

Demonstração. Vamos fazer a demonstração por indução. A existência do subespaço $\mathbb{N}_{m-1}$ é uma consequência imediata do Teorema 12.4. Suponhamos que os subespaços $\mathbb{N}_{m-1}, \mathbb{N}_{m-2}, \ldots, \mathbb{N}_{m-k}$, $k < m$, já estão construídos e verificam as primeiras k igualdades (12.3). Como $\mathbb{N}_{m-i} \subset \mathbb{M}_{m-i+1}$, $i = \overline{1, k}$, temos

$$\{0\} = A^{m-i+1}\mathbb{N}_{m-i} = A^{m-k}AA^{k-i}\mathbb{N}_{m-i}.$$

Logo, $AA^{k-i}\mathbb{N}_{m-i} \subset \mathbb{M}_{m-k}$ e, portanto,

$$A\left(\bigoplus_{i=1}^{k} A^{k-i}\mathbb{N}_{m-i}\right) \subset \mathbb{M}_{m-k}.$$

Mostremos que

$$A\left(\bigoplus_{i=1}^{k} A^{k-i}\mathbb{N}_{m-i}\right) = \bigoplus_{i=1}^{k} A^{k-i+1}\mathbb{N}_{m-i}.$$

Obviamente temos

$$A\left(\bigoplus_{i=1}^{k} A^{k-i}\mathbb{N}_{m-i}\right) = \sum_{i=1}^{k} A^{k-i+1}\mathbb{N}_{m-i}.$$

Sejam $x_{m-i} \in \mathbb{N}_{m-i}$, $i = \overline{1,k}$, tais que $0 = \sum_{i=1}^{k} A^{k-i+1}x_{m-i}$. Então, $0 = A\sum_{i=1}^{k} A^{k-i}x_{m-i}$ e como $k < m$ temos $\sum_{i=1}^{k} A^{k-i}x_{m-i} \in \mathbb{M}_1 \subset \mathbb{M}_{m-k}$. Da k-ésima igualdade (12.3), temos $0 = \sum_{i=1}^{k} A^{k-i}x_{m-i}$. Logo, $A^{k-i}x_{m-i} = 0$, $i = \overline{1,k}$. Aplicando A a estas igualdades obtemos $A^{k-i+1}x_{m-i+1} = 0$, $i = \overline{1,k}$, ou seja,

$$\sum_{i=1}^{k} A^{k-i+1}\mathbb{N}_{m-i} = \bigoplus_{i=1}^{k} A^{k-i+1}\mathbb{N}_{m-i}.$$

Mostremos que

$$\mathbb{M}_{m-k-1} \bigcap A\bigoplus_{i=1}^{k} A^{k-i}\mathbb{N}_{m-i} = \{0\}.$$

Sejam $x_{m-i} \in \mathbb{N}_{m-i}$, $i = \overline{1,k}$, e

$$0 = A^{m-k-1}A\sum_{i=1}^{k} A^{k-i}x_{m-i} = A^{m-k}\sum_{i=1}^{k} A^{k-i}x_{m-i},$$

o que implica que

$$\sum_{i=1}^{k} A^{k-i}x_{m-i} \in \mathbb{M}_{m-k}.$$

Logo,

$$\sum_{i=1}^{k} A^{k-i}x_{m-i} = 0 \quad \text{e} \quad A\sum_{i=1}^{k} A^{k-i}x_{m-i} = 0.$$

Obviamente, temos

$$\mathbb{M}_{m-k-1} + A\sum_{i=1}^{k} A^{k-i}\mathbb{N}_{m-i} = \mathbb{M}_{m-k-1} \oplus A^{k}\mathbb{N}_{m-1} \oplus \cdots \oplus A\mathbb{N}_{m-k} \subset \mathbb{M}_{m-k}.$$

