Álgebra Linear

M. Esmeralda Sousa Dias

Versão final: Mar. 2012
1ª versão: Jan. 2009
2ª versão: Fev. 2011

À memória do Professor

José Sousa Pinto

Prefácio

Este trabalho baseia-se na experiência adquirida de ensino da álgebra linear ao longo de vários anos, enquanto docente desta disciplina no Instituto Superior Técnico (Lisboa). Ele destina-se a servir como texto de apoio aos alunos dos cursos de Engenharia, de Física e de Matemática, bem como a todos os estudantes das áreas onde a álgebra linear é instrumental para efeitos de modelação matemática de problemas de carácter científico ou tecnológico.

A principal diferença entre este texto e a maioria dos manuais de língua portuguesa existentes no mercado consiste numa abordagem integrada da álgebra linear, de modo a conter além dos conceitos básicos fundamentais alguns tópicos complementares que entretanto ganharam especial relevo ao nível das aplicações. As matérias aqui tratadas são hoje indispensáveis ao estudante que pretenda dotar-se de conhecimentos suficientemente abrangentes de álgebra linear, tendo por horizonte uma preparação teórica sólida que o habilite a prosseguir estudos pós-graduados nas variadas áreas onde esta disciplina é manifestamente imprescindível.

Ao contrário do que é habitual encontrar-se nos textos didácticos de álgebra linear ao nível dos primeiros anos de um curso universitário, de uma forma geral optámos por demonstrar a quase totalidade dos resultados enunciados neste livro, com raras excepções onde se apresentam apenas esboços de demonstração. Por conseguinte, em princípio, o estudante que deseje efectuar um estudo aprofundado das matérias versadas não necessitará de recorrer a livros de texto mais especializados. O livro é auto-contido, sendo que em cada capítulo se procurou explanar com suficiente detalhe os temas centrais da álgebra linear, tratando depois algumas das suas aplicações. Os conceitos e definições fundamentais são sistematicamente acompanhados de exemplos ilustrativos de modo a facilitar a sua compreensão, sendo incluídos ao longo do texto vários exercícios resolvidos. Para não se quebrar o ritmo da exposição com detalhe excessivo, são ainda propostos outros exercícios que se destinam a enunciar resultados úteis em tópicos subsequentes. São também enfatizados os aspectos geométricos da álgebra linear, apelando aos conhecimentos prévios em dimensões dois e três.

A matéria ultrapassa largamente o que é possível ser leccionado num semestre. No entanto a apresentação está estruturada de modo a facilitar alguma flexibilidade na selecção de tópicos, sem grande quebra de continuidade. Um curso semestral típico cobre as matérias básicas do Capítulos 1 (sistemas de equações lineares e álgebra matricial); Capítulo 2 (determinantes); Capítulo 3 (espaços lineares); Capítulo 4 (valores e vectores próprios); Capítulo 5 (ortogonalidade e produto interno); Capítulo 6 (funções lineares); Capítulo 7 (matrizes ortogonais, unitárias, simétricas e hermitianas) e Capítulo 8 (forma canónica de Jordan).

Entre os temas que são pouco comuns ou habitualmente relegados para cursos posteriores, destacamos: decomposição em valores singulares; grupos de matrizes e suas álgebras de Lie; a exponencial de matrizes; resolução de equações diferenciais lineares e forma canónica de Jordan.

Com frequência o estudante de um primeiro ano de um curso universitário não dispõe de maturidade matemática para lidar confortavelmente com um certo nível de abstracção. Para colmatar essa dificuldade optou-se por abordar a álgebra linear sob o ponto de vista matricial, evitando o formalismo mais abstracto das aplicações lineares. Todavia procurou-se introduzir, de forma lenta e progressiva, a abstracção necessária a uma boa compreensão global da álgebra linear, indispensável ao tratamento de aplicações concretas com alguma relevância.

Nos exercícios apresentados no final de cada capítulo incluem-se frequentemente questões do tipo verdadeiro ou falso, as quais se destinam a proporcionar ao estudante a oportunidade de testar o seu nível de compreensão e assimilação dos conceitos fundamentais, bem como estabelecer relações entre tópicos distintos. As soluções dos problemas com numeração ímpar são apresentadas no final do livro.

Lisboa, 18 de Fevereiro de 2011

Agradecimentos:

A preparação deste texto beneficiou em primeiro lugar da interacção com os meus alunos ao longo dos últimos 10 anos. Devo-lhes um agradecimento pelo interesse crítico com que seguiram as sucessivas versões deste texto.

A todos os colegas que comigo leccionaram esta disciplina e cujos comentários influenciaram decisivamente a organização das matérias, quero expressar também a minha gratidão.

Agradecimentos muito particulares são devidos à Professora Lina Oliveira pelas inúmeras sugestões suscitadas pela leitura da primeira versão, e ainda o ter chamado a minha atenção para um artigo aqui referenciado. Aos Professores

Mário Graça, Ana Maria Martins, José Manuel Ferreira e Heitor Pina, agradeço igualmente os comentários efectuados em versões preliminares do texto.

Quero em especial destacar o meu reconhecimento pelo incansável e excelente trabalho de revisão efectuado pelo meu colega Professor Gustavo Granja que, pelo seu porfiado rigor, comentários e sugestões pertinentes, muito contribuiu para a melhoria deste texto. Os erros e/ou omissões que subsistam são da minha inteira responsabilidade.

À minha família agradeço o apoio incondicional e a compreensão pelas inúmeras horas que este texto lhes subtraiu.

Lisboa, 26 de Março de 2012

Conteúdo

Capítulo 1

Sistemas de Equações Lineares e Matrizes

A resolução de sistemas de equações lineares é uma tarefa comum a diversas áreas de aplicação da matemática, sendo frequente a ocorrência de sistemas lineares com um grande número de equações e de incógnitas em física, química, astronomia, engenharia, economia, etc. Do ponto de vista teórico, a resolução de um sistema de equações lineares não é difícil. Em particular, sistemas com um número reduzido de equações podem ser resolvidos sem recurso a um qualquer método específico. Porém, quando o número de equações é significativo, é natural utilizar-se um procedimento sistemático que possa ser facilmente programável.

Na primeira década do século XIX, Carl Friedrich Gauss[1] apresentou um método geral para a resolução de sistemas de equações lineares designado modernamente por *método de eliminação de Gauss*. Trata-se de um algoritmo bastante simples, o qual permanece como um dos marcos fundamentais da álgebra linear, dada a sua relevância quer do ponto de vista teórico quer do ponto de vista computacional. Neste capítulo, esse método é utilizado não apenas para resolver sistemas lineares, mas também como suporte à introdução de conceitos fundamentais, servindo de motivação para o estudo de matrizes e do espaço $\mathbb{R}^n$.

Neste capítulo são definidas as operações básicas com matrizes à custa de operações com vectores coluna. As operações de adição de vectores e de multiplicação de um vector por um número real são introduzidas como generalizações das operações bem conhecidas com vectores do plano e do espaço. Em $\mathbb{R}^n$ trata-

[1] Johann Carl Friedrich Gauss (ou Gauß), (1777-1855), matemático alemão com inúmeras contribuições em diversas áreas da matemática e física.

se com algum detalhe o conceito de combinação linear de vectores e conjunto gerador, dando especial ênfase aos aspectos geométricos no caso de vectores do plano e do espaço. Estes conceitos são fundamentais, não só na introdução de certas operações com matrizes mas também em estudos subsequentes como, por exemplo, no estudo de espaços lineares (Capítulo 3).

Na Secção 1.4 estudam-se matrizes invertíveis e descreve-se um método para determinar a inversa de uma matriz quadrada, conhecido pela designação de método de eliminação de Gauss-Jordan[2]. Estabelecem-se ainda relações entre matrizes invertíveis e a sua característica, bem como entre matrizes invertíveis e soluções de sistemas lineares.

Finalmente, tendo em vista utilizações posteriores, abordamos outros tópicos, como seja a factorização LU de uma matriz, bem como alguns aspectos relacionados com matrizes particionadas em blocos.

1.1 Sistemas de equações lineares

No contexto deste livro, quando nos referimos a números ou a *escalares* estamos a subentender tratar-se de números reais ou complexos. Comecemos por definir o que se entende por equação linear e por solução.

Definição 1.1. Uma *equação linear* nas *variáveis* (ou incógnitas) $x_1, x_2, \ldots, x_n$ é uma equação do tipo

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = d, \tag{1.1}$$

onde $a_1, \ldots, a_n$ e d são constantes (reais ou complexas).

Os números $a_1, \ldots, a_n$ são designados por *coeficientes* da equação (1.1), e d é designado por *segundo membro* ou *termo independente* dessa equação.

Exemplo 1.1.

$2x - 3y + 5z - 1 = 0$	é uma equação linear nas variáveis x, y, z.
$2x - y^2 + 5z - 1 = 0$	não é uma equação linear.
$\sqrt{2}x + y = 1$	é uma equação linear nas variáveis x, y.
$\sqrt{2x} + y = 1$	não é uma equação linear.
$2x - \operatorname{sen} y = 0$	não é uma equação linear.

[2]Wilhelm Jordan, (1842-1899), geodesista alemão.

♦

Uma solução da equação (1.1) é uma sequência de n números $s_1, s_2, \ldots, s_n$ que satisfaz a equação quando se substitui $x_1 = s_1, x_2 = s_2, \ldots, x_n = s_n$. Nomeadamente, quando se verifica a igualdade

$$a_1 s_1 + a_2 s_2 + \cdots + a_n s_n = d.$$

Esta solução é habitualmente designada pelo n-uplo $(s_1, s_2, \ldots, s_n)$.

Um sistema de equações lineares

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = d_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = d_2 \\ \qquad\qquad\qquad\vdots \\ a_{p1}x_1 + a_{p2}x_2 + \cdots + a_{pn}x_n = d_p \end{cases} \tag{1.2}$$

tem solução $(s_1, s_2, \ldots, s_n)$ se este n-uplo é solução de todas as equações do sistema.

Definição 1.2. O n-uplo $(s_1, s_2, \ldots, s_n)$ é solução da equação linear (1.1) se

$$a_1 s_1 + a_2 s_2 + \cdots + a_n s_n = d.$$

O n-uplo $(s_1, s_2, \ldots, s_n)$ é solução do sistema de equações lineares (1.2) se é solução de todas as equações do sistema.

Ao conjunto de todas as soluções de um sistema chamamos *solução geral*.

Um sistema de equações lineares também se designa simplesmente por *sistema linear*.

Exemplo 1.2. O par $(x_1, x_2) = (1, 2)$ é solução do sistema

$$\begin{cases} x_1 - x_2 = -1 \\ -x_1 + 2x_2 = 3. \end{cases} \tag{1.3}$$

Porém, o par $(2, 1)$ não é uma solução (verifique).

♦

Relembremos que uma recta do plano é definida por uma equação linear em duas variáveis (isto é, $ax + by = d$). Por exemplo, as equações do sistema linear (1.3) definem duas rectas que designamos respectivamente por l_1 e l_2. Uma

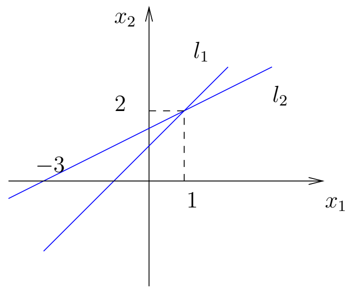


Figura 1.1: O sistema (1.3) tem solução única $(1, 2)$.

solução deste sistema é um ponto comum a l_1 e l_2. Na Figura 1.1 encontram-se representadas estas rectas bem como a (única) solução do sistema.

Geometricamente a solução geral de um sistema de duas equações a duas incógnitas é o conjunto dos pontos de intersecção de duas rectas do plano. Assim, a solução geral de sistemas de duas equações a duas incógnitas é de um dos seguintes três tipos: (i) solução única (as duas rectas intersectam-se num único ponto) e portanto a solução geral é um conjunto com um único elemento; (ii) não existe solução (as rectas são paralelas), pelo que a solução geral é o conjunto vazio; (iii) existem infinitas soluções (as rectas são coincidentes) pelo que a solução geral tem infinitos elementos.

No caso de um sistema linear com um número qualquer de equações e de incógnitas, distinguimos os casos que se seguem.

Um sistema de equações lineares diz-se:

1. *Possível e determinado* se tem uma única solução;
2. *Impossível* se não tem soluções;
3. *Indeterminado* se tem infinitas soluções.

Nota 1. *Como se mostra na Proposição 1.4 (pág. 43), um sistema linear com mais do que uma solução possui uma infinidade de soluções. Conclui-se portanto que os três casos acima esgotam todas as possibilidades.*

1.2 Método de eliminação de Gauss

Para resolver sistemas com várias equações e incógnitas é conveniente usar um algoritmo, ou seja, um procedimento sistemático. O método de eliminação de Gauss é um algoritmo de eleição e baseia-se na substituição de um dado sistema por outro (de mais fácil resolução) que possui exactamente as mesmas soluções. Este novo sistema é obtido aplicando três tipos de operações que eliminam de forma sistemática as variáveis do sistema. As operações permitidas, e que passamos a designar por *operações elementares*, são:

> Operações elementares sobre as equações de um sistema
>
> 1. Trocar a ordem das equações.
> 2. Multiplicar uma equação por uma constante não nula.
> 3. Substituir uma equação pela sua soma com um múltiplo de outra equação.

Note-se que a aplicação de uma operação elementar a um sistema produz um sistema com as mesmas soluções, isto é, um sistema equivalente.

A escolha das operações elementares a aplicar num sistema tem como objectivo eliminar incógnitas das equações por forma a ser possível determinar a solução geral do sistema por substituição regressiva (isto é, determinar as soluções da última equação, substituir essas soluções na penúltima equação e determinar as suas soluções, continuar este processo até à primeira equação). No exemplo seguinte ilustramos quais os critérios habitualmente utilizados na escolha das operações a aplicar.

Exemplo 1.3. Apliquemos operações elementares ao sistema

$$\begin{cases} \phantom{-x_1 + {}} x_2 - 3x_3 = 1 \\ x_1 - 5x_2 + x_3 = 4 \\ -x_1 + 4x_2 - 2x_3 = 1, \end{cases}$$

por forma a obter um sistema com as seguintes características: a terceira equação é uma equação somente na variável x_3, a segunda uma equação nas variáveis x_2 e x_3, e a primeira equação uma equação nas variáveis x_1, x_2 e x_3.

Como a primeira equação não possui a incógnita x_1, vamos trocar a segunda equação com a primeira equação

$$\begin{cases} \phantom{-x_1 + 4x_2 -{}} x_2 - 3x_3 = 1 \\ x_1 - 5x_2 + x_3 = 4 \\ -x_1 + 4x_2 - 2x_3 = 1 \end{cases} \xrightarrow{\text{trocar equação 1 com a 2}} \begin{cases} x_1 - 5x_2 + x_3 = 4 \\ \phantom{-x_1 + 4x_2 -{}} x_2 - 3x_3 = 1 \\ -x_1 + 4x_2 - 2x_3 = 1 \end{cases}$$

A segunda equação já é uma equação nas variáveis x_2 e x_3, como pretendíamos. Teremos agora de eliminar x_1 e x_2 da terceira equação. Para eliminar x_1 da terceira equação, vamos substituir a terceira equação pela sua soma com a primeira

$$\begin{cases} x_1 - 5x_2 + x_3 = 4 \\ \phantom{-x_1 + 4x_2 -{}} x_2 - 3x_3 = 1 \\ -x_1 + 4x_2 - 2x_3 = 1 \end{cases} \xrightarrow{\text{substituir equação 3 pela sua soma com a equação 1}} \begin{cases} x_1 - 5x_2 + x_3 = 4 \\ \phantom{x_1 - 5{}} x_2 - 3x_3 = 1 \\ - x_2 - x_3 = 5 \end{cases}$$

Para eliminar x_2 da terceira equação sem introduzir x_1, temos que usar a segunda equação. Nomeadamente, substituindo a terceira equação pela sua soma com a segunda obtemos o seguinte sistema

$$\begin{cases} x_1 - 5x_2 + x_3 = 4 \\ \phantom{x_1 - 5{}} x_2 - 3x_3 = 1 \\ - x_2 - x_3 = 5 \end{cases} \xrightarrow{\text{substituir equação 3 pela sua soma com a equação 2}} \begin{cases} x_1 - 5x_2 + x_3 = 4 \\ \phantom{x_1 - 5{}} x_2 - 3x_3 = 1 \\ - 4x_3 = 6 \end{cases}$$

Podemos agora determinar a solução do sistema por substituição regressiva. Resolvendo a última equação temos $x_3 = \frac{-3}{2}$. Substituindo este valor na segunda equação obtemos $x_2 = \frac{-7}{2}$ e, finalmente, substituindo x_3 e x_2 na primeira equação obtemos $x_1 = 4 + 5 \times \left(\frac{-7}{2}\right) + \frac{3}{2} = -12$.

O sistema tem solução única:

$$(x_1, x_2, x_3) = \left(-12, \frac{-7}{2}, \frac{-3}{2}\right).$$

♦

É claro do exemplo anterior que a informação essencial de um sistema se encontra nos seus coeficientes e nos termos do segundo membro de cada equação. Esta informação pode ser descrita de forma sucinta construindo um quadro de números dispostos em linhas e colunas onde os coeficientes das variáveis de uma dada equação se situam numa linha, e os coeficientes de uma dada variável do sistema se encontram numa coluna. Este quadro, que se designa por *matriz dos coeficientes* do sistema, permite identificar os coeficientes de cada incógnita em qualquer das equações do sistema. Por exemplo, para o sistema do último exemplo temos

$$\begin{cases} x_2 - 3x_3 = 1 \\ x_1 - 5x_2 + x_3 = 4 \\ -x_1 + 4x_2 - 2x_3 = 1 \end{cases} \rightsquigarrow \begin{bmatrix} 0 & 1 & -3 \\ 1 & -5 & 1 \\ -1 & 4 & -2 \end{bmatrix}$$ (**Matriz dos coeficientes do sistema**).

Na matriz dos coeficientes deste sistema, os coeficientes das variáveis da 1ª equação situam-se na 1ª linha, da 2ª equação na segunda linha e da 3ª equação na terceira linha. Os coeficientes da variável x_1 situam-se na 1ª coluna, da variável x_2 na segunda coluna, e da variável x_3 na terceira coluna.

Como as operações elementares sobre as equações de um sistema também afectam o segundo membro de cada equação, é conveniente aumentar em uma coluna (correspondente ao segundo membro das equações) a matriz dos coeficientes do sistema. A esta matriz chamamos *matriz aumentada* do sistema. Por exemplo,

$$\begin{cases} x_2 - 3x_3 = 1 \\ x_1 - 5x_2 + x_3 = 4 \\ -x_1 + 4x_2 - 2x_3 = 1 \end{cases} \rightsquigarrow \left[\begin{array}{ccc|c} 0 & 1 & -3 & 1 \\ 1 & -5 & 1 & 4 \\ -1 & 4 & -2 & 1 \end{array}\right]$$ (**Matriz aumentada do sistema**).

Note-se que a matriz aumentada de um sistema determina completamente o sistema no sentido em que dada uma tal matriz podemos escrever um sistema correspondente.

Passamos a designar por *matriz* qualquer quadro de números dispostos por linhas e colunas, e por *entradas* os números que aparecem neste quadro.

Exemplo 1.4. O sistema linear nas incógnitas x, y, z e w correspondente à matriz aumentada

$$\left[\begin{array}{cccc|c} 1 & 2 & -5 & 1 & 3 \\ 3 & -1 & 1 & -2 & 5 \\ 4 & -2 & 7 & -3 & 6 \end{array}\right] \quad \text{é} \quad \begin{cases} x + 2y - 5z + w = 3 \\ 3x - y + z - 2w = 5 \\ 4x - 2y + 7z - 3w = 6. \end{cases}$$

♦

As operações elementares sobre as equações de um sistema traduzem-se imediatamente em operações sobre as linhas da matriz aumentada do sistema.

Operações elementares sobre as linhas de uma matriz

1. Trocar linhas.
2. Multiplicar uma linha por uma constante não nula.
3. Substituir uma linha pela sua soma com outra linha multiplicada por uma constante.

Exemplo 1.5. Apliquemos operações elementares sobre as linhas da matriz aumentada do sistema do Exemplo 1.3, a fim de reproduzirmos as operações ali efectuadas.

$$\begin{cases} x_2 - 3x_3 = 1 \\ x_1 - 5x_2 + x_3 = 4 \\ -x_1 + 4x_2 - 2x_3 = 1 \end{cases} \quad \rightsquigarrow \quad \left[\begin{array}{ccc|c} 0 & 1 & -3 & 1 \\ 1 & -5 & 1 & 4 \\ -1 & 4 & -2 & 1 \end{array}\right].$$

Abreviaremos as operações elementares usando a seguinte notação: L_i designa a linha i da matriz; $L_i \leftrightarrow L_j$ designa a troca da linha i com a linha j; αL_i designa a substituição da linha i por essa linha multiplicada por $\alpha \neq 0$; $L_i + \alpha L_j$ designa a substituição da linha i pela sua soma com a linha j multiplicada pelo escalar α. Note-se que nesta notação a linha que é substituída aparece em primeiro lugar na operação indicada, pelo que o significado de $L_i + \alpha L_j$ é distinto de $\alpha L_j + L_i$.

$$\left[\begin{array}{ccc|c} 0 & 1 & -3 & 1 \\ 1 & -5 & 1 & 4 \\ -1 & 4 & -2 & 1 \end{array}\right] \xrightarrow{L_1 \leftrightarrow L_2} \left[\begin{array}{ccc|c} 1 & -5 & 1 & 4 \\ 0 & 1 & -3 & 1 \\ -1 & 4 & -2 & 1 \end{array}\right] \xrightarrow{L_3 + L_1} \left[\begin{array}{ccc|c} 1 & -5 & 1 & 4 \\ 0 & 1 & -3 & 1 \\ 0 & -1 & -1 & 5 \end{array}\right]$$

$$\xrightarrow{L_3 + L_2} \left[\begin{array}{ccc|c} 1 & -5 & 1 & 4 \\ 0 & 1 & -3 & 1 \\ 0 & 0 & -4 & 6 \end{array}\right]$$

Escrevendo o sistema correspondente à matriz (aumentada) resultante, obtemos exactamente o sistema final do Exemplo 1.3. Como operações elementares não alteram as soluções de um sistema, o sistema obtido possui o mesmo conjunto de soluções que o sistema de partida. ◆

Notação:

$L_i \leftrightarrow L_j$	trocar a linha i com a linha j.
αL_i	multiplicar a linha i por $\alpha \neq 0$.
$L_i + \alpha L_j$	substituir a linha i pela soma da linha i com a linha j multiplicada pelo escalar α.

Nota 2. *Para indicar a passagem de uma matriz a outra obtida por meio de uma certa operação elementar sobre as linhas, usou-se uma seta e não o sinal de igualdade. Como veremos adiante, a igualdade de matrizes tem outro significado.*

A questão que naturalmente se coloca é a de saber quando devemos parar de aplicar operações elementares, ou seja, qual deve ser a forma da matriz (aumentada) final. A definição que se segue responde a essa questão.

Definição 1.3. Matriz em escada por linhas

Uma matriz diz-se que está em *escada por linhas*, ou simplesmente em *escada*, se satisfaz as seguintes propriedades:

a) Não possui linhas nulas seguidas de linhas não nulas.

b) A primeira entrada não nula de cada linha encontra-se numa coluna à direita da coluna a que pertence a primeira entrada não nula da linha imediatamente anterior.

Numa matriz em escada por linhas chamamos *pivô* à primeira entrada não nula de cada linha.

Uma matriz diz-se em escada por linhas na *forma reduzida* se está em escada por linhas e além disso satisfaz as condições:

- todos os pivôs são iguais a 1;
- cada pivô é a única entrada não nula da coluna respectiva.

Nota 3. *A definição de matriz em escada por linhas implica que uma tal matriz tenha todas as entradas nulas por baixo de cada pivô.*

Eis um exemplo esquemático de uma matriz em escada por linhas:

$$\begin{bmatrix} 0 & \blacksquare & * & * & * & * & * & * & * & * \\ 0 & 0 & 0 & \blacksquare & * & * & * & * & * & * \\ 0 & 0 & 0 & 0 & \blacksquare & * & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \blacksquare & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \blacksquare & * \end{bmatrix},$$

onde $\blacksquare$ indica um número diferente de zero, 0 indica um zero e $*$ pode ser qualquer número. A correspondente matriz em escada por linhas na forma reduzida é

$$\begin{bmatrix} 0 & 1 & * & 0 & 0 & * & * & 0 & 0 & * \\ 0 & 0 & 0 & 1 & 0 & * & * & 0 & 0 & * \\ 0 & 0 & 0 & 0 & 1 & * & * & 0 & 0 & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & * \end{bmatrix}.$$

Exemplo 1.6. Ilustremos a noção de matriz em escada com alguns exemplos.

- $\begin{bmatrix} 1 & 5 & 0 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$ não está em escada por linhas, visto que tem uma linha nula seguida de uma linha não nula.

- $\begin{bmatrix} 1 & 5 & 0 & 2 \\ 0 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$ não está em escada por linhas, uma vez que a primeira entrada não nula da terceira linha se encontra numa coluna à esquerda da coluna a que pertence a primeira entrada não nula da segunda linha.

- $\begin{bmatrix} 2 & 0 & 4 & 6 & 4 \\ 0 & 0 & 3 & 4 & 5 \\ 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$ está em escada por linhas.

- $\begin{bmatrix} 2 & 5 & 4 & 6 \\ 0 & 0 & 3 & 4 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 2 \end{bmatrix}$ não está em escada por linhas, já que por baixo da primeira entrada não nula da segunda linha há entradas não nulas.

♦

Aplicando operações elementares a uma matriz não nula podemos sempre obter uma matriz em escada na forma reduzida. Se uma matriz está em escada por linhas, a sua forma reduzida obtém-se dividindo cada linha pelo pivô dessa linha, e anulando as entradas acima de cada pivô usando operações elementares. Por exemplo,

$$\underbrace{\begin{bmatrix} 2 & 4 & 6 & 0 & 2 \\ 0 & 1 & 1 & 2 & -2 \\ 0 & 0 & 0 & 2 & 4 \end{bmatrix}}_{\text{escada por linhas}} \underset{L_1/2}{\overset{L_3/2}{\longrightarrow}} \begin{bmatrix} 1 & 2 & 3 & 0 & 1 \\ 0 & 1 & 1 & 2 & -2 \\ 0 & 0 & 0 & 1 & 2 \end{bmatrix} \overset{L_2-2L_3}{\longrightarrow} \begin{bmatrix} 1 & 2 & 3 & 0 & 1 \\ 0 & 1 & 1 & 0 & -6 \\ 0 & 0 & 0 & 1 & 2 \end{bmatrix} \longrightarrow$$

$$\overset{L_1-2L_2}{\longrightarrow} \underbrace{\begin{bmatrix} 1 & 0 & 1 & 0 & 13 \\ 0 & 1 & 1 & 0 & -6 \\ 0 & 0 & 0 & 1 & 2 \end{bmatrix}}_{\text{escada por linhas na forma reduzida}}.$$

Verifica-se facilmente que não é única a matriz em escada resultante da aplicação de operações elementares a uma dada matriz. Por exemplo, se alterarmos a sequência das operações elementares efectuadas na matriz do Exemplo 1.5 obtemos uma matriz em escada distinta:

$$\left[\begin{array}{ccc|c} 0 & 1 & -3 & 1 \\ 1 & -5 & 1 & 4 \\ -1 & 4 & -2 & 1 \end{array}\right] \overset{L_1\leftrightarrow L_3}{\longrightarrow} \left[\begin{array}{ccc|c} -1 & 4 & -2 & 1 \\ 1 & -5 & 1 & 4 \\ 0 & 1 & -3 & 1 \end{array}\right] \overset{L_2+L_1}{\longrightarrow} \left[\begin{array}{ccc|c} -1 & 4 & -2 & 1 \\ 0 & -1 & -1 & 5 \\ 0 & 1 & -3 & 1 \end{array}\right]$$

$$\overset{L_3+L_2}{\longrightarrow} \left[\begin{array}{ccc|c} -1 & 4 & -2 & 1 \\ 0 & -1 & -1 & 5 \\ 0 & 0 & -4 & 6 \end{array}\right].$$

Para a matriz do exemplo referido obtivemos as seguintes matrizes em escada (distintas)

$$\left[\begin{array}{ccc|c} 1 & -5 & 1 & 4 \\ 0 & 1 & -3 & 1 \\ 0 & 0 & -4 & 6 \end{array}\right] \quad \text{e} \quad \left[\begin{array}{ccc|c} -1 & 4 & -2 & 1 \\ 0 & -1 & -1 & 5 \\ 0 & 0 & -4 & 6 \end{array}\right].$$

No entanto, a matriz em escada na forma reduzida correspondente a estas duas matrizes é a matriz

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 12 \\ 0 & 1 & 0 & -7/2 \\ 0 & 0 & 1 & -3/2 \end{array}\right].$$

No Teorema 1.2 (pág. 43), mostra-se que dada uma matriz a respectiva forma reduzida em escada por linhas *é única*. A unicidade da matriz em escada na forma reduzida é um resultado com consequências importantes. Nomeadamente, uma vez que a forma reduzida de uma matriz em escada R tem o mesmo número de pivôs que R, a unicidade da matriz em escada na forma reduzida permite concluir que o número de pivôs de uma matriz em escada obtida de uma matriz A é independente do número e da sequência de operações elementares utilizadas. Definimos a *característica* da matriz A como sendo o número de pivôs de uma qualquer matriz em escada obtida de A por meio de operações elementares sobre as suas linhas.

Definição 1.4. Característica

Chama-se *característica* de uma matriz A ao número de linhas não nulas de uma qualquer matriz em escada obtida de A por meio de operações elementares sobre as suas linhas. Designamos a característica de A por:

$$\operatorname{car} A.$$

A terminologia anglo-saxónica para a característica de uma matriz é "rank".

Exemplo 1.7. (a) Consideremos a matriz aumentada $\left[A \mid b\right]$ do Exemplo 1.5, e a respectiva matriz em escada obtida nesse exemplo:

$$\left[A \mid b\right] = \left[\begin{array}{ccc|c} 0 & 1 & -3 & 1 \\ 1 & -5 & 1 & 4 \\ -1 & 4 & -2 & 1 \end{array}\right] \xrightarrow{\text{Operações elementares}} \left[\begin{array}{ccc|c} 1 & -5 & 1 & 4 \\ 0 & 1 & -3 & 1 \\ 0 & 0 & -4 & 6 \end{array}\right].$$

Como a matriz em escada por linhas possui 3 pivôs, têm-se que $\operatorname{car}\left[A \mid b\right] = 3$. Podemos também concluir que, neste caso, a característica da matriz aumentada $\left[A \mid b\right]$ é igual à característica da matriz A dos coeficientes do sistema, isto é

$$\operatorname{car}\left[A \mid b\right] = 3 \quad \text{e} \quad \operatorname{car} A = 3.$$

Como observámos no Exemplo 1.3, o sistema correspondente é possível e determinado. De facto, como veremos já no próximo exemplo, a igualdade $\operatorname{car}\left[A \mid b\right] = \operatorname{car} A$ é necessária para que o sistema seja possível.

(b) Consideremos agora a matriz aumentada $\left[A \mid b\right] = \left[\begin{array}{cc|c} 1 & -2 & 5 \\ -2 & 4 & 4 \end{array}\right]$. Calculemos a sua característica.

$$\left[A \mid b\right] = \left[\begin{array}{cc|c} 1 & -2 & 5 \\ -2 & 4 & 4 \end{array}\right] \xrightarrow{L_2+2L_1} \left[\begin{array}{cc|c} 1 & -2 & 5 \\ 0 & 0 & 14 \end{array}\right].$$

Donde se conclui que

$$\operatorname{car}\left[A \mid b\right] = 2 \quad \text{e} \quad \operatorname{car} A = 1.$$

Neste caso, a característica da matriz aumentada $[A|b]$ é diferente da característica de A. Como o sistema inicial tem as mesmas soluções que o sistema correspondente à matriz em escada obtida, ou seja, que o sistema

$$\begin{cases} x - 2y = 5 \\ 0 = 14, \end{cases}$$

o sistema original é impossível. ♦

Os exemplos (a) e (b) fazem prever que a existência ou não de solução para um determinado sistema linear está dependente do facto de ser ou não válida a igualdade $\operatorname{car} A = \operatorname{car}[A|b]$ (ver Proposição 1.1 adiante).

Algoritmo para redução de uma matriz a uma matriz em escada

Tendo em vista a construção de um algoritmo facilmente programável, é conveniente efectuar as operações elementares segundo uma determinada ordem, a fim de obter uma matriz em escada. Tal algoritmo recebe usualmente a designação de *método de eliminação de Gauss*. Ilustramos esse algoritmo através do cálculo de uma matriz em escada por linhas para a matriz

$$\begin{bmatrix} 0 & 2 & -6 & 4 & 1 & 2 \\ 3 & 9 & 5 & 1 & 0 & 1 \\ 3 & 7 & 2 & -1 & 2 & 1 \end{bmatrix}.$$

Passo 1:

Selecionar a primeira coluna não nula da matriz. Se o primeiro elemento dessa coluna for não nulo, esse elemento será designado por pivô[a]. Caso contrário, efectua-se troca de linhas colocando um elemento não nulo na primeira linha dessa coluna.

[a] Embora tenhamos definido pivô apenas para matrizes em escada, neste algoritmo usamos a mesma designação de pivô para os elementos que vão ser precisamente os pivôs da matriz em escada final.

$$\begin{bmatrix} 0 & 2 & -6 & 4 & 1 & 2 \\ 3 & 9 & 5 & 1 & 0 & 1 \\ 3 & 7 & 2 & -1 & 2 & 1 \end{bmatrix} \xrightarrow{L_1 \leftrightarrow L_2} \begin{bmatrix} 3 & 9 & 5 & 1 & 0 & 1 \\ 0 & 2 & -6 & 4 & 1 & 2 \\ 3 & 7 & 2 & -1 & 2 & 1 \end{bmatrix}$$

Passo 2:

Usar operações elementares para obter zeros abaixo do pivô selecionado no passo anterior.

$$\begin{bmatrix} 3 & 9 & 5 & 1 & 0 & 1 \\ 0 & 2 & -6 & 4 & 1 & 2 \\ 3 & 7 & 2 & -1 & 2 & 1 \end{bmatrix} \xrightarrow{L_3-L_1} \begin{bmatrix} 3 & 9 & 5 & 1 & 0 & 1 \\ 0 & 2 & -6 & 4 & 1 & 2 \\ 0 & -2 & -3 & -2 & 2 & 0 \end{bmatrix}$$

Passo 3:

Repetir os passos 1 e 2 para a matriz que se obtém suprimindo a linha e coluna que contêm o pivô.

$$\begin{bmatrix} 3 & 9 & 5 & 1 & 0 & 1 \\ 0 & 2 & -6 & 4 & 1 & 2 \\ 0 & -2 & -3 & -2 & 2 & 0 \end{bmatrix} \xrightarrow{L_3+L_2} \begin{bmatrix} 3 & 9 & 5 & 1 & 0 & 1 \\ 0 & 2 & -6 & 4 & 1 & 2 \\ 0 & 0 & -9 & 2 & 3 & 2 \end{bmatrix}$$

pivôs

A matriz final é uma matriz em escada por linhas.

A aplicação do método de eliminação de Gauss para a resolução de sistemas é a seguir resumido.

Método de Eliminação de Gauss

A resolução de um sistema de equações lineares pelo método de eliminação de Gauss consiste em:

1. Escrever a matriz aumentada do sistema.

2. Por meio de operações elementares sobre as linhas da matriz aumentada, obter uma matriz em escada de acordo com o algoritmo anteriormente descrito.

3. Escrever o sistema correspondente à matriz (aumentada) em escada. Resolver este sistema por substituição (regressiva) começando pela última equação.

Ilustremos o método de eliminação de Gauss através de alguns exemplos.

Exemplo 1.8. Usemos o método de eliminação de Gauss para resolver o sistema seguinte.

$$\begin{cases} x_1 - 3x_2 - x_3 \phantom{{}-4x_4} = -2 \\ x_2 + x_3 - 4x_4 = 1 \\ \phantom{x_1 - 3x_2 +{}} x_3 + 9x_4 = 4. \end{cases}$$

A matriz aumentada do sistema

$$\left[\begin{array}{cccc|c} 1 & -3 & -1 & 0 & -2 \\ 0 & 1 & 1 & -4 & 1 \\ 0 & 0 & 1 & 9 & 4 \end{array}\right] \tag{1.4}$$

já está em escada por linhas. Logo, resolvendo a última equação em ordem a x_3, vem $x_3 = 4 - 9x_4$. Resolvendo a segunda equação em ordem a x_2 e substituindo x_3, obtém-se $x_2 = 1 - x_3 + 4x_4 = -3 + 13x_4$. Substituindo x_2 e x_3 na primeira equação, resulta

$$x_1 = -2 + 3\,(-3 + 13x_4) + (4 - 9x_4) = -7 + 30x_4.$$

O sistema possui infinitas soluções e a sua solução geral é o conjunto

$$\{(x_1, x_2, x_3, x_4) : \ x_1 = -7 + 30x_4, \ \ x_2 = -3 + 13x_4, \ \ x_3 = 4 - 9x_4 \text{ e } x_4 \in \mathbb{R}\} =$$

$$= \{(-7 + 30x_4, -3 + 13x_4, 4 - 9x_4, x_4) : x_4 \in \mathbb{R}\}.$$

Uma vez que podemos atribuir livremente valores a x_4, esta variável parametriza o conjunto das soluções. Ou seja, a cada valor do parâmetro x_4 corresponde uma solução do sistema. ♦

Exemplo 1.9. Usemos o método de eliminação de Gauss para resolver o sistema

$$\begin{cases} x - y \phantom{{}+z} = 2 \\ x \phantom{{}-2y} + z = 1 \\ -2x + 2y \phantom{{}+z} = 2. \end{cases}$$

Reduza-se a matriz aumentada do sistema a uma matriz em escada.

$$\left[\begin{array}{ccc|c} 1 & -1 & 0 & 2 \\ 1 & 0 & 1 & 1 \\ -2 & 2 & 0 & 2 \end{array}\right] \xrightarrow{L_2 - L_1} \left[\begin{array}{ccc|c} 1 & -1 & 0 & 2 \\ 0 & 1 & 1 & -1 \\ -2 & 2 & 0 & 2 \end{array}\right] \xrightarrow{L_3 + 2L_1} \left[\begin{array}{ccc|c} 1 & -1 & 0 & 2 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 6 \end{array}\right].$$

O sistema correspondente a esta matriz em escada é

$$\begin{cases} x - y = 2 \\ \phantom{x-{}} y + z = -1 \\ \phantom{x-y+{}} 0 = 6. \end{cases}$$

Da última equação concluímos que o sistema é impossível. Saliente-se que a característica da matriz aumentada do sistema é 3 enquanto a característica da matriz dos coeficientes do sistema é 2. ♦

1.2.1 A solução geral de um sistema

Como vimos, um sistema linear ou é possível ou é impossível. No caso de ser possível ou possui solução única e diz-se *possível e determinado*, ou tem um número infinito de soluções sendo por isso *indeterminado* (ver Proposição 1.4 na página 43). Uma condição necessária e suficiente[3] para um sistema ser possível é que a matriz aumentada do sistema e a matriz dos coeficientes do sistema tenham a mesma característica. Enunciamos este resultado na proposição a seguir.

Proposição 1.1. Seja $[A|b]$ a matriz aumentada de um sistema de equações lineares. O sistema é possível se e só se

$$\operatorname{car}[A|b] = \operatorname{car} A.$$

Demonstração. [4] Comecemos por mostrar que se $\operatorname{car}[A|b] > \operatorname{car}(A)$ o sistema é impossível. Se $\operatorname{car}[A|b] > \operatorname{car}(A)$, o método de eliminação de Gauss aplicado a $[A|b]$ dá origem a uma matriz em escada que tem uma linha com todas as entradas nulas excepto a entrada mais à direita. O sistema correspondente à matriz em escada obtida possui uma equação do tipo $0 = k$ com $k \neq 0$, por conseguinte o respectivo sistema é impossível. Como o sistema original e o sistema correspondente à matriz em escada por linhas têm as mesmas soluções, o sistema original é impossível.

[3] Uma proposição P diz-se uma condição necessária e suficiente para a proposição Q se são satisfeitas as implicações: $P \implies Q$ e $Q \implies P$. Quando se verifica esta dupla implicação, ou seja, "se P, então Q" e "se Q, então P", resume-se este facto numa única proposição, a saber, "P se e só se Q". A expressão "P se e só se Q" (ou abreviadamente "P sse Q") pode também escrever-se na forma $P \iff Q$, traduzindo que P é equivalente a Q.

[4] Assume-se implicitamente nesta demonstração que a característica de uma matriz é um número bem definido, resultado que segue do Teorema 1.2. Não há contudo qualquer erro lógico em demonstrar esta proposição antes desse teorema, uma vez que a prova do teorema é independente desta demonstração.

Para a implicação recíproca, atenda-se a que a característica da matriz aumentada é sempre maior ou igual do que a característica da matriz dos coeficientes. Como por hipótese o sistema é possível, a característica da matriz aumentada não pode ser maior que a característica da matriz dos coeficientes. Logo, se o sistema é possível, então $\mathrm{car}[A|b] = \mathrm{car}(A)$.

□

A solução geral de um sistema é o conjunto de todas as soluções do sistema. Assim, um sistema impossível (sem soluções) tem para solução geral o conjunto vazio, enquanto que um sistema possível e determinado tem para solução geral um conjunto com um único elemento. Quando um sistema é indeterminado, isto é com infinitas soluções, a descrição do conjunto das soluções é mais complexa.

No Exemplo 1.8 o sistema é indeterminado e optámos por resolver a última equação em ordem a x_3 (isto é, $x_3 = 4 - 9x_4$), obtendo x_3 em função de x_4. As soluções do sistema obtêm-se atribuindo um valor arbitrário a x_4 e substituindo esse valor nas igualdades obtidas para as restantes variáveis. Dizemos então que x_4 é uma variável *livre* e que as restantes são *dependentes*. No entanto, poderíamos ter optado por resolver a última equação em ordem a x_4, obtendo-se x_4 dependente de x_3, nomeadamente $x_4 = 4/9 - x_3/9$. Neste caso, o conjunto solução geral seria descrito em termos de x_3, ou seja, x_3 poderia ser qualquer número real, e as restantes variáveis dependeriam de x_3. Este exemplo mostra que há várias maneiras de descrever a solução geral, conforme as variáveis que escolhamos para descrever (ou *parametrizar*) os elementos do conjunto das soluções.

É habitual introduzir-se alguma nomenclatura que corresponde a uma escolha criteriosa das variáveis usadas na parametrização do conjunto das soluções de um sistema. Relembremos que na notação matricial, cada coluna da matriz dos coeficientes de um sistema está associada a uma variável. No Exemplo 1.8 a solução geral foi descrita em termos da variável x_4, nomeadamente a solução geral é o conjunto

$$\{(-7 + 30x_4, -3 + 13x_4, 4 - 9x_4, x_4) : x_4 \in \mathbb{R}\}.$$

Observando a matriz em escada em (1.4), verifica-se que a variável x_4 é a variável associada à coluna dessa matriz que não contém qualquer pivô e que x_1, x_2 e x_3 são variáveis associadas a colunas com um pivô.

Definição 1.5. Numa matriz em escada por linhas obtida por eliminação de Gauss da matriz dos coeficientes de um sistema possível, designamos por:

i) variáveis *livres* ou *independentes*, as variáveis correspondentes a colunas sem pivô;

ii) variáveis *dependentes*, as variáveis correspondentes a colunas com pivô.

Chamamos *grau de indeterminação* de um sistema (possível) ao número de variáveis livres.

Note-se que o grau de indeterminação de um sistema possível é igual ao número de variáveis do conjunto das soluções que podem tomar qualquer valor, ou seja, o número de *parâmetros* necessários para descrever a solução geral.

Como o número de colunas de uma matriz em escada é igual à soma do número de colunas sem pivô com número de colunas com pivô, da definição anterior e da definição de característica de uma matriz (Definição 1.4), seguem os resultados que resumimos na próxima proposição.

Proposição 1.2. Seja A a matriz dos coeficientes de um sistema possível com n incógnitas. Então,

1) $\text{car}(A) =$ número de variáveis dependentes.

2) $n - \text{car}(A) =$ número de variáveis livres = grau de indeterminação do sistema.

Demonstração. Os itens 1) e 2) são imediatos usando as definições de característica e de variáveis livres. □

Exemplo 1.10. No Exemplo 1.8 temos $\text{car}[A|b] = \text{car}[A]$ e o grau de indeterminação do sistema é 1 (a matriz A só tem uma coluna sem pivô, ou seja, o sistema correspondente tem uma variável livre).

No Exemplo 1.9, como $\text{car}[A|b] > \text{car}(A)$, o sistema é impossível.

No Exemplo 1.2 verifica-se que uma matriz (aumentada) em escada é

$$\left[\begin{array}{cc|c} 1 & -1 & -1 \\ 0 & 1 & 2 \end{array}\right].$$

Logo, $\text{car}[A|b] = \text{car}[A] = 2$ e não existem variáveis livres, uma vez que o número de variáveis do sistema é igual ao número de pivôs. Por conseguinte, o sistema é possível e determinado. ♦

1.3 Álgebra de matrizes

Vimos que o uso de matrizes na resolução de sistemas representa uma economia de notação considerável. No entanto, a utilização de matrizes é útil num contexto muito mais vasto. Em particular, a manipulação algébrica de matrizes permite demonstrar facilmente resultados importantes, e tem aplicações que vão muito além da resolução de sistemas. Nesta secção definimos as operações usuais com matrizes e estabelecemos certas identificações envolvendo matrizes e vectores.

Comecemos por precisar alguns conceitos já usados.

Definição 1.6. Uma *matriz* do tipo $p \times n$ é um quadro de números com p linhas e n colunas. Cada número na matriz é designado por *entrada*.

Se $p \neq n$ a matriz diz-se *rectangular* e se $p = n$ a matriz diz-se *quadrada* e de *ordem* n.

As matrizes são habitualmente designadas por letras maiúsculas, por exemplo A, e cada entrada pela correspondente letra minúscula indexada por dois índices que representam a posição da entrada na matriz. Assim, a_{ij} é o número que se encontra na linha i e na coluna j da matriz A. Por exemplo, a matriz

$$A = \begin{bmatrix} 2 & 3 & 4 \\ 1 & 0 & 3 \end{bmatrix}$$

é uma matriz com duas linhas e três colunas e portanto é do tipo 2×3. A entrada situada na primeira linha e na segunda coluna é $a_{12} = 3$, enquanto que a entrada na segunda linha e na primeira coluna é $a_{21} = 1$. O primeiro índice em a_{ij} designa a linha e o segundo a coluna.

Uma matriz A do tipo $p \times n$ (com p linhas e n colunas), de entradas a_{ij} é representada por

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pn} \end{bmatrix} = [a_{ij}]_{\substack{i=1,\ldots,p \\ j=1,\ldots,n}}$$

Exemplo 1.11. A matriz $A = [a_{ij}]_{\substack{i=1,2,3 \\ j=1,2}}$ definida por

$$a_{ij} = \begin{cases} i-1 & \text{se } i > j \\ 0 & \text{se } i = j \\ j+2 & \text{se } i < j, \end{cases}$$

é a seguinte matriz 3×2:

$$A = \begin{bmatrix} 0 & 4 \\ 1 & 0 \\ 2 & 2 \end{bmatrix}.$$

♦

Numa matriz quadrada A do tipo $n \times n$, chamamos *diagonal principal* de A à diagonal constituída pelas entradas $a_{11}, a_{22}, \ldots, a_{nn}$.

Definição 1.7. A uma matriz com uma única coluna (isto é, do tipo $p \times 1$) chamamos *vector coluna*. As entradas de um vector coluna são designadas por *componentes* ou *coordenadas*.

Os vectores coluna são denotados por letras minúsculas a cheio, e por vezes com uma seta em cima.

Até ao início da próxima secção, onde se clarificam as relações entre vectores (geométricos) e vectores coluna, abreviamos vector coluna para vector.

Eis um exemplo de um vector,

$$\mathbf{v} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}, \text{ ou } \quad \vec{\mathbf{v}} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}, \quad \text{é um vector cuja segunda componente é 3.}$$

Soluções de equações lineares podem ser identificadas com vectores. Nomeadamente, uma solução $(s_1, \ldots, s_n)$ da equação linear $a_1x_1 + a_2x_2 + \cdots + a_nx_n = d$ pode escrever-se como o vector

$$\mathbf{s} = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix}$$

que satisfaz a igualdade $a_1s_1 + a_2s_2 + \cdots + a_ns_n = d$.

Sempre que conveniente, também escrevemos $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ para designar o vector coluna

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

Nota 4. *a) Os parênteses habitualmente utilizados para delimitar as entradas de uma matriz são parênteses rectos ou curvos. As chavetas e os traços verticais não se usam para este efeito pois estão reservados para denotar os conceitos de conjunto e de determinante (a estudar no próximo capítulo).*

b) Por vezes usaremos a designação de tamanho quando nos referimos ao tipo da matriz.

c) Noutros livros o termo vector é também usado para designar matrizes com uma única linha mas aqui segue-se outra convenção.

Se encararmos uma matriz como um conjunto (ordenado) de vectores coluna é natural definirmos operações com matrizes baseando-as em operações com vectores coluna. Esta é a estratégia que seguiremos na definição de operações com matrizes, sendo por isso necessário começar por estudar operações com vectores colunas.

Como veremos de seguida, um vector coluna de duas ou três componentes reais representa um vector geométrico, respectivamente, no plano e no espaço. Um vector com $n > 3$ componentes (isto é, um vector de $\mathbb{R}^n$) generaliza os vectores geométricos. As operações com vectores coluna serão definidas como generalizações das correspondentes operações com vectores do plano ou do espaço.

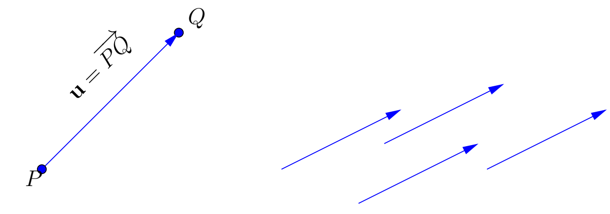
1.3.1 Vectores de $\mathbb{R}^n$

Na representação de certas grandezas físicas tais como a velocidade, a aceleração ou a força, é necessário ter em conta não só a intensidade mas também uma direcção e sentido segundo os quais essa grandeza se manifesta. Este tipo de grandezas físicas são grandezas representadas por vectores.

Nesta secção introduzimos geometricamente a noção de vector, estabelecemos relações entre vectores, pontos, e vectores em sistemas de coordenadas. As noções estudadas no plano e no espaço são depois generalizadas para vectores de $\mathbb{R}^n$.

Vectores e geometria

Dois pontos (distintos) P e Q, definem um segmento de recta. Tomando P como ponto inicial e Q como ponto final, o segmento de recta fica orientado no sentido de P para Q. A notação $\overrightarrow{PQ}$ designa esse segmento orientado.



(a) (b)

Figura 1.2: (a) O vector $\mathbf{u}$ é representado pelo segmento orientado $\vec{PQ}$. (b) Vectores equivalentes.

Segmentos orientados paralelos, com o mesmo sentido e o mesmo comprimento dizem-se *segmentos orientados equivalentes*. Isto é, dois segmentos orientados são equivalentes se coincidem quando se desloca um deles, paralelamente a si próprio, por forma a que os pontos iniciais dos dois segmentos orientados coincidam (ver Figura 1.2-(b)).

Frequentemente interessa considerar uma grandeza vectorial, independente do ponto do espaço em que a grandeza se manifesta, sendo apenas relevantes a sua intensidade e sentido. Uma maneira de formalizar este conceito consiste em definir um *vector*, ou *vector livre*, como o conjunto de *todos* os segmentos orientados equivalentes a um dado segmento orientado. O *vector nulo* é definido como o conjunto dos segmentos degenerados (segmentos com ponto inicial e final coincidentes).

Quando se representa um vector por um segmento orientado particular (trata-se de representar um conjunto por um dos seus elementos) diz-se que o vector se encontra aplicado no ponto inicial do segmento orientado que o representa (ver Figura 1.2-(a)).

A seguir definimos a soma de dois vectores bem como a multiplicação de um vector por um número real.

Definição 1.8. A soma $\mathbf{u}+\mathbf{v}$ de dois vectores $\mathbf{u}$ e $\mathbf{v}$ é o vector representado pelo segmento orientado obtido da seguinte forma: considerem-se representantes dos vectores $\mathbf{u}$ e $\mathbf{v}$ tais que o ponto inicial do segmento orientado que representa o vector $\mathbf{v}$ coincide com ponto final do segmento orientado que representa $\mathbf{u}$, o vector $\mathbf{u}+\mathbf{v}$ é o vector representado pelo segmento orientado cujo ponto inicial é o ponto inicial do representante de $\mathbf{u}$, e o ponto final é o ponto final do representante de $\mathbf{v}$ (ver Figura 1.3).

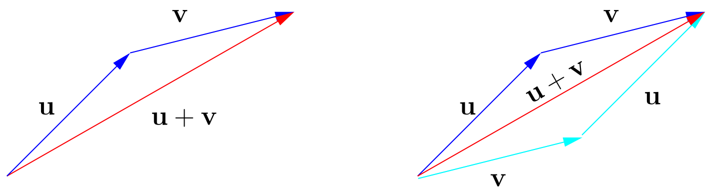


Figura 1.3: A soma $\mathbf{u}+\mathbf{v}$ e $\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$.

Na Figura 1.3 verificamos geometricamente que a adição de vectores é comutativa, isto é,

$$\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}.$$

Definição 1.9. Se $\mathbf{u}$ é um vector não nulo e k um número real, então $k\mathbf{u}$ é o vector representado pelo segmento orientado cujo comprimento é $|k|$ vezes o comprimento do segmento orientado que representa $\mathbf{u}$, e o sentido é o desse segmento se $k>0$, e é o oposto se $k<0$. Definimos $k\mathbf{u}$ como sendo o vector nulo (ou zero) $\mathbf{0}$ se $k=0$ ou $\mathbf{u}$ é o vector nulo.

Na Figura 1.4 é ilustrada a multiplicação de um vector por um número real.

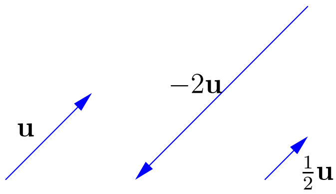


Figura 1.4: O produto de um número real por um vector.

Como consequência da definição de soma de vectores e das propriedades da multiplicação por um escalar, o simétrico $(-\mathbf{v})$ de um vector $\mathbf{v}$ é o vector que se obtém multiplicando $\mathbf{v}$ pelo escalar (-1).

Podemos definir diferença de dois vectores como sendo a seguinte soma:

$$\mathbf{u}-\mathbf{v}=\mathbf{u}+(-\mathbf{v}).$$

Como se ilustra na Figura 1.5, a construção geométrica de $\mathbf{u}-\mathbf{v}$ pode efectuar-se do seguinte modo: considerem-se representantes de $\mathbf{u}$ e de $\mathbf{v}$ com o mesmo ponto inicial, o vector $\mathbf{u}-\mathbf{v}$ é representado pelo segmento orientado cujo ponto inicial

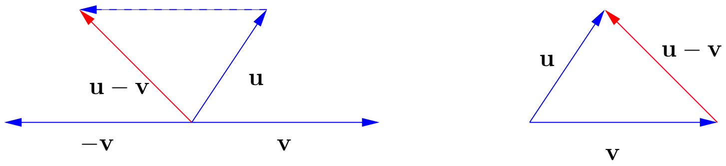


Figura 1.5: O vector diferença $\mathbf{u} - \mathbf{v}$.

é o ponto terminal do representante de $\mathbf{v}$ e cujo ponto final é o ponto final do segmento orientado que representa $\mathbf{u}$ (ver Figura 1.5).

Fixando um ponto O no espaço, um qualquer vector livre $\mathbf{u}$ pode sempre representar-se por um (e um só) segmento orientado cujo ponto inicial é o ponto O. Em particular, o vector nulo é representado pelo ponto O. Existe portanto uma correspondência biunívoca entre vectores (livres) e segmentos orientados com ponto inicial O. Além disso, uma vez fixado um ponto O do espaço, pode estabelecer-se uma correspondência biunívoca entre vectores (livres) e pontos do espaço. Esta correspondência é estabelecida do seguinte modo: ao vector $\mathbf{u}$ representado pelo segmento orientado $\overrightarrow{OP}$ corresponde o ponto P; reciprocamente, ao ponto P corresponde o conjunto de segmentos orientados equivalentes a $\overrightarrow{OP}$. Assim, fixado um ponto do espaço, temos a correspondência biunívoca

$$\text{Vectores} \longleftrightarrow \text{Pontos.}$$

Vectores em sistemas de coordenadas

Num plano podemos fixar um referencial, considerando para tal um ponto que se toma como origem e duas rectas não paralelas (eixos coordenados) que se intersectam na origem. Estas rectas são munidas duma unidade de comprimento e de um sentido.

Um ponto do plano fica completamente caracterizado se forem especificadas as suas coordenadas, isto é, se for dado o par ordenado (a, b) que indica que o ponto se projecta (paralelamente aos eixos coordenados) nos eixos coordenados em pontos que se situam, respectivamente, a a e b unidades (positivas ou negativas) da origem.

Cada vector livre do plano pode representar-se por um único segmento orientado cujo ponto inicial é a origem O do referencial fixado.

A correspondência biunívoca entre pontos do plano e vectores, permite identificar cada ponto do plano de coordenadas (a, b) com o segmento orientado cujo ponto inicial é a origem do referencial e cujo ponto final é (a, b). Ou seja, cada ponto pode identificar-se com um vector aplicado na origem do referencial. Este vector é designado por *vector de posição* do ponto (Figura 1.6).

Denotamos pontos por letras maiúsculas e os respectivos vectores de posição pela correspondente letra minúscula (a cheio). Isto é, para o ponto P o vector de posição respectivo é denotado por $\mathbf{p}$.

Existe portanto uma identificação natural entre o conjunto de vectores do plano e o conjunto dos vectores coluna de duas componentes reais, o qual se designa por $\mathbb{R}^2$. Isto é,

$$\mathbb{R}^2 = \left\{ \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} : x_1 \in \mathbb{R} \text{ e } x_2 \in \mathbb{R} \right\} = \{(x_1, x_2) : x_1 \in \mathbb{R} \text{ e } x_2 \in \mathbb{R}\}.$$

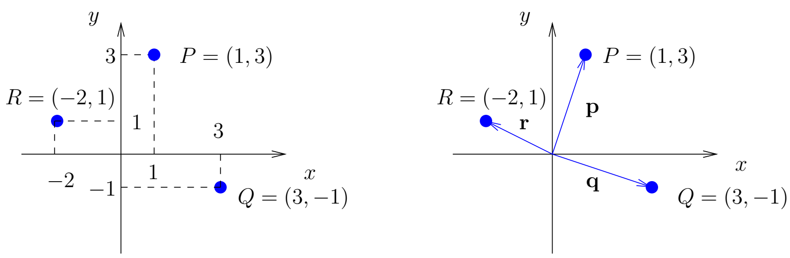


Figura 1.6: Representação de pontos e respectivos vectores de posição no referencial de eixos x e y.

Mediante a identificação descrita acima entre vectores do plano e vectores coluna, a soma de vectores do plano corresponde à soma componente a componente de vectores coluna. Ou seja,

$$\mathbf{u} + \mathbf{v} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \end{bmatrix}, \qquad \mathbf{u} \in \mathbb{R}^2,\ \mathbf{v} \in \mathbb{R}^2.$$

Geometricamente, o vector soma $\mathbf{u} + \mathbf{v}$ é o vector de posição de um ponto que é o vértice de um paralelogramo cujos outros vértices são a origem e os pontos correspondentes aos vectores de posição $\mathbf{u}$ e $\mathbf{v}$ (ver Figura 1.7).

A multiplicação de um vector $\mathbf{u}$ de $\mathbb{R}^2$ por um número real k corresponde à multiplicação de cada componente do vector coluna por k. Isto é,

$$k\mathbf{u} = k \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} ku_1 \\ ku_2 \end{bmatrix}, \ k \in \mathbb{R}.$$

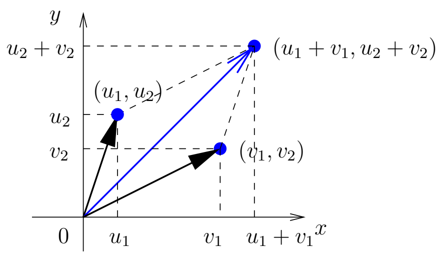


Figura 1.7: Adição de vectores de $\mathbb{R}^2$.

Na Figura 1.8 ilustra-se a representação da multiplicação de um vector de $\mathbb{R}^2$ por um real.

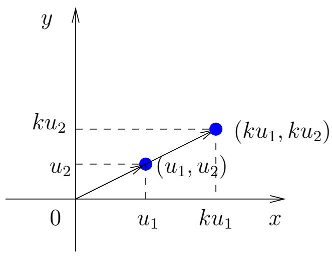


Figura 1.8: Multiplicação de vectores de $\mathbb{R}^2$ pelo escalar $k > 0$.

Vectores de $\mathbb{R}^3$

Pontos do espaço podem ser representados por ternos ordenados de números reais, desde que esteja fixado um referencial. Um referencial no espaço é obtido fixando um ponto do espaço que se toma como origem, e três rectas (eixos coordenados) distintas que se intersectam na origem. Em cada eixo coordenado escolhe-se ainda um sentido e uma unidade de comprimento.

Um terno ordenado (u_1, u_2, u_3) de números reais representa um ponto do espaço cujas coordenadas no referencial em causa valem, respectivamente, u_1 unidades do primeiro eixo, u_2 unidades do segundo eixo e u_3 unidades do terceiro. Um ponto P do espaço de coordenadas (u_1, u_2, u_3) pode ser identificado com o vector de posição $\mathbf{p} = (u_1, u_2, u_3)$, representado pelo segmento orientado cujo ponto inicial é a origem e cujo ponto final é o ponto P (ver Figura 1.9).

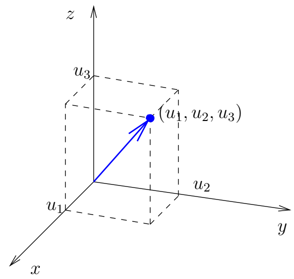


Figura 1.9: Representação de um vector de $\mathbb{R}^3$ no referencial de eixos x, y, z.

O conjunto de todos os vectores coluna de três componentes reais designa-se por $\mathbb{R}^3$, isto é,

$$\begin{aligned} \mathbb{R}^3 &= \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} : x_1 \in \mathbb{R}, x_2 \in \mathbb{R} \text{ e } x_3 \in \mathbb{R} \right\} \\ &= \{(x_1, x_2, x_3) : x_1, x_2, x_3 \in \mathbb{R}\}. \end{aligned}$$

De forma inteiramente análoga ao caso de vectores de $\mathbb{R}^2$, não é difícil verificar que as definições de igualdade de vectores, adição de vectores e multiplicação de um vector por um número real, se traduzem da forma que se segue. Para quaisquer vectores $\mathbf{u} = (u_1, u_2, u_3)$ e $\mathbf{v} = (v_1, v_2, v_3)$ de $\mathbb{R}^3$ tem-se:

- $\mathbf{u} = \mathbf{v}$ se e só se $u_1 = v_1$, $u_2 = v_2$ e $u_3 = v_3$.
- $\mathbf{u} + \mathbf{v} = (u_1 + v_1, u_2 + v_2, u_3 + v_3)$.
- $k\mathbf{u} = (ku_1, ku_2, ku_3)$ para $k \in \mathbb{R}$.

Por vezes é necessário determinar as coordenadas de um vector $\overrightarrow{PQ}$ conhecendo-se as coordenadas dos pontos P e de Q. Por exemplo, sabendo que as coordenadas dos pontos P e Q são respectivamente (p_1, p_2, p_3) e (q_1, q_2, q_3), pretende-se as coordenadas do vector $\overrightarrow{PQ}$ representado na Figura 1.10. Designando por O a origem do referencial, tem-se

$$\overrightarrow{OP} + \overrightarrow{PQ} = \overrightarrow{OQ}.$$

Os vectores $\overrightarrow{OP}$ e $\overrightarrow{OQ}$ são os vectores de posição $\mathbf{p}$ e $\mathbf{q}$, respectivamente, dos pontos P e Q. Logo,

$$\begin{aligned}\overrightarrow{PQ} &= \overrightarrow{OQ} - \overrightarrow{OP} = \mathbf{q} - \mathbf{p} \\ &= (q_1, q_2, q_3) - (p_1, p_2, p_3) = (q_1 - p_1, q_2 - p_2, q_3 - p_3).\end{aligned}$$

As coordenadas do vector pretendido são portanto as coordenadas do vector $\mathbf{q}-\mathbf{p}$.

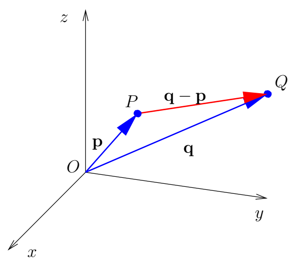


Figura 1.10: $\vec{PQ} = \mathbf{q} - \mathbf{p}$.

Vectores de $\mathbb{R}^n$

Por analogia com os casos $n = 2$ e $n = 3$, definimos $\mathbb{R}^n$ (n inteiro positivo) como o conjunto dos vectores coluna de n componentes reais. Nomeadamente

$$\begin{aligned}\mathbb{R}^n &= \left\{ \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} : \, x_i \in \mathbb{R}, \text{ para } i = 1, \ldots, n \right\} \\ &= \{(x_1, x_2, \ldots, x_n) : \, x_i \in \mathbb{R} \text{ para } i = 1, \ldots, n\}\,.\end{aligned}$$

O vector de $\mathbb{R}^n$ com todas as componentes iguais a zero chama-se o *vector nulo* (ou *vector zero*), que designamos por $\mathbf{0}$. A adição de vectores de $\mathbb{R}^n$ e a multiplicação destes vectores por números reais definem-se de modo análogo à definição apresentada para $n = 2$ e $n = 3$. Nomeadamente, para

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n \qquad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \in \mathbb{R}^n,$$

definimos

$$\mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix} \quad \text{e} \quad k\mathbf{x} = \begin{bmatrix} kx_1 \\ kx_2 \\ \vdots \\ kx_n \end{bmatrix} \text{ para } k \in \mathbb{R}. \tag{1.5}$$

Com base nas propriedades da adição e multiplicação de números reais, é fácil verificar as propriedades que a seguir enunciamos.

Propriedades algébricas de $\mathbb{R}^n$

Para quaisquer vectores $\mathbf{u}, \mathbf{v}, \mathbf{w}$ de $\mathbb{R}^n$ e números reais k e l, tem-se:

(i) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$

(ii) $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$

(iii) $\mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u}$

(iv) $\mathbf{u} + (-\mathbf{u}) = -\mathbf{u} + \mathbf{u} = \mathbf{0}$
onde $-\mathbf{u}$ designa $(-1)\mathbf{u}$

(v) $k(\mathbf{u} + \mathbf{v}) = k\mathbf{u} + k\mathbf{v}$

(vi) $(k + l)\mathbf{u} = k\mathbf{u} + l\mathbf{u}$

(vii) $k(l\mathbf{u}) = (kl)\mathbf{u}$

(viii) $1\mathbf{u} = \mathbf{u}$

1.3.2 Combinação linear de vectores

Os conceitos de combinação linear de vectores e de conjunto gerado por um conjunto de vectores desempenham um papel fundamental em toda a álgebra linear. Em particular, o produto de uma matriz A por um vector (coluna) $\mathbf{x}$ será definido como uma combinação linear dos vectores coluna de A. Nesta secção introduzimos estas noções para vectores de $\mathbb{R}^n$, enfatizando os seus aspectos geométricos no caso de vectores do plano ou do espaço.

Definição 1.10. Diz-se que o vector $\mathbf{y} \in \mathbb{R}^n$ é uma *combinação linear* dos vectores $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p$ de $\mathbb{R}^n$, se existem p números reais $c_1, c_2, \ldots, c_p$ tais que

$$\mathbf{y} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_p\mathbf{u}_p.$$

Aos escalares $c_1, c_2, \ldots, c_p$ chamamos *coeficientes* da combinação linear.

Designamos o conjunto de todas as combinações lineares de $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p$ por

$$\operatorname{Span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots \mathbf{u}_p\} = \{c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_p\mathbf{u}_p : \ c_1, \ldots, c_p \in \mathbb{R}\}.$$

A este subconjunto de $\mathbb{R}^n$ chamamos *conjunto gerado*[5]por, ou *expansão linear* de, $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$.

Note-se que o conjunto de todas as combinações lineares de um vector $\mathbf{v}$, é o conjunto de todos os vectores da forma $\alpha\mathbf{v}$, onde α é um escalar real. Dois vectores $\mathbf{u}$ e $\mathbf{v}$ tais que $\mathbf{u} = \alpha\mathbf{v}$, com α um escalar real, recebem a designação de vectores *colineares*.

Exemplifiquemos agora, com vectores de $\mathbb{R}^2$ e $\mathbb{R}^3$, as noções que acabámos de definir.

Exemplo 1.12. Na Figura 1.11 ilustram-se algumas combinações lineares dos vectores $\mathbf{u}_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$ e $\mathbf{u}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$.

Os conjuntos $\operatorname{Span}\{\mathbf{u}_1\}$ e $\operatorname{Span}\{\mathbf{u}_2\}$ são, respectivamente, os conjuntos

$$\operatorname{Span}\{\mathbf{u}_1\} = \{\alpha\mathbf{u}_1 : \ \alpha \in \mathbb{R}\}, \quad \operatorname{Span}\{\mathbf{u}_2\} = \{\alpha\mathbf{u}_2 : \alpha \in \mathbb{R}\}.$$

Geometricamente, estes conjuntos são rectas que passam pela origem e têm, respectivamente, a direcção de $\mathbf{u}_1$ e de $\mathbf{u}_2$.

O conjunto gerado por dois vectores é o conjunto das somas de todos os múltiplos de um vector com todos os múltiplos de outro vector. Para os vectores $\mathbf{u}_1$ e $\mathbf{u}_2$ o conjunto $\operatorname{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$ é o plano. Assim, de acordo com a Figura 1.11, tem-se:

$$\begin{array}{lll} \operatorname{Span}\{\mathbf{u}_1, \mathbf{u}_2\} = \mathbb{R}^2, & \operatorname{Span}\{\mathbf{u}_1\} = l_1, & \operatorname{Span}\{\mathbf{u}_2\} = l_2 \\ \mathbf{w} \in \operatorname{Span}\{\mathbf{u}_1, \mathbf{u}_2\}, & \mathbf{w} \notin \operatorname{Span}\{\mathbf{u}_1\}, & \mathbf{w} \notin \operatorname{Span}\{\mathbf{u}_2\}. \end{array}$$

♦

[5]A designação “Span” para conjunto gerado é a designação habitualmente utilizada em língua inglesa.

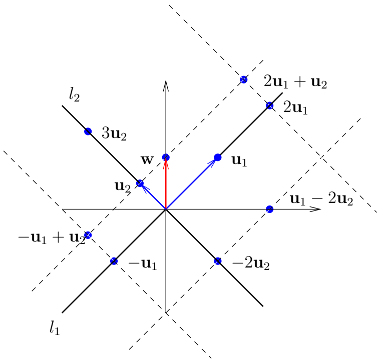


Figura 1.11: Combinação linear de vectores de $\mathbb{R}^2$. O vector $\mathbf{w}$ é $\mathbf{w} = \frac{1}{2}\mathbf{u}_1 + \mathbf{u}_2$.