Pelo Teorema 12.4 existe um subespaço $\mathbb{N}_{m-k-1}$ tal que

$$\mathbb{M}_{m-k} = \mathbb{M}_{m-k-1} \oplus A^{k}\mathbb{N}_{m-1} \oplus \cdots \oplus A\mathbb{N}_{m-k} \oplus N_{m-k-1}.$$

■

Escolhemos uma base $e_j^{(i)}$, $i = \overline{1,n_j}$, em $\mathbb{N}_j$. Sejam $e_{jl}^{(i)} = A^l e_j^{(i)}$, $l = \overline{0,j}$. Obviamente, cada vetor do subespaço $A^l\mathbb{N}_j$ pode ser representado como uma combinação linear dos vetores $e_{jl}^{(i)}$. Mostremos que estes vetores são linearmente independentes. Com efeito, se $\sum_{i=1}^{n_j} \alpha_i e_{jl}^{(i)} = 0$, então $A^l\sum_{i=1}^{n_j} \alpha_i e_j^{(i)} = 0$. Logo, $\sum_{i=1}^{n_j} \alpha_i e_j^{(i)} \in \mathbb{M}_l \subset \mathbb{M}_j$,

visto que $l \leqslant j$. Por outro lado $\sum_{i=1}^{n_j} \alpha_i e_j^{(i)} \in \mathbb{N}_j$. Como $\mathbb{M}_j \cap \mathbb{N}_j = \{0\}$, temos $\sum_{i=1}^{n_j} \alpha_i e_j^{(i)} = 0$ o que implica $\alpha_i = 0$, $i = \overline{1, n_j}$. Portanto, os vetores $e_{jl}^{(i)}$, $i = \overline{1, n_j}$, formam uma base no subespaço $A^l\mathbb{N}_j$. Logo, os vetores $e_{jl}^{(i)}$, $i = \overline{1, n_j}$, $l = \overline{0, j}$, $j = \overline{0, m-1}$, formam uma base em $\mathbb{X}$. Representemos os vetores desta base na forma da seguinte tabela:

$$\begin{array}{ccccccccccc}
e_{m-1}^{(1)} & \cdots & e_{m-1}^{(n_{m-1})} & & & & & & & & \\
Ae_{m-1}^{(1)} & \cdots & Ae_{m-1}^{(n_{m-1})} & e_{m-2}^{(1)} & \cdots & e_{m-2}^{(n_{m-2})} & & & & & \\
\vdots & \ddots & \vdots & \vdots & & \vdots & \ddots & & & & \\
\vdots & \ddots & \vdots & \vdots & & \vdots & & \ddots & & & \\
A^{m-1}e_{m-1}^{(1)} & \cdots & A^{m-1}e_{m-1}^{(n_{m-1})} & A^{m-2}e_{m-2}^{(1)} & \cdots & A^{m-2}e_{m-2}^{(n_{m-2})} & \cdots & e_0^1 & \cdots & e_0^{(n_0)}
\end{array}$$

Cada coluna desta tabela determina um subespaço invariante da aplicação A. As primeiras n_{m-1} colunas determinam subespaços de dimensão m. As n_{m-2} colunas seguintes determinam subespaços de dimensão $m-1$, etc. Finalmente, as últimas n_0 colunas determinam subespaços da dimensão um. Escrevamos a matriz da aplicação A nesta base. Consideremos os elementos de uma coluna da tabela

$$A^j e_j^{(i)}, A^{j-1}e_j^{(i)}, \ldots, Ae_j^{(i)}, e_j^{(i)},$$

escritos nesta ordem. A aplicação A transforma o primeiro vetor em zero, o segundo é transformado no primeiro, e assim sucessivamente. Portanto, a matriz da restrição da aplicação A ao subespaço gerado pelos vetores $A^l e_j^{(i)}$, $l = \overline{0, j}$, nesta base tem a forma

$$\begin{pmatrix}
0 & 1 & 0 & \cdots & 0 \\
0 & 0 & 1 & \ddots & \vdots \\
0 & 0 & 0 & \ddots & 0 \\
\vdots & \vdots & \vdots & \ddots & 1 \\
0 & 0 & 0 & \cdots & 0
\end{pmatrix}. \tag{12.4}$$