Exemplo 1.13. Consideremos os vectores

$$\mathbf{u}_1 = \begin{bmatrix} 1 \\ 3 \\ -2 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -2 \\ 4 \\ 1 \end{bmatrix} \quad \text{e} \quad \mathbf{w} = \begin{bmatrix} -4 \\ 18 \\ -1 \end{bmatrix}.$$

Pretendemos saber se $\mathbf{w}$ é ou não combinação linear de $\mathbf{u}_1$ e $\mathbf{u}_2$. Ou seja, se existem números reais c_1 e c_2 tais que

$$\mathbf{w} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 \iff \begin{bmatrix} -4 \\ 18 \\ -1 \end{bmatrix} = c_1 \begin{bmatrix} 1 \\ 3 \\ -2 \end{bmatrix} + c_2 \begin{bmatrix} -2 \\ 4 \\ 1 \end{bmatrix}.$$

Isto é,

$$\begin{bmatrix} -4 \\ 18 \\ -1 \end{bmatrix} = \begin{bmatrix} c_1 \\ 3c_1 \\ -2c_1 \end{bmatrix} + \begin{bmatrix} -2c_2 \\ 4c_2 \\ c_2 \end{bmatrix} = \begin{bmatrix} c_1 - 2c_2 \\ 3c_1 + 4c_2 \\ -2c_1 + c_2 \end{bmatrix}.$$

Como dois vectores são iguais se as componentes correspondentes são iguais, a igualdade acima é equivalente ao sistema

$$\begin{cases} c_1 - 2c_2 = -4 \\ 3c_1 + 4c_2 = 18 \\ -2c_1 + c_2 = -1. \end{cases}$$

Usando o método de eliminação de Gauss para resolver este sistema, temos

$$\left[\begin{array}{cc|c} 1 & -2 & -4 \\ 3 & 4 & 18 \\ -2 & 1 & -1 \end{array}\right] \xrightarrow{L_2-3L_1} \left[\begin{array}{cc|c} 1 & -2 & -4 \\ 0 & 10 & 30 \\ -2 & 1 & -1 \end{array}\right]$$

$$\xrightarrow{L_3+2L_1} \left[\begin{array}{cc|c} 1 & -2 & -4 \\ 0 & 10 & 30 \\ 0 & -3 & -9 \end{array}\right] \xrightarrow{L_3+\frac{3}{10}L_2} \left[\begin{array}{cc|c} 1 & -2 & -4 \\ 0 & 10 & 30 \\ 0 & 0 & 0 \end{array}\right].$$

Como o sistema

$$\begin{cases} c_1 - 2c_2 = -4 \\ 10c_2 = 30 \end{cases}$$

tem solução $c_1 = 2$ e $c_2 = 3$, podemos dizer que $\mathbf{w}$ é combinação linear de $\mathbf{u}_1$ e $\mathbf{u}_2$. De facto,

$$\begin{bmatrix} -4 \\ 18 \\ -1 \end{bmatrix} = 2\begin{bmatrix} 1 \\ 3 \\ -2 \end{bmatrix} + 3\begin{bmatrix} -2 \\ 4 \\ 1 \end{bmatrix} \iff \mathbf{w} = 2\mathbf{u}_1 + 3\mathbf{u}_2.$$

Ou seja, o vector $\mathbf{w}$ pertence ao conjunto gerado por $\mathbf{u}_1$ e $\mathbf{u}_2$, isto é, $\mathbf{w} \in \operatorname{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$.

Note-se que o conjunto gerado por um único vector não nulo de $\mathbb{R}^3$ é uma recta que passa pela origem e tem a direcção do vector, enquanto que o conjunto gerado pelos vectores $\mathbf{u}_1$ e $\mathbf{u}_2$ é um plano que contém a origem. As rectas que passam pela origem e têm as direcções de $\mathbf{u}_1$ e de $\mathbf{u}_2$ pertencem a este plano, bem como o vector $\mathbf{w}$ (ver Figura 1.12).

♦

Exemplo 1.14. Consideremos os vectores

$$\mathbf{u}_1 = (1, 1), \quad \mathbf{u}_2 = (-2, -2) \quad \text{e} \quad \mathbf{u}_3 = (1, 2).$$

Os vectores $\mathbf{u}_1$ e $\mathbf{u}_2$ são colineares ($\mathbf{u}_1$ é proporcional a $\mathbf{u}_2$). Por conseguinte, o conjunto gerado por $\mathbf{u}_1$ e $\mathbf{u}_2$ é a recta que passa pela origem e tem a direcção destes vectores.

Os vectores $\mathbf{u}_1$ e $\mathbf{u}_3$ não são colineares (nem os vectores $\mathbf{u}_2$ e $\mathbf{u}_3$), logo o conjunto gerado por estes vectores é $\mathbb{R}^2$. Isto é,

$$\operatorname{Span}\{\mathbf{u}_1, \mathbf{u}_3\} = \mathbb{R}^2, \text{ bem como } \operatorname{Span}\{\mathbf{u}_2, \mathbf{u}_3\} = \mathbb{R}^2.$$

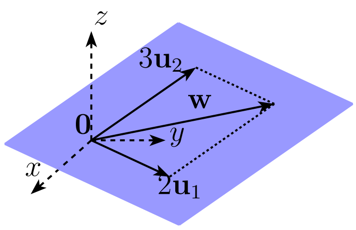


Figura 1.12: Os vectores $\mathbf{u}_1$ e $\mathbf{u}_2$ geram um plano e o vector $\mathbf{w}$ pertence a esse plano.

O conjunto gerado pelos três vectores $\mathbf{u}_1, \mathbf{u}_2$ e $\mathbf{u}_3$ é igualmente todo o $\mathbb{R}^2$. Ou seja,

$$\mathrm{Span}\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\} = \mathbb{R}^2.$$

♦

Dos exemplos e definições apresentadas anteriormente podemos tirar as conclusões que se seguem. Para vectores não nulos, temos:

- Um vector de $\mathbb{R}^2$ (ou de $\mathbb{R}^3$) gera uma recta que passa pela origem e tem a direcção do vector.

- Dois vectores colineares de $\mathbb{R}^2$ (ou de $\mathbb{R}^3$) geram uma recta que passa pela origem e tem a direcção dos vectores.

- Dois vectores não colineares de $\mathbb{R}^2$ geram o plano $\mathbb{R}^2$. Se forem vectores de $\mathbb{R}^3$ geram um plano que passa pela origem e que contém as rectas que passam pela origem e cujas direcções são as dos vectores dados.

- Três ou mais vectores de $\mathbb{R}^2$, em que pelo menos dois deles não são colineares geram $\mathbb{R}^2$.

- Três ou mais vectores de $\mathbb{R}^3$, em que pelo menos três deles não são colineares dois a dois, geram $\mathbb{R}^3$.

Exercício 1.1. Mostrar que para quaisquer vectores $\mathbf{u}$ e $\mathbf{v}$ de $\mathbb{R}^n$ se verifica

$$\mathrm{Span}\{\mathbf{u}, \mathbf{v}\} = \mathrm{Span}\{\mathbf{u}, \alpha\mathbf{u} + \beta\mathbf{v}\}, \text{ para todos os } \alpha \in \mathbb{R} \text{ e } \beta \in \mathbb{R} \setminus \{0\}.$$

▲

1.3.3 A equação matricial $A\mathbf{x} = \mathbf{b}$

Nesta secção mostra-se que um sistema de p equações lineares a n incógnitas reais pode escrever-se na forma matricial como $A\mathbf{x} = \mathbf{b}$, onde A é uma matriz $p \times n$, $\mathbf{x}$ é um vector de $\mathbb{R}^n$ e $\mathbf{b}$ é um vector de $\mathbb{R}^p$. Para tal necessitamos de definir o produto $A\mathbf{x}$. Iremos definir este produto como uma combinação linear das colunas de A.

Definição 1.11. Seja A a matriz $A = [a_{ij}]_{\substack{i=1,\dots,p \\ j=1,\dots,n}}$ e $\mathbf{x}$ um vector de n componentes.

O *produto de A por* $\mathbf{x}$ é a combinação linear dos vectores coluna de A que tem para coeficientes as componentes de $\mathbf{x}$. Ou seja,

$$\begin{aligned} A\mathbf{x} &= \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\ &= \underbrace{x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{p1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{p2} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{pn} \end{bmatrix}}_{\text{combinação linear das colunas de } A}. \end{aligned}$$

Nota 5.

1. *O produto $A\mathbf{x}$ só está definido se o número de colunas de A for igual ao número de componentes de $\mathbf{x}$ (isto é, o número de colunas de A é igual ao número de linhas da matriz coluna $\mathbf{x}$).*

2. *De acordo com a definição acima, se A é do tipo $p \times n$ e $\mathbf{x}$ é do tipo $n \times 1$, então $A\mathbf{x}$ é um vector do tipo $p \times 1$.*

Da definição anterior, e usando as definições de produto de um escalar por um vector e de adição de vectores, temos

$$\begin{aligned} A\mathbf{x} &= x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{p1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{p2} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{pn} \end{bmatrix} \\ &= \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ \vdots \\ a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n \\ \vdots \\ a_{p1}x_1 + a_{p2}x_2 + \cdots + a_{pn}x_n \end{bmatrix}. \end{aligned} \tag{1.6}$$

Nota 6. *A entrada da linha i do vector $A\mathbf{x}$ é igual à soma dos produtos das entradas da linha i de A pelas componentes correspondentes de $\mathbf{x}$.*

Exemplo 1.15. (a)

$$\underbrace{\begin{bmatrix} 1 & 3 & 4 & 2 \\ 2 & 4 & 5 & 1 \end{bmatrix}}_{2\times 4} \underbrace{\begin{bmatrix} 1 \\ -2 \\ 3 \\ 4 \end{bmatrix}}_{4\times 1} = 1\begin{bmatrix} 1 \\ 2 \end{bmatrix} - 2\begin{bmatrix} 3 \\ 4 \end{bmatrix} + 3\begin{bmatrix} 4 \\ 5 \end{bmatrix} + 4\begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1\times 1 - 2\times 3 + 3\times 4 + 4\times 2 \\ 1\times 2 - 2\times 4 + 3\times 5 + 4\times 1 \end{bmatrix} = \underbrace{\begin{bmatrix} 15 \\ 13 \end{bmatrix}}_{2\times 1}.$$

(b) $$\underbrace{\begin{bmatrix} 1 & 3 \\ 5 & 1 \\ 0 & 2 \\ 7 & 4 \end{bmatrix}}_{4\times 2} \underbrace{\begin{bmatrix} 1 \\ -2 \end{bmatrix}}_{2\times 1} = 1\begin{bmatrix} 1 \\ 5 \\ 0 \\ 7 \end{bmatrix} - 2\begin{bmatrix} 3 \\ 1 \\ 2 \\ 4 \end{bmatrix} = \underbrace{\begin{bmatrix} -5 \\ 3 \\ -4 \\ -1 \end{bmatrix}}_{4\times 1}.$$

♦

Exemplo 1.16. Considere-se o sistema de equações lineares

$$\begin{cases} x_1 + 3x_2 + 4x_3 + 2x_4 = 5 \\ 2x_1 + 4x_2 + 5x_3 + x_4 = 3. \end{cases}$$

Este sistema pode ser escrito como uma igualdade entre dois vectores coluna de $\mathbb{R}^2$, nomeadamente

$$\begin{bmatrix} x_1 + 3x_2 + 4x_3 + 2x_4 \\ 2x_1 + 4x_2 + 5x_3 + x_4 \end{bmatrix} = \begin{bmatrix} 5 \\ 3 \end{bmatrix} = \mathbf{b}.$$

Porém, o lado esquerdo da equação anterior pode ser escrito como uma combinação linear de vectores de $\mathbb{R}^2$, isto é

$$\begin{aligned} \begin{bmatrix} x_1 + 3x_2 + 4x_3 + 2x_4 \\ 2x_1 + 4x_2 + 5x_3 + x_4 \end{bmatrix} &= x_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + x_2 \begin{bmatrix} 3 \\ 4 \end{bmatrix} + x_3 \begin{bmatrix} 4 \\ 5 \end{bmatrix} + x_4 \begin{bmatrix} 2 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 3 & 4 & 2 \\ 2 & 4 & 5 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = A\mathbf{x}. \end{aligned}$$

Conclui-se que o sistema pode ser escrito na forma matricial $A\mathbf{x} = \mathbf{b}$, onde A é a matriz dos coeficientes do sistema e $\mathbf{b}$ é o vector coluna dos termos independentes. ♦

A igualdade (1.6) permite escrever qualquer sistema de equações lineares na forma $A\mathbf{x} = \mathbf{b}$, onde A é a matriz dos coeficientes do sistema, $\mathbf{x}$ é o vector cujas componentes são as variáveis do sistema, e $\mathbf{b}$ é o vector coluna cujas componentes são o segundo membro das equações do sistema. Por conseguinte, dizer que um sistema de equações lineares é possível é dizer que existe um vector $\mathbf{x}$ que satisfaz a igualdade $A\mathbf{x} = \mathbf{b}$. Ou seja, o sistema $A\mathbf{x} = \mathbf{b}$ é possível se e só se o vector $\mathbf{b}$ é uma combinação linear das colunas de A. Na proposição seguinte enunciamos este resultado para referência futura.

Proposição 1.3. Um sistema de equações lineares $A\mathbf{x} = \mathbf{b}$ é possível se e só se $\mathbf{b}$ é uma combinação linear das colunas de A.

Quando no sistema $A\mathbf{x} = \mathbf{b}$ o vector $\mathbf{b}$ é o vector nulo, dizemos que o sistema é um *sistema homogéneo*. Um sistema homogéneo é sempre possível já que, admite pelo menos a solução nula.

Exemplo 1.17. Façamos a discussão do sistema homogéneo $A_\alpha \mathbf{x} = \mathbf{0}$ em termos do parâmetro real α, onde

$$A_\alpha = \begin{bmatrix} 1 & 1 & -1 \\ 4 & \alpha^2 & -4 \\ 2 & 2 & \alpha \end{bmatrix}.$$

Apliquemos o método de eliminação de Gauss à matriz dos coeficientes do sistema.

$$\begin{bmatrix} 1 & 1 & -1 \\ 4 & \alpha^2 & -4 \\ 2 & 2 & \alpha \end{bmatrix} \xrightarrow{L_2-4L_1} \begin{bmatrix} 1 & 1 & -1 \\ 0 & \alpha^2-4 & 0 \\ 2 & 2 & \alpha \end{bmatrix} \xrightarrow{L_3-2L_1} \begin{bmatrix} 1 & 1 & -1 \\ 0 & \alpha^2-4 & 0 \\ 0 & 0 & \alpha+2 \end{bmatrix}.$$

A característica de A_α é

$$\operatorname{car}(A_\alpha) = \begin{cases} 1 & \text{se } \alpha = -2 \\ 2 & \text{se } \alpha = 2 \\ 3 & \text{se } \alpha \in \mathbb{R} \setminus \{-2, 2\}. \end{cases}$$

De acordo com a Proposição 1.2 (pág. 18), concluimos que para $\alpha \in \mathbb{R} \setminus \{-2, 2\}$ o sistema só possui a solução nula (não existem incógnitas livres), enquanto que para $\alpha = 2$ o sistema tem grau de indeterminação 1, e para $\alpha = -2$ tem grau de indeterminação 2.

O sistema $A_\alpha \mathbf{x} = \mathbf{0}$ é equivalente ao sistema

$$\begin{cases} x \quad +y \qquad -z = 0 \\ \quad (\alpha^2-4)y \qquad = 0 \\ \qquad (\alpha+2)z = 0. \end{cases} \tag{1.7}$$

Quando $\alpha = 2$ o sistema reduz-se à primeira e terceira equações. Resolvendo esse sistema obtém-se $z = 0$ e $x = -y$. Consequentemente, para $\alpha = 2$ a solução geral é

$$\begin{aligned} \big\{(x,y,z) \in \mathbb{R}^3 : x = -y \text{ e } z = 0\big\} &= \{(-y, y, 0) : \, y \in \mathbb{R}\} \\ &= \{y(-1,1,0) : \, y \in \mathbb{R}\} = \operatorname{Span}\{(-1,1,0)\}. \end{aligned}$$

Para $\alpha = -2$ o sistema (1.7) reduz-se à primeira equação, donde $x = -y + z$. Neste caso a solução geral é

$$\begin{aligned} \big\{(x,y,z) \in \mathbb{R}^3 : x = -y + z\big\} &= \{(-y+z, y, z) : \, y, z \in \mathbb{R}\} \\ &= \{y(-1,1,0) + z(1,0,1) : \, y, z \in \mathbb{R}\} \\ &= \operatorname{Span}\{(-1,1,0), (1,0,1)\}. \end{aligned}$$

♦

1.3.4 Operações com matrizes

Já definimos as operações de adição de vectores coluna, multiplicação de um escalar por um vector e multiplicação de uma matriz por um vector coluna. Baseados nestas operações, definimos agora operações sobre matrizes com várias colunas encarando para tal uma matriz como um conjunto ordenado de vectores coluna. Posteriormente, estudamos as propriedades das operações que definimos a seguir.

Sejam A e B duas matrizes do mesmo tipo, por exemplo $p \times n$,

$$A = [a_{ij}]_{\substack{i=1,\dots,p \\ j=1,\dots,n}} \quad \text{e} \quad B = [b_{ij}]_{\substack{i=1,\dots,p \\ j=1,\dots,n}} .$$

- A e B são *iguais* se as entradas respectivas são iguais, isto é, $a_{ij} = b_{ij}$ para todo $i = 1, \dots, p$ e $j = 1, \dots, n$.

- A soma $A + B$ é a matriz $p \times n$ cujas entradas são as somas das entradas de A com as respectivas entradas de B. Tal significa que a entrada na linha i e na coluna j de $A + B$ é $a_{ij} + b_{ij}$, para todo $i = 1, \dots, p$ e $j = 1, \dots, n$.

- A multiplicação de uma matriz A, por um escalar α é a matriz cujas entradas são as entradas de A multiplicadas por α.

Resumindo:

Sejam $A = [a_{ij}]$ e $B = [b_{ij}]$ matrizes do tipo $p \times n$ e α um escalar. Então,

$$A + B = [a_{ij} + b_{ij}]_{\substack{i=1,\dots,p \\ j=1,\dots,n}} \tag{1.8}$$

$$\alpha A = [\alpha a_{ij}]_{\substack{i=1,\dots,p \\ j=1,\dots,n}} . \tag{1.9}$$

Nota 7. *A adição de matrizes só está definida quando as matrizes são do mesmo tipo.*

Seguidamente definimos multiplicação de matrizes. Se considerarmos uma matriz B como um conjunto ordenado de vectores coluna, é natural definir o produto AB como o conjunto das colunas que são o produto de A pelas colunas de B, respeitando a ordem de B.

Definição 1.12. Seja A uma matriz $p \times n$ e B uma matriz $n \times k$ com colunas $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_k$.

O produto AB é a matriz $p \times k$ cujas colunas são $A\mathbf{b}_1, A\mathbf{b}_2, \ldots, A\mathbf{b}_k$. Isto é,

$$AB = A \begin{bmatrix} | & | & & | \\ \mathbf{b}_1 & \mathbf{b}_2 & \ldots & \mathbf{b}_k \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ A\mathbf{b}_1 & A\mathbf{b}_2 & \ldots & A\mathbf{b}_k \\ | & | & & | \end{bmatrix}.$$

Note-se que o produto AB só está definido se o número de colunas de A é igual ao número de linhas de B. Além disso, a matriz AB é uma matriz com o mesmo número de linhas de A e o mesmo número de colunas de B.

$$\begin{array}{ccccc} A & & B & = & AB. \\ p \times n & & n \times k & & p \times k \end{array}$$

No caso particular da matriz B ser $n \times 1$ (ou seja, um vector coluna) a definição do produto AB coincide com a definição anteriormente apresentada (como não podia deixar de ser).

Por razões práticas convém deduzir como se calcula uma entrada específica do produto de duas matrizes. Para tal comecemos com dois exemplos.

Exemplo 1.18. Calculemos

$$A\mathbf{u} = \begin{bmatrix} 2 & 3 & 4 \\ 1 & 3 & -2 \end{bmatrix} \begin{bmatrix} 5 \\ 6 \\ 7 \end{bmatrix}.$$

Por definição de produto de uma matriz por um vector, o produto $A\mathbf{u}$ é a matriz 2×1 dada por

$$A\mathbf{u} = A \begin{bmatrix} 5 \\ 6 \\ 7 \end{bmatrix} = 5 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + 6 \begin{bmatrix} 3 \\ 3 \end{bmatrix} + 7 \begin{bmatrix} 4 \\ -2 \end{bmatrix} = \begin{bmatrix} 2 \times 5 + 3 \times 6 + 4 \times 7 \\ 1 \times 5 + 3 \times 6 - 2 \times 7 \end{bmatrix} = \begin{bmatrix} 56 \\ 9 \end{bmatrix}.$$

♦

O exemplo anterior ilustra como podemos obter as entradas de uma certa coluna do produto de duas matrizes. Nomeadamente, se $C = AB$, a entrada c_{ij} de C obtém-se multiplicando as entradas da linha i de A pelas entradas correspondentes da coluna j de B e somando os produtos obtidos.

Exemplo 1.19. Considerem-se as matrizes

$$A = \begin{bmatrix} 2 & 3 & 1 \\ 0 & 4 & 6 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 2 & 3 & 1 \\ -1 & 0 & 2 & 3 \\ 5 & 2 & 3 & 1 \end{bmatrix}.$$

Como A é uma matriz 2×3 e B é 3×4, o produto AB está definido e AB é uma matriz do tipo 2×4. Para determinar a entrada na linha 2 e na coluna 3 da matriz AB escolhemos a linha 2 de A e a coluna 3 de B, multiplicamos as entradas da referida linha pelas entradas respectivas da coluna selecionada, e somamos. Ou seja,

$$A = \begin{bmatrix} 2 & 3 & 1 \\ 0 & 4 & 6 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 1 \\ -1 & 0 & 2 & 3 \\ 5 & 2 & 3 & 1 \end{bmatrix} = \begin{bmatrix} \square & \square & \square & \square \\ \square & \square & \boxed{26} & \square \end{bmatrix} = AB.$$

$$\uparrow$$

$$\boxed{0 \times 3 + 4 \times 2 + 6 \times 3 = 26}$$

Por exemplo, a entrada na linha 1 e na coluna 4 de AB é

$$\boxed{2 \times 1 + 3 \times 3 + 1 \times 1 = 12}$$

$$\downarrow$$

$$A = \begin{bmatrix} 2 & 3 & 1 \\ 0 & 4 & 6 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 1 \\ -1 & 0 & 2 & 3 \\ 5 & 2 & 3 & 1 \end{bmatrix} = \begin{bmatrix} \square & \square & \square & \boxed{12} \\ \square & \square & \square & \square \end{bmatrix}.$$

Calculando as restantes entradas de AB, obtemos

$$AB = \begin{bmatrix} 4 & 6 & 15 & 12 \\ 26 & 12 & 26 & 18 \end{bmatrix}.$$

♦

Para $A = [a_{ij}]_{\substack{i=1,\ldots,p \\ j=1,\ldots,n}}$, $B = [b_{ij}]_{\substack{i=1,\ldots,n \\ j=1,\ldots,k}}$ e $AB = C = [c_{ij}]_{\substack{i=1,\ldots,p \\ j=1,\ldots,k}}$, isto é

$$
\begin{aligned}
AB &= \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pn} \end{bmatrix} \begin{bmatrix} b_{11} & \cdots & b_{1j} & \cdots & b_{1k} \\ b_{21} & \cdots & b_{2j} & \cdots & b_{2k} \\ \vdots & \vdots & & \vdots & \\ b_{n1} & \cdots & b_{nj} & \cdots & b_{nk} \end{bmatrix} \\
&= \begin{bmatrix} c_{11} & \cdots & c_{1j} & \cdots & c_{1k} \\ \vdots & & \vdots & & \vdots \\ c_{i1} & \cdots & c_{ij} & \cdots & c_{ik} \\ \vdots & & \vdots & & \vdots \\ c_{p1} & \cdots & c_{pj} & \cdots & c_{pk} \end{bmatrix},
\end{aligned}
$$

tem-se $c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj} = \sum_{r=1}^{n} a_{ir}b_{rj}$. Resumindo:

> Se $A = [a_{ij}]$ é do tipo $p \times n$ e $B = [b_{ij}]$ do tipo $n \times k$, a entrada na linha i e na coluna j da matriz $AB = C = [c_{ij}]$ é dada por
>
> $$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj} = \sum_{r=1}^{n} a_{ir}b_{rj}. \tag{1.10}$$

Propriedades das operações com matrizes

A adição de matrizes reais (resp. complexas) goza das propriedades comutativa e associativa como facilmente se deduz usando as propriedades comutativa e associativa dos números reais (resp. complexos) e a definição (1.8) de adição de matrizes. Contudo, a multiplicação de matrizes **não goza** em geral da propriedade comutativa, ou seja, pode ter-se $AB \neq BA$ mesmo no caso em que ambos os produtos AB e BA estão definidos. Por exemplo, para

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} 2 & 4 \\ -1 & 5 \end{bmatrix},$$

tem-se

$$AB = \begin{bmatrix} 0 & 14 \\ 5 & 17 \end{bmatrix} \quad \text{e} \quad BA = \begin{bmatrix} 14 & 8 \\ 14 & 3 \end{bmatrix}.$$

Logo, $AB \neq BA$. Quando $AB = BA$ dizemos que A e B *comutam*.

Passamos a enunciar algumas propriedades das operações com matrizes.

Teorema 1.1. Sejam A, B e C matrizes para as quais as operações abaixo indicadas estão definidas, e α, β escalares. São válidas as seguintes propriedades:

(a) $A + B = B + A$	**Comutatividade da adição**
(b) $A + (B + C) = (A + B) + C$	**Associatividade da adição**
(c) $A(BC) = (AB)C$	**Associatividade da multiplicação**
(d) $A(B + C) = AB + AC$	**Distributividade à esquerda**
(e) $(B + C)A = BA + CA$	**Distributividade à direita**
(f) $\alpha(A + B) = \alpha A + \alpha B$	
(g) $(\alpha + \beta)A = \alpha A + \beta A$	
(h) $\alpha(AB) = (\alpha A)B = A(\alpha B)$	

As alíneas (a), (b), (f), (g) e (h) são imediatas usando propriedades das operações com números reais ou complexos. A seguir apresentamos apenas a demonstração da alínea (d), demonstração esta que, no entanto, é bem ilustrativa dos procedimentos a seguir para as outras alíneas.

Demonstração. (d): Para que $A(B + C)$ seja igual a $AB + AC$, as matrizes B e C deverão ter o mesmo tamanho e o número de colunas de A deverá ser igual ao número de linhas de B e de C. Suponha-se que $A = [a_{ij}]$ tem n colunas e que as matrizes $B = [b_{ij}]$ e $C = [c_{ij}]$ têm n linhas. De modo a facilitar a compreensão da demonstração, usaremos simultaneamente a notação d_{ij} e $[D]_{ij}$ para designar a entrada ij da matriz D.

Pretendemos mostrar que as entradas homólogas de $A(B+C)$ e de $AB+AC$ são iguais, ou seja, que é satisfeita a igualdade

$$[A(B + C)]_{ij} = [AB]_{ij} + [AC]_{ij},$$

para quaisquer valores de i e de j. Usando (1.8) e (1.10) obtemos

$$\begin{aligned}
[A(B + C)]_{ij} &= a_{i1}(b_{1j} + c_{1j}) + a_{i2}(b_{2j} + c_{2j}) + \cdots + a_{in}(b_{nj} + c_{nj}) \\
&= (a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}) + (a_{i1}c_{1j} + a_{i2}c_{2j} + \cdots + a_{in}c_{nj}) \\
&= [AB]_{ij} + [AC]_{ij}.
\end{aligned}$$

□

Nota 8. *Para quaisquer números reais a e b é válida a propriedade $ab = 0$ se e só se $a = 0$ ou $b = 0$. Este resultado não é em geral válido para o produto de matrizes, como se verifica no exemplo que apresentamos a seguir.*

Exemplo 1.20. Considere-se $A = \begin{bmatrix} 0 & 0 \\ 1 & 2 \end{bmatrix}$ e $B = \begin{bmatrix} -2 & 2 \\ 1 & -1 \end{bmatrix}$. O produto AB é igual à matriz nula

$$AB = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},$$

porém, nem A nem B são matrizes nulas. ♦

As propriedades algébricas das matrizes são essenciais na demonstração de um grande número de resultados. Por exemplo, as propriedades das operações com matrizes habilitam-nos agora a apresentar a demonstração de duas proposições anteriormente referidas. Sem estas propriedades a prova destas proposições seria com certeza muito mais extensa.

A proposição seguinte diz-nos que um sistema possível ou tem solução única ou uma infinidade de soluções.

Proposição 1.4. Se um sistema de equações lineares admite mais do que uma solução, então possui uma infinidade de soluções.

Demonstração. Sejam $\mathbf{u}$ e $\mathbf{v}$ soluções distintas do sistema $A\mathbf{x} = \mathbf{b}$. O vector $\mathbf{y} = \mathbf{u} - \mathbf{v}$ é uma solução não nula do sistema homogéneo $A\mathbf{x} = \mathbf{0}$, já que

$$A\mathbf{y} = A(\mathbf{u} - \mathbf{v}) = A\mathbf{u} - A\mathbf{v} = \mathbf{b} - \mathbf{b} = \mathbf{0}.$$

Para qualquer escalar α, o vector $\mathbf{w} = \mathbf{u} + \alpha\mathbf{y}$ é solução do sistema $A\mathbf{x} = \mathbf{b}$ já que

$$A\mathbf{w} = A(\mathbf{u} + \alpha\mathbf{y}) = A\mathbf{u} + \alpha A\mathbf{y} = \mathbf{b} + \alpha\mathbf{0} = \mathbf{b}.$$

Logo, se $\mathbf{u}$ e $\mathbf{v}$ são soluções distintas do sistema $A\mathbf{x} = \mathbf{b}$, então qualquer vector da forma $\mathbf{w} = \mathbf{u} + \alpha\mathbf{y}$, com $\alpha \in \mathbb{R}$ e $\mathbf{y} = \mathbf{u} - \mathbf{v}$, é solução do sistema. Ou seja, há uma infinidade de soluções $\mathbf{w} = \mathbf{u} + \alpha\mathbf{y}$ (uma para cada α fixado em $\mathbb{R}$). □

Relembremos que na génese da noção de característica de uma matriz, apresentada na Secção 1.2 (pág. 12), encontra-se a *unicidade da forma reduzida em escada por linhas de uma matriz*. O teorema que enunciamos a seguir estabelece este resultado.

Teorema 1.2. A forma reduzida em escada por linhas de uma matriz é única.

Demonstração. [6] Seja A uma matriz do tipo $p \times n$. Usemos indução[7]sobre n.

Para $n = 1$, a matriz A tem uma única coluna e portanto a correspondente forma reduzida em escada por linhas é obviamente única. Considere-se como hipótese de indução que para qualquer matriz do tipo $p \times (n-1)$ a respectiva forma reduzida em escada é única.

Seja A' a matriz que se obtém de A suprimindo a última coluna. Note-se que qualquer sequência de operações elementares sobre A que reduza A a uma matriz em escada na forma reduzida também reduz A' a uma matriz em escada por linhas na forma reduzida. Assim, pela hipótese de indução, se B e C são formas reduzidas em escada por linhas da matriz A, estas matrizes diferem quando muito na coluna número n.

Suponha-se que B e C são matrizes em escada na forma reduzida, obtidas de A, e que (por absurdo) $B \neq C$. Pretendemos chegar a uma contradição. Sendo $B \neq C$, existe um inteiro k tal que a linha k de B é distinta da linha k de C. Seja $\mathbf{u}$ uma qualquer solução do sistema $B\mathbf{x} = \mathbf{0}$. Como B e C são matrizes obtidas de A por meio de operações elementares sobre as suas linhas, os sistemas $B\mathbf{x} = \mathbf{0}$ e $C\mathbf{x} = \mathbf{0}$ têm as mesmas soluções. Dado que $B\mathbf{u} = \mathbf{0} = C\mathbf{u}$, da propriedade distributiva segue $(B - C)\mathbf{u} = \mathbf{0}$. Como B e C têm as primeiras $(n-1)$ colunas iguais (por hipótese de indução) a matriz $(B-C)$ tem as primeiras $(n-1)$ colunas nulas, e portanto a componente k do vector $(B - C)\mathbf{u}$ é $(b_{kn} - c_{kn})u_n = 0$. Esta equação possui como única solução $u_n = 0$ já que, por hipótese, $b_{kn} \neq c_{kn}$. Logo, qualquer solução $\mathbf{u}$ dos sistemas $B\mathbf{x} = \mathbf{0}$ e $C\mathbf{x} = \mathbf{0}$ tem a $n^{\text{ésima}}$ componente nula. Assim sendo, a coluna número n de B e de C tem um pivô (igual a 1), caso contrário x_n seria uma variável livre dos sistemas $B\mathbf{x} = \mathbf{0}$ e $C\mathbf{x} = \mathbf{0}$. Como as primeiras $(n-1)$ colunas de B e de C são iguais, o pivô da coluna n de B e de C aparece na mesma linha (a última). Logo, por definição de matriz em escada por linhas na forma reduzida, tem-se $B = C$, o que é uma contradição. $\square$

Matriz Identidade

A matriz identidade de ordem n é uma matriz quadrada $n \times n$ cujas entradas da diagonal principal são todas iguais a 1 e as restantes entradas são nulas. Por

[6] Esta demonstração é apresentada por T. Yuster em [13].

[7] O método de indução matemática é um método de demonstração usado frequentemente quando se pretende provar que uma certa proposição é verdadeira para todos os números naturais n. Uma versão deste método é a seguinte: (i) mostrar que a proposição é válida para o primeiro valor de n; (ii) supor que a proposição é válida para $n = r$ (esta hipótese é conhecida como hipótese de indução); (iii) deduzir de (ii) que a proposição é válida para $n = r + 1$.

exemplo,

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad I_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

são respectivamente a matriz identidade de ordem $2, 3$, e 4.

Definição 1.13. A *matriz identidade de ordem* n é uma matriz quadrada $n \times n$ cujas entradas na diagonal principal são todas iguais a 1 e as outras entradas iguais a zero.

A matriz identidade de ordem n será designada I_n, ou simplesmente por I quando no contexto for claro qual a ordem da matriz. Isto é,

$$I_n = [\delta_{ij}]_{i,j=1,\dots,n} \quad \text{com} \quad \delta_{ij} = \begin{cases} 0 & \text{se } i \neq j \\ 1 & \text{se } i = j \end{cases}$$

As entradas da matriz identidade são tradicionalmente denotadas pelo símbolo δ_{ij}. Este símbolo é conhecido por *delta de Kronecker*[8] e tem o significado especificado na definição anterior.

A partir da definição de multiplicação de matrizes, é fácil mostrar (basta usar (1.10)) que o produto de uma matriz A por uma matriz identidade é sempre igual a A. Ilustremos este facto com o exemplo seguinte.

Exemplo 1.21. Seja $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix}$.

Apenas podemos multiplicar A à esquerda pela matriz identidade de ordem 3, e à direita pela de ordem 4.

$$I_3 A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix} = A$$

e

$$A I_4 = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix} = A.$$

[8] Leopold Kronecker (1823-1891), matemático alemão.

♦

A matriz identidade é um exemplo de uma matriz *diagonal*. Isto é, uma matriz quadrada com todas as entradas iguais a zero excepto possivelmente as entradas da diagonal principal. Por exemplo, a matriz

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

é uma matriz diagonal.

Transposta de uma matriz

A *transposta* de uma matriz A do tipo $p \times n$, é a matriz $n \times p$ cujas linhas são as colunas de A, pela mesma ordem em que aparecem em A. Designamos por A^T a matriz transposta de A. Por exemplo, para

$$A = \begin{bmatrix} 1 & 2 & 4 & 5 \\ 2 & 3 & 5 & 6 \end{bmatrix} \quad \text{tem-se} \quad A^T = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 4 & 5 \\ 5 & 6 \end{bmatrix}.$$

Como as linhas de A^T são as colunas de A, pela mesma ordem, tem-se:

$$\text{se } A = [a_{ij}]_{\substack{i=1,\ldots,p \\ j=1,\ldots,n}} \text{ e } A^T = [a'_{ij}]_{\substack{i=1,\ldots,n \\ j=1,\ldots,p}}, \text{ então } a'_{ij} = a_{ji}.$$

Enunciamos a seguir algumas propriedades da transposta de uma matriz.

Propriedades da matriz transposta

Sejam A e B matrizes para as quais as operações abaixo indicadas estão definidas. Verificam-se as igualdades:

(a) $(A^T)^T = A$;

(b) $(A + B)^T = A^T + B^T$;

(c) $(AB)^T = B^T A^T$.

Demonstração. Deixamos como exercício a prova de (a) e (b) e mostramos somente (c). Para tal, considere-se

$$A = [a_{ij}]_{\substack{i=1,\dots p \\ j=1,\dots n}} \quad \text{e} \quad B = [b_{ij}]_{\substack{i=1,\dots n \\ j=1,\dots k}}.$$

A matriz AB é do tipo $p \times k$ e portanto $(AB)^T$ é do tipo $k \times p$. Da mesma forma, $B^T A^T$ é do tipo $k \times p$. Designemos por $\left[(AB)^T\right]_{ij}$ a entrada ij de $(AB)^T$ e por $[B^T A^T]_{ij}$ a entrada ij de $B^T A^T$. Assim, como as linhas da transposta de uma matriz são as colunas da matriz, tem-se

$$\left[(AB)^T\right]_{ij} = [AB]_{ji} = a_{j1}b_{1i} + a_{j2}b_{2i} + \cdots + a_{jn}b_{ni}, \tag{1.11}$$

onde na última igualdade usámos a igualdade (1.10), da definição de produto de matrizes.

Para calcular $[B^T A^T]_{ij}$ designemos por b'_{ij} a entrada ij de B^T e por a'_{ij} a de A^T. É claro que

$$b'_{ij} = b_{ji} \quad \text{e} \quad a'_{ij} = a_{ji}.$$

Usando de novo (1.10) temos

$$\left[B^T A^T\right]_{ij} = b'_{i1}a'_{1j} + b'_{i2}a'_{2j} + \cdots + b'_{in}a'_{nj} = b_{1i}a_{j1} + b_{2i}a_{j2} + \cdots + b_{ni}a_{jn}.$$

Comparando a última igualdade com (1.11) fica mostrado o resultado enunciado na alínea (c).

□

Definição 1.14. Uma matriz A diz-se *simétrica* se $A = A^T$ e *anti-simétrica* se $A = -A^T$.

É de realçar que a definição anterior implica que matrizes simétricas e anti-simétricas sejam matrizes quadradas. De facto, a igualdade $A = \pm A^T$ só é possível se A é quadrada. Além disso, como as entradas da diagonal principal de A e A^T são as mesmas, uma matriz anti-simétrica deverá ter estas entradas iguais a zero, uma vez que satisfaz a igualdade $A = -A^T$. A designação de matriz simétrica está relacionada com a simetria que este tipo de matrizes possui relativamente à diagonal principal, conforme se verifica nos exemplos apresentados a seguir.

Exemplo 1.22. No sentido de clarificar a noção de matriz simétrica e anti-simétrica apresentamos alguns exemplos.

$$\text{A matriz} \quad \begin{bmatrix} 2 & 5 & 3 \\ 5 & 1 & 0 \\ 3 & 0 & 2 \end{bmatrix} \qquad \text{é simétrica.}$$

$$\text{A matriz} \quad \begin{bmatrix} 0 & 5 & -3 \\ -5 & 0 & 1 \\ 3 & -1 & 0 \end{bmatrix} \qquad \text{é anti-simétrica.}$$

$$\text{A matriz} \quad \begin{bmatrix} 1 & 5 & -3 \\ -5 & 0 & 1 \\ 3 & -1 & 0 \end{bmatrix} \qquad \text{não é anti-simétrica (nem simétrica).}$$

♦

Exemplo 1.23. Dada uma matriz A do tipo $p \times n$, as matrizes $A^T A$ e AA^T são simétricas, respectivamente, do tipo $n \times n$ e $p \times p$. De facto, usando as propriedades da matriz transposta atrás enunciadas, resultam as seguintes igualdades:

$$(A^T A)^T \overset{(c)}{=} A^T (A^T)^T \overset{(a)}{=} A^T A \quad \text{e} \quad (AA^T)^T \overset{(c)}{=} (A^T)^T A^T \overset{(a)}{=} AA^T.$$

♦

Exercício 1.2. Seja A uma matriz quadrada.

a) Mostrar que $(A + A^T)$ é simétrica e $(A - A^T)$ é anti-simétrica.
b) Usar a alínea anterior para escrever A como a soma de uma matriz simétrica com uma anti-simétrica.

1.4 Inversa de uma matriz

Nesta secção apresentamos a definição de matriz inversa de uma matriz (quadrada) e estudamos as suas propriedades, nomeadamente a relação entre a existência de inversa e a característica da matriz. O cálculo da inversa de uma matriz é efectuado recorrendo ao método de eliminação de Gauss-Jordan, método este que pode ser descrito em termos de multiplicação sucessiva por certas matrizes ditas elementares. Mostra-se ainda que a matriz dos coeficientes de um sistema, com igual número de equações e de incógnitas, é invertível se e só se o sistema tem uma única solução.

Definição 1.15. Uma matriz quadrada A do tipo $n \times n$, diz-se *invertível*, ou *não singular*, se existe uma matriz B do tipo $n \times n$, tal que

$$AB = I_n = BA. \tag{1.12}$$

Se B é uma matriz que satisfaz a igualdade anterior dizemos que B é uma *inversa* de A.

Uma matriz que não seja invertível diz-se uma *matriz singular*.

Exemplo 1.24. A matriz

$$A = \begin{bmatrix} 1 & 10 & 3 \\ 0 & 0 & 0 \\ 1 & 3 & 2 \end{bmatrix}$$

não é invertível. Para verificar este facto considere-se uma matriz B do tipo 3×3 de entradas b_{ij} e efectue-se o produto AB. Como a segunda linha de AB é nula, este produto nunca pode ser igual à matriz identidade de terceira ordem.

♦

Nota 9. *O exemplo anterior ilustra o facto mais geral: qualquer matriz quadrada que possua uma linha nula é uma matriz singular. Este resultado pode ser provado usando o mesmo procedimento do exemplo anterior e a expressão* (1.10).

A unicidade da matriz inversa é estabelecida no próximo teorema.

Teorema 1.3. A inversa de uma matriz, quando existe, é única.

Demonstração. Suponha-se que B e C são inversas de A. Isto é,

$$AB = BA = I \quad \text{e} \quad AC = CA = I.$$

Uma vez que $BA = I = AC$, tem-se

$$BAC = B(AC) = BI = B,$$

e

$$BAC = (BA)C = IC = C.$$

Logo, $B = C$. □

Passamos a designar a matriz inversa de uma matriz invertível A por:

$$A^{-1}$$

Proposição 1.5. Se $ad - bc \neq 0$, a inversa da matriz

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

é dada por

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Demonstração. Deixamos como exercício verificar que $AA^{-1} = I_2$ e $A^{-1}A = I_2$. □

Exemplo 1.25. A matriz $\begin{bmatrix} 2 & 1 \\ 4 & 4 \end{bmatrix}$ é invertível e a sua inversa é

$$A^{-1} = \frac{1}{4} \begin{bmatrix} 4 & -1 \\ -4 & 2 \end{bmatrix} = \begin{bmatrix} 1 & \frac{-1}{4} \\ -1 & \frac{1}{2} \end{bmatrix}.$$

♦

Verifiquemos agora que o produto de matrizes invertíveis ainda é invertível.

Proposição 1.6. Se A e B são matrizes invertíveis da mesma ordem, então AB é invertível e a sua inversa é $B^{-1}A^{-1}$. Ou seja,

$$(AB)^{-1} = B^{-1}A^{-1}. \tag{1.13}$$

Demonstração. Como a inversa de uma matriz é única, basta-nos mostrar que $B^{-1}A^{-1}$ é a inversa de AB.

$$\begin{aligned} (AB)(B^{-1}A^{-1}) &= A(BB^{-1})A^{-1} && \text{(associatividade da multiplicação)} \\ &= AIA^{-1} && \text{(já que } B^{-1} \text{ é a inversa de } B\text{)} \\ &= AA^{-1} \\ &= I. \end{aligned}$$

Da mesma forma,

$$(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}B = I.$$

□

Potências de uma matriz

Para qualquer inteiro não negativo k e qualquer matriz (quadrada) A, definimos A^k como o produto de k factores todos iguais a A

$$A^k = \underbrace{AA\cdots A}_{k \text{ factores}}$$

sendo a expressão A^0 interpretada como a matriz identidade.

Se A é uma matriz invertível, definimos potências negativas de A em termos de potências positivas de A^{-1}. Isto é,

$$A^{-k} = (A^{-1})^k = \underbrace{A^{-1}A^{-1}\cdots A^{-1}}_{k \text{ factores}}$$

A proposição seguinte sumariza alguns resultados sobre matrizes invertíveis e potências destas matrizes.

Proposição 1.7. Seja A uma matriz invertível. São válidas as proposições seguintes.

1) A^{-1} é invertível e $\left(A^{-1}\right)^{-1} = A$.

2) A^k é invertível e $\left(A^k\right)^{-1} = \left(A^{-1}\right)^k = A^{-k}$.

3) Para qualquer escalar não nulo α, a matriz αA é invertível e $(\alpha A)^{-1} = \dfrac{1}{\alpha}A^{-1}$.

Demonstração. As provas das alíneas 1) e 2) são deixadas como exercício. Façamos a prova de 3). Como

$$(\alpha A)\left(\frac{1}{\alpha}A^{-1}\right) = \alpha\frac{1}{\alpha}AA^{-1} = I,$$

e $\left(\dfrac{1}{\alpha}A^{-1}\right)(\alpha A) = \alpha\dfrac{1}{\alpha}A^{-1}A = I$, segue o resultado. □

Como já referimos, o produto de matrizes não é em geral comutativo. Assim, a potência

$$(A+B)^2 = (A+B)(A+B) = A^2 + AB + BA + B^2,$$

é em geral diferente de $A^2+2AB+B^2$ (excepto se as matrizes A e B comutarem). Sugerimos a procura de um contra-exemplo.

1.4.1 Matrizes elementares e cálculo da inversa

Nesta secção apresentamos um algoritmo para calcular a inversa de uma matriz. Subjacente a este algoritmo está o facto de ser possível obter a matriz identidade aplicando operações elementares à matriz invertível dada. Cada operação elementar (ver definição na página 8) aplicada a uma matriz vai traduzir-se na multiplicação dessa matriz por uma matriz dita elementar.

Definição 1.16. Um matriz *elementar* é uma matriz quadrada $n \times n$ que é obtida da identidade I_n por meio de uma única operação elementar.

Exemplo 1.26. Exemplos de matrizes elementares:

(a) $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ Troca da segunda linha de I_3 com a primeira linha.

(b) $\begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}$ A primeira linha de I_2 foi substituída pela sua soma com a 2ª linha multiplicada por 3. ♦

(c) $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ A segunda linha de I_4 foi multiplicada por 2.

Quando uma matriz A é multiplicada à *esquerda* por uma matriz elementar, a matriz produto é a matriz que se obtém de A efectuando a mesma operação elementar efectuada em I para obter a matriz elementar em causa. A proposição seguinte enuncia este facto. A sua demonstração é deixada como exercício.

Proposição 1.8. Se E é uma matriz elementar, então EA é a matriz que se obtém de A efectuando a mesma operação elementar que permite obter E a partir da identidade.

Ilustremos esta proposição através de um exemplo.

Exemplo 1.27. Consideremos a matriz

$$A = \begin{bmatrix} 1 & 3 & 5 & 7 & 2 \\ 5 & 3 & 1 & 5 & 8 \\ 10 & 2 & 7 & 2 & 6 \end{bmatrix},$$

e a matriz elementar

$$E = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

obtida de I_2 por:

$$I_2 \overset{L_2+2L_1}{\longrightarrow} E.$$

A matriz E resulta de I_2 substituindo a segunda linha pela soma desta linha com a primeira linha multiplicada por 2 . Calculando EA, verificamos que

$$A = \begin{bmatrix} 1 & 3 & 5 & 7 & 2 \\ 5 & 3 & 1 & 5 & 8 \\ 10 & 2 & 7 & 2 & 6 \end{bmatrix} \overset{L_2+2L_1}{\longrightarrow} \begin{bmatrix} 1 & 3 & 5 & 7 & 2 \\ 7 & 9 & 11 & 19 & 12 \\ 10 & 2 & 7 & 2 & 6 \end{bmatrix} = EA.$$

♦

Nota 10. *A Proposição 1.8 só é válida para multiplicação à esquerda. De facto, se uma matriz A é multiplicada à direita por uma matriz elementar E, a matriz AE é a matriz que se obtém de A efectuando a operação elementar correspondente a E sobre as colunas de A. Por exemplo, para $\alpha \neq 0$ e*

$$A = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix},\ E = \begin{bmatrix} 1 & 0 \\ 0 & \alpha \end{bmatrix} \textit{ tem-se } EA = \begin{bmatrix} 1 & 3 \\ 2\alpha & 4\alpha \end{bmatrix} \textit{ e } AE = \begin{bmatrix} 1 & 3\alpha \\ 2 & 4\alpha \end{bmatrix}.$$

Como veremos, qualquer matriz elementar E é invertível já que, se E é obtida da identidade por meio de uma operação elementar, então efectuando sobre E a operação elementar "inversa" obtemos a identidade. Ou seja, existe sempre uma matriz elementar (correspondente à operação elementar "inversa") que quando multiplicada (à esquerda) por E dá a identidade (e esta matriz é a inversa de E). No exemplo seguinte ilustramos este facto bem como o que se entende por operação elementar "inversa".

Exemplo 1.28. Calculemos a inversa das matrizes elementares E seguintes.

(a) $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \overset{3L_2}{\longrightarrow} \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} \overset{\frac{1}{3}L_2}{\longrightarrow} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

$$E = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} \quad \text{e} \quad E^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{3} \end{bmatrix}.$$

(b) $\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix} \xrightarrow{L_2+2L_1} \begin{bmatrix}1&0&0\\2&1&0\\0&0&1\end{bmatrix} \xrightarrow{L_2-2L_1} \begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}$

$$E=\begin{bmatrix}1&0&0\\2&1&0\\0&0&1\end{bmatrix} \quad \text{e} \quad E^{-1}=\begin{bmatrix}1&0&0\\-2&1&0\\0&0&1\end{bmatrix}.$$

(c) $\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix} \xrightarrow{L_3\leftrightarrow L_1} \begin{bmatrix}0&0&1\\0&1&0\\1&0&0\end{bmatrix} \xrightarrow{L_3\leftrightarrow L_1} \begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}$

$$E=\begin{bmatrix}0&0&1\\0&1&0\\1&0&0\end{bmatrix} \quad \text{e} \quad E^{-1}=\begin{bmatrix}0&0&1\\0&1&0\\1&0&0\end{bmatrix}.$$

Para cada uma das alíneas anteriores, calculando os produtos EE^{-1} e $E^{-1}E$ confirma-se que é obtida a matriz identidade. ♦

No quadro seguinte resumimos o que entendemos por operação elementar inversa.

Operação elementar	Operação elementar inversa
$L_i \leftrightarrow L_j$	$L_i \leftrightarrow L_j$
$\alpha L_i \ \ (\alpha \neq 0)$	$\frac{1}{\alpha}L_i \ \ (\alpha \neq 0)$
$L_i + \alpha L_j$	$L_i - \alpha L_j$

Uma matriz elementar que é obtida da matriz identidade por troca de duas das suas linhas é um exemplo de uma matriz de *permutação*[9]. Como se deduz facilmente, a inversa de uma matriz elementar de permutação coincide com a própria matriz (ver Exemplo1.28-(c)).

Proposição 1.9. Qualquer matriz elementar é invertível e a sua inversa é uma matriz elementar.

[9]Uma matriz de permutação é uma matriz que se obtém da identidade efectuando um número finito de trocas de linhas.

Demonstração. Sendo E uma matriz elementar, esta matriz é obtida da identidade efectuando uma operação elementar. Designe-se por E_1 a matriz que se obtém da identidade efectuando a operação inversa da que permitiu obter E. Então, pela Proposição 1.8, temos

$$E_1 E = I.$$

De igual modo, como E é uma matriz elementar, também se verifica $EE_1 = I$. Portanto, a matriz elementar E_1 é a inversa de E. □

Como veremos, dada uma matriz invertível A, podemos realizar um número finito de operações elementares sobre as linhas de A até obter a matriz identidade. Isto significa que a matriz em escada na forma reduzida de uma matriz invertível é a matriz identidade. Para reduzir uma matriz invertível à matriz identidade por meio de operações elementares aplicamos o método conhecido por *método de eliminação de Gauss-Jordan*. Este método consiste na aplicação do método de eliminação de Gauss até obter uma matriz em escada com todos os pivôs iguais a 1, prosseguindo-se com a eliminação das entradas acima da diagonal principal usando os mesmos procedimentos do algoritmo de eliminação de Gauss, mas começando este algoritmo com a última coluna da matriz. Ilustramos no próximo exemplo os diversos passos do método de eliminação de Gauss- Jordan.

Exemplo 1.29. Vejamos como obter a matriz identidade por meio de operações elementares para a seguinte matriz invertível A.

$$A = \begin{bmatrix} 1 & -1 & 2 \\ 2 & 0 & 1 \\ -1 & 1 & 2 \end{bmatrix} \underset{L_3+L_1}{\overset{L_2-2L_1}{\longrightarrow}} \begin{bmatrix} 1 & -1 & 2 \\ 0 & 2 & -3 \\ 0 & 0 & 4 \end{bmatrix} \underset{1/2L_2}{\overset{1/4L_3}{\longrightarrow}} \begin{bmatrix} 1 & -1 & 2 \\ 0 & 1 & \frac{-3}{2} \\ 0 & 0 & 1 \end{bmatrix}$$

$$\overset{L_2+3/2L_3}{\longrightarrow} \begin{bmatrix} 1 & -1 & 2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \overset{L_1-2L_3}{\longrightarrow} \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \overset{L_1+L_2}{\longrightarrow} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I.$$

As operações realizadas na primeira linha da expressão anterior correspondem ao método de eliminação de Gauss, e as operações na segunda linha eliminam as entradas acima da diagonal principal começando com a última coluna da matriz em escada por linhas já obtida.

Realizámos 7 operações elementares. A cada operação elementar corresponde uma matriz elementar. Designemos por E_i a matriz elementar correspondente à operação elementar número i[10]. Para as operações elementares efectuadas, as matrizes E_i são:

$$E_1 = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \qquad E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{4} \end{bmatrix}$$

$$E_4 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad E_5 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \frac{3}{2} \\ 0 & 0 & 1 \end{bmatrix} \qquad E_6 = \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$E_7 = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Da Proposição 1.8, concluimos

$$E_7E_6E_5E_4E_3E_2E_1A = I. \tag{1.14}$$

Como qualquer matriz elementar é invertível (Proposição 1.9), e o produto de matrizes invertíveis é invertível (Proposição 1.6), a matriz $(E_7E_6E_5E_4E_3E_2E_1)$ é invertível. Logo, multiplicando a igualdade (1.14) à esquerda por $(E_7E_6E_5E_4E_3E_2E_1)^{-1}$ obtemos

$$A = (E_7E_6E_5E_4E_3E_2E_1)^{-1}I = (E_7E_6E_5E_4E_3E_2E_1)^{-1}. \tag{1.15}$$

Assim, pelo primeiro item da Proposição 1.7, a matriz A é invertível e

$$A^{-1} = E_7E_6E_5E_4E_3E_2E_1 = \begin{bmatrix} \frac{-1}{8} & \frac{1}{2} & \frac{-1}{8} \\ \frac{-5}{8} & \frac{1}{2} & \frac{3}{8} \\ \frac{1}{4} & 0 & \frac{1}{4} \end{bmatrix}.$$

Pode confirmar-se que a matriz anterior é de facto a inversa de A calculando o produto de A por essa matriz.

[10] As operações que aparecem sobrepostas na primeira linha da aplicação do método podem ser realizadas por qualquer ordem. Em particular, isso significa que as matrizes elementares correspondentes comutam.

A igualdade (1.15) é equivalente a

$$A = (E_7E_6E_5E_4E_3E_2E_1)^{-1} = E_1^{-1}E_2^{-1}E_3^{-1}E_4^{-1}E_5^{-1}E_6^{-1}E_7^{-1},$$

onde na última igualdade aplicámos (1.13). Como a inversa de uma matriz elementar é ainda uma matriz elementar, a expressão anterior dá a matriz invertível A como um produto de matrizes elementares. ♦

Algumas propriedades que vimos no exemplo anterior caracterizam as matrizes invertíveis. Esta afirmação é parte do conteúdo do teorema seguinte.

Teorema 1.4. Sendo A uma matriz quadrada, são equivalentes as afirmações:

(i) A é invertível.

(ii) O sistema $A\mathbf{x} = \mathbf{b}$ tem solução única.

(iii) É possível aplicar operações elementares sobre as linhas de A e obter a matriz identidade.

(iv) A matriz A pode exprimir-se como um produto de matrizes elementares.

Demonstração. Provemos a seguinte sequência de implicações: $(i) \Rightarrow (ii) \Rightarrow (iii) \Rightarrow (iv) \Rightarrow (i)$.

$(i) \Rightarrow (ii)$: Se A é invertível, o sistema $A\mathbf{x} = \mathbf{b}$ tem solução única uma vez que, multiplicando (à esquerda) a igualdade $A\mathbf{x} = \mathbf{b}$ por A^{-1} se obtém $\mathbf{x} = A^{-1}\mathbf{b}$.

$(ii) \Rightarrow (iii)$: Seja $\mathbf{x} = (c_1, \ldots, c_n)$ a única solução de $A\mathbf{x} = \mathbf{b}$. Assim, a matriz aumentada deste sistema pode ser reduzida (por meio de operações elementares sobre as suas linhas) à matriz

$$\left[\begin{array}{cccc|c} 1 & 0 & \cdots & 0 & c_1 \\ 0 & 1 & \cdots & 0 & c_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & c_n \end{array}\right].$$

Isto é, existe uma sequência de operações elementares que reduzem A à matriz identidade.

$(iii) \Rightarrow (iv)$: Se é possível aplicar operações elementares sobre as linhas de A e obter a matriz identidade, então existe um número finito de matrizes elementares, sejam $E_1, \ldots, E_k$, tal que

$$E_k \cdots E_2E_1A = I.$$

Como qualquer matriz elementar é invertível (Proposição 1.9) e o produto de matrizes invertíveis é invertível (Proposição 1.6), podemos multiplicar a igualdade anterior, à esquerda, por $(E_k \cdots E_2 E_1)^{-1}$, obtendo-se

$$A = (E_k \cdots E_2 E_1)^{-1} I = E_1^{-1} E_2^{-1} \cdots E_k^{-1}.$$

Uma vez que a inversa de uma matriz elementar ainda é uma matriz elementar (Proposição 1.9), a igualdade anterior exprime A como um produto de matrizes elementares.

$(iv) \Rightarrow (i)$: Se A é igual ao produto de matrizes elementares, a matriz A é invertível visto que cada matriz elementar é invertível, e o produto de matrizes invertíveis ainda é invertível. $\square$

Método de eliminação de Gauss-Jordan e cálculo da inversa

O Teorema 1.4 diz-nos que uma matriz invertível pode reduzir-se à matriz identidade por meio de operações elementares sobre as suas linhas. Assim, se A é invertível existem matrizes elementares $E_1, \ldots, E_k$ tais que $E_k \cdots E_2 E_1 A = I$. Esta igualdade é equivalente a

$$A^{-1} = E_k \cdots E_2 E_1 I. \tag{1.16}$$

Como o produto EA é a matriz que se obtém de A efectuando a operação elementar que permitiu obter E da identidade, a igualdade (1.16) diz-nos que a inversa de A se pode obter realizando sobre a matriz identidade as mesmas operações elementares (e pela mesma ordem) que reduzem A à matriz identidade. Esquematicamente,

$$\begin{bmatrix} A & | & I \end{bmatrix} \xrightarrow{\text{Operações elementares}} \begin{bmatrix} I & | & A^{-1} \end{bmatrix}.$$

Exemplo 1.30. Determinemos a inversa de

$$A = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & 1 \\ -1 & 1 & 2 \end{bmatrix}.$$

Coloquemos a matriz identidade de terceira ordem a par da matriz A, e efectuemos operações elementares sobre $[A \,|\, I]$ até obtermos a matriz $[I \,|\, A^{-1}]$.

$$\left[\begin{array}{ccc|ccc} 1 & -1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ -1 & 1 & 2 & 0 & 0 & 1 \end{array}\right] \underset{L_3+L_1}{\overset{L_2-L_1}{\longrightarrow}} \left[\begin{array}{ccc|ccc} 1 & -1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & -1 & 1 & 0 \\ 0 & 0 & 2 & 1 & 0 & 1 \end{array}\right]$$

$$\xrightarrow{1/2L_3} \left[\begin{array}{ccc|ccc} 1 & -1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & -1 & 1 & 0 \\ 0 & 0 & 1 & \frac{1}{2} & 0 & \frac{1}{2} \end{array}\right] \xrightarrow{L_2-L_3} \left[\begin{array}{ccc|ccc} 1 & -1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & \frac{-3}{2} & 1 & \frac{-1}{2} \\ 0 & 0 & 1 & \frac{1}{2} & 0 & \frac{1}{2} \end{array}\right]$$

$$\xrightarrow{L_1+L_2} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{-1}{2} & 1 & \frac{-1}{2} \\ 0 & 1 & 0 & \frac{-3}{2} & 1 & \frac{-1}{2} \\ 0 & 0 & 1 & \frac{1}{2} & 0 & \frac{1}{2} \end{array}\right].$$

Logo, a matriz inversa de A é

$$A^{-1} = \begin{bmatrix} \frac{-1}{2} & 1 & \frac{-1}{2} \\ \frac{-3}{2} & 1 & \frac{-1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}.$$

♦

Algumas consequências da invertibilidade de uma matriz

Um sistema de equações lineares com igual número de equações e incógnitas possui matriz dos coeficientes quadrada, e portanto podemos falar da invertibilidade desta matriz. O Teorema 1.4 diz-nos que um tal sistema é possível e determinado se e só se a matriz dos coeficientes do sistema é invertível. Além disso, a demonstração desse teorema estabelece que a solução (única) se obtém multiplicando o vector correspondente ao termo independente do sistema pela inversa da matriz dos coeficientes do sistema. Para referência futura enunciemos este resultado.

Proposição 1.10. Seja $\mathbf{b}$ um vector qualquer. Um sistema $A\mathbf{x} = \mathbf{b}$ cuja matriz dos coeficientes é quadrada, tem solução única se e só se A é invertível. Além disso, sendo A invertível a solução do sistema $A\mathbf{x} = \mathbf{b}$ é

$$\mathbf{x} = A^{-1}\mathbf{b}.$$

Em particular, um sistema homogéneo $A\mathbf{x} = \mathbf{0}$ tem solução única $\mathbf{x} = \mathbf{0}$ se e só se a matriz A é invertível.

Demonstração. Pelo Teorema 1.4, um sistema $A\mathbf{x} = \mathbf{b}$ cuja matriz dos coeficientes é quadrada, tem solução única se e só se A é invertível. Se A^{-1} existe, podemos multiplicar $A\mathbf{x} = \mathbf{b}$, à esquerda, por A^{-1} e obtemos

$$A^{-1}A\mathbf{x} = A^{-1}\mathbf{b} \Longleftrightarrow \mathbf{x} = A^{-1}\mathbf{b}.$$

Se $\mathbf{b} = \mathbf{0}$, como o produto de uma qualquer matriz pelo vector nulo é o vector nulo, a solução anterior é $\mathbf{x} = \mathbf{0}$. $\square$

Nota 11. *Na demonstração anterior não faz sentido multiplicar $A\mathbf{x} = \mathbf{b}$ à direita por A^{-1}, já que tal multiplicação não está definida (o vector $\mathbf{b}$ é do tipo $n \times 1$ e A^{-1} é do tipo $n \times n$).*

Exemplo 1.31. Consideremos o sistema

$$\begin{cases} 2x_1 + x_2 = 6 \\ 4x_1 + 4x_2 = 8. \end{cases}$$

A matriz dos coeficientes do sistema é exactamente a matriz A do Exemplo 1.25, na página 50. Nesse exemplo calculámos a inversa de A, e portanto o sistema tem solução única dada por

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = A^{-1}\begin{bmatrix} 6 \\ 8 \end{bmatrix} = \begin{bmatrix} 1 & \frac{-1}{4} \\ -1 & \frac{1}{2} \end{bmatrix}\begin{bmatrix} 6 \\ 8 \end{bmatrix} = \begin{bmatrix} 4 \\ -2 \end{bmatrix}.$$

♦

De acordo com a definição de matriz invertível, uma matriz A é invertível se existir uma matriz B que satisfaz as condições:

$$AB = I \quad \text{e} \quad BA = I.$$

Na proposição seguinte mostramos que é suficiente que uma das condições se verifique para a outra ser automaticamente válida (e portanto A ser invertível).

Proposição 1.11. Seja A uma matriz quadrada.

a) Se B é uma matriz que satisfaz $BA = I$ então B é a inversa de A.

b) Se B é uma matriz que satisfaz $AB = I$ então B é a inversa de A.

Demonstração. Mostremos o item a). Comecemos por provar que se $BA = I$ então A é invertível. Para tal, basta mostar que a única solução do sistema homogéneo $A\mathbf{x} = \mathbf{0}$ é a solução nula (conforme Proposição 1.10). Seja $\mathbf{u}$ uma qualquer solução de $A\mathbf{x} = \mathbf{0}$. Multiplicando ambos os membros da igualdade $A\mathbf{u} = \mathbf{0}$ por B, obtém-se

$$BA\mathbf{u} = B\mathbf{0} \iff I\mathbf{u} = \mathbf{0} \iff \mathbf{u} = \mathbf{0}.$$

Como $\mathbf{u}$ é uma qualquer solução de $A\mathbf{x} = \mathbf{0}$, este sistema apenas admite a solução nula, e portanto a Proposição 1.10 garante que a matriz A é invertível.

Sendo A invertível e $BA = I$, multiplicando (à direita) ambos os membros desta igualdade por A^{-1}, tem-se

$$BAA^{-1} = A^{-1} \iff BI = A^{-1} \iff B = A^{-1}.$$

Para demonstrar o item b) usamos o item a) trocando os papéis de A e B. Assim, se $AB = I$ o item a) diz-nos que A é a inversa de B, e portanto

$$A = B^{-1} \iff A^{-1} = B.$$

□

A Proposição 1.6 diz que se A e B são matrizes (do mesmo tipo) invertíveis, então o produto AB é invertível. No teorema a seguir mostramos que é condição necessária e suficiente para a invertibilidade do produto AB que ambas as matrizes A e B sejam invertíveis.

Teorema 1.5. Sejam A e B matrizes quadradas da mesma ordem. A matriz produto AB é invertível se e só se as matrizes A e B são invertíveis. Além disso, a inversa do produto é dada por

$$(AB)^{-1} = B^{-1}A^{-1}.$$

Demonstração. Tendo em conta a Proposição 1.6, resta mostrar que se AB é uma matriz invertível, então as matrizes A e B são invertíveis.

Sendo a matriz AB invertível, por definição de inversa existe uma matriz D tal que

$$D(AB) = I \quad \text{e} \quad (AB)D = I.$$

Como o produto de matrizes é associativo, as igualdades anteriores são equivalentes a

$$(DA)B = I \quad \text{e} \quad A(BD) = I.$$

Usando, respectivamente, os itens a) e b) da Proposição 1.11, tem-se

$$\begin{aligned}(DA)B = I &\Longrightarrow B \text{ é invertível}\\ A(BD) = I &\Longrightarrow A \text{ é invertível.}\end{aligned}$$

□

Finalizamos esta secção enunciando algumas características de matrizes reais invertíveis.

Proposição 1.12. Seja A uma matriz real do tipo $n \times n$. São equivalentes as afirmações seguintes.

a) A é invertível.

b) A tem característica n.

c) Os vectores coluna de A geram $\mathbb{R}^n$.

Demonstração. Esta prova será realizada mostrando as seguintes relações: $a) \Longleftrightarrow b)$, $a) \Rightarrow c)$ e $c) \Rightarrow a)$.

$a) \Longleftrightarrow b)$: Uma matriz é invertível se e só se a sua forma reduzida em escada por linhas é a identidade. Consequentemente, uma matriz do tipo $n \times n$ é invertível se e só se tem característica igual a n.

$a) \Rightarrow c)$: Como A é invertível, o sistema $A\mathbf{x} = \mathbf{b}$ é possível, qualquer que seja o vector $\mathbf{b}$ de $\mathbb{R}^n$. Assim, pela Proposição 1.3, qualquer vector de $\mathbb{R}^n$ é uma combinação linear dos vectores colunas de A. Isto significa que qualquer vector de $\mathbb{R}^n$ pertence ao conjunto gerado pelas colunas de A.

$c) \Rightarrow a)$: Como qualquer vector de $\mathbb{R}^n$ é uma combinação linear das colunas de A, os vectores

$$\mathbf{e}_1 = \begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix}, \mathbf{e}_2 = \begin{bmatrix}0\\1\\\vdots\\0\end{bmatrix}, \ldots, \mathbf{e}_n = \begin{bmatrix}0\\0\\\vdots\\1\end{bmatrix} \in \mathbb{R}^n$$

são combinações lineares das colunas de A. Isto significa, que existem vectores $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ tais que

$$A\mathbf{x}_1 = \mathbf{e}_1,\ A\mathbf{x}_2 = \mathbf{e}_2,\ \ldots,\ A\mathbf{x}_n = \mathbf{e}_n.$$

Seja B a matriz cujas colunas são os vectores $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$, por esta ordem. Usando a Definição 1.12 de produto de matrizes, tem-se

$$\begin{aligned} AB &= A \begin{bmatrix} | & | & & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \ldots & \mathbf{x}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ A\mathbf{x}_1 & A\mathbf{x}_2 & \ldots & A\mathbf{x}_n \\ | & | & & | \end{bmatrix} \\ &= \begin{bmatrix} | & | & & | \\ \mathbf{e}_1 & \mathbf{e}_2 & \ldots & \mathbf{e}_n \\ | & | & & | \end{bmatrix} = I_n. \end{aligned}$$

Da Proposição 1.11, segue que B é a inversa de A.

$\square$

1.5 Matrizes triangulares

Matrizes ditas triangulares surgem nos mais diversos contextos. Em particular, a aplicação do método de eliminação de Gauss a uma matriz quadrada produz uma matriz em escada que é triangular superior.

Definição 1.17. Uma matriz quadrada com todas as entradas abaixo das entradas da diagonal principal iguais a zero é designada por matriz *triangular superior*.

Uma matriz quadrada com todas as entradas acima das entradas da diagonal principal iguais a zero diz-se *triangular inferior*.

Uma matriz A do tipo $n \times n$ que seja triangular superior e triangular inferior, isto é, tal que todas as entradas não nulas estão na diagonal principal, diz-se uma matriz *diagonal* e denota-se

$$A = \operatorname{diag}(a_{11}, a_{22}, \ldots, a_{nn}).$$

Um exemplo de uma matriz diagonal é a matriz identidade $I_n = \operatorname{diag}(1, 1, \ldots, 1)$.

A definição de matriz triangular exprime-se em termos das entradas da matriz da forma que se segue.