A mesma forma tem a matriz da aplicação A considerada nos subespaços gerados pelas outras colunas da tabela. Donde, na base construída, a matriz da aplicação A tem a forma diagonal de blocos

$$A = \begin{pmatrix}
A_{1,m-1} & & & & & & & & \\
& A_{2,m-1} & & & 0 & & & & \\
& & \ddots & & & & & & \\
& & & A_{n_{m-1},m-1} & & & & & \\
& & & & \ddots & & & & \\
& & & & & A_{1,0} & & & \\
& & 0 & & & & \ddots & & \\
& & & & & & & A_{n_0,0}
\end{pmatrix},$$

onde todos os blocos têm a forma (12.4).

12.6. Álgebra de polinómios

Para demonstrar o Teorema de Jordan vamos precisar de um resultado importante ligado às propriedades de polinómios. Já sabemos que o espaço $\mathcal{L}(\mathbb{X},\mathbb{X})$ é uma álgebra (veja a Secção 5.4). Os polinómios

$$P(\lambda) = \sum_{k=0}^{m} a_k \lambda^k,$$

também formam uma álgebra, isto é, um espaço linear cujos elementos podem ser multiplicados, e esta operação, denotada por "$\cdot$", é associativa e distributiva. A existência dos elementos inversos não é necessária.

Vamos, agora, definir um dos conceitos mais importantes da Álgebra.

Definição 12.4. Ideal

A um subespaço $\mathcal{J}$ da álgebra $\mathcal{A}$ chamamos *ideal* se $x \cdot a \in \mathcal{J}$ para todos $x \in \mathcal{J}$ e $a \in \mathcal{A}$.

Seja $\hat{Q}(\lambda)$ um polinómio. Obviamente o conjunto dos polinómios que têm a forma $P(\lambda)\hat{Q}(\lambda)$, onde $P(\lambda)$ é um polinómio, é um ideal na álgebra de polinómios. Mostremos que todos os ideais nesta álgebra têm esta estrutura.

Teorema 12.6

Seja $\mathcal{J}$ um ideal na álgebra de polinómios $\mathcal{A}$. Então, existe um polinómio $\hat{Q}(\lambda) \in \mathcal{A}$ tal que $\mathcal{J} = \{P(\lambda)\hat{Q}(\lambda) : P(\lambda) \in \mathcal{A}\}$. O polinómio $\hat{Q}(\lambda)$ está determinado a menos de um multiplicador.

Demonstração. Seja $\hat{Q}(\lambda)$ um polinómio de $\mathcal{J}$ que tem o grau menor possível. Então, qualquer polinómio $Q(\lambda) \in \mathcal{J}$ pode ser representado na forma

$$Q(\lambda) = P(\lambda)\hat{Q}(\lambda) + R(\lambda),$$

onde $R(\lambda)$ é um polinómio de grau menor do que o grau do polinómio $\hat{Q}(\lambda)$, o resultado da divisão de $Q(\lambda)$ por $\hat{Q}(\lambda)$. Como $Q(\lambda) \in \mathcal{J}$ e $P(\lambda)\hat{Q}(\lambda) \in \mathcal{J}$, temos

$$R(\lambda) = Q(\lambda) - P(\lambda)\hat{Q}(\lambda) \in \mathcal{J}. \tag{12.5}$$

O grau de $R(\lambda)$ é menor do que o grau de $\hat{Q}(\lambda)$, o que é uma contradição com (12.5). Logo, $R(\lambda) \equiv 0$ e $\mathcal{J} = \{P(\lambda)\hat{Q}(\lambda) : P(\lambda) \in \mathcal{A}\}$.