Seja $A = [a_{ij}]_{i,j=1,\ldots,n}$.

i) A é triangular superior $\iff a_{ij} = 0$ para todo $i > j$.

ii) A é triangular inferior $\iff a_{ij} = 0$ para todo $i < j$.

iii) A é diagonal $\iff a_{ij} = 0$ para todo $i \neq j$.

Exemplo 1.32. A matriz

$$\begin{bmatrix} 1 & 7 & 5 & 2 \\ 0 & 8 & 2 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

é triangular superior e

$$\begin{bmatrix} 1 & 0 & 0 \\ 2 & 8 & 0 \\ 1 & 4 & 1 \end{bmatrix}$$

é triangular inferior. ♦

A matriz transposta de uma matriz triangular superior é uma matriz triangular inferior e vice-versa.

Uma matriz triangular é invertível se e só se a sua diagonal principal não tem entradas nulas. De facto, como a característica de uma matriz triangular é igual ao número de entradas não nulas da sua diagonal principal, pela Proposição 1.12 (página 62) uma matriz triangular é invertível se e só se todas as entradas da diagonal principal são distintas de zero. No Exemplo 1.32 a primeira matriz não é invertível enquanto que a segunda matriz é.

Nota 12. *Se uma matriz não é triangular podem existir zeros na diagonal principal e a matriz ser invertível. Por exemplo, a matriz*

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

é invertível (é uma matriz de permutação).

O algoritmo apresentado para o cálculo da inversa (método de eliminação de Gauss-Jordan) permite concluir que a inversa de uma matriz triangular inferior é triangular inferior, e a inversa de uma matriz triangular superior é triangular superior.

Deixamos como exercício mostrar que o produto de matrizes triangulares superiores (resp. inferiores) é ainda uma matriz triangular superior (resp. inferior). Para tal, basta usar a expressão (1.10) para o cálculo da entrada ij da matriz produto.

Resumimos no quadro seguinte os resultados que acabámos de referir.

Propriedades das matrizes triangulares

(a) A transposta de uma matriz triangular superior (resp. inferior) é triangular inferior (resp. superior).

(b) O produto de matrizes triangulares superiores (resp. inferiores) é triangular superior (resp. inferior).

(c) Uma matriz triangular é invertível se e só se as entradas da diagonal principal são todas diferentes de zero.

(d) A inversa de uma matriz triangular superior (resp. inferior) é uma matriz triangular superior (resp. inferior).

1.5.1 Factorização LU

Uma factorização de uma matriz A é uma igualdade que exprime A como um produto de uma ou mais matrizes. A factorização $A = LU$ exprime a matriz quadrada A como o produto de uma matriz triangular inferior L (do inglês "lower triangular") por uma matriz triangular superior U (do inglês "upper triangular"). Quando as matrizes triangulares L e U gozam das propriedades adicionais que especificaremos a seguir, dizemos que $A = LU$ é uma factorização LU de A.

A factorização LU desempenha um papel fundamental na classificação de matrizes simétricas (estudo efectuado no Capítulo 7) e está na base de algoritmos vantajosos do ponto de vista computacional para a resolução de sistemas lineares. Por exemplo, suponha-se que é dada uma factorização $A = LU$, com L e U invertíveis, e que pretendemos resolver um sistema linear $A\mathbf{x} = \mathbf{b}$. A solução deste sistema pode ser determinada da seguinte forma:

- O sistema $A\mathbf{x} = \mathbf{b}$ é reescrito na forma
$$LU\mathbf{x} = \mathbf{b}. \tag{1.17}$$
- Defina-se o vector
$$\mathbf{y} = U\mathbf{x}. \tag{1.18}$$
- Usando (1.18), o sistema (1.17) é equivalente ao sistema triangular $L\mathbf{y} = \mathbf{b}$. Resolva-se este sistema para obter $\mathbf{y}$.
- Substituindo $\mathbf{y}$ em (1.18) e resolvendo o sistema triangular correspondente, obtém-se a solução do sistema inicial $A\mathbf{x} = \mathbf{b}$.

O processo acabado de descrever substitui o sistema $A\mathbf{x} = \mathbf{b}$ por dois sistemas ($L\mathbf{y} = \mathbf{b}$ e $U\mathbf{x} = \mathbf{y}$) cujas matrizes dos coeficientes são triangulares. Tal representa uma economia computacional considerável, em particular quando se pretende resolver vários sistemas com a mesma matriz dos coeficientes A e termos independentes distintos.

No exemplo seguinte ilustramos a aplicação deste método.

Exemplo 1.33. Consideremos a seguinte factorização LU de A

$$A = \begin{bmatrix} 4 & 1 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} 4 & 1 \\ 0 & \frac{1}{2} \end{bmatrix}.$$

Podemos obter a solução do sistema

$$\begin{cases} 4x_1 + x_2 = 3 \\ 2x_1 + x_2 = 1, \end{cases}$$

resolvendo os sistemas

$$L\mathbf{y} = \mathbf{b} \iff \begin{cases} y_1 \qquad\quad = 3 \\ \frac{1}{2}y_1 + y_2 = 1 \end{cases} \quad \text{e} \quad U\mathbf{x} = \mathbf{y} \iff \begin{cases} 4x_1 + \;\; x_2 = y_1 \\ \qquad\; \frac{1}{2}x_2 = y_2. \end{cases}$$

A solução do sistema $L\mathbf{y} = \mathbf{b}$ é $y_1 = 3$ e $y_2 = -\frac{1}{2}$. Substituindo o vector $\mathbf{y}$ obtido em $U\mathbf{x} = \mathbf{y}$, a solução do sistema é $x_1 = 1$ e $x_2 = -1$. ♦

Apresentamos a seguir o que entendemos por uma factorização LU.

Factorização LU

Diz-se que a matriz A (invertível) admite uma factorização LU se $A = LU$, em que

- L é uma matriz triangular inferior com todas as entradas da diagonal principal iguais a 1.
- U é uma matriz triangular superior com todas as entradas da diagonal principal diferentes de zero.

Vejamos como calcular uma factorização LU para uma matriz invertível A no caso em que é possível aplicar o método de eliminação de Gauss sem troca de linhas. Como A é quadrada, o método de eliminação de Gauss produz uma matriz em escada U que é triangular superior e que possui na diagonal principal os pivôs.

Como estamos supondo que não efectuamos troca de linhas, não podemos encontrar durante o processo nenhuma entrada nula na diagonal principal, caso contrário seria necessário usar troca de linhas para obter U. Logo, U tem entradas na diagonal principal diferentes de zero.

Operações elementares sobre as linhas de A traduzem-se na multiplicação sucessiva (à esquerda) por matrizes elementares. Assim, existe um número finito de matrizes elementares $E_1, E_2, \dots, E_k$, tais que

$$E_k \cdots E_2 E_1 A = U.$$

Como não foram efectuadas troca de linhas, as matrizes $E_1, \dots, E_k$ são triangulares inferiores com todas as entradas na diagonal principal iguais a 1. Logo, o produto $E_k \cdots E_2 E_1$ é uma matriz triangular inferior com 1's na diagonal principal. Por outro lado, como matrizes elementares são invertíveis, o seu produto é invertível, tendo-se portanto

$$E_k \cdots E_2 E_1 A = U \iff A = (E_k \cdots E_2 E_1)^{-1} U. \tag{1.19}$$

A matriz $(E_k \cdots E_2 E_1)^{-1}$ é a inversa de uma matriz triangular inferior pelo que também é triangular inferior. Podemos assim tomar para L a matriz

$$L = (E_k \cdots E_2 E_1)^{-1} = E_1^{-1} E_2^{-1} \cdots E_k^{-1}.$$

A matriz L é igual a um produto de matrizes (elementares) triangulares inferiores com 1's na diagonal principal. Consequentemente, L tem todas as entradas na diagonal principal iguais a 1. Ou seja, a equação (1.19) exprime A na forma LU.

Acabámos de mostrar o resultado que enunciamos a seguir.

Teorema 1.6. Se a aplicação do método de eliminação de Gauss, sem troca de linhas, a uma matriz quadrada A produz uma matriz triangular superior U com todas as entradas na diagonal principal não nulas, então A admite a factorização $A = LU$, onde L é uma matriz triangular inferior com todas as entradas da diagonal principal iguais a 1.

Nota 13. *Para que uma matriz admita uma factorização LU é necessário que a matriz seja invertível, uma vez que* L *e* U *são invertíveis. A condição de invertibilidade da matriz não é contudo uma condição suficiente. No Exercício 7.13 (pág. 398) indica-se uma matriz invertível que não admite uma factorização* LU.

Refira-se ainda que a Proposição 7.9 na página 398, fornece uma condição necessária e suficiente para a existência de uma factorização LU.

Exemplo 1.34. Verifiquemos que a matriz

$$A = \begin{bmatrix} 2 & 2 & 2 \\ 6 & 0 & 3 \\ 4 & 1 & 1 \end{bmatrix}$$

admite uma factorização LU e calculemos os respectivos factores L e U.

$$A = \begin{bmatrix} 2 & 2 & 2 \\ 6 & 0 & 3 \\ 4 & 1 & 1 \end{bmatrix} \xrightarrow{L_2 - 3L_1} \begin{bmatrix} 2 & 2 & 2 \\ 0 & -6 & -3 \\ 4 & 1 & 1 \end{bmatrix} \quad \rightsquigarrow \quad E_1 = \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 2 & 2 \\ 0 & -6 & -3 \\ 4 & 1 & 1 \end{bmatrix} \xrightarrow{L_3 - 2L_1} \begin{bmatrix} 2 & 2 & 2 \\ 0 & -6 & -3 \\ 0 & -3 & -3 \end{bmatrix} \quad \rightsquigarrow \quad E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix}.$$

$$\begin{bmatrix} 2 & 2 & 2 \\ 0 & -6 & -3 \\ 0 & -3 & -3 \end{bmatrix} \xrightarrow{L_3 - 1/2 L_2} \begin{bmatrix} 2 & 2 & 2 \\ 0 & -6 & -3 \\ 0 & 0 & \frac{-3}{2} \end{bmatrix} = U \quad \rightsquigarrow \quad E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \frac{-1}{2} & 1 \end{bmatrix}$$

Logo, $E_3E_2E_1A = U$ e $L = (E_3E_2E_1)^{-1}$. Efectuando o produto $L = E_1^{-1}E_2^{-1}E_3^{-1}$, tem-se

$$L = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 2 & \frac{1}{2} & 1 \end{bmatrix} \quad \text{e} \quad U = \begin{bmatrix} 2 & 2 & 2 \\ 0 & -6 & -3 \\ 0 & 0 & \frac{-3}{2} \end{bmatrix}.$$

♦

A matriz L obtida no exemplo anterior é uma matriz triangular inferior cujas entradas da diagonal principal são iguais a 1 e as entradas abaixo da diagonal principal são os simétricos dos valores (ou *multiplicadores*) usados no método de eliminação de Gauss para anular essa entrada. Esta propriedade da matriz L é verificada para qualquer factorização LU. De facto, atendendo à forma particular da matriz inversa de uma matriz elementar (que não seja de permutação), conclui-se facilmente que o produto $L = (E_k \cdots E_2E_1)^{-1}$ tem sempre a forma observada no exemplo anterior. Resumindo,

Se A admite uma factorização LU a matriz L é uma matriz triangular inferior com todas as entradas da diagonal principal iguais a 1, e cada entrada abaixo da diagonal principal é igual ao simétrico do escalar usado no método de eliminação de Gauss para anular essa entrada. Estes escalares designam-se por *multiplicadores*.

Deixamos como exercício mostrar que se uma matriz é factorizável na forma LU, a factorização é única.

Exercício 1.3. Mostre que a factorização $A = LU$ é única.

Sugestão: Supor que L_1U_1 e L_2U_2 são duas factorizações LU de A. Usar o facto dos factores L e U serem matrizes triangulares invertíveis para verificar que $L_2^{-1}L_1 = I = U_2U_1^{-1}$.

▲

Para finalizar esta secção, refira-se que se A é invertível é possível que durante o processo de eliminação de Gauss se encontre uma entrada nula na diagonal principal que pode ser removida por troca de linhas. Mostra-se neste caso que existe uma matriz P obtida da matriz identidade I efectuando uma sucessão de trocas de linhas tal que $PA = LU$. O procedimento para obter a matriz (de permutação) P sai do âmbito deste livro. Remetemos o leitor interessado nesta questão, ou na eficiência computacional dos algoritmos aqui estudados para a resolução de sistemas lineares, para um livro de álgebra linear numérica como por exemplo [12].

1.6 Partição de matrizes em blocos

Muitas vezes é conveniente encarar uma matriz como uma matriz cujas entradas são matrizes mais pequenas chamadas blocos. Matrizes em blocos aparecem em várias áreas de aplicação da álgebra linear, sendo que a partição em blocos está quase sempre relacionada com a estrutura do modelo em estudo, o qual é frequentemente traduzido através de um sistema linear. Neste texto, em particular nalgumas demonstrações, usamos operações com matrizes em blocos. Por isso, nesta secção fazemos uma breve referência a estas operações. Comecemos por dar um exemplo de uma partição em blocos de uma matriz do tipo 3×6. Seja

$$A = \left[\begin{array}{ccc|cc|c} 2 & 1 & 3 & 4 & 1 & 2 \\ 5 & 3 & 2 & 7 & 8 & 1 \\ \hline -1 & 3 & -2 & 4 & 5 & 1 \end{array}\right]. \tag{1.20}$$

A matriz A pode ser vista como uma matriz em blocos do tipo 2×3

$$A = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{bmatrix},$$

onde os blocos são respectivamente as matrizes

$$A_{11} = \begin{bmatrix} 2 & 1 & 3 \\ 5 & 3 & 2 \end{bmatrix}, \quad A_{12} = \begin{bmatrix} 4 & 1 \\ 7 & 8 \end{bmatrix}, \qquad A_{13} = \begin{bmatrix} 2 \\ 1 \end{bmatrix},$$

$$A_{21} = \begin{bmatrix} -1 & 3 & -2 \end{bmatrix}, \quad A_{22} = \begin{bmatrix} 4 & 5 \end{bmatrix}, \quad \text{e } A_{23} = \begin{bmatrix} 1 \end{bmatrix}.$$

Todas as operações com matrizes em blocos são definidas como se as submatrizes (blocos) fossem as entradas de uma matriz. Por exemplo, se A e B são do mesmo tipo e estão particionadas por forma a que os blocos correspondentes sejam do mesmo tipo, a soma $A+B$ é ainda uma matriz em blocos cujos blocos são a soma dos blocos de A com os blocos homólogos de B. A multiplicação de uma matriz em blocos por um escalar pode também ser efectuada bloco a bloco.

Para obter o produto AB, onde A e B são matrizes em blocos, é necessário que a partição de A por colunas coincida com a partição de B por linhas. Por exemplo, a partição da matriz A em (1.20) foi realizada usando 3, 2 e 1 colunas. Logo, o produto AB só está definido se a partição de B for realizada em 3, 2 e 1 linhas, como é o caso da matriz

$$B = \left[\begin{array}{cc} 1 & 7 \\ 2 & 3 \\ 4 & 2 \\ \hline 2 & 3 \\ 1 & 4 \\ \hline 8 & 1 \end{array}\right], \quad \text{isto é,} \quad B = \begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix}.$$

O produto AB é efectuado aplicando a fórmula utilizada para definir o produto de matrizes usando nessa fórmula os blocos em vez das entradas. Mais precisamente, para a matriz A dada em (1.20), tem-se

$$AB = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix} = \begin{bmatrix} A_{11}B_1 + A_{12}B_2 + A_{13}B_3 \\ A_{21}B_1 + A_{22}B_2 + A_{23}B_3 \end{bmatrix},$$

com

$$A_{11}B_1 + A_{12}B_2 + A_{13}B_3 = \begin{bmatrix} 2 & 1 & 3 \\ 5 & 3 & 2 \end{bmatrix} \begin{bmatrix} 1 & 7 \\ 2 & 3 \\ 4 & 2 \end{bmatrix} + \begin{bmatrix} 4 & 1 \\ 7 & 8 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix} + \begin{bmatrix} 2 \\ 1 \end{bmatrix} \begin{bmatrix} 8 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 16 & 23 \\ 19 & 48 \end{bmatrix} + \begin{bmatrix} 9 & 16 \\ 22 & 53 \end{bmatrix} + \begin{bmatrix} 16 & 2 \\ 8 & 1 \end{bmatrix} = \begin{bmatrix} 41 & 41 \\ 49 & 102 \end{bmatrix},$$

$$A_{21}B_1 + A_{22}B_2 + A_{23}B_3 = \begin{bmatrix} -1 & 3 & -2 \end{bmatrix} \begin{bmatrix} 1 & 7 \\ 2 & 3 \\ 4 & 2 \end{bmatrix} + \begin{bmatrix} 4 & 5 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix} + \begin{bmatrix} 1 \end{bmatrix} \begin{bmatrix} 8 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} -3 & -2 \end{bmatrix} + \begin{bmatrix} 13 & 32 \end{bmatrix} + \begin{bmatrix} 8 & 1 \end{bmatrix} = \begin{bmatrix} 18 & 31 \end{bmatrix}.$$

Assim, a matriz em blocos AB é

$$AB = \left[\begin{array}{cc} 41 & 41 \\ 49 & 102 \\ \hline 18 & 31 \end{array}\right].$$

Sugere-se que verifique que a matriz obtida coincide com a matriz produto AB que se obtém aplicando a fórmula (1.10) na página 41.

Uma matriz quadrada em blocos, da forma

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1r} \\ \mathbf{0} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & A_{kr} \end{bmatrix},$$

onde os blocos A_{ii} são matrizes quadradas e $\mathbf{0}$ denota matrizes nulas, diz-se uma *matriz triangular superior por blocos*. De modo análogo se define *matriz triangular inferior por blocos* como uma matriz em que os blocos acima da diagonal são nulos. Uma matriz quadrada A diz-se uma *matriz diagonal por blocos* se os únicos blocos não nulos são os blocos (necessariamente quadrados) na diagonal.

Uma matriz triangular por blocos não é necessariamente uma matriz triangular, como se pode observar pelos exemplos que se seguem.

Exemplo 1.35. Exemplos de matrizes triangulares por blocos.

a) $\left[\begin{array}{cc|c} 2 & 1 & 0 \\ 3 & 2 & 0 \\ \hline 0 & 0 & 1 \end{array}\right]$ é uma matriz diagonal por blocos.

b) $B = \left[\begin{array}{cc|ccc} 2 & 1 & 5 & 0 & 2 \\ 3 & 2 & 3 & 0 & 3 \\ \hline 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 2 & 0 & 1 \end{array}\right]$ é uma matriz triangular superior por blocos.

c) $C = \left[\begin{array}{cc|cc} 2 & 1 & 0 & 0 \\ 3 & 1 & 0 & 0 \\ \hline 1 & 2 & 1 & 2 \end{array}\right]$ não é uma matriz triangular por blocos uma vez que o bloco C_{22} não é uma matriz quadrada (a matriz C também não é quadrada).

♦

Consideremos agora a matriz triangular por blocos $A = \begin{bmatrix} A_{11} & A_{12} \\ \mathbf{0} & A_{22} \end{bmatrix}$, em que as submatrizes A_{11} e A_{22} são quadradas respectivamente $p \times p$ e $q \times q$. Pretende-se saber em que condições esta matriz é invertível, e nesse caso, calcular a sua inversa.

Usando a definição de inversa, a matriz A^{-1} pode ser vista como uma matriz em blocos B tal que $AB = I$. Isto é,

$$AB = \begin{bmatrix} A_{11} & A_{12} \\ \mathbf{0} & A_{22} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} I_p & \mathbf{0} \\ \mathbf{0} & I_q \end{bmatrix},$$

onde I_p e I_q são as matrizes identidade de ordens p e q respectivamente. Efectuando este produto, obtém-se

$$\begin{aligned} A_{11}B_{11} + A_{12}B_{21} &= I_p \\ A_{11}B_{12} + A_{12}B_{22} &= \mathbf{0} \\ A_{22}B_{21} &= \mathbf{0} \\ A_{22}B_{22} &= I_q. \end{aligned} \tag{1.21}$$

A (última) igualdade $A_{22}B_{22} = I_q$ é válida se e só se A_{22} for invertível (Proposição 1.11, pág. 60). Logo, $B_{22} = A_{22}^{-1}$. Multiplicando, à esquerda, a penúltima equação por A_{22}^{-1} obtém-se a matriz nula $B_{21} = \mathbf{0}$. Substituindo $B_{21} = \mathbf{0}$ na primeira equação vem $A_{11}B_{11} = I_p$. Esta equação é satisfeita se e só se A_{11} for invertível. Assim, se as matrizes A_{11} e A_{22} admitem inversa, a segunda equação de (1.21) é equivalente a

$$A_{11}B_{12} = -A_{12}B_{22} = -A_{12}A_{22}^{-1} \iff B_{12} = -A_{11}^{-1}A_{12}A_{22}^{-1}.$$

Por conseguinte, A é invertível se e só se as matrizes A_{11} e A_{22} são invertíveis, e a inversa de A é

$$A^{-1} = \begin{bmatrix} A_{11}^{-1} & -A_{11}^{-1}A_{12}A_{22}^{-1} \\ \\ \mathbf{0} & A_{22}^{-1} \end{bmatrix}. \tag{1.22}$$

Não é difícil verificar que a prova realizada para uma matriz triangular com apenas 2 blocos na diagonal se generaliza a qualquer matriz triangular por blocos com um número superior de blocos. Concluindo,

Uma matriz triangular por blocos é invertível se e só se os blocos na diagonal (principal) são matrizes invertíveis.

Exercício 1.4. Calcular a inversa da matriz B do Exemplo 1.35 usando a expressão (1.22). Confirmar o resultado aplicando o método de eliminação de Gauss-Jordan a B.

♦

Exercícios

1. Quais das seguintes equações são lineares em x, y e z?

a) $x + 5y - \sqrt{2}z = 1$.
b) $x + 5y - \sqrt{2z} = 1$.
c) $x = -2y + \pi$.
d) $x + 5y + \cos z = 0$.
e) $x^{-2} + y - 3z = -3$.

2. Determine a equação da parábola, $y = ax^2+bx+c$, que passa pelos pontos $(1,3)$, $(2,4)$ e $(-1,-1)$.

3. Use os resultados do problema anterior para determinar o polinómio p, de grau menor ou igual a dois, que toma os valores indicados na tabela seguinte.

x	1	2	-1
$p(x)$	3	4	-1

4. Diga para que valores de α e de β o sistema seguinte é possível.

$$\begin{cases} -2x + 10y = \alpha \\ 8x - 40y = \beta. \end{cases}$$

5. Considere as rectas $x - y = 10$ e $x + 2y = 4$. Diga se são verdadeiras ou falsas as afirmações seguintes.

a) O ponto $(100, 90)$ é ponto de intersecção das rectas.
b) O ponto $(100, 90)$ pertence a uma das rectas e não é ponto de intersecção das duas rectas.
c) As rectas são paralelas.

6. Sendo i a unidade imaginária (isto é, $i^2 = -1$), diga se $(z_1, z_2) = (2-3i, 1-i)$ é ou não solução do sistema

$$\begin{cases} z_1 - z_2 = 1-2i \\ (1+i)z_2 = 2. \end{cases}$$

7. Considere uma função definida por $f(\mathbf{x}) = A\mathbf{x}$, que aplica o vector $\mathbf{u} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ no vector $\begin{bmatrix} 3 \\ 2 \end{bmatrix}$ e o vector $\mathbf{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ no vector $\begin{bmatrix} 5 \\ 6 \end{bmatrix}$.

a) Sem determinar a matriz A calcule
$$f(\mathbf{u} - 2\mathbf{v}).$$
b) Determine a matriz A e use-a para calcular $f\left(\begin{bmatrix} 3 \\ 2 \end{bmatrix}\right)$ e $f\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right)$.

8. Considere as seguintes matrizes aumentadas (em escada por linhas). Resolva os respectivos sistemas de equações lineares.

a) $\left[\begin{array}{ccc|c} 1 & -3 & 4 & 7 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 1 & 5 \end{array}\right]$ b) $\left[\begin{array}{cccc|c} 1 & 0 & 8 & -5 & 6 \\ 0 & 1 & 4 & -9 & 3 \\ 0 & 0 & 1 & 1 & 2 \end{array}\right]$

9. Quais das seguintes matrizes 3×3 são matrizes em escada por linhas? Indique as respectivas características.

$a)$ $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ $b)$ $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

$c)$ $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$ $d)$ $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$

$e)$ $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ $f)$ $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

$g)$ $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ $h)$ $\begin{bmatrix} 0 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

$i)$ $\begin{bmatrix} 2 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$ $j)$ $\begin{bmatrix} 2 & 1 & 0 \\ 0 & -1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$.

10. Determine a característica de cada uma das matrizes seguintes (i denota a unidade imaginária $\sqrt{-1}$):

a) $\begin{bmatrix} 1 & 1 & 3 \\ 1 & 1 & 4 \\ 1 & 1 & 5 \end{bmatrix}$.

b) $\begin{bmatrix} -1 & 1+2i \\ -3+i & 5+5i \end{bmatrix}$.

11. Considere as matrizes reais A e $\mathbf{b}$:

$$A = \begin{bmatrix} 1 & -1 & 1 \\ 4 & -4 & \alpha^2 \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} 0 \\ \alpha+2 \end{bmatrix}.$$

a) Determine a característica da matriz A e da matriz aumentada $[A \,|\, \mathbf{b}]$ em função do parâmetro $\alpha \in \mathbb{R}$.

b) Use os resultados da alínea anterior para determinar a natureza (em função de α) os sistemas $A\mathbf{x} = \mathbf{b}$, indicando em cada caso a solução geral.

12. Construa uma matriz aumentada para um sistema de equações lineares cuja solução geral seja:

a) $\{(3, 1, 5)\}$.

b) $\{(3, t, 5) :\ t \in \mathbb{R}\}$.

c) $\{(2y + z, y, z) :\ y, z \in \mathbb{R}\}$.

d) $\{(x, 2x, 4x) :\ x \in \mathbb{R}\}$.

13. Resolva cada um dos sistemas de equações lineares, utilizando o método de eliminação de Gauss.

a) $\begin{cases} x + y + 2z = 8 \\ -x - 2y + 3z = 1 \\ 3x - 7y + 4z = 10. \end{cases}$

b) $\begin{cases} x - y + 2z - w = -1 \\ 2x + y - 2z - 2w = -2 \\ -x + 2y - 4z + w = 1. \end{cases}$

14. Faça a discussão de cada um dos seguintes sistemas de equações lineares nas variáveis x, y, z em função dos respectivos parâmetros, e determine a solução geral em cada caso.

a) $\begin{cases} 2x - y = -3 \\ az = 0 \\ x - 5y - 5z = b. \end{cases}$

b) $\begin{cases} \alpha x + \beta z = 2 \\ \alpha x + \alpha y + 4z = 4 \\ \alpha y + 2z = \beta \end{cases}$

c) $\begin{cases} -2z = 0 \\ cy + 4z = d \\ 4x + 5y - 2z = -2. \end{cases}$

d) $\begin{cases} x + y + z = 4 \\ z = 2 \\ (a^2 - 4)z = a - 2. \end{cases}$

15. Escreva a matriz $A = [a_{ij}]_{i,j=1,\ldots,4}$ definida por:

a) $a_{ij} = \begin{cases} 1 & \text{se } i = j \\ -1 & \text{se } i = j + 1 \\ 0 & \text{caso contrário.} \end{cases}$

b) $a_{ij} = i^2$.

c) $a_{ij} = \begin{cases} -a_{ji} & \text{para todo } i, j \\ j & \text{para } j > i. \end{cases}$

16. Complete a igualdade:

$$\begin{bmatrix} 10 & 200 & 0.5 \\ 3 & 1 & 1 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 20 \\ -1 \\ \square \end{bmatrix} = \begin{bmatrix} 1 \\ \square \\ \square \end{bmatrix}.$$

17. Seja B uma matriz 3×3 cuja 3ª coluna é $\mathbf{b} = \begin{bmatrix} a \\ 2 \\ c \end{bmatrix}$ e $A = \begin{bmatrix} 5 & 2 & 3 \\ 1 & 0 & 1 \\ 3 & 2 & -1 \end{bmatrix}$. Complete as afirmações:

a) A 3ª coluna de AB é ....

b) $A\mathbf{b}$ é uma combinação linear das colunas de A com coeficientes.......................

18. Considere os vectores:

$$\mathbf{u}_1 = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix},$$

$$\mathbf{u}_3 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}.$$

a) Verifique se $\mathbf{b}$ é ou não combinação linear de $\mathbf{u}_1, \mathbf{u}_2$ e $\mathbf{u}_3$, e em caso afirmativo indique os coeficientes da combinação linear.

b) Seja A a matriz cujas colunas são os vectores $\mathbf{u}_1, \mathbf{u}_2$ e $\mathbf{b}$, por esta ordem. Usando a alínea anterior indique um vector $\mathbf{w} = \begin{bmatrix} \alpha \\ \beta \\ -1 \end{bmatrix}$ e um número real γ, tal que $A\mathbf{w} = \gamma \mathbf{u}_3$.

19. Complete a igualdade

$$\begin{bmatrix} -10 & 1 \\ \square & \square \end{bmatrix} \begin{bmatrix} \square \\ \square \end{bmatrix} = \begin{bmatrix} \square \\ 3 \end{bmatrix},$$

por forma a que resolver esta equação seja equivalente à resolução do problema:

- Determinar a equação da recta, $y = mx+b$, que passa pelos pontos $(-10, 1)$ e $(2, 3)$.

Indique ainda a solução do problema.

20. Considere os vectores $\mathbf{u} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$ e $\mathbf{v} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$. No gráfico 1.13 represente os vectores: $\mathbf{u}, \mathbf{v}, \mathbf{u} + 2\mathbf{v}, 2\mathbf{u} + \mathbf{v}$ e $\mathbf{u} - 2\mathbf{v}$.

21. Calcule, se possível, os produtos AB e BA.

a) $A = \begin{bmatrix} 1 & 4 & \sqrt{2} \\ -2 & 1 & \sqrt{3} \end{bmatrix}$ e $B = \begin{bmatrix} 1 & 2 & \pi \\ \sqrt{3} & -1 & 2 \\ 0 & 1 & -1 \end{bmatrix}$

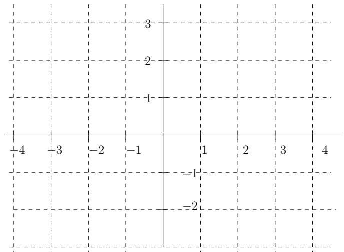


Figura 1.13: Gráfico do problema 20.

b) $A = \begin{bmatrix} 1+\mathrm{i} & -\mathrm{i} \end{bmatrix}$ e $B = \begin{bmatrix} 1-\mathrm{i} \\ 2-3\mathrm{i} \end{bmatrix}$ onde i designa a unidade imaginária ($i^2 = -1$).

22. Para cada par de matrizes A e B abaixo indicadas determinar, caso estejam definidas, as matrizes $A + 2B$, $A - B$, A^2, B^2, AB e BA.

a) $A = \begin{bmatrix} 1 & 0 & 1 \\ 2 & -1 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 3 & 1 \\ 4 & 1 \\ 1 & 0 \end{bmatrix}$.

b) $A = \begin{bmatrix} 1 & 0 & 1 \\ 2 & -1 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 3 & 4 & 1 \\ 1 & 1 & 0 \end{bmatrix}$.

c) $A = [2]$ e $B = \begin{bmatrix} 3 \\ 4 \\ 1 \end{bmatrix}$.

d) $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$.

e) $A = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$.

23. Diga se são verdadeiras ou falsas as afirmações:

a) Para $\mathbf{u} = (1, 2)$ e $\mathbf{v} = (-5, -10)$,

$\mathrm{Span}\{\mathbf{u},\mathbf{v}\}$ é a recta definida por $y = 2x$.

b) Para $\mathbf{u} = (1,2)$ e $\mathbf{v} = (3,-6)$, $\mathrm{Span}\{\mathbf{u},\mathbf{v}\}$ é a recta definida por $y = 2x$.

c) Sendo $\mathbf{u} = (1,0,1)$, $\mathbf{v} = (2,1,1)$ e $\mathbf{w} = (1,1,0)$, então $\mathrm{Span}\{\mathbf{u},\mathbf{v},\mathbf{w}\}$ é um plano.

d) Sendo $\mathbf{u} = (1,0,1)$, $\mathbf{v} = (2,1,1)$ e $\mathbf{w} = (1,1,0)$, então $\mathrm{Span}\{\mathbf{u},\mathbf{v},\mathbf{w}\} = \mathrm{Span}\{\mathbf{u},\mathbf{v}\}$.

24. Para $\mathbf{u}_1 = (1,1,1)$, $\mathbf{u}_2 = (1,-1,0)$, $\mathbf{u}_3 = (0,0,-1)$ e $\mathbf{u}_4 = (2,0,1)$, diga quais das afirmações seguintes são verdadeiras.

a) $\mathbf{u}_3$ é uma combinação linear de $\mathbf{u}_1$ e $\mathbf{u}_2$.

b) $\mathbf{u}_4$ é uma combinação linear de $\mathbf{u}_1$, $\mathbf{u}_2$ e $\mathbf{u}_3$.

c) $\mathrm{Span}\{\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3,\mathbf{u}_4\} = \mathrm{Span}\{\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_3\} = \mathbb{R}^3$.

d) $\mathrm{Span}\{\mathbf{u}_1,\mathbf{u}_2\} = \mathrm{Span}\{\mathbf{u}_1,\mathbf{u}_2,\mathbf{u}_4\}$.

25. Obtenha uma fórmula para A^n, onde A é a matriz:

$$a)\ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad b)\ \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

$$c)\ \begin{bmatrix} \mathrm{i} & 0 \\ 0 & -\mathrm{i} \end{bmatrix} \qquad d)\ \begin{bmatrix} \cos\alpha & -\operatorname{sen}\alpha \\ \operatorname{sen}\alpha & \cos\alpha \end{bmatrix}$$

26. Para $A = \begin{bmatrix} 1 & 2 & 0 \\ 3 & 2 & 1 \end{bmatrix}$ e $B = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 0 & 1 \end{bmatrix}$, calcule: AB, $(AB)^T$, B^TA^T e A^TB^T.

27. Determine a matriz simétrica A que satisfaz as expressões indicadas, onde

$$\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}.$$

a) $\mathbf{x}^TA\mathbf{x} = x^2 - 4xy + 3y^2$.

b) $\mathbf{x}^TA\mathbf{x} = 5x^2 + 5xy + 4y^2$.

28. Para cada uma das matrizes $A = [a_{ij}]_{i,j=1,2}$ diga se são simétricas ou anti-simétricas.

a) $a_{ij} = \begin{cases} 2-i-j & \text{se } i < j \\ i-j & \text{se } i \geq j \end{cases}$.

b) $a_{ij} = \begin{cases} j & \text{se } i \geq j \\ -i & \text{se } i < j \end{cases}$.

29. Indique o valor lógico das afirmações seguintes.

a) Uma matriz rectangular pode ser uma matriz simétrica.

b) Se uma matriz quadrada tem as entradas da diagonal principal todas nulas, então é anti-simétrica.

c) A matriz AA^T é simétrica somente se A é uma matriz quadrada.

d) Se A e B são matrizes simétricas da mesma ordem, então AB é uma matriz simétrica.

e) Qualquer matriz diagonal é uma matriz simétrica e anti-simétrica.

30. Sempre que possível, dê exemplos de matrizes quadradas A e B que verifiquem as condições abaixo indicadas. Caso não seja possível, justifique porquê.

a) $AB = BA \neq 0$.

b) $AB \neq BA$.

c) $AB = 0$ e $BA \neq 0$.

d) $(A+B)^2 = A^2 + B^2 + 2AB$.

e) $AB = 0$ e $A \neq 0$ e $B \neq 0$.

31. A matriz $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ é uma matriz elementar. Sendo B uma matriz 2×2, a matriz BA obtém-se de B:

a) trocando a 1ª linha com a 2ª linha .

b) trocando a 1ª coluna com a 2ª coluna .

c) substituindo a 2ª linha de B pela soma da 2º linha com a 1ª multiplicada por 2.

d) substituindo a 2ª coluna de B pela soma da 2ª coluna com a 1ª multiplicada por 2.

32. Mostre que a matriz

$$A = \begin{bmatrix} 0 & a & 0 & 0 & 0 \\ b & 0 & c & 0 & 0 \\ 0 & d & 0 & e & 0 \\ 0 & 0 & f & 0 & g \\ 0 & 0 & 0 & h & 0 \end{bmatrix}$$

não é invertível para quaisquer entradas a, b, c, d, e, f, g, h reais.

33. Considere a matriz:

$$A = \begin{bmatrix} 1 & 0 & 0 \\ -5 & 0 & 1 \\ 0 & -2 & 0 \end{bmatrix}.$$

a) Determine matrizes elementares E_i tais que $E_k \cdots E_1 A = I$.

b) Escreva A^{-1} como um produto de k matrizes elementares.

c) Escreva A como um produto de k matrizes elementares.

34. Calcule, sempre que existir, a matriz inversa de cada uma das matrizes:

a) $\begin{bmatrix} k_1 & 0 & 0 & 0 \\ 0 & k_2 & 0 & 0 \\ 0 & 0 & k_3 & 0 \\ 0 & 0 & 0 & k_4 \end{bmatrix}$.

b) $\begin{bmatrix} 0 & 0 & 0 & k_1 \\ 0 & 0 & k_2 & 0 \\ 0 & k_3 & 0 & 0 \\ k_4 & 0 & 0 & 0 \end{bmatrix}$.

em que k_1, k_2, k_3 e k_4 são reais.

35. Utilizando o método de Gauss-Jordan, calcule, sempre que existir, a matriz inversa de cada uma das seguintes matrizes.

a) $\begin{bmatrix} 3 & 4 & -1 \\ 1 & 0 & 3 \\ 2 & 5 & -4 \end{bmatrix}$ b) $\begin{bmatrix} -1 & 3 & -4 \\ 2 & 4 & 0 \\ -4 & 2 & -9 \end{bmatrix}$

c) $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}$ d) $\begin{bmatrix} 2 & 6 & 6 \\ 2 & 7 & 6 \\ 2 & 7 & 7 \end{bmatrix}$

e) $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 1 & 3 & 5 & 0 \\ 1 & 3 & 5 & 7 \end{bmatrix}$.

36. Sendo A e B matrizes invertíveis, diga quais das matrizes seguintes são invertíveis. Se essa matriz for sempre invertível indique uma expressão para a inversa, caso contrário dê um exemplo de matrizes A e B tais que a matriz resultante não é ínvertível.

$a)AB \qquad b)A+B \qquad c)A^{-1}B$

37. Sejam A e B matrizes quadradas tais que $AB = I$. Calcule a matriz $BA^2 - A$.

38. Considere o sistema não homogéneo $A\mathbf{x} = \mathbf{b}$, onde A é $n \times n$. Diga quais das afirmações seguintes são verdadeiras.

a) Se $\mathbf{b}$ é uma combinação linear das colunas de A, o sistema é possível.

b) Se o sistema é possível, a característica de A é n.

c) Se A é invertível, o sistema pode ser indeterminado.

d) Se A é invertível, a característica de A é menor do que n.

39. Considere o sistema homogéneo $A\mathbf{x} = \mathbf{0}$, onde A é $n \times p$. Diga quais das afirmações seguintes são verdadeiras.

a) A característica de A e a característica da matriz aumentada do sistema podem ser diferentes.
b) Se $n = p$, então o sistema nunca é indeterminado.
c) A solução nula é a única solução do sistema, quando $n = p$.
d) Se $n > p$, então a característica de A é maior do que p.
e) Se $n > p$ e a característica de A é igual a p, então o sistema é indeterminado.

40. Para A uma matriz quadrada e $\mathbf{b}$ um vector não nulo, diga qual das afirmações seguintes é verdadeira.

a) Se $\mathbf{x}$ é solução do sistema $A\mathbf{u} = \mathbf{0}$ e $\mathbf{y}$ solução do sistema $A\mathbf{u} = \mathbf{b}$, então $(\mathbf{x} - \mathbf{y})$ é solução de $A\mathbf{u} = \mathbf{b}$.
b) Se $\mathbf{x}$ é solução do sistema $A\mathbf{u} = \mathbf{0}$ e $\mathbf{y}$ solução do sistema $A\mathbf{u} = \mathbf{b}$, então $(\mathbf{x} - \mathbf{y})$ é solução de $A\mathbf{u} = -\mathbf{b}$.
c) Se $\mathbf{x}$ é uma solução de $A\mathbf{u} = \mathbf{0}$, então necessariamente $\mathbf{x} = \mathbf{0}$.
d) Se A não é invertível, então $\mathbf{x} = \mathbf{0}$ é a única solução de $A\mathbf{u} = \mathbf{0}$.

41. Sendo $A^{-1} = \begin{pmatrix} 1 & 0 & 3 \\ 4 & 5 & 0 \\ 0 & 1 & 0 \end{pmatrix}$ a matriz inversa de A, a solução do sistema $A\mathbf{x} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$ é:

a) $(3, 4, 0)$. b) $(4, 4, 0)$.
c) $(4, 9, 1)$. d) $(2, 9, 1)$.

42. Determine, se posssível, a factorização LU das matrizes:

a) $A = \begin{bmatrix} 2 & 1 & 2 \\ 1 & 4 & 1 \\ 0 & 3 & 1 \end{bmatrix}$

b) $A = \begin{bmatrix} 3 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 2 & 0 \end{bmatrix}$.

43. Indique o valor lógico das afirmações seguintes:

a) Uma matriz admite uma factorização LU se e só se é invertível.
b) Se A admite uma factorização LU, o sistema $A\mathbf{x} = \mathbf{b}$ pode ser indeterminado.
c) Se é satisfeita a igualdade
$$\begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} A = \begin{bmatrix} 2 & 1 \\ 0 & -1 \end{bmatrix},$$
a matriz A admite necessariamente uma factorização LU.
d) Se é satisfeita a igualdade
$$\begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} A = \begin{bmatrix} 2 & 1 \\ 0 & -1 \end{bmatrix},$$
a matriz A não admite uma factorização LU.

44. Admitindo as hipóteses necessárias sobre os tamanhos das matrizes por forma a que as igualdades seguintes se verifiquem, determine as matrizes X, Y e Z em termos de A, B e C. (A matriz I designa a identidade).

a) $\begin{bmatrix} A & B \\ C & 0 \end{bmatrix} \begin{bmatrix} I & 0 \\ X & Y \end{bmatrix} = \begin{bmatrix} 0 & I \\ Z & 0 \end{bmatrix}$.

b) $\begin{bmatrix} X & 0 & 0 \\ Y & 0 & I \end{bmatrix} \begin{bmatrix} A & Z \\ 0 & 0 \\ B & I \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$.

45. Seja $A = \begin{bmatrix} B & 0 \\ C & D \end{bmatrix}$ uma matriz triangular inferior por blocos. Supondo que A é invertível e que B e D são matrizes quadradas, respectivamente $p \times p$ e $q \times q$, determine A^{-1}.

46. Use o exercício anterior para calcular a inversa da matriz

$$A = \begin{bmatrix} 1 & 3 & 0 & 0 & 0 \\ 2 & 5 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 4 & 2 \\ 0 & 2 & 0 & 2 & 5 \end{bmatrix}.$$

Capítulo 2

Determinantes

O determinante é um número que se associa a uma matriz quadrada. Neste capítulo começa-se por definir determinante, deduzindo-se a seguir algumas das suas propriedades. Duas dessas propriedades merecem especial destaque. A primeira estabelece uma relação entre o determinante de uma matriz A e o determinante da matriz que resulta de A após aplicação do método de eliminação de Gauss. A outra propriedade diz-nos que é possível utilizar o determinante de uma matriz como teste de invertibilidade (uma matriz é invertível se e só se o seu determinante é diferente de zero).

Na Secção 2.3 é apresentado o chamado *desenvolvimento de Laplace* para o cálculo do determinante de uma matriz. Com base no desenvolvimento de Laplace deduz-se uma fórmula para o cálculo da inversa de uma matriz envolvendo determinantes, bem como a *regra de Cramer* para resolver sistemas possíveis e determinados.

A interpretação geométrica do conceito de determinante é diferida para o Capítulo 5, onde veremos que o módulo do determinante de uma matriz 2×2 (resp. 3×3) corresponde à área de um paralelogramo (resp. volume de um paralelipípedo) definido pelos vectores coluna da matriz.

2.1 Definição de determinante

Comecemos por apresentar alguns resultados preliminares sobre permutações, necessários para a compreensão da definição de determinante. O leitor interessado em aprofundar a teoria das permutações poderá consultar Cohn [4].

Uma *permutação* do conjunto X é uma função bijectiva[1] σ de X em si próprio. Uma permutação σ do conjunto $\{1, 2, \ldots, n\}$ pode ser representada por uma matriz com duas linhas, em que na primeira linha aparecem os números $1, 2, \ldots, n$ e na segunda linha as imagens $\sigma(1), \sigma(2), \ldots, \sigma(n)$. Isto é,

$$\sigma = \begin{pmatrix} 1 & 2 & \cdots & n \\ \sigma(1) & \sigma(2) & \ldots & \sigma(n) \end{pmatrix}.$$

Uma outra representação da permutação σ é considerando o n-uplo:

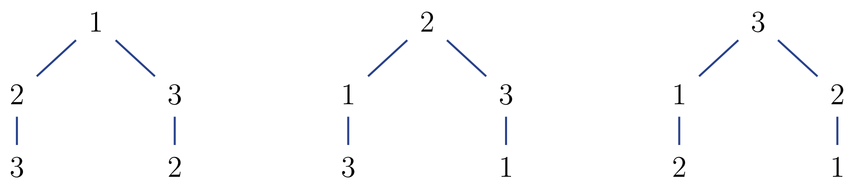
$$\sigma = (\sigma(1), \sigma(2), \ldots, \sigma(n)) = (\sigma_1, \sigma_2, \ldots, \sigma_n).$$

Passamos a adoptar esta notação para σ. Saliente-se que dada a bijectividade de σ, os números $1, 2, \ldots, n$ ocorrem sem repetições em $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n)$.

Exemplo 2.1. a) É óbvio que só existem 2 permutações de $\{1, 2\}$, nomeadamente $(1, 2)$ e $(2, 1)$. Logo, o conjunto Π de todas as permutações de $\{1, 2\}$ é

$$\Pi = \{(1, 2), (2, 1)\}.$$

b) Para determinar o conjunto das permutações de $\{1, 2, 3\}$, considere-se o esquema seguinte correspondente às possíveis escolhas de $1, 2$ e 3 para a primeira, segunda e terceira posições de uma permutação de $\{1, 2, 3\}$.



$(1, 2, 3)$ $(1, 3, 2)$ $(2, 1, 3)$ $(2, 3, 1)$ $(3, 1, 2)$ $(3, 2, 1)$

É claro que não existem outras permutações de $\{1, 2, 3\}$ para além das indicadas. O conjunto de todas as permutações de $\{1, 2, 3\}$ é

$$\Pi = \{(1, 2, 3), (1, 3, 2), (2, 1, 3), (2, 3, 1), (3, 1, 2), (3, 2, 1)\}.$$

♦

[1] A função $\sigma : X \to X$ é bijectiva se aplica elementos distintos de X em elementos distintos, e todo o elemento de X é imagem por σ de algum elemento de X (se necessitar, consulte o Capítulo 6).

O cardinal do conjunto de todas as permutações de $\{1, 2, \ldots, n\}$ é facilmente calculado, uma vez que temos n escolhas possíveis para a primeira componente do n-uplo, pelo que a seguir apenas dispomos de $(n-1)$ escolhas para a segunda componente, $(n-2)$ para a terceira , etc. Por conseguinte, o cardinal do conjunto Π de todas as permutações de $\{1, 2, \ldots, n\}$ é $n!$ (n factorial):

$$n(n-1)(n-2)\cdots 2 \cdot 1 = n!.$$

Dada uma permutação $(j_1, j_2, \ldots, j_n)$ de $\{1, 2, \ldots, n\}$ dizemos que ocorreu uma *inversão* sempre que um inteiro na permutação é seguido de um inteiro menor. A permutação $(1, 2, \ldots, n)$ tem zero inversões e é designada por *permutação identidade*.

Definimos o *número total de inversões* em $(j_1, j_2, \ldots, j_n)$ como sendo a soma dos números obtidos da seguinte forma: determinar o número de inteiros menores que j_1 que lhe sucedem na permutação; determinar o número de inteiros menores que j_2 que lhe sucedem na permutação; continuar esta contagem para $j_3, \ldots, j_{n-1}$.

Exemplo 2.2. Determinar o número total de inversões das permutações seguintes.

$$(a)\ (3, 4, 1, 2, 6, 5, 7) \qquad (b)\ (1, 3, 2, 4) \qquad (c)\ (1, 4, 3, 9, 8, 2, 5, 7, 6).$$

(a) O número total de inversões é: $2 + 2 + 0 + 0 + 1 + 0 = 5$.
(b) O número total de inversões é: $0 + 1 + 0 = 1$.
(c) O número total de inversões é: $0 + 2 + 1 + 5 + 4 + 0 + 0 + 1 = 13$.

◆

A composição de permutações de $\{1, 2, \ldots, n\}$, bem como a inversa de uma permutação, são ainda permutações de $\{1, 2, \ldots, n\}$. É habitual designar-se a composição de permutações por *produto* de permutações. Por exemplo se $\sigma = (2, 4, 1, 3)$ e $\pi = (1, 4, 3, 2)$ são permutações de $\{1, 2, 3, 4\}$, a composição $\sigma\pi$ e a inversa σ^{-1} de σ são, respectivamente, $\sigma\pi = (2, 3, 1, 4)$ e $\sigma^{-1} = (3, 1, 4, 2)$.

Chama-se *transposição* a uma permutação que se obtém trocando a posição de dois números na permutação identidade $(1, 2, \ldots, n)$, deixando os restantes fixos. Por exemplo, a permutação $(3, 2, 1, 4)$ é uma transposição. Note-se que o número total de inversões de uma transposição é um número ímpar.

Mostra-se que qualquer permutação $\sigma = (j_1, j_2, \ldots, j_n)$ é igual ao produto de um número finito de transposições. Embora este número de transposições não seja único, a sua paridade é única (ver [4]). Assim, uma permutação σ diz-se par (resp. ímpar) se for possível escrever σ como um produto de um número par (resp. ímpar) de transposições. Uma permutação par diz-se que tem sinal $+1$ e uma permutação ímpar que tem sinal -1. Podemos desde já enunciar que

Toda a permutação muda de sinal quando se trocam entre si duas quaisquer das suas componentes.

O número total de inversões de uma permutação σ corresponde a um possível número de transposições sucessivas que permitem obter σ a partir da permutação identidade. Por conseguinte, adoptamos a seguinte definição de paridade.

Definição 2.1. Uma permutação diz-se *par* se o respectivo número total de inversões é um número par, e *ímpar* se esse número é ímpar.

O *sinal* de uma permutação é $+1$ se a permutação é par e -1 se é ímpar. Designamos por $\operatorname{sign}\sigma$ o sinal da permutação σ.

Definição de determinante.

Comecemos por definir o conceito de produto elementar de entradas de uma matriz quadrada. Um *produto elementar de entradas* de uma matriz A do tipo $n \times n$, é um produto de n entradas da matriz no qual não existem dois factores provenientes da mesma linha ou da mesma coluna da matriz.

Um produto elementar da matriz $A = [a_{ij}]_{i,j=1,\ldots,n}$ pode escrever-se na forma:

$$a_{1j_1}a_{2j_2}\cdots a_{nj_n}, \tag{2.1}$$

onde $(j_1, j_2, \ldots, j_n)$ é uma permutação de $\{1, 2, \ldots, n\}$. O facto de $(j_1, j_2, \ldots, j_n)$ ser uma permutação corresponde precisamente ao facto dos factores no produto (2.1) pertencerem a colunas distintas. De igual modo, podemos escrever um produto elementar de entradas de A na forma

$$a_{i_1 1}a_{i_2 2}\cdots a_{i_n n}, \tag{2.2}$$

onde $(i_1, i_2, \ldots, i_n)$ é uma permutação de $\{1, 2, \ldots, n\}$.

Exemplo 2.3. Determinar todos os produtos elementares das matrizes

$$(a)\quad \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \qquad\qquad (b)\quad \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

Para a matriz da alínea (a) só existem dois produtos elementares, nomeadamente $a_{11}a_{22}$ e $a_{12}a_{21}$. Estes dois produtos elementares são da forma indicada em (2.1), já que usando o conjunto das permutações de $\{1, 2\}$, determinado no Exemplo 2.1-(a), temos

Permutação	Produto elementar de A
$(1,2)$	$a_{11}a_{22}$
$(2,1)$	$a_{12}a_{21}$

Para a matriz da alínea (b) o número de produtos elementares é igual ao número de permutações de $\{1,2,3\}$. Assim, usando a expressão (2.1) e o conjunto Π de todas as permutações de $\{1,2,3\}$, determinado no Exemplo 2.1-(b), tem-se

Π	Produto elementar de A
$(1,2,3)$	$a_{11}a_{22}a_{33}$
$(1,3,2)$	$a_{11}a_{23}a_{32}$
$(2,1,3)$	$a_{12}a_{21}a_{33}$
$(2,3,1)$	$a_{12}a_{23}a_{31}$
$(3,1,2)$	$a_{13}a_{21}a_{32}$
$(3,2,1)$	$a_{13}a_{22}a_{31}$

♦

Definimos *sinal de um produto elementar* como sendo o sinal da permutação que lhe está associada. Ou seja, o sinal do produto elementar $a_{1j_1}a_{2j_2}\cdots a_{nj_n}$ é o sinal de $(j_1, j_2, \ldots, j_n)$, enquanto que o sinal de $a_{i_1 1}a_{i_2 2}\cdots a_{i_n n}$ é o de $(i_1, i_2, \ldots, i_n)$.

Definição 2.2. O determinante de uma matriz A do tipo $n \times n$, é igual à soma de todos os seus produtos elementares multiplicados pelo sinal respectivo. Ou seja, para $A = [a_{ij}]$ o determinante de A é

$$\det(A) = \sum_{\substack{\sigma \in \Pi \\ \sigma = (j_1, j_2, \ldots, j_n)}} \operatorname{sign}(\sigma)\, a_{1j_1}a_{2j_2}\cdots a_{nj_n}, \tag{2.3}$$

onde Π designa o conjunto das permutações de $\{1,2,\ldots,n\}$.

O determinante de uma matriz A é designado por $\det(A)$ ou por $|A|$.

Os cálculos efectuados nos exemplos anteriores permitem obter o determinante de uma matriz 2×2 e 3×3.

Usando a definição de determinante, para a matriz $A = [a_{ij}]_{i,j=1,2}$ temos

Π	Paridade	Produto elementar com sinal
$(1,2)$	par	$a_{11}a_{22}$
$(2,1)$	ímpar	$-a_{12}a_{21}$

Logo,

$$\det(A) = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}. \tag{2.4}$$

Da mesma forma, para a matriz $A = [a_{ij}]_{i,j=1,2,3}$ tem-se

Π	Paridade	Produto elementar com sinal
$(1,2,3)$	par	$a_{11}a_{22}a_{33}$
$(1,3,2)$	ímpar	$-a_{11}a_{23}a_{32}$
$(2,1,3)$	ímpar	$-a_{12}a_{21}a_{33}$
$(2,3,1)$	par	$a_{12}a_{23}a_{31}$
$(3,1,2)$	par	$a_{13}a_{21}a_{32}$
$(3,2,1)$	ímpar	$-a_{13}a_{22}a_{31}$

Assim,

$$\begin{aligned} \det(A) = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} =& a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - \\ & - a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33}. \end{aligned} \tag{2.5}$$

Podemos usar uma mnemónica[2] para a expressão do determinante de uma matriz 3×3: os produtos elementares positivos são os produtos das entradas da diagonal principal e das entradas situadas nos vértices de dois triângulos com um lado paralelo à diagonal principal (a vermelho na Figura 2.1), enquanto que os de sinal negativo são os produtos das entradas da diagonal oposta à diagonal principal e das entradas situadas nos vértices de dois triângulos com um lado paralelo a essa diagonal (a azul na Figura 2.1).

Note-se que, quando $\sigma = (p_1, p_2, \dots, p_n)$ varia no conjunto Π de todas as permutações de $\{1, 2, \dots, n\}$, o conjunto de todos os produtos elementares da forma $\operatorname{sign}(\sigma)\, a_{1p_1}a_{2p_2} \cdots a_{np_n}$ é igual ao conjunto de todos os produtos da forma $\operatorname{sign}(\sigma) a_{p_1 1}a_{p_2 2} \cdots a_{p_n n}$. Por conseguinte, o determinante da matriz $A = [a_{ij}]_{i,j=1,\dots,n}$ pode ser igualmente escrito na forma

$$\det(A) = \sum_{\substack{\sigma \in \Pi \\ \sigma = (i_1, i_2, \dots, i_n)}} \operatorname{sign}(\sigma)\, a_{i_1 1}a_{i_2 2} \cdots a_{i_n n}. \tag{2.6}$$

[2]Esta mnemónica é conhecida pela designação de regra de Sarrus (Pierre Frédéric Sarrus, 1798-1861, matemático francês).

Figura 2.1: Para uma matriz 3×3: a vermelho os produtos elementares com sinal positivo e a azul os produtos elementares com sinal negativo.

Exercício 2.1. Repita o cálculo do determinante de uma matriz A do tipo 3×3, usando agora a representação (2.2) para os produto elementares. Verifique ainda que quando σ varia no conjunto das permutações de $\{1, 2, 3\}$ os produtos elementares com sinal que obtém é igual ao conjunto anteriormente obtido.

2.2 Propriedades do determinante

A partir da definição de determinante vamos deduzir algumas propriedades. Como o determinante de uma matriz $n \times n$ é a soma de todos os produtos elementares com o respectivo sinal, e um produto elementar é um produto de n entradas da matriz extraídas de linhas e colunas distintas, podemos concluir:

- Se uma matriz tem uma linha de zeros o seu determinante é igual a zero, já que os produtos elementares são todos nulos. De facto, cada produto elementar tem um factor pertencente à linha nula.

- Como os factores de um produto elementar correspondem a uma escolha de n entradas da matriz de tal forma que não haja duas da mesma linha nem da mesma coluna, numa matriz triangular apenas um dos produtos elementares poderá ser não nulo. Este é o produto das entradas da diagonal principal. Logo, o determinante de uma matriz triangular é igual ao produto das entradas da diagonal principal.

- Se a matriz B é obtida da matriz A multiplicando uma linha de A por uma constante k, por exemplo a linha i, então

$$\det B = \sum_{\substack{\sigma \in \Pi \\ \sigma = (j_1, j_2, \ldots, j_n)}} \operatorname{sign}(\sigma)\, a_{1j_1} a_{2j_2} \cdots (k a_{ij_i}) \cdots a_{nj_n}$$

$$= k \sum_{\substack{\sigma \in \Pi \\ \sigma = (j_1, j_2, \ldots, j_n)}} \operatorname{sign}(\sigma)\, a_{1j_1} a_{2j_2} \cdots a_{ij_i} \cdots a_{nj_n} = k \det A.$$

Se A é uma matriz de ordem n, então $\det(kA) = k^n \det(A)$.

- Como uma permutação muda de sinal se trocarmos a posição de duas das suas componentes, o determinante de uma matriz muda de sinal se trocarmos duas colunas da matriz. De facto, se a matriz B é obtida de A por troca da coluna p com a coluna r tem-se

$$\det B = \sum_{\substack{\sigma \in \Pi \\ \sigma = (j_1, \ldots, j_n)}} \operatorname{sign}(\sigma)\, a_{1j_1} \cdots a_{pj_p} \cdots a_{rj_r} \cdots a_{nj_n}$$

$$= \sum_{\substack{\sigma \in \Pi \\ \sigma = (j_1, \ldots, j_n)}} -\operatorname{sign}(\sigma)\ a_{1j_1} \cdots a_{rj_r} \cdots a_{pj_p} \cdots a_{nj_n} = -\det A.$$

- Uma matriz A e a sua transposta, A^T, possuem os mesmos produtos elementares. Além disso, se $A^T = [a'_{ij}]$ e $A = [a_{ij}]$ tem-se $a'_{1j_1} a'_{2j_2} \cdots a'_{nj_n} = a_{j_1 1} a_{j_2 2} \cdots a_{j_n n}$. Por conseguinte, os sinais dos produtos elementares de A^T são iguais aos sinais dos produtos elementares de A. Assim, usando as expressões (2.3) e (2.6), tem-se $\det(A) = \det(A^T)$.

Uma vez que $\det(A) = \det(A^T)$, todas as propriedades válidas para as linhas de uma matriz são igualmente válidas substituindo a palavra linha por coluna. Resumimos no quadro que se segue as propriedades do determinante já deduzidas.

Propriedades do determinante

Seja A uma matriz $n \times n$.

P1. $\det(A) = \det(A^T)$.

P2. Se uma matriz tem uma linha (ou coluna) nula, então o seu determinante é igual a zero.

P3. O determinante de uma matriz triangular é igual ao produto das entradas da diagonal principal. Em particular, o determinante da matriz identidade é igual a 1.

P4. Trocando duas linhas (ou colunas) de uma matriz, muda o sinal do determinante.

P5. Se uma linha (ou coluna) de uma matriz for multiplicada por uma constante k, o determinante vem multiplicado por k. Em particular,

$$\det(kA) = k^n \det(A). \tag{2.7}$$

Exercício 2.2. Mostrar que $\det(-I_2) = 1$ e $\det(-I_3) = -1$, onde I_2 e I_3 são respectivamente as matrizes identidade de ordem 2 e 3. ▲

As propriedades P4 e P5 dizem respeito ao comportamento do determinante de uma matriz relativamente a duas das três operações elementares sobre as linhas da matriz. Vamos deduzir novas propriedades com o objectivo de determinar o comportamento do determinante relativamente à operação elementar em falta.

P6. Se uma matriz tem duas linhas (ou colunas) iguais, então o seu determinante é igual a zero.

P7. Se A, B e C são matrizes $n \times n$ que diferem apenas na linha (resp. coluna) número i, e a linha (resp. coluna) i de C é igual à soma da linha (resp. coluna) i de A com a linha (resp. coluna) i de B, então $\det(C) = \det(A) + \det(B)$. Isto é,

$$\begin{vmatrix} a_{11} & \cdots & a_{nn} \\ \vdots & \cdots & \vdots \\ a_{i1} + a'_{i1} & \cdots & a_{in} + a'_{in} \\ \vdots & \cdots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix} = \begin{vmatrix} a_{11} & \cdots & a_{nn} \\ \vdots & \cdots & \vdots \\ a_{i1} & \cdots & a_{in} \\ \vdots & \cdots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix} + \begin{vmatrix} a_{11} & \cdots & a_{nn} \\ \vdots & \cdots & \vdots \\ a'_{i1} & \cdots & a'_{in} \\ \vdots & \cdots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix}.$$

P8. Se B é uma matriz obtida de A substituindo uma linha (resp. coluna) de A pela sua soma com outra linha (resp. coluna) multiplicada por uma constante, então $\det(B) = \det(A)$.

Demonstração. P6: Suponha-se que a matriz A tem a linha i igual à linha r. Se trocarmos a linha i de A com a linha r obtemos de novo a matriz A. Assim, usando a propriedade P4, temos

$$\det(A) = -\det(A) \iff \det(A) = 0.$$

P7: Seja

$$C = \begin{bmatrix} a_{11} & \cdots & a_{nn} \\ \vdots & \cdots & \vdots \\ a_{i1} + a'_{i1} & \cdots & a_{in} + a'_{in} \\ \vdots & \cdots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}.$$

Por definição de determinante, tem-se

$$
\begin{aligned}
\det(C) &= \sum_{\substack{\sigma \in \Pi \\ \sigma = (j_1, j_2, \ldots, j_n)}} \operatorname{sign}(\sigma)\, a_{1j_1} a_{2j_2} \cdots (a_{ij_i} + a'_{ij_i}) \cdots a_{nj_n} \\
&= \sum_{\substack{\sigma \in \Pi \\ \sigma = (j_1, j_2, \ldots, j_n)}} \operatorname{sign}(\sigma)\, a_{1j_1} a_{2j_2} \cdots a_{ij_i} \cdots a_{nj_n} \\
&+ \sum_{\substack{\sigma \in \Pi \\ \sigma = (j_1, j_2, \ldots, j_n)}} \operatorname{sign}(\sigma)\, a_{1j_1} a_{2j_2} \cdots a'_{ij_i} \cdots a_{nj_n} = \det(A) + \det(B).
\end{aligned}
$$

P8: Considere-se a matriz $A = [a_{ij}]_{i,j=1,\ldots,n}$ e a matriz B que se obtém de A substituindo a linha i pela sua soma com a linha r multiplicada por k. Então,

$$
\det(B) = \begin{vmatrix}
a_{11} & a_{12} & \cdots\cdots & a_{1n} \\
\vdots & \vdots & & \vdots \\
a_{r1} & a_{r2} & \cdots\cdots & a_{rn} \\
\vdots & \vdots & & \vdots \\
a_{i1} + ka_{r1} & a_{i2} + ka_{r2} & \cdots\cdots & a_{in} + ka_{rn} \\
\vdots & \vdots & & \vdots \\
a_{n1} & a_{n2} & \cdots\cdots & a_{nn}
\end{vmatrix}
$$

$$
\overset{P7,P5}{=} \begin{vmatrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
\vdots & \vdots & & \vdots \\
a_{r1} & a_{r2} & \cdots & a_{rn} \\
\vdots & \vdots & & \vdots \\
a_{i1} & a_{i2} & \cdots & a_{in} \\
\vdots & \vdots & & \vdots \\
a_{n1} & a_{n2} & \cdots & a_{nn}
\end{vmatrix} + k \begin{vmatrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
\vdots & \vdots & & \vdots \\
a_{r1} & a_{r2} & \cdots & a_{rn} \\
\vdots & \vdots & & \vdots \\
a_{r1} & a_{r2} & \cdots & a_{rn} \\
\vdots & \vdots & & \vdots \\
a_{n1} & a_{n2} & \cdots & a_{nn}
\end{vmatrix}
$$

$$
\overset{P6}{=} \det(A) + 0.
$$

□

As propriedades P5 e P7 traduzem o que se designa por *propriedade de linearidade* do determinante em cada linha (ou coluna) da matriz. De facto, se encararmos o determinante de uma matriz A como uma função que ao conjunto das linhas (resp. das colunas) de A faz corresponder o valor do determinante de A, as propriedades P5 e P7 abreviam-se dizendo que esta função é linear[3] em cada linha (resp. coluna) quando se conservam as outras linhas (resp. colunas) fixas. O estudo de funções lineares é efectuado no Capítulo 6.

Exemplo 2.4. Usando apenas as propriedades do determinante até agora enunciadas, vamos calcular o determinante de uma matriz 2×2. Pela propriedade P7 podemos escrever o determinante de uma matriz 2×2 como a soma de 4 determinantes.

$$\begin{aligned}
\begin{vmatrix} a & b \\ c & d \end{vmatrix} &= \begin{vmatrix} a & 0 \\ c & d \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & d \end{vmatrix} && \text{por P7} \\
&= \begin{vmatrix} a & 0 \\ c & 0 \end{vmatrix} + \begin{vmatrix} a & 0 \\ 0 & d \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & 0 \end{vmatrix} + \begin{vmatrix} 0 & b \\ 0 & d \end{vmatrix} && \text{por P7} \\
&= 0 + ad - bc + 0 = ad - bc && \text{por P2, P3 e P4.}
\end{aligned}$$

Na antepenúltima igualdade há dois determinantes que valem zero pois as matrizes têm uma coluna nula; o segundo determinante é igual ao produto das entradas da diagonal principal (determinante de uma matriz diagonal); no terceiro determinante se trocarmos duas linhas da matriz obtemos uma matriz diagonal e portanto o determinante da matriz inicial é igual a $(-bc)$.

O resultado obtido é obviamente igual à expressão (2.4). ♦

Nota 14. *A propriedade P7 não diz que o determinante da soma de duas quaisquer matrizes seja igual à soma dos determinantes. Deixamos como exercício encontrar duas matrizes quadradas A e B que verifiquem*

$$\det(A + B) \neq \det(A) + \det(B).$$

Exercício 2.3. Mostrar que se uma matriz tem duas linhas proporcionais, então o seu determinante é zero.

▲

As propriedades P4, P5 e P8 referem-se ao comportamento do determinante de uma matriz relativamente a operações elementares sobre as suas linhas. Estas

[3] Uma função de várias variáveis que seja linear em cada uma das variáveis diz-se uma *função multilinear*.

propriedades podem resumir-se do seguinte modo:

$$
\begin{aligned}
A \xrightarrow{L_i \leftrightarrow L_j} B &\implies \det(B) = -\det(A) \\
A \xrightarrow{\alpha L_i} B &\implies \det(B) = \alpha \det(A) \\
A \xrightarrow{L_i + \alpha L_j} B &\implies \det(B) = \det(A),
\end{aligned} \tag{2.8}
$$

onde L_i e L_j são linhas (distintas) da matriz A.

Tendo em conta estas propriedades, é imediato o cálculo do determinante de matrizes elementares.

E obtida de $\mathbf{I_n}$ por	**Propriedade**	$\det(\mathbf{E})$
Troca de duas linhas	P4	$\det(E) = -1$
Multiplicação de uma linha por $\alpha \neq 0$	P5	$\det(E) = \alpha$
Substituição de uma linha pela sua soma com outra linha multiplicada pelo escalar α	P8	$\det(E) = 1$

Relembremos que se E é uma matriz elementar, a matriz EA é a matriz que se obtém de A efectuando sobre A a operação elementar que permitiu obter E da identidade (conforme Proposição 1.8). Assim, tendo em conta as relações (2.8) e o valor do determinante de uma matriz elementar E podemos enunciar a proposição que se segue.

Proposição 2.1. Se E é uma matriz elementar da mesma ordem de A, então

$$\det(EA) = \det(E)\det(A).$$

Determinante e método de eliminação de Gauss

As propriedades P4, P5 e P8 permitem relacionar o determinante de uma matriz A com o determinante de uma matriz B obtida de A usando o método de eliminação de Gauss. Em particular, se na aplicação do método de eliminação de Gauss não for efectuada a operação "multiplicação de uma linha por $\alpha \neq 0$", então $\det(B) = \pm\det(A)$. Neste caso, o sinal -1 ocorre sempre que o número de troca de linhas efectuado for ímpar, e o sinal $+1$ se for par. Caso não haja troca de linhas, nem multiplicação de linhas por um escalar, então $\det(A) = \det(B)$. Em resumo: o determinante de uma matriz B obtida de A por aplicação do método de eliminação de Gauss é um múltiplo (não nulo) do determinante de A.

No exemplo que apresentamos a seguir aplicamos o método de eliminação de Gauss-Jordan para obter a inversa de uma matriz e usamos as propriedades P4, P5 e P8 para relacionar o determinante da matriz com o da sua inversa.

Exemplo 2.5. Relacionemos o determinante da matriz (invertível)

$$A = \begin{bmatrix} 0 & 0 & 3 \\ 1 & 1 & 0 \\ 1 & 2 & 0 \end{bmatrix}$$

com o da sua inversa.