Seja $\tilde{Q}(\lambda)$ um outro polinómio que verifica $\mathcal{J} = \{P(\lambda)\tilde{Q}(\lambda) : P(\lambda) \in \mathcal{A}\}$. Logo, existem polinómios $\hat{P}(\lambda)$ e $\tilde{P}(\lambda)$ tais que $\tilde{Q}(\lambda) = \hat{P}(\lambda)\hat{Q}(\lambda)$ e $\hat{Q}(\lambda) = \tilde{P}(\lambda)\tilde{Q}(\lambda)$. Daqui vemos que os polinómios $\hat{P}(\lambda)$ e $\tilde{P}(\lambda)$ não dependem de λ, isto é, são números.

■

A identidade de Bézout, bem conhecida para os números primos entre si, é também válida para os polinómios.

Teorema 12.7

Sejam $Q_k(\lambda)$, $k = \overline{1,m}$, polinómios, em que pelo menos um deles é diferente de zero. Suponhamos que o grau do máximo divisor comum destes polinómios é zero. Então, existem polinómios $\hat{P}_k(\lambda)$, $k = \overline{1,m}$, tais que

$$\sum_{k=1}^{m} \hat{P}_k(\lambda) Q_k(\lambda) = 1. \tag{12.6}$$

Demonstração. Consideremos o conjunto

$$\mathcal{J} = \left\{ \sum_{k=1}^{m} P_k(\lambda) Q_k(\lambda) : P_k(\lambda) \in \mathcal{A} \right\}.$$

Obviamente, $\mathcal{J}$ é um ideal em $\mathcal{A}$. Pelo Teorema 12.6, existe um polinómio $\hat{Q}(\lambda) \in \mathcal{J}$ tal que $\mathcal{J} = \{P(\lambda)\hat{Q}(\lambda) : P(\lambda) \in \mathcal{A}\}$. Logo, existem polinómios $R_k(\lambda)$, $k = \overline{1,m}$, tais que $Q_k(\lambda) = R_k(\lambda)\hat{Q}(\lambda)$, $k = \overline{1,m}$. Como o grau do máximo divisor comum dos polinómios $Q_k(\lambda)$ é igual a zero, então o grau do polinómio $\hat{Q}(\lambda)$ é zero. Como $\hat{Q}(\lambda) \in \mathcal{J}$, existem polinómios $\hat{P}_k(\lambda)$, $k = \overline{1,m}$, tais que

$$\hat{Q}(\lambda) = \sum_{k=1}^{m} \hat{P}_k(\lambda) Q_k(\lambda). \tag{12.7}$$

O polinómio $\hat{Q}(\lambda)$ é uma constante diferente de zero. (Caso contrário, o ideal $\mathcal{J}$ é zero.) Dividindo (12.7) por esta constante, obtemos (12.6). ■

Para a construção que vamos fazer a seguir é muito importante a seguinte observação. Seja $A \in \mathcal{L}(\mathbb{X}, \mathbb{X})$. A cada polinómio

$$P(\lambda) = \sum_{k=0}^{m} a_k \lambda^k,$$

corresponde a aplicação linear

$$P(A) = \sum_{k=0}^{m} a_k A^k \in \mathcal{L}(\mathbb{X}, \mathbb{X}). \tag{12.8}$$

Se $P_1(\lambda)$ e $P_2(\lambda)$ são dois polinómios, então à soma (produto) destes polinómios corresponde a soma (produto) das aplicações correspondentes. Portanto, os teoremas aqui demonstrados são válidos para os polinómios de aplicações do tipo (12.8). É exatamente nesta forma que vamos utilizar os resultados desta secção.

12.7. Polinómio anulador de uma aplicação linear

Sejam $A \in \mathcal{L}(\mathbb{X}, \mathbb{X})$ e $Q(\lambda)$ um polinómio. Para o estudo da estrutura de uma aplicação linear é muito importante a seguinte definição.

Definição 12.5. Polinómio anulador

Dizemos que um polinómio $Q(\lambda)$ é polinómio *anulador* da aplicação A se $Q(A) = 0$.