Apliquemos o método de eliminação de Gauss-Jordan para obter A^{-1}.

$$[A|I] = \left[\begin{array}{ccc|ccc} 0 & 0 & 3 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 2 & 0 & 0 & 0 & 1 \end{array}\right] \overset{L_1 \leftrightarrow L_3}{\longrightarrow} [B|J] = \left[\begin{array}{ccc|ccc} 1 & 2 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 3 & 1 & 0 & 0 \end{array}\right]$$

$$\overset{L_2 - L_1}{\longrightarrow} [C|K] = \left[\begin{array}{ccc|ccc} 1 & 2 & 0 & 0 & 0 & 1 \\ 0 & -1 & 0 & 0 & 1 & -1 \\ 0 & 0 & 3 & 1 & 0 & 0 \end{array}\right] \overset{-L_2}{\longrightarrow} [D|L] = \left[\begin{array}{ccc|ccc} 1 & 2 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & -1 & 1 \\ 0 & 0 & 3 & 1 & 0 & 0 \end{array}\right]$$

$$\overset{\frac{1}{3}L_3}{\longrightarrow} [E|M] = \left[\begin{array}{ccc|ccc} 1 & 2 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & \frac{1}{3} & 0 & 0 \end{array}\right] \overset{L_1 - 2L_2}{\longrightarrow} [I|A^{-1}] = \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 2 & -1 \\ 0 & 1 & 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & \frac{1}{3} & 0 & 0 \end{array}\right].$$

Calculando $\det(A)$ e $\det(A^{-1})$ mediante a expressão (2.5) obtemos $\det(A^{-1}) = \frac{1}{3}$ e $\det(A) = 3$. Confirmemos este resultado usando as propriedades do determinante.

- $\det J = -\det I = -1$ (por P4).
- $\det K = \det J = -1$ (por P8).
- $\det L = -\det K = 1$ e $\det M = \frac{1}{3}\det L = \frac{1}{3}$ (por P5).
- $\det(A^{-1}) = \det M = \frac{1}{3}$ (por P8).

♦

Como veremos seguidamente o determinante de uma matriz proporciona um teste de invertibilidade da matriz.

P9. Uma matriz A é invertível se e só se $\det(A) \neq 0$.

Demonstração. Se a matriz quadrada A é invertível, a aplicação do método de eliminação de Gauss-Jordan reduz A à matriz identidade. Concluímos assim que o determinante da matriz A é um múltiplo não nulo de 1 e portanto diferente de zero.

Para mostrar a implicação recíproca, suponha-se que $\det(A) \neq 0$. Seja U a matriz triangular superior que se obtém de A por aplicação do método de eliminação de Gauss. O determinante de U é um múltiplo não nulo de $\det(A)$, portanto $\det(U) \neq 0$. Como a matriz em escada U é triangular superior, o seu determinante é igual ao produto das entradas da diagonal principal, portanto a matriz U não tem linhas nulas. Logo, $\text{car}(A) = \text{car}(U) = n$. Sendo $\text{car}(A) = n$, a Proposição 1.12 (pág. 62) garante que A é invertível.

□

P10. $\det(AB) = \det(A)\det(B)$.

Demonstração. Recorde-se que o produto AB é invertível se e só se as matrizes A e B são invertíveis (Teorema 1.5, pág. 61). Logo, se uma das matrizes, por exemplo A, não é invertível o produto AB não é invertível e $\det(AB)$=0 e $\det(A) = 0$, pelo que o resultado fica provado para o caso de não invertibilidade de uma das matrizes.

Suponha-se agora que A e B são ambas invertíveis. Pelo Teorema 1.4-(iv), página 57, uma matriz invertível pode exprimir-se como um produto de matrizes elementares. Seja $A = E_1 \cdots E_r$ com E_i matrizes elementares. Pela Proposição 2.1 a propriedade do enunciado é válida quando um dos factores é uma matriz elementar. Logo,

$$\begin{aligned}
\det(AB) &= \det(E_1 \cdots E_r B) = \det(E_1)\det(E_2 \cdots E_r B) \\
&= \det(E_1)\det(E_2)\det(E_3 \cdots E_r B) = \cdots = \det(E_1 \cdots E_r)\det(B) \\
&= \det(A)\det(B).
\end{aligned}$$

□

Finalizamos esta secção enunciando duas proposições que nos serão úteis mais tarde.

Proposição 2.2. Seja A é uma matriz quadrada. O sistema homogéneo $A\mathbf{x} = \mathbf{0}$ possui soluções não nulas se e só se $\det(A) = 0$.

Demonstração. A Proposição 1.10 (pág. 59) garante que $\mathbf{x} = \mathbf{0}$ é solução única do sistema $A\mathbf{x} = \mathbf{0}$ se e só se A é invertível. Como $A\mathbf{x} = \mathbf{0}$ tem sempre a solução nula, conclui-se da propriedade P9 que existem soluções não nulas se e só se $\det(A) = 0$. □

Proposição 2.3. Se A é invertível, então $\det\left(A^{-1}\right) = \dfrac{1}{\det A}$.

Demonstração.

$$1 \overset{P3}{=} \det(I) = \det(A^{-1}A) \overset{P10}{=} \det(A^{-1})\det A.$$

A igualdade anterior é equivalente a $\det\left(A^{-1}\right) = \dfrac{1}{\det A}$, já que $\det(A) \neq 0$. □

2.3 Desenvolvimento de Laplace

O desenvolvimento de Laplace[4] permite-nos calcular de forma recursiva o determinante de uma qualquer matriz. Baseados neste desenvolvimento deduzimos duas fórmulas, uma para o cálculo da inversa e outra, conhecida por regra de Cramer, para o cálculo de soluções de certos sistemas de equações lineares.

Uma vez que a dedução da fórmula de Laplace a partir da definição de determinante requer algumas manipulações algébricas cujas expressões podem parecer complicadas, optámos por apresentar um exemplo a partir do qual é fácil deduzir o desenvolvimento de Laplace no caso geral.

Exemplo 2.6. Relembremos a expressão do determinante de uma matriz 3×3 obtida em (2.5).

$$\begin{aligned}\det(A) = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} =& a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - \\ & - a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33}.\end{aligned}$$

Pondo em evidência nesta expressão as entradas da primeira linha (por exemplo) da matriz A, obtemos

$$\det(A) \quad = a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}).$$

[4]Pierre-Simon (1749 -1827), marquês de Laplace, matemático, astrónomo e físico francês.

Nesta expressão os valores entre parênteses são iguais ao determinante de uma matriz 2×2. Assim,

$$\begin{aligned} \det(A) \quad &= a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ &= (-1)^{1+1}a_{11}M_{11} + (-1)^{1+2}a_{12}M_{12} + (-1)^{1+3}a_{13}M_{13}. \end{aligned} \tag{2.9}$$

Da expressão anterior, conclui-se que o determinante da matriz é a soma dos produtos das entradas da primeira linha por um determinante de uma matriz de ordem inferior. Os determinantes M_{ij} nessa expressão são os determinantes das matrizes que se obtêm da matriz dada eliminando a linha i e a coluna j. Por exemplo, M_{12} é obtido eliminando a linha 1 e a coluna 2 de A.

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \rightsquigarrow M_{12} = \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix}.$$

Se tivéssemos escolhido pôr em evidência as entradas de uma outra linha (ou mesmo de uma coluna) resultaria uma expressão do mesmo tipo para o determinante da matriz. De facto, o que obtivemos em (2.9) é conhecido como o desenvolvimento de Laplace segundo a primeira linha da matriz A.

♦

Exercício 2.4. Usar o mesmo procedimento do último exemplo para obter uma expressão para $\det(A)$ análoga a (2.9) onde figurem agora as entradas da segunda coluna.

▲

Como verificamos na definição seguinte, os determinantes M_{ij} do Exemplo 2.6 recebem uma designação particular.

Definição 2.3. Se A é uma matriz quadrada, o *menor-ij*, ou *menor da entrada* a_{ij}, é o determinante da matriz que se obtém de A suprimindo a linha i e a coluna j. Este menor denota-se por M_{ij}.

Ao número $C_{ij} = (-1)^{i+j}M_{ij}$ chamamos *cofactor-ij*, ou *cofactor da entrada* a_{ij}.

Em certos textos é usada a designação de *complemento algébrico* para cofactor.

A expressão (2.9) do determinante de A pode reescrever-se na forma

$$\det(A) = a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13},$$

indicando que este determinante é a soma dos produtos das entradas da primeira linha de A pelos respectivos cofactores.

A fórmula de Laplace que enunciamos a seguir permite calcular o determinante de uma matriz como a soma de produtos das entradas de uma qualquer linha (ou de uma qualquer coluna) da matriz pelos respectivos cofactores. Esta fórmula exprime o determinante de uma matriz à custa de determinantes de matrizes de ordem inferior.

Fórmula de Laplace para o cálculo do determinante

Seja

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1j} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nj} & \cdots & a_{nn} \end{bmatrix}.$$

Sendo C_{ij} o cofactor da entrada a_{ij}, o desenvolvimento de Laplace

a) ao longo da linha i de A é

$$\det(A) = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in}; \tag{2.10}$$

b) ao longo da coluna j de A é

$$\det(A) = a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj}. \tag{2.11}$$

Usamos as expressões 'ao longo da linha...' e 'segundo a linha...' com o mesmo significado.

Exemplo 2.7. Apliquemos a fórmula de Laplace para calcular o determinante da matriz

$$A = \begin{bmatrix} 1 & 0 & 4 & 2 \\ -2 & 0 & 0 & 0 \\ 10 & 3 & -2 & -1 \\ 3 & 2 & 6 & 3 \end{bmatrix}.$$

Observando que a terceira coluna é o dobro da quarta, o determinante desta matriz é zero (ver Exercício 2.3). Vamos confirmar este resultado usando o desenvolvimento de Laplace.

Da análise das expressões (2.10) e (2.11) é evidente que para efectuarmos o desenvolvimento de Laplace é vantajoso escolher uma linha (ou coluna) da matriz que tenha o maior número de zeros. Neste caso, a segunda linha de A. Assim,

$$\begin{aligned}
\det(A) &= -2C_{21} = -2\,(-1)^{1+2}M_{21} \quad \text{(usando a 2ª linha de } A\text{)}\\
&= -2(-1)^{1+2}\begin{vmatrix} 0 & 4 & 2 \\ 3 & -2 & -1 \\ 2 & 6 & 3 \end{vmatrix}\\
&= 2\left(-3\begin{vmatrix} 4 & 2 \\ 6 & 3 \end{vmatrix} + 2\begin{vmatrix} 4 & 2 \\ -2 & -1 \end{vmatrix}\right) \quad \text{(usando a 1ª coluna da matriz } 3\times 3\text{)}\\
&= -6\times 0 + 4\times 0 = 0.
\end{aligned}$$

♦

O exemplo anterior ilustra como aplicações sucessivas da fórmula de Laplace permitem reduzir o cálculo do determinante de uma matriz de qualquer ordem ao cálculo de determinantes de matrizes 2×2 (ou até de matrizes 1×1).

2.3.1 Matriz adjunta e cálculo de A^{-1}

O desenvolvimento de Laplace segundo uma linha (ou coluna) de uma matriz fornece o valor do determinante da matriz em termos dos cofactores e das entradas da linha (ou coluna) escolhida. Deduzimos agora a expressão da inversa de uma matriz em termos do seu determinante e de uma matriz construída com os cofactores das entradas da matriz.

Definição 2.4. Dada uma matriz A do tipo $n \times n$, chama-se *matriz adjunta* de A à matriz transposta da matriz dos cofactores de A. Designando por $\operatorname{adj}(A)$ a matriz adjunta de A, tem-se

$$\operatorname{adj}(A) = [C_{ij}]^T_{i,j=1,\ldots,n} = \begin{bmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \vdots & \vdots & \cdots & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{bmatrix},$$

onde C_{ij} designa o cofactor da entrada a_{ij} de A.

Por vezes, usamos $cof(A) = [C_{ij}]_{i,j=1,\ldots,n}$ para designar a matriz dos cofactores de A.

Exemplo 2.8. Determinar a matriz adjunta da matriz $A = \begin{bmatrix} 1 & 2 & 0 \\ 3 & 2 & 1 \\ 0 & 3 & 4 \end{bmatrix}$.

Usando a definição de cofactor de uma determinada entrada, vem

$$C_{11} = \begin{vmatrix} 2 & 1 \\ 3 & 4 \end{vmatrix} = 5 \qquad C_{12} = -\begin{vmatrix} 3 & 1 \\ 0 & 4 \end{vmatrix} = -12 \qquad C_{13} = \begin{vmatrix} 3 & 2 \\ 0 & 3 \end{vmatrix} = 9$$

$$C_{21} = -\begin{vmatrix} 2 & 0 \\ 3 & 4 \end{vmatrix} = -8 \qquad C_{22} = \begin{vmatrix} 1 & 0 \\ 0 & 4 \end{vmatrix} = 4 \qquad C_{23} = -\begin{vmatrix} 1 & 2 \\ 0 & 3 \end{vmatrix} = -3$$

$$C_{31} = \begin{vmatrix} 2 & 0 \\ 2 & 1 \end{vmatrix} = 2 \qquad C_{32} = -\begin{vmatrix} 1 & 0 \\ 3 & 1 \end{vmatrix} = -1 \qquad C_{33} = \begin{vmatrix} 1 & 2 \\ 3 & 2 \end{vmatrix} = -4$$

Logo,

$$\text{adj}(A) = \begin{bmatrix} 5 & -12 & 9 \\ -8 & 4 & -3 \\ 2 & -1 & -4 \end{bmatrix}^T = \begin{bmatrix} 5 & -8 & 2 \\ -12 & 4 & -1 \\ 9 & -3 & -4 \end{bmatrix}.$$

♦

Antes de prosseguirmos, mostremos que é nula a soma dos produtos das entradas de uma linha pelos correspondentes cofactores de uma outra linha. Provamos este facto para o caso de uma matriz 3×3, embora a demonstração apresentada indique claramente qual o procedimento a seguir no caso geral de uma matriz $n \times n$.

Consideremos a matriz

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

Usando o desenvolvimento de Laplace segundo a primeira linha de A, temos

$$\det(A) = a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13}.$$

Suponha-se que em vez de multiplicar as entradas da 1ª linha de A pelos respectivos cofactores, se multiplicam estas entradas pelos cofactores das entradas de uma outra linha, por exemplo da 3ª linha. Ou seja,

$$a_{11}C_{31} + a_{12}C_{32} + a_{13}C_{33}. \tag{2.12}$$

Para mostrar que esta expressão vale zero construa-se uma matriz A' cujas primeira e segunda linhas são iguais às de A e a terceira linha é igual à primeira linha de A, isto é,

$$A' = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{11} & a_{12} & a_{13} \end{bmatrix}.$$

Os cofactores C'_{3j} das entradas da terceira linha de A', são iguais aos cofactores das entradas da terceira linha de A (já que as duas primeiras linhas de A e A' são iguais), ou seja,

$$C'_{31} = C_{31} \quad C'_{32} = C_{32} \quad C'_{33} = C_{33}.$$

O determinante de A' é igual a zero visto que A' tem duas linhas iguais. Por outro lado, usando o desenvolvimento de Laplace segundo a terceira linha de A' , tem-se

$$0 = \det(A') = a_{11}C'_{31} + a_{12}C'_{32} + a_{13}C'_{33} = a_{11}C_{31} + a_{12}C_{32} + a_{13}C_{33}.$$

Ou seja, a expressão (2.12) é igual a zero, como pretendíamos mostrar.

A soma dos produtos das entradas da linha i de $A = [a_{ij}]$ pelos cofactores das entradas da linha $j \neq i$ é igual a zero. Isto é,

$$0 = a_{i1}C_{j1} + a_{i2}C_{j2} + \cdots + a_{in}C_{jn} = \sum_{k=1}^{n} a_{ik}C_{jk} \quad \text{para} \quad i \neq j \tag{2.13}$$

Exercício 2.5. Mostrar que a expressão (2.13) é válida para qualquer matriz A do tipo $n \times n$. ▲

A matriz adjunta vai desempenhar um papel importante no cálculo da inversa de uma matriz. De facto, calculando o produto $A\,\mathrm{adj}(A) = [b_{ij}]_{i,j=1,\dots n}$,

$$A\,\mathrm{adj}(A) = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ C_{j1} & C_{j2} & \cdots & C_{jn} \\ \vdots & \vdots & \vdots & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} \end{bmatrix}^T$$

$$= \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} C_{11} & \cdots & C_{j1} & \cdots & C_{n1} \\ \vdots & & \vdots & & \vdots \\ C_{1i} & \cdots & C_{ji} & \cdots & C_{ni} \\ \vdots & & \vdots & & \vdots \\ C_{1n} & \cdots & C_{jn} & \cdots & C_{nn} \end{bmatrix}.$$

tem-se

$$b_{ij} = a_{i1}C_{j1} + a_{i2}C_{j2} + \cdots + a_{in}C_{jn} = \sum_{k=1}^{n} a_{ik}C_{jk} \quad \text{para todo } i, j = 1, \dots, n.$$

Notemos agora que:

- Se $i = j$, então $b_{ii} = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in} = \det(A)$ (pela fórmula de Laplace (2.10)).

- Se $i \neq j$, então b_{ij} é a soma dos produtos de entradas da linha i pelos cofactores de uma outra linha. Portanto, de (2.13) segue que $b_{ij} = 0$ se $i \neq j$.

Conclui-se portanto que

$$A\,\mathrm{adj}(A) = [b_{ij}]_{i,j=1,\dots,n} = \det(A)I.$$

Em resumo,

Qualquer matriz quadrada A satisfaz a igualdade

$$A\,\mathrm{adj}(A) = \det(A)\,I. \tag{2.14}$$

Suponha-se agora que A é invertível. Multiplicando a expressão (2.14), à esquerda, pela matriz A^{-1}, obtemos

$$\operatorname{adj}(A) = \det(A)A^{-1} \iff A^{-1} = \frac{1}{\det(A)}\operatorname{adj}(A).$$

Note-se que a equivalência anterior resulta do facto de $\det(A) \neq 0$ sempre que A é invertível. Obtemos assim o teorema que enunciamos a seguir referente à expressão da inversa de uma matriz A em termos da matriz adjunta.

Teorema 2.1. Se A é uma matriz invertível, a sua inversa é

$$A^{-1} = \frac{1}{\det(A)}\operatorname{adj}(A). \tag{2.15}$$

Exemplo 2.9. Calculemos a inversa de $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. A matriz A é invertível se e só se $\det(A) = ad - bc \neq 0$ e neste caso, usando (2.15), temos

$$\begin{aligned} A^{-1} &= \frac{1}{\det(A)}\operatorname{adj}(A) = \frac{1}{ad-bc}\begin{bmatrix} d & -c \\ -b & a \end{bmatrix}^T \\ &= \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}, \end{aligned}$$

expressão que já conhecíamos da Proposição 1.5. ♦

Exemplo 2.10. Calculemos a inversa da matriz do Exemplo 2.8, ou seja de

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 3 & 2 & 1 \\ 0 & 3 & 4 \end{bmatrix}.$$

No exemplo referido calculámos a matriz adjunta de A, e portanto usando (2.15) temos

$$A^{-1} = \frac{1}{\det(A)}\operatorname{adj}(A) = \frac{1}{\det(A)}\begin{bmatrix} 5 & -8 & 2 \\ -12 & 4 & -1 \\ 9 & -3 & -4 \end{bmatrix} = \frac{-1}{19}\begin{bmatrix} 5 & -8 & 2 \\ -12 & 4 & -1 \\ 9 & -3 & -4 \end{bmatrix},$$

uma vez que $\det(A) = 8 - 3 - 24 = -19$ (por exemplo, pela regra de Sarrus). ♦

Exercício 2.6. Use a expressão (2.15) para mostrar

$$(A^{-1})^T = (A^T)^{-1}.$$

▲

2.3.2 Regra de Cramer

A regra de Cramer[5] permite obter uma fórmula para a solução de um sistema $A\mathbf{x} = \mathbf{b}$, quando a matriz A é quadrada e invertível. Embora de valor computacional muito reduzido, esta fórmula pode ter interesse nomeadamente no estudo do comportamento das componentes da solução de um sistema $A\mathbf{x} = \mathbf{b}$ em função de variações no segundo membro $\mathbf{b}$.

Dado um sistema $A\mathbf{x} = \mathbf{b}$, com o mesmo número de equações que incógnitas, sabemos que o sistema tem solução única se e só se A é invertível, ou seja, se e só se $\det(A) \neq 0$. Neste caso, a solução é dada por $\mathbf{x} = A^{-1}\mathbf{b}$ (ver Proposição 1.10, pág. 59). A regra de Cramer usa a expressão (2.15) de A^{-1} para calcular a solução do sistema. De facto, se A é invertível, a solução de $A\mathbf{x} = \mathbf{b}$ é

$$\mathbf{x} = A^{-1}\mathbf{b} = \frac{1}{\det(A)} \operatorname{adj}(A)\, \mathbf{b} = \frac{1}{\det(A)} [C_{ij}]^T \mathbf{b}.$$

Tomando $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ e $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$, da expressão anterior resulta

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_j \\ \vdots \\ x_n \end{bmatrix} = \frac{1}{\det(A)} \begin{bmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ \vdots & \vdots & \cdots & \vdots \\ C_{1j} & C_{2j} & \cdots & C_{nj} \\ \vdots & \vdots & \cdots & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.$$

Desta expressão, podemos concluir que cada componente x_j da solução $\mathbf{x}$ é dada

[5]Esta regra, publicada em 1750, recebe o nome de Gabriel Cramer (1704–1752), embora já em 1748 esta regra tivesse sido publicada por Colin Maclaurin.

por

$$\begin{aligned}
x_j &= \frac{1}{\det(A)}(b_1C_{1j} + b_2C_{2j} + \cdots + b_nC_{nj}) \\
&= \frac{1}{\det(A)} \begin{vmatrix} a_{11} & a_{12} & \cdots & b_1 & \cdots & a_{nn} \\ a_{21} & a_{22} & \cdots & b_2 & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & b_n & \cdots & a_{nn} \end{vmatrix} \quad \text{(usando a igualdade (2.11)).} \\
& \qquad\qquad\qquad\qquad\quad \uparrow \\
& \qquad\qquad\qquad\quad\;\; \text{coluna } j
\end{aligned}$$

O determinante que aparece no numerador da expressão da componente x_j da solução $\mathbf{x}$ é o determinante da matriz que se obtém de A substituindo a coluna j de A pela coluna $\mathbf{b}$ do termo independente do sistema.

Proposição 2.4. Regra de Cramer

Se $A\mathbf{x} = \mathbf{b}$ é um sistema de n equações a n incógnitas e $\det(A) \neq 0$, então o sistema tem solução (única) $\mathbf{x} = (x_1, \ldots, x_n)$, dada por

$$x_1 = \frac{\det A_1}{\det A}, \; x_2 = \frac{\det A_2}{\det A}, \; \ldots, \; x_n = \frac{\det A_n}{\det A},$$

onde A_j é a matriz que se obtém de A substituindo a coluna j de A pela coluna $\mathbf{b}$.

Exemplo 2.11. Utilizemos a regra de Cramer para resolver o sistema

$$\begin{cases} x_1 - 2x_2 = -5 \\ x_1 + \;\; x_2 = \;\; 3. \end{cases}$$

Comecemos por calcular o determinante da matriz A dos coeficientes do sistema.

$$\det(A) = \begin{vmatrix} 1 & -2 \\ 1 & 1 \end{vmatrix} = 1 + 2 = 3.$$

Usando a regra de Cramer resulta

$$x_1 = \frac{1}{3}\begin{vmatrix} -5 & -2 \\ 3 & 1 \end{vmatrix} = \frac{1}{3} \qquad \text{e} \qquad x_2 = \frac{1}{3}\begin{vmatrix} 1 & -5 \\ 1 & 3 \end{vmatrix} = \frac{8}{3}.$$

A solução do sistema é $(x_1, x_2) = \frac{1}{3}(1, 8)$. ♦

Exercícios

1. Determine o sinal dos seguintes produtos elementares da matriz $A = [a_{ij}]_{i,j=1,\cdots,5}$.
a) $a_{13}a_{25}a_{31}a_{42}a_{54}$; b) $a_{21}a_{42}a_{13}a_{54}a_{35}$;
c) $a_{14}a_{23}a_{32}a_{45}a_{51}$.

2. Use operações elementares sobre as linhas da matriz para verificar as igualdades:

a) $\begin{vmatrix} 0 & 0 & 0 & d \\ 0 & 0 & c & e \\ 0 & b & f & g \\ a & h & i & j \end{vmatrix} = abcd$

b) $\begin{vmatrix} 0 & 0 & 0 & 0 & k \\ 0 & 0 & 0 & d & l \\ 0 & 0 & c & e & m \\ 0 & b & f & g & p \\ a & h & i & j & u \end{vmatrix} = abcdk.$

3. Seja $A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$. Considere por hipótese que $\det(A) = -7$. Calcule:

a) $\det(3A)$ b) $\det(2A^{-1})$
c) $\det((2A)^{-1})$
d) $\det \begin{bmatrix} a & g & d \\ b & h & e \\ c & i & f \end{bmatrix}$.

4. Sem calcular explicitamente o determinante, mostre que para $x = 0$ e $x = 2$ é satisfeita a equação

$$\begin{vmatrix} x & x^2 & 2 \\ 1 & 2 & 1 \\ 0 & 0 & -3 \end{vmatrix} = 0.$$

5. Sem calcular explicitamente o determinante, mostre que

$$\begin{vmatrix} b+c & c+a & b+a \\ a & b & c \\ 1 & 1 & 1 \end{vmatrix} = 0.$$

6. Use a propriedade de linearidade do determinante para escrever

$$\begin{vmatrix} a_1 + b_1 & c_1 + d_1 \\ a_2 + b_2 & c_2 + d_2 \end{vmatrix}$$

como uma soma de quatro determinantes de matrizes cujas entradas não são expressas em termos de adições.

7. Diga, justificando, se é ou não verdadeira a igualdade

$$\det(A + B) = \det(A) + \det(B).$$

8. Sem calcular os determinantes, mostre as igualdades seguintes.

a) $\begin{vmatrix} a_1 & b_1 & a_1 + b_1 + c_1 \\ a_2 & b_2 & a_2 + b_2 + c_2 \\ a_3 & b_3 & a_3 + b_3 + c_3 \end{vmatrix} = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}$

b) $\begin{vmatrix} a_1 + b_1 & a_1 - b_1 & c_1 \\ a_2 + b_2 & a_2 - b_2 & c_2 \\ a_3 + b_3 & a_3 - b_3 & c_3 \end{vmatrix} = -2 \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}$.

9. Quais os valores de k para os quais a matriz A é singular?

a) $A = \begin{bmatrix} 1 & 2 & 4 \\ 3 & 1 & 6 \\ k & 3 & 2 \end{bmatrix}$

b) $A = \begin{bmatrix} k-3 & -2 \\ -2 & k-2 \end{bmatrix}$

10. Considere a matriz

$$M = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

a) Calcule o determinante de M.
b) Calcule $\det(2M), \det(2M^{-1})$ e $\det((2M)^{-1})$.
c) Determine a entrada $(1,4)$ da matriz M^{-1}.

11. Seja $A = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 3 & 2 \\ 0 & 1 & -2 \end{bmatrix}$. Verifique que A é invertível e calcule:

a) $\det(2A^{-1})$ b) $\det(A^2(2A^{-1}))$
c) $\det(A^T(\operatorname{tr} A)A)^{-1}$, onde $\operatorname{tr} A$ designa o traço de A, ou seja, a soma das entradas da diagonal principal de A.

12. A matriz dos cofactores da matriz A é

$$cof(A) = \begin{bmatrix} 1 & 2 \\ -1 & 3 \end{bmatrix}.$$

a) Use a fórmula $A\,(cof(A))^T = \det(A)I$ para calcular $\det(A)$.
b) Determine a matriz inversa de A.

13. A matriz dos cofactores da matriz A é

$$cof(A) = \begin{bmatrix} 1 & 2 & 2 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{bmatrix}.$$

a) Use a fórmula $A\,(cof(A))^T = \det(A)I$ para achar os possíveis valores para $\det(A)$.
b) Justifique por que razão A é invertível e calcule a entrada (3,2) da inversa de A.

14. Use o desenvolvimento de Laplace para calcular os determinantes das matrizes seguintes.

$$A = \begin{bmatrix} 2 & 5 & 1 & 0 \\ 0 & 3 & 2 & 1 \\ 1 & -1 & 2 & 0 \\ -1 & 0 & 3 & 2 \end{bmatrix}$$

$$B = \begin{bmatrix} 0 & 5 & 1 & 0 & 2 \\ 0 & 3 & 2 & 1 & -1 \\ 1 & 0 & 2 & 0 & 0 \\ -1 & 0 & 3 & 2 & 1 \\ 1 & 2 & 3 & 4 & 5 \end{bmatrix}.$$

Além disso, calcule a inversa de A e de B sem utilizar o método de eliminação de Gauss-Jordan.

15. Resolva as equações seguintes.

a) $\begin{vmatrix} 1 & x & 1 \\ 0 & -1 & 1 \\ 1 & 0 & 2 \end{vmatrix} = 0$

b) $\begin{vmatrix} x & x & x & x \\ x & 1 & x & x \\ x & x & 1 & x \\ x & x & x & 1 \end{vmatrix} = 0.$

16. Resolva os seguintes sistemas de equações lineares utilizando a regra de Cramer:

a) $\begin{cases} x_1 - 2x_2 = 4 \\ 2x_1 - x_2 = -2 \end{cases}$

b) $\begin{cases} x_1 - 3x_2 + x_3 = 4 \\ 2x_1 - x_2 = -2 \\ 4x_1 - 3x_3 = -2 \end{cases}$

17. Considere A e B duas matrizes quadradas de ordem 3 e a seguinte lista de afirmações.

I) $\det(AB) = \det(BA)$.
II) Se $\det A = 0$ e $\det B = 0$, então $\det(A+B) = 0$.

III) $\det(2AB) = 8 \det\,(AB)$.

A lista completa de afirmações correctas é:

a) I e II
b) I e III
c) II e III
d) I e II e III

18. Considere o sistema $(A - \lambda I)\mathbf{x} = \mathbf{0}$, onde λ é um escalar, I é a matriz identidade, e A é a matriz

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$

a) Determine os valores de λ para os quais o sistema tem soluções não nulas.
b) Para cada valor de λ encontrado na alínea anterior, determine uma solução não nula do sistema correspondente.

19. Indique o valor lógico das afirmações:

a) Existem matrizes A e B, do tipo 2×2, tais que

$$A^{-1}B^{-1}AB = \begin{bmatrix} 7 & 4 \\ 2 & 1 \end{bmatrix}.$$

b) $\det(A - 5I) = \det(A^T - 5I)$.
c) Se $B = SAS^{-1}$, então $\det B = \det A$.
d) Se $A^2 = A$, então $\det A = 0$ ou $\det A = 1$.

Capítulo 3

Espaços Lineares

A noção de espaço linear reveste-se de importância fundamental nos mais diversos ramos da matemática, permitindo unificar o tratamento matemático de estruturas distintas. Como veremos no final deste capítulo, a noção abstracta de espaço linear permite atribuir a mesma estrutura a conjuntos de vectores de $\mathbb{R}^n$, a conjuntos de funções (tais como polinómios, funções exponenciais ou trigonométricas), bem como a conjuntos de matrizes.

Um espaço linear é um conjunto no qual estão definidas duas operações: (i) uma, sobre pares de elementos do conjunto (análoga à adição de vectores de $\mathbb{R}^n$); (ii) outra, de escalares por elementos do conjunto (tal como a multiplicação de um escalar real por um vector de $\mathbb{R}^n$). Se estas duas operações verificarem as propriedades enunciadas na página 29 para as operações de adição de vectores de $\mathbb{R}^n$ e de multiplicação destes vectores por escalares reais, então dizemos que o conjunto é um espaço linear (ou espaço vectorial).

Uma vez que nos espaços lineares $\mathbb{R}^2$ e $\mathbb{R}^3$ é possível utilizar uma abordagem geométrica para ilustrar muitos dos conceitos fundamentais relativos a espaços lineares, adoptamos como estratégia pedagógica o uso do espaço linear $\mathbb{R}^n$ como modelo. Assim, na Secção 3.2 são apresentados os quatro subespaços fundamentais de $\mathbb{R}^n$ (associados a uma matriz real), os quais são depois utilizados para introduzir várias noções como as de subespaço e dimensão de um espaço linear. Estes conceitos são depois definidos e estudados para espaços lineares gerais.

As noções fundamentais de independência linear, base e dimensão de um espaço linear são estudadas na Secção 3.3. Os resultados principais desta secção, ilustrados com exemplos em $\mathbb{R}^2$ e $\mathbb{R}^3$, apoiam-se em grande parte em resultados previamente obtidos para sistemas lineares. Mostra-se que, fixada uma base ordenada num espaço linear real de dimensão n, existe uma correspondência biunívoca entre os vectores desse espaço e vectores de $\mathbb{R}^n$. Esta correspondência associa a cada vector do espaço linear o vector das coordenadas na base ordenada fixada. Na Secção 3.4 relacionam-se os vectores das coordenadas em bases distintas através de uma matriz designada por matriz de mudança de base.

Na última secção deste capítulo apresentamos exemplos de espaços lineares distintos de $\mathbb{R}^n$. Para cada um desses exemplos exploramos as noções de base e dimensão bem como outros conceitos desenvolvidos ao longo do capítulo.

3.1 Definição de espaço linear

Embora neste texto só consideremos espaços lineares reais ou complexos, ou seja, espaços em que os escalares são números reais ou complexos, poderíamos considerar escalares pertencentes a um qualquer outro *corpo*[1]. Eis a definição de espaço linear sobre um corpo $\mathbb{K}$, onde $\mathbb{K} = \mathbb{R}$ ou $\mathbb{K} = \mathbb{C}$.

[1]Um corpo é um conjunto com duas operações binárias designadas por adição e multiplicação, gozando das propriedades: comutatividade, associatividade, distributividade da multiplicação em relação à adição, existência de elemento neutro para a adição e para a multiplicação, existência de simétricos e existência de inverso para qualquer elemento diferente de zero.

Definição 3.1. Espaço linear

Seja W um conjunto não vazio e $\mathbb{K}$ um corpo. Suponha-se que está definida uma operação sobre pares de elementos de W designada por *adição* e denotada por $+$, e outra operação, que denominamos por *multiplicação por escalares*, que a cada $\alpha \in \mathbb{K}$ e a cada $u \in W$ associa um um elemento que denotamos por $\alpha \cdot u$, ou simplesmente por αu.

Dizemos que W, munido destas duas operações, é um *espaço linear* ou *espaço vectorial*, sobre o corpo $\mathbb{K}$, se as operações de adição e multiplicação por escalares satisfazem os axiomas que a seguir se enunciam, onde $u, v, w \in W$ e $\alpha, \beta \in \mathbb{K}$:

A1 **Fecho da adição:** $u + v \in W$.

A2 **Fecho da multiplicação por escalares:** $\alpha \cdot u \in W$.

A3 **Comutatividade da adição:** $u + v = v + u$.

A4 **Associatividade da adição:** $(u + v) + w = u + (v + w)$.

A5 **Existência de elemento neutro da adição:** existe um elemento em W, designado por 0 (zero), tal que $0 + u = u$.

A6 **Existência de simétrico:** para cada elemento $u \in W$ existe um elemento $u' \in W$ tal que $u + u' = 0$.

A7 **Associatividade da multiplicação por escalares:** $(\alpha\beta) \cdot u = \alpha(\beta \cdot u)$.

A8 **Distributividade:** $\alpha \cdot (u + v) = \alpha \cdot u + \alpha \cdot v$ e $(\alpha + \beta) \cdot u = \alpha \cdot u + \beta \cdot u$.

A9 **Existência de identidade:** sendo 1 a identidade de $\mathbb{K}$, tem-se $1 \cdot u = u$.

Os elementos de um espaço linear são designados por *vectores*. Quando o corpo $\mathbb{K}$ é $\mathbb{R}$ diz-se que W é um *espaço linear real*, e no caso em que $\mathbb{K} = \mathbb{C}$ diz-se um *espaço linear complexo*.

O fecho da adição significa que a adição é uma operação binária, isto é, uma operação que a cada par (u, v) de elementos de W associa um (e um só) elemento de W.

Se um conjunto W está munido de uma operação fechada, o conjunto diz-se fechado em relação à operação. Os termos "a operação é fechada em W" ou "W é fechado para a operação" têm o mesmo significado: o resultado da operação é sempre um elemento de W.