Cada aplicação A tem pelo menos um polinómio anulador.

Teorema 12.8

Seja $A \in L(\mathbb{X}, \mathbb{X})$. Então, existe um polinómio anulador de A.

Demonstração. O espaço $\mathcal{L}(\mathbb{X}, \mathbb{X})$ tem dimensão n^2. Logo, quaisquer n^2+1 elementos deste espaço são linearmente dependentes. Portanto, existem números $c_k \in \mathbb{C}$, $k = \overline{0, n^2}$, tais que, $\sum_{k=0}^{n^2} |c_k| > 0$, e $\sum_{k=0}^{n^2} c_k A^k = 0$, isto é, o polinómio $Q(\lambda) = \sum_{k=0}^{n^2} c_k \lambda^k$ é um anulador da aplicação A. ■

Obviamente, o conjunto de todos os anuladores de uma aplicação $A \in L(\mathbb{X}, \mathbb{X})$ é um ideal na álgebra dos polinómios. Pelo Teorema 12.6, existe um polinómio anulador $\hat{Q}(\lambda)$, determinado a menos de um multiplicador, tal que todos os anuladores têm a forma $P(\lambda)\hat{Q}(\lambda)$, onde $P(\lambda)$ é um polinómio. Ao polinómio $\hat{Q}(\lambda)$, que tem o grau mínimo possível entre os anuladores da aplicação A, chamamos *polinómio anulador mínimo*.

Teorema 12.9

Suponhamos que o anulador $Q(\lambda)$ da aplicação $A \in \mathcal{L}(\mathbb{X}, \mathbb{X})$ se representa na forma $Q(\lambda) = Q_1(\lambda)Q_2(\lambda)$, e o grau do máximo divisor comum dos polinómios $Q_1(\lambda)$ e $Q_2(\lambda)$ é igual a zero. Então, existem subespaços $\mathbb{X}_1$ e $\mathbb{X}_2$ tais que

1. $A\mathbb{X}_1 \subset \mathbb{X}_1$ e $A\mathbb{X}_2 \subset \mathbb{X}_2$,
2. $Q_1(\lambda)$ é anulador da restrição da aplicação A ao subespaço $\mathbb{X}_2$ e $Q_2(\lambda)$ é anulador da restrição da aplicação A ao subespaço $\mathbb{X}_1$,
3. $\mathbb{X} = \mathbb{X}_1 \oplus \mathbb{X}_2$.

Demonstração. Pelo Teorema 12.7, existem polinómios $P_1(\lambda)$ e $P_2(\lambda)$ tais que

$$P_1(\lambda)Q_1(\lambda) + P_2(\lambda)Q_2(\lambda) = 1.$$

Graças à correspondência entre os polinómios e polinómios de aplicações mencionada na secção anterior, temos a igualdade

$$P_1(A)Q_1(A) + P_2(A)Q_2(A) = I. \tag{12.9}$$

Sejam $\mathbb{X}_k = Q_k(A)\mathbb{X}$, $k = 1, 2$. É fácil ver que

$$A\mathbb{X}_k = AQ_k(A)\mathbb{X} = Q_k(A)A\mathbb{X} \subset Q_k(A)\mathbb{X} = \mathbb{X}_k, k = 1, 2.$$

Sejam $x_k \in \mathbb{X}_k$, $k = 1, 2$. Então, existem $y_k \in \mathbb{X}$ tais que $x_k = Q_k(A)y_k$, $k = 1, 2$. Logo,

$$Q_2(A)x_1 = Q_2(A)Q_1(A)y_1 = Q(A)y_1 = 0$$

e

$$Q_1(A)x_2 = Q_1(A)Q_2(A)y_2 = Q(A)y_2 = 0.$$

Finalmente, para todo $x \in \mathbb{X}$, da igualdade (12.9) temos

$$x = Q_1(A)P_1(A)x + Q_2(A)P_2(A)x \in \mathbb{X}_1 + \mathbb{X}_2.$$

Seja $x \in \mathbb{X}_1 \cap \mathbb{X}_2$. Então, $Q_1(A)x = Q_2(A)x = 0$. Da igualdade (12.9) obtemos

$$x = P_1(A)Q_1(A)x + P_2(A)Q_2(A)x = 0.$$

Portanto, temos $\mathbb{X} = \mathbb{X}_1 \oplus \mathbb{X}_2$. ■

12.8. Forma canónica de Jordan

Pelo Teorema Fundamental da Álgebra o polinómio anulador $P(\lambda)$ da aplicação A pode ser representado na forma

$$P(\lambda) = \prod_{k=1}^{K} (\lambda - \lambda_k)^{r_k}, \tag{12.10}$$

onde os números λ_k, $k = \overline{1, K}$, são todos diferentes. Aplicando várias vezes o Teorema 12.9 obtemos o resultado seguinte.

Teorema 12.10

Se o polinómio anulador da aplicação $A : \mathbb{X} \to \mathbb{X}$ tem a forma (12.10), então existem subespaços $\mathbb{X}_k$, $k = \overline{1, K}$, tais que

1. $A\mathbb{X}_k \subset \mathbb{X}_k, \quad k = \overline{1, K}$,
2. $\mathbb{X} = \bigoplus_{k=1}^{K} \mathbb{X}_k$ e
3. $(A - \lambda_k I)^{r_k}\mathbb{X}_k = \{0\}, \quad k = \overline{1, K}$.

Deste teorema vemos que as aplicações $B_k = A - \lambda_k I$, $k = \overline{1, K}$, são nilpotentes nos subespaços $\mathbb{X}_k$, $k = \overline{1, K}$. Como nos subespaços $\mathbb{X}_k$, $k = \overline{1, K}$, temos $A = B_k + \lambda_k I$ e existe uma base onde as aplicações B_k têm matrizes diagonais por blocos com blocos do tipo (12.4), chegamos ao resultado seguinte.

Teorema 12.11. (Jordan)

Existe uma base (base de Jordan) tal que a matriz da aplicação A nesta base tem a forma canónica de Jordan

$$\begin{pmatrix}
\lambda_1 & 1 & 0 & \cdots & 0 & & & & & & & & & & \\
0 & \lambda_1 & 1 & \ddots & \vdots & & & & & & & & & & \\
0 & 0 & \ddots & \ddots & 0 & & & 0 & & & & & 0 & & \\
\vdots & \ddots & \ddots & \ddots & 1 & & & & & & & & & & \\
0 & \cdots & 0 & 0 & \lambda_1 & & & & & & & & & & \\
 & & & & & \lambda_1 & 1 & 0 & \cdots & 0 & & & & & \\
 & & & & & 0 & \lambda_1 & 1 & \ddots & \vdots & & & & & \\
 & & 0 & & & 0 & 0 & \ddots & \ddots & 0 & & & 0 & & \\
 & & & & & \vdots & \ddots & \ddots & \ddots & 1 & & & & & \\
 & & & & & 0 & \cdots & 0 & 0 & \lambda_1 & & & & & \\
 & & & & & & & & & & \ddots & & & & \\
 & & & & & & & & & & & \lambda_k & 1 & 0 & \cdots & 0 \\
 & & & & & & & & & & & 0 & \lambda_k & 1 & \ddots & \vdots \\
 & & 0 & & & & & 0 & & & & 0 & 0 & \ddots & \ddots & 0 \\
 & & & & & & & & & & & \vdots & \ddots & \ddots & \ddots & 1 \\
 & & & & & & & & & & & 0 & \cdots & 0 & 0 & \lambda_k \\
 & & & & & & & & & & & & & & & \ddots
\end{pmatrix}.$$

O polinómio característico da aplicação A não depende da base (veja a Secção 7.3). Calculando o polinómio característico na base de Jordan obtemos

$$\det(A - \lambda I) = \prod_{k=1}^{K} (\lambda_k - \lambda)^{l_{1k} + \cdots + l_{n_k k}},$$

onde l_{ik} são as dimensões dos blocos na forma canónica de Jordan que correspondem ao valor próprio λ_k. O anulador mínimo da aplicação A tem a forma

$$Q(\lambda) = \prod_{k=1}^{K} (\lambda - \lambda_k)^{l_k},$$

onde l_k é a dimensão máxima dos blocos na forma canónica de Jordan que correspondem ao valor próprio λ_k. Portanto, $Q(\lambda)$ é divisor do polinómio característico. Isto demonstra o teorema seguinte.

Teorema 12.12. (Cayley-Hamilton)

O polinómio característico da aplicação A, $\Delta(\lambda) = \det(A - \lambda I)$, é um anulador desta aplicação.

12.9. Solução geral da equação diferencial de coeficientes constantes

Seja $A \in \mathcal{M}_{n\times n}$. Consideremos a equação diferencial

$$\dot{x} = Ax. \tag{12.11}$$

A aplicação A tem K valores próprios λ_k, $k = \overline{1,K}$. A cada um destes valores, na forma canónica de Jordan, correspondem n_k blocos e cada bloco com o índice $i = \overline{1,n_k}$, tem dimensão l_{ik}. Pelo Teorema de Jordan em $\mathbb{X}$ existe uma base formada pelos vetores $v_l^{(ik)}$, onde $k = \overline{1,K}$, $i = \overline{1,n_k}$, $l = \overline{1,l_{ik}}$, que verificam as igualdades

$$\begin{aligned}
&(A-\lambda_k I)v_1^{(ik)} = 0,\\
&(A-\lambda_k I)v_2^{(ik)} = v_1^{(ik)},\\
&\quad\vdots\\
&(A-\lambda_k I)v_{l_{ik}}^{(ik)} = v_{l_{ik}-1}^{(ik)},
\end{aligned}$$

ou, na forma equivalente,

$$\begin{aligned}
&Av_1^{(ik)} = \lambda_k v_1^{(ik)},\\
&Av_2^{(ik)} = \lambda_k v_2^{(ik)} + v_1^{(ik)},\\
&\quad\vdots\\
&Av_{l_{ik}}^{(ik)} = \lambda_k v_{l_{ik}}^{(ik)} + v_{l_{ik}-1}^{(ik)}.
\end{aligned}$$

Os vetores $v_1^{(ik)}$ são os vetores próprios da aplicação A correspondentes aos valores próprios λ_k, e os vetores $v_l^{(ik)}$, $l > 1$, são conhecidos como *vetores associados* ao vetor próprio $v_1^{(ik)}$.

Teorema 12.13

A solução geral da equação (12.11) tem a forma

$$x(t) = \sum_{k=1}^{K} \exp(\lambda_k t) \sum_{i=1}^{n_k} \sum_{l=1}^{l_{ik}} C_l^{(ik)} \left(\frac{v_1^{(ik)}}{(l-1)!} t^{l-1} + \frac{v_2^{(ik)}}{(l-2)!} t^{l-2} + \ldots + \frac{v_{l-1}^{(ik)}}{1!} t + v_l^{(ik)} \right),$$

onde $C_l^{(ik)}$, $k = \overline{1,K}$, $i = \overline{1,n_k}$, $l = \overline{1,l_{ik}}$, são constantes.

Demonstração. Derivando $x(t)$, obtemos

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}x(t) &= \sum_{k=1}^{K}\lambda_k \exp(\lambda_k t)\sum_{i=1}^{n_k}\sum_{l=1}^{l_{ik}} C_l^{(ik)}\left(\frac{v_1^{(ik)}}{(l-1)!}t^{l-1}+\frac{v_2^{(ik)}}{(l-2)!}t^{l-2}+\cdots+v_l^{(ik)}\right)\\
&\quad+\sum_{k=1}^{K}\exp(\lambda_k t)\sum_{i=1}^{n_k}\sum_{l=1}^{l_{ik}} C_l^{(ik)}\left(\frac{v_1^{(ik)}}{(l-2)!}t^{l-2}+\frac{v_2^{(ik)}}{(l-3)!}t^{l-3}+\cdots+v_{l-1}^{(ik)}\right)\\
&= \sum_{k=1}^{K}\exp(\lambda_k t)\sum_{i=1}^{n_k}\sum_{l=1}^{l_{ik}} C_l^{(ik)}\left(\frac{\lambda_k v_1^{(ik)}}{(l-1)!}t^{l-1}+\frac{\lambda_k v_2^{(ik)}+v_1^{(ik)}}{(l-2)!}t^{l-2}+\cdots+\lambda_k v_l^{(ik)}+v_{l-1}^{(ik)}\right)\\
&= A\left(\sum_{k=1}^{K}\exp(\lambda_k t)\sum_{i=1}^{n_k}\sum_{l=1}^{l_{ik}} C_l^{(ik)}\left(\frac{v_1^{(ik)}}{(l-1)!}t^{l-1}+\frac{v_2^{(ik)}}{(l-2)!}t^{l-2}+\cdots+v_l^{(ik)}\right)\right)\\
&= Ax(t).
\end{aligned}$$

■

Consideremos um exemplo.

Exemplo 12.1. Vamos encontrar a solução geral do sistema $\dot{v}=Av$, $v=\begin{pmatrix}x\\y\\z\end{pmatrix}\in\mathbb{R}^3$, onde $A=\begin{pmatrix}1&-1&1\\1&1&-1\\0&-1&2\end{pmatrix}$. O polinómio característico da matriz

$$\det(A-\lambda I)=-\lambda^3+4\lambda^2-5\lambda+2=-(\lambda-1)^2(\lambda-2)$$

tem raízes $\lambda=1$ de multiplicidade 2 e $\lambda=2$. Resolvendo os sistemas $Av=v$ e $Av=2v$, encontramos o vetor próprio $v_1=\begin{pmatrix}1\\1\\1\end{pmatrix}$ que corresponde ao valor próprio $\lambda=1$ e o vetor próprio $v_2=\begin{pmatrix}1\\0\\1\end{pmatrix}$ que corresponde ao valor próprio $\lambda=2$. Não há outros vetores próprios. Procuremos um vetor w associado ao vetor próprio v_1. A equação $v_1+\lambda w=Aw$, ou seja,

$$\begin{pmatrix}1\\1\\1\end{pmatrix}=\begin{pmatrix}-y+z\\x-z\\-y+z\end{pmatrix},$$

tem, por exemplo, a solução $w=\begin{pmatrix}3\\1\\2\end{pmatrix}$. Os vetores v_1, v_2 e w formam uma base em $\mathbb{R}^3$. A solução geral do sistema $\dot{v}=Av$ é

$$v=\begin{pmatrix}x\\y\\z\end{pmatrix}=C_1\begin{pmatrix}1\\1\\1\end{pmatrix}\exp t+C_2\left(t\begin{pmatrix}1\\1\\1\end{pmatrix}+\begin{pmatrix}3\\1\\2\end{pmatrix}\right)\exp t+C_3\begin{pmatrix}1\\0\\1\end{pmatrix}\exp(2t).$$

□

Índice remissivo

Neste livro, escrito para os estudantes de cursos de Matemática Aplicada, Física e Engenharia, a ênfase é dada ao aspeto geométrico da Álgebra Linear. Embora a maior parte do seu conteúdo seja voltada para quem inicia o estudo desta disciplina, leitores de níveis mais avançados também encontrarão aqui ligações importantes entre a Álgebra Linear e outras áreas da Matemática.

UMinho Editora

Universidade do Minho

ISBN 978-989-9074-77-4

9 789899 074774 >