Nota 15. *Na definição de espaço linear deixámos de usar letra a cheio para denotar vectores. Continuaremos no entanto a usar essa notação quando em causa estão vectores de $\mathbb{R}^n$ (ou de $\mathbb{C}^n$).*

Como consequência lógica dos axiomas da Definição 3.1 resultam várias propriedades das operações definidas num espaço linear, algumas das quais são enunciadas na proposição que se segue.

Proposição 3.1. Seja W um espaço linear sobre o corpo $\mathbb{K}$. São válidas as afirmações:

a) O vector zero 0 de W é único.
b) Para cada $u \in W$, o simétrico de u é único.
c) $\alpha \cdot 0 = 0, \quad$ para todo $\alpha \in \mathbb{K}$.
d) Designando por $-u$ o simétrico de u, tem-se

$$(-\alpha)u = \alpha(-u) = -(\alpha u) \quad \text{para todo } \alpha \in \mathbb{K} \text{ e todo } u \in W.$$

e) $\alpha u \neq 0 \quad$ para todo $\alpha \neq 0 \in \mathbb{K}$ e $u \neq 0 \in W$.

Demonstração. Realizamos apenas a prova dos dois primeiros itens deixando os restantes como exercício.

a) Suponha-se que $w \in W$ verifica

$$u + w \overset{\text{A3}}{=} w + u = u, \quad \text{para todo } u \in W.$$

Em particular (para $u = 0$), tem-se $0 + w = 0$ e portanto, por A5, $w = 0$.

b) Admitamos que u' e w são vectores simétricos de u, isto é

$$u + u' = 0 \quad \text{e} \quad u + w = 0.$$

Logo,

$$w + (u + u') = w + 0 \overset{\text{A3}}{=} 0 + w \overset{\text{A5}}{=} w$$

e

$$(w + u) + u' \overset{\text{A3}}{=} (u + w) + u' = 0 + u' \overset{\text{A5}}{=} u'.$$

A associatividade da adição garante que as duas expressões anteriores são iguais, e portanto $w = u'$.

$\square$

O simétrico de um vector u passará a ser designado por $(-u)$.

Como vimos no Capítulo 1, em $\mathbb{R}^n$ as operações de adição e multiplicação por escalares reais definidas por

$$\mathbf{x}+\mathbf{y} = (x_1+y_1, x_2+y_2, \ldots, x_n+y_n) \quad \text{e} \quad \alpha\mathbf{x} = (\alpha x_1, \alpha x_2, \ldots, \alpha x_n), \tag{3.1}$$

onde $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, $\mathbf{y} = (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$ e $\alpha \in \mathbb{R}$, satisfazem os axiomas na definição de espaço linear. Em particular, o zero de $\mathbb{R}^n$, isto é o elemento de $\mathbb{R}^n$ que adicionado a qualquer vector $\mathbf{x} \in \mathbb{R}^n$ é igual a $\mathbf{x}$, é o vector $\mathbf{0} = (0, 0, \ldots, 0) \in \mathbb{R}^n$. Por isso, $\mathbb{R}^n$ munido destas operações é um espaço linear real.

Um outro exemplo de espaço linear é o conjunto das matrizes com as operações de adição de matrizes e multiplicação de matrizes por um escalar tal como foram definidas no primeiro capítulo. As propriedades destas operações (ver propriedades (a), (b), (f) e (g) listadas no Teorema 1.1, pág. 42) garantem que o conjunto das matrizes reais do tipo $p \times n$ é um espaço linear real quando munido dessas operações. Neste espaço linear, a matriz nula (isto é, com todas as entradas iguais a zero) é o elemento neutro da adição e o simétrico de uma matriz A é a matriz $-A$.

À semelhança da definição de $\mathbb{R}^n$ como o conjunto dos n-uplos de números reais, define-se $\mathbb{C}^n$ como sendo o conjunto dos n-uplos de números complexos. Isto é,

$$\mathbb{C}^n = \{(z_1, z_2, \ldots, z_n) : \; z_i \in \mathbb{C}, i = 1, \ldots, n\}.$$

De forma inteiramente análoga às definições de adição de vectores de $\mathbb{R}^n$ e multiplicação de um escalar por um vector de $\mathbb{R}^n$, definem-se estas operações em $\mathbb{C}^n$. Nomeadamente, para $\mathbf{z}, \mathbf{w} \in \mathbb{C}^n$ e $\alpha \in \mathbb{C}$, definimos

$$\mathbf{z}+\mathbf{w} = (z_1+w_1, z_2+w_2, \ldots, z_n+w_n) \quad \text{e} \quad \alpha\mathbf{z} = (\alpha z_1, \alpha z_2, \ldots, \alpha z_n), \tag{3.2}$$

Como a soma de dois complexos é um complexo e o produto de dois complexos ainda é um complexo, a adição e a multiplicação por escalares complexos são operações fechadas em $\mathbb{C}^n$. O elemento neutro da adição é o n-uplo $(0+0i, 0+0i, \ldots, 0+0i)$. Os restantes axiomas de espaço linear são consequências imediatas das propriedades da adição e multiplicação de números complexos. Assim, $\mathbb{C}^n$ munido das operações (3.2) é um exemplo de um espaço linear complexo.

Consideremos agora um exemplo de um conjunto que não é um espaço linear. Seja W o conjunto dos pares ordenados de números reais cuja segunda componente é igual a 1, isto é, $W = \{(a, 1) : \; a \in \mathbb{R}\}$. Considerem-se as operações usuais de adição e multiplicação por escalares definidas em $\mathbb{R}^2$. Uma vez que para quaisquer elementos $(a, 1)$ e $(b, 1)$ de W e $\alpha \in \mathbb{R}$ se tem

$$(a, 1) + (b, 1) = (a + b, 2) \notin W \quad \text{e} \quad \alpha(a, 1) = (\alpha a, \alpha) \notin W \text{ para } \alpha \neq 1,$$

as operações de adição e multiplicação por escalares não são fechadas em W. Portanto W não é um espaço linear.

3.2 Os quatro subespaços fundamentais

Abordamos nesta secção as noções de subespaço e de dimensão para certos conjuntos de $\mathbb{R}^n$ (ou de $\mathbb{C}^n$) associados a uma matriz. Estes conjuntos recebem a designação de *subespaços fundamentais* (de $\mathbb{R}^n$ ou de $\mathbb{C}^n$). Os quatro subespaços fundamentais associados a uma matriz A são: o núcleo de uma matriz A e da sua transposta, e o espaço das colunas de A e da sua transposta. A designação de subespaços fundamentais, introduzida por G. Strang [11], sublinha o papel importante que estes subespaços desempenham em álgebra linear, como terá o leitor oportunidade de comprovar ao longo deste livro.

3.2.1 Núcleo de uma matriz

Definição 3.2. Núcleo de uma matriz

O *núcleo* da matriz A, que designamos por $N(A)$, é o conjunto das soluções do sistema homogéneo $A\mathbf{x} = \mathbf{0}$. Em particular, se A é real e do tipo $p \times n$, o núcleo de A é

$$N(A) = \{\mathbf{x} \in \mathbb{R}^n : \ A\mathbf{x} = \mathbf{0}\}.$$

A designação anglo-saxónica para núcleo de uma matriz é ”kernel”, usando-se a notação $\operatorname{Ker}(A)$ para $N(A)$.

O núcleo de uma matriz nunca é um conjunto vazio já que um sistema homogéneo tem pelo menos a solução nula $\mathbf{x} = \mathbf{0}$. Assim, $\mathbf{0} \in N(A)$ qualquer que seja a matriz A. Além disso, se A é uma matriz quadrada temos: (i) se A é invertível, então $N(A) = \{\mathbf{0}\}$ (cf. Teorema 1.4, pág. 57); (ii) se A não é invertível, o sistema $A\mathbf{x} = \mathbf{0}$ é indeterminado, pelo que o núcleo de A tem um número infinito de elementos (cf. Proposição 2.2, pág. 95).

Exemplo 3.1. Determinemos um conjunto gerador para o núcleo de $A = \begin{bmatrix} 2 & 1 & 1 \\ 5 & 4 & 1 \end{bmatrix}$.

O núcleo de A é a solução geral do sistema homogéneo

$$\begin{cases} 2x + \ \ y + z = \ 0 \\ 5x + 4y + z = 0. \end{cases}$$

Aplicando o método de eliminação de Gauss à matriz A, temos

$$\begin{bmatrix} 2 & 1 & 1 \\ 5 & 4 & 1 \end{bmatrix} \overset{L_2-\frac{5}{2}L_1}{\longrightarrow} \begin{bmatrix} 2 & 1 & 1 \\ 0 & \frac{3}{2} & -\frac{3}{2} \end{bmatrix}.$$

O sistema dado tem a mesma solução geral que o sistema

$$\begin{cases} 2x + \ y + \ z = 0 \\ \quad\ \ \frac{3}{2}y - \frac{3}{2}z = 0. \end{cases}$$

Logo, da última equação resulta $y = z$, e substituindo na primeira equação vem $x = -z$. Ou seja,

$$\begin{aligned} N(A) &= \left\{(x, y, z) \in \mathbb{R}^3 : x = -z \text{ e } y = z\right\} = \{(-z, z, z) : z \in \mathbb{R}\} \\ &= \{z(-1, 1, 1) : z \in \mathbb{R}\} = \operatorname{Span}\{(-1, 1, 1)\}, \end{aligned}$$

onde na última igualdade usámos a definição de conjunto gerador (Definição 1.10, pág. 30). O núcleo da matriz A é o conjunto gerado pelo vector $(-1, 1, 1)$, isto é, o conjunto de todos os múltiplos do vector $(-1, 1, 1)$. Geometricamente, $N(A)$ é uma recta do espaço que passa pela origem e tem a direcção do vector $(-1, 1, 1)$. ♦

Exemplo 3.2. Determinemos o núcleo da matriz

$$A_\alpha = \begin{bmatrix} 1 & 2 & -1 \\ -\alpha & -2 & \alpha \\ \alpha & 2\alpha & -\alpha \end{bmatrix}$$

em termos do parâmetro real α.

Como a terceira linha de A_α é um múltiplo da primeira linha, $\det(A_\alpha) = 0$ para qualquer α. Logo, $N(A_\alpha) \neq \{(0, 0, 0)\}$ (ver Proposição 2.2, pág 95).

Apliquemos o método de eliminação de Gauss para decidir qual a característica de A_α.

$$\begin{bmatrix} 1 & 2 & -1 \\ -\alpha & -2 & \alpha \\ \alpha & 2\alpha & -\alpha \end{bmatrix} \underset{L_3-\alpha L_1}{\overset{L_2+\alpha L_1}{\longrightarrow}} \begin{bmatrix} 1 & 2 & -1 \\ 0 & -2+2\alpha & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Assim, para $\alpha = 1$ (isto é, $2\alpha - 2 = 0$) tem-se $\operatorname{car}(A_1) = 1$, e para $\alpha \neq 1$ a característica é $\operatorname{car}(A_\alpha) = 2$.

Da Proposição 1.2 segue que o sistema $A_1\mathbf{x} = \mathbf{0}$ tem grau de indeterminação 2 uma vez que $\operatorname{car}(A_1) = 1$. Da matriz em escada por linhas obtida, conclui-se que $A_1\mathbf{x} = \mathbf{0}$ é equivalente a um sistema com uma única equação, cuja solução é $x = -2y + z$. Logo,

$$\begin{aligned} N(A_1) &= \{(-2y+z, y, z) : \; y, z \in \mathbb{R}\} = \{y(-2,1,0) + z(1,0,1) : \; y, z \in \mathbb{R}\} \\ &= \operatorname{Span}\{(-2,1,0), (1,0,1)\}. \end{aligned}$$

Para $\alpha \neq 1$ o sistema $A_\alpha\mathbf{x} = \mathbf{0}$ tem grau de indeterminação 1, e as suas soluções são as soluções do sistema

$$\begin{cases} x + \quad 2y - z = 0 \\ \quad (-2+2\alpha)y \quad = 0. \end{cases}$$

A solução deste sistema é $y = 0$ e $x = z$. Logo, para $\alpha \neq 1$ o núcleo de A_α é

$$N(A_\alpha) = \{(z, 0, z) : \; z \in \mathbb{R}\} = \{z(1,0,1) : \; z \in \mathbb{R}\} = \operatorname{Span}\{(1,0,1)\}.$$

Geometricamente, $N(A_1)$ é um plano que contém a origem e os vectores $(-2,1,0)$ e $(1,0,1)$, enquanto que para $\alpha \neq 1$ o núcleo de A_α é uma recta que passa pela origem e tem a direcção de $(1,0,1)$. Relembre-se que em $\mathbb{R}^3$ dois vectores não colineares geram um plano, enquanto que um vector (não nulo) gera uma recta (ver Capítulo 1). ♦

No exemplo anterior vimos que, quando $\alpha \neq 1$ bastava um vector (não nulo) para gerar o núcleo da matriz A_α, enquanto que para $\alpha = 1$ são necessários dois vectores para gerar o núcleo de A_α. Em ambos os casos, o número de vectores necessários para gerar o núcleo de A_α é igual ao grau de indeterminação do sistema $A_\alpha\mathbf{x} = \mathbf{0}$. Como veremos em secções subsequentes o grau de indeterminação de $A\mathbf{x} = \mathbf{0}$ é sempre igual ao menor número de vectores necessários para gerar $N(A)$. Por isso, ao grau de indeterminação de um sistema $A\mathbf{x} = \mathbf{0}$ também chamamos *dimensão do núcleo* de A, a qual é designada por $\dim N(A)$. Quando $N(A) = \{\mathbf{0}\}$ convencionamos que a dimensão de $N(A)$ é zero. O conceito de dimensão de um espaço linear (não necessariamente o núcleo de uma matriz) será estudado na Secção 3.3, onde mostramos que de facto a dimensão do espaço linear $N(A)$ é precisamente o grau de indeterminação do sistema $A\mathbf{x} = \mathbf{0}$ (ver Teorema 3.5, pág. 143).

Dos resultados obtidos para sistemas de equações lineares (possíveis), sabemos que a característica da matriz dos coeficientes de um sistema é igual ao número de variáveis dependentes, enquanto que o grau de indeterminação é igual ao número de variáveis livres (ver Proposição 1.2, pág. 18). Podemos assim concluir a seguinte relação fundamental entre a característica de A e a dimensão do núcleo de A:

$$\operatorname{car}(A) + \dim N(A) = \text{número de colunas de } A. \tag{3.3}$$

Enunciamos a seguir uma propriedade do núcleo que permitirá estabelecer que o núcleo de uma matriz real (resp. complexa) $p \times n$ é um espaço linear de $\mathbb{R}^n$ (resp. de $\mathbb{C}^n$).

Proposição 3.2. Seja A uma matriz e $N(A)$ o seu núcleo. Tem-se:

a) $N(A)$ é fechado para a adição:

$$\text{para quaisquer } \mathbf{v}, \mathbf{w} \in N(A) \Longrightarrow \mathbf{v} + \mathbf{w} \in N(A).$$

b) $N(A)$ é fechado para a multiplicação por escalares:

$$\text{para qualquer } \mathbf{v} \in N(A) \text{ e qualquer } \alpha \text{ escalar} \Longrightarrow \alpha\mathbf{v} \in N(A).$$

Demonstração. Da definição de núcleo de uma matriz, tem-se

$$\mathbf{v} \in N(A) \Longleftrightarrow \mathbf{v} \text{ é solução do sistema } A\mathbf{x} = \mathbf{0} \Longleftrightarrow A\mathbf{v} = \mathbf{0}$$
$$\mathbf{w} \in N(A) \Longleftrightarrow \mathbf{w} \text{ é solução do sistema } A\mathbf{x} = \mathbf{0} \Longleftrightarrow A\mathbf{w} = \mathbf{0}.$$

Adicionando $A\mathbf{v} = \mathbf{0}$ a $A\mathbf{w} = \mathbf{0}$, obtemos

$$\mathbf{0} = A\mathbf{v} + A\mathbf{w} = A(\mathbf{v} + \mathbf{w}) \Longleftrightarrow \mathbf{v} + \mathbf{w} \in N(A).$$

Multiplicando por um escalar α a expressão $A\mathbf{v} = \mathbf{0}$, vem

$$\mathbf{0} = \alpha A\mathbf{v} = A(\alpha\mathbf{v}) \Longleftrightarrow \alpha\mathbf{v} \in N(A).$$

□

A proposição anterior fornece um exemplo de um conjunto fechado para a adição e multiplicação por escalares.

A um subconjunto não vazio de um espaço linear que seja fechado para as operações definidas no espaço linear chama-se *subespaço* linear.

Definição 3.3. Subespaço linear
Um subconjunto não vazio V de um espaço linear W diz-se um *subespaço linear* (ou vectorial) de W, se V é fechado para as operações de adição e multiplicação por escalares definidas em W.

A definição de subespaço e a Proposição 3.2 permitem concluir:

Corolário 3.1. O núcleo de uma matriz real (resp. complexa) A do tipo $p \times n$ é um subespaço linear de $\mathbb{R}^n$ (resp. de $\mathbb{C}^n$).

Nota 16.

1. *Se V é um subespaço linear de W, então contém o vector zero. De facto, sendo V fechado para a adição de vectores e para a multiplicação por escalares, tem-se que para qualquer $x \in V$, o vector $-x$ e o vector $x+(-x) = 0$ pertencem ambos a V (usámos o facto de que $(-x) = (-1)x$, conforme Proposição 3.1).*

2. *Como o nome indica, um subespaço V de um espaço linear de W é ele próprio um espaço linear com a restrição das operações de W. Note-se que as operações em V são as operações definidas em W e portanto gozam das propriedades necessárias para que V seja um espaço linear.*

Exemplo 3.3. São subespaços lineares de $\mathbb{R}^2$:

- $\{(0,0)\}$.
- Rectas que passam pela origem.
- $\mathbb{R}^2$.

São subespaços lineares de $\mathbb{R}^3$:

- $\{(0,0,0)\}$.
- Rectas que passam pela origem.
- Planos que contêm a origem.
- $\mathbb{R}^3$.

♦

Exemplo 3.4. Verifiquemos que o conjunto gerado por um subconjunto não vazio V de $\mathbb{R}^n$ (resp. $\mathbb{C}^n$), é um subespaço linear de $\mathbb{R}^n$ (resp. $\mathbb{C}^n$).

Relembremos que $\mathrm{Span}(V)$ é o conjunto de todas as combinações lineares de elementos de V.

(i) Como a soma de duas combinações lineares de elementos de V ainda é uma combinação linear de elementos de V, conclui-se que $\mathrm{Span}(V)$ é fechado para a adição.

(ii) Como a multiplicação de uma combinação linear de elementos de V por um escalar real (resp. complexo) ainda é uma combinação linear de elementos de V, conclui-se que $\mathrm{Span}(V)$ é fechado para a multiplicação por escalares.

Assim,

Se V é um subconjunto não vazio de $\mathbb{R}^n$ (resp. de $\mathbb{C}^n$), então $\mathrm{Span}(V)$ é um subespaço linear de $\mathbb{R}^n$ (resp. de $\mathbb{C}^n$).

♦

Exemplo 3.5. O subconjunto $V = \{(1+x, x) : x \in \mathbb{R}\}$ de $\mathbb{R}^2$ não é um subespaço linear de $\mathbb{R}^2$ uma vez que não contém $(0, 0)$. Facilmente se verifica que V não é fechado, por exemplo, para a adição. ♦

Exercício 3.1. Dados dois subespaços U e V de um espaço linear W mostre que:

1. A intersecção $U \cap V$ é um subespaço de W.

2. O conjunto $U + V = \{x \in W : x = u + v, \text{ para alguns } u \in U \text{ e } v \in V\}$ é um subespaço de W.

3. A união $U \cup V$ não é em geral um subespaço de W. (Encontre um contra-exemplo em $\mathbb{R}^2$).

▲

3.2.2 Espaço das linhas e das colunas de uma matriz

Consideremos a matriz real A do tipo $p \times n$

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pn} \end{bmatrix}. \tag{3.4}$$

Esta matriz tem p linhas que são vectores de $\mathbb{R}^n$, nomeadamente

$$\mathbf{l}_1 = (a_{11}, a_{12}, \ldots, a_{1n}),\ \ \mathbf{l}_2 = (a_{21}, a_{22}, \ldots, a_{2n}), \ldots, \mathbf{l}_p = (a_{p1}, a_{p2}, \ldots, a_{pn}).$$

Além disso, a matriz A tem n colunas que pertencem a $\mathbb{R}^p$, nomeadamente

$$\mathbf{c}_1 = (a_{11}, a_{21}, \ldots, a_{p1}),\ \ \mathbf{c}_2 = (a_{12}, a_{22}, \ldots, a_{p2}), \ldots, \mathbf{c}_n = (a_{1n}, a_{2n}, \ldots, a_{pn}).$$

Definição 3.4. Chama-se *espaço das linhas* de uma matriz A ao espaço linear gerado pelas linhas de A, e *espaço das colunas* de A ao espaço linear gerado pelas colunas de A. Designamos os espaços das linhas e das colunas de A, respectivamente, por $EL(A)$ e $EC(A)$.

Para a matriz real A em (3.4), temos

$$EL(A) = \operatorname{Span}(\{\mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_p\}) \subseteq \mathbb{R}^n \text{ e } EC(A) = \operatorname{Span}(\{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n\}) \subseteq \mathbb{R}^p.$$

No Exemplo 3.4 vimos que o conjunto gerado por um subconjunto não vazio de $\mathbb{R}^k$ é um subespaço linear de $\mathbb{R}^k$. Assim, os espaços $EC(A)$ e $EL(A)$ de uma matriz real A do tipo $p \times n$ são subespaços lineares, respectivamente, de $\mathbb{R}^p$ e de $\mathbb{R}^n$.

Baseando-nos em resultados dos capítulos 1 e 2 podemos enunciar de imediato uma proposição relativa aos subespaços fundamentais de uma matriz (quadrada) real invertível.

Proposição 3.3. Se A é uma matriz real invertível de ordem n, então

$$N(A) = N(A^T) = \{\mathbf{0}\} \quad \text{e} \quad \mathbb{R}^n = EC(A) = EL(A).$$

Demonstração. A Proposição 1.10 (pág. 59) garante que um sistema homogéneo tem solução única $\mathbf{x} = \mathbf{0}$ se e só se a matriz dos coeficientes do sistema é invertível. Como o determinante de A e de A^T são iguais, as matrizes A e A^T são invertíveis e portanto $N(A) = N(A^T) = \{\mathbf{0}\}$.

Uma vez que uma matriz real é invertível se e só se os seus vectores coluna geram $\mathbb{R}^n$ (ver Proposição 1.12, pág. 62), e o espaço das colunas de uma matriz é precisamente o espaço gerado pelas suas colunas, temos $\mathbb{R}^n = EC(A)$. De igual modo, como A^T é invertível tem-se $\mathbb{R}^n = EC(A^T) = EL(A)$. □

Exemplo 3.6. Consideremos a matriz rectangular

$$A = \begin{bmatrix} 2 & 4 & 6 \\ -1 & -2 & -3 \end{bmatrix}.$$

É óbvio que os três vectores coluna $(2,-1),(4,-2)$ e $(6,3)$ são colineares, e portanto o espaço das colunas de A é uma recta em $\mathbb{R}^2$. Logo, apenas precisamos de um vector para gerar $EC(A)$:

$$EC(A) = \text{Span}\{(2,-1),(4,-2),(6,-3)\} = \text{Span}\{(2,-1)\} \subset \mathbb{R}^2.$$

Da mesma forma, os vectores linha $(2,4,6)$ e $(-1,-2,-3)$ são colineares e portanto

$$EL(A) = \text{Span}\{(-1,-2,-3)\} = \text{Span}\{(-1,-2,-3),(2,4,6)\} \subset \mathbb{R}^3.$$

Tanto $EL(A)$ como $EC(A)$ são rectas, embora estejam em espaços diferentes. O espaço das linhas de A é uma recta no espaço $\mathbb{R}^3$, e o espaço das colunas de A é uma recta no plano $\mathbb{R}^2$. Na Figura 3.1 encontram-se representados os espaços das linhas e colunas de A.

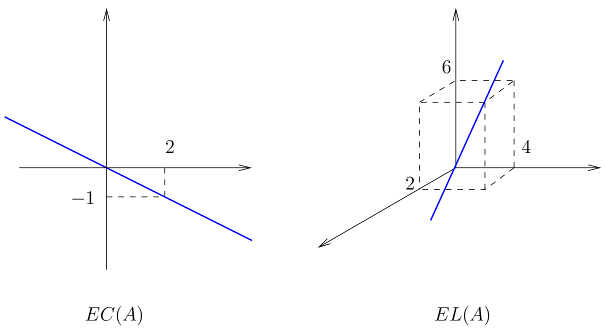


Figura 3.1: Espaço das linhas, $EL(A)$, e espaço das colunas, $EC(A)$, da matriz A do Exemplo 3.6.

♦

Sabemos, desde o primeiro capítulo, que qualquer subespaço não nulo de $\mathbb{R}^n$ (ou de $\mathbb{C}^n$) admite vários conjuntos geradores. O nosso objectivo agora é encontrar conjuntos geradores de um subespaço V de $\mathbb{R}^n$ (ou de $\mathbb{C}^n$) com o menor número possível de elementos. Uma vez que se trata de subespaços de $\mathbb{R}^n$ (ou $\mathbb{C}^n$) podemos sempre construir uma matriz colocando os vectores de V nas linhas (ou nas colunas), reduzindo-se o problema a determinar conjuntos geradores dos espaços das linhas (ou das colunas) de uma matriz com o menor número possível de vectores.

Comecemos por relembrar (ver Exercício 1.1, pág. 34) que se substituirmos um (ou mais) vector de um conjunto gerador $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ do subespaço $V \subset \mathbb{R}^n$ por uma combinação linear (não nula) de vectores de S, o novo conjunto ainda gera V. Por exemplo, substituindo $\mathbf{u}_k$ por $\mathbf{w} = \alpha_1\mathbf{u}_1 + \cdots + \alpha_k\mathbf{u}_k$ com $\alpha_k \neq 0$, tem-se

$$V = \operatorname{Span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\} = \operatorname{Span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{k-1}, \mathbf{w}\},$$

uma vez que

$$\beta_1\mathbf{u}_1+\cdots+\beta_{k-1}\mathbf{u}_{k-1}+\beta_k\mathbf{w} = (\beta_1+\beta_k\alpha_1)\mathbf{u}_1+\cdots+(\beta_{k-1}+\beta_k\alpha_{k-1})\mathbf{u}_{k-1}+\beta_k\alpha_k\mathbf{u}_k.$$

Assim, como as operações elementares usadas pelo método de eliminação de Gauss consistem em troca de linhas e substituição de linhas por combinações lineares de linhas, o método de eliminação de Gauss não altera o espaço das linhas de uma matriz. Ou seja, se R é uma matriz obtida de A por aplicação do método de eliminação de Gauss, tem-se $EL(A) = EL(R)$. Em particular, se R é uma matriz em escada obtida de A por eliminação de Gauss, então um conjunto gerador do espaço das linhas de A é constituído pelas linhas não nulas de R.

De facto, o método de eliminação de Gauss "elimina" linhas que sejam combinações lineares de linhas que a precedem. Esta propriedade de podermos retirar de um conjunto gerador de um subespaço um vector que seja combinação linear dos outros vectores sem alterar o espaço gerado pelo conjunto dado, é completamente geral. Verifiquemos esta afirmação.

Seja $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k, \mathbf{v}\}$ um conjunto gerador do subespaço V de $\mathbb{R}^n$ tal que o vector $\mathbf{v}$ é uma combinação linear dos vectores $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ (distintos de $\mathbf{v}$). Isto é, tal que $\mathbf{v} = \beta_1\mathbf{u}_1 + \cdots + \beta_k\mathbf{u}_k$, para certos escalares $\beta_1, \ldots, \beta_k$. Mostremos que o conjunto $S' = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ ainda gera V, ou seja, $V = \operatorname{Span} S = \operatorname{Span} S'$. Considere-se um qualquer vector $\mathbf{x}$ de V, isto é,

$$\mathbf{x} = \alpha_1\mathbf{u}_1 + \ldots + \alpha_k\mathbf{u}_k + \alpha_{k+1}\mathbf{v}. \tag{3.5}$$

Como $\mathbf{v} = \beta_1\mathbf{u}_1 + \cdots + \beta_k\mathbf{u}_k$, de (3.5) obtemos

$$\begin{aligned}\mathbf{x} &= \alpha_1\mathbf{u}_1 + \cdots + \alpha_k\mathbf{u}_k + \alpha_{k+1}(\beta_1\mathbf{u}_1 + \cdots + \beta_k\mathbf{u}_k)\\ &= (\alpha_1 + \alpha_{k+1}\beta_1)\mathbf{u}_1 + \cdots + (\alpha_k + \alpha_{k+1}\beta_k)\mathbf{u}_k \in \operatorname{Span} S'.\end{aligned}$$

Como $\mathbf{x}$ é um qualquer vector de V, tem-se $\operatorname{Span} S = \operatorname{Span} S' = V$.

O método de eliminação de Gauss também não altera o núcleo de uma matriz uma vez que, se R é uma matriz em escada obtida de A por aplicação do método de eliminação de Gauss, os sistemas $A\mathbf{x} = \mathbf{0}$ e $R\mathbf{x} = \mathbf{0}$ são equivalentes (isto é, têm as mesmas soluções). Podemos assim enunciar a seguinte proposição.

Proposição 3.4. O método de eliminação de Gauss não altera o espaço das linhas nem o núcleo de uma matriz. Isto é, se R é uma matriz em escada obtida de A por eliminação de Gauss, então

$$N(A) = N(R) \qquad \text{e} \qquad EL(A) = EL(R) = \text{Span}\{\mathbf{l}_1, \ldots, \mathbf{l}_k\},$$

onde $\mathbf{l}_1, \ldots, \mathbf{l}_k$ são as linhas não nulas (as linhas que contêm pivôs) de R.

Utilizemos a Proposição 3.4 para determinar um conjunto gerador para o espaço das linhas e das colunas da matriz A do Exemplo 3.6. Aplicando o método de eliminação de Gauss, obtemos

$$A = \begin{bmatrix} 2 & 4 & 6 \\ -1 & -2 & -3 \end{bmatrix} \xrightarrow{L_2 + \frac{1}{2}L_1} \begin{bmatrix} 2 & 4 & 6 \\ 0 & 0 & 0 \end{bmatrix} = R.$$

Logo

$$EL(A) = \text{Span}\{(2, 4, 6)\} = EL(R).$$

Para determinarmos um conjunto gerador do espaço das colunas de A usamos o método de eliminação de Gauss para a matriz transposta de A, uma vez que as colunas de A são as linhas de A^T. De facto, aplicando eliminação de Gauss a A^T, obtemos

$$A^T = \begin{bmatrix} 2 & -1 \\ 4 & -2 \\ 6 & -3 \end{bmatrix} \underset{L_3 - 3L_1}{\xrightarrow{L_2 - 2L_1}} \begin{bmatrix} 2 & -1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Assim,

$$EC(A) = EL(A^T) = \text{Span}\{(2, -1)\}.$$

Neste exemplo conclui-se que:

1) As linhas de R que contêm pivôs geram $EL(A)$, embora as colunas de R que contêm pivôs não gerem $EC(A)$. De facto, a coluna $(2, 0)$ de R (que contém o pivô) gera o eixo dos xx enquanto que $EC(A) = \text{Span}\{(2, -1)\}$ é uma recta que passa pela origem e tem a direcção do vector $(2, -1)$. No entanto, podemos observar que a coluna de A correspondente à coluna de R que contém o pivô, isto é, o vector$(2, -1)$, gera $EC(A)$.

2) O menor número de vectores necessários para gerar $EL(A)$ é igual à característica da matriz A. Além disso, o menor número de vectores necessários para gerar $EC(A)$ é também igual à característica de A.

Certas conclusões deste exemplo são válidas para qualquer matriz como veremos na próxima secção.

Passamos a chamar *dimensão do espaço das linhas* de uma matriz A à característica da matriz. A dimensão do espaço das linhas de A é designada por $\dim EL(A)$.

A igualdade (3.3) pode assim ser reescrita na forma:

Se A é uma matriz $p \times n$, então

$$\dim EL(A) + \dim N(A) = n = \text{ número colunas de } A.$$

3.3 Independência linear, bases e dimensão

As noções a abordar nesta secção são cruciais na teoria dos espaços lineares. No sentido de facilitar a sua compreensão recorremos frequentemente aos espaços lineares $\mathbb{R}^2$ e $\mathbb{R}^3$ para obter uma interpretação geométrica.

No Capítulo 1 vimos que para gerar $\mathbb{R}^2$ são precisos no mínimo dois vectores, porém não servem dois quaisquer vectores já que, por exemplo, dois vectores colineares não geram $\mathbb{R}^2$ mas sim uma recta. Da mesma forma, para gerar $\mathbb{R}^3$ são necessários 3 vectores que não sejam coplanares. Em ambos os casos é necessário escolher vectores que designamos por vectores linearmente independentes. A questão que se coloca é a de saber qual é o menor número de vectores necessários para gerar um dado espaço linear. A este número chama-se dimensão do espaço linear.

Como veremos, fixada uma base num espaço linear real (resp. complexo) de dimensão finita n, estabelece-se uma correspondência biunívoca entre vectores desse espaço e vectores de $\mathbb{R}^n$ (resp. de $\mathbb{C}^n$). Esta correspondência biunívoca desempenha um papel fundamental no estudo dos espaços lineares de dimensão finita, como o leitor poderá reconhecer ao longo deste texto.

Subjacente às noções de independência linear, base e dimensão encontra-se a noção de combinação linear já introduzida no Capítulo 1 para vectores do espaço linear $\mathbb{R}^n$. Estendemos agora as definições de combinação linear e de conjunto gerador a um espaço linear qualquer.

Definição 3.5. Seja W um espaço linear sobre o corpo $\mathbb{K}$ e $S = \{v_1, v_2, \ldots, v_p\} \subseteq W$.

i) Chama-se *combinação linear* dos vectores $v_1, v_2, \ldots, v_p$ a um vector da forma
$$\alpha_1 v_1 + \cdots + \alpha_p v_p,$$
onde $\alpha_1, \ldots, \alpha_p$ são escalares de $\mathbb{K}$. Aos escalares α_i chamamos *coeficientes da combinação* linear.

ii) O conjunto S é *linearmente independente* se a única solução da equação
$$\alpha_1 v_1 + \cdots + \alpha_p v_p = 0 \tag{3.6}$$
é $\alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$. Quando S não é linearmente independente diz-se que é *linearmente dependente*.

iii) Ao conjunto de todas as combinações lineares de elementos de S chamamos *espaço gerado* por, ou *expansão linear de*, S, que designamos por $\mathrm{Span}(S)$. Ou seja,
$$\mathrm{Span}\, S = \{\alpha_1 v_1 + \cdots + \alpha_p v_p : \ \alpha_1, \alpha_2, \ldots, \alpha_p \in \mathbb{K}\}.$$

Quando um conjunto de vectores é linearmente independente (resp. linearmente dependente) também nos referimos aos vectores do conjunto como vectores linearmente independentes (resp. linearmente dependentes).

Nota 17. *1. Faz-se notar que se S é um subconjunto de um espaço linear W, então $\mathrm{Span}(S)$ é um subespaço linear de W. A demonstração desta afirmação é exactamente igual à prova apresentada no Exemplo 3.4 para o caso de subconjuntos de $\mathbb{R}^n$.*

2. Saliente-se que no membro direito da equação (3.6)*, o vector 0 é o vector zero do espaço linear em causa.*

3. Neste livro só consideramos espaços lineares de dimensão finita e por isso apenas apresentámos a definição de independência linear para um conjunto finito. Um conjunto infinito de vectores diz-se linearmente dependente se algum seu subconjunto finito é linearmente dependente, caso contrário diz-se linearmente independente. A maioria dos resultados que apresentamos para espaços lineares de dimensão finita são igualmente válidos para espaços lineares de dimensão infinita, com alterações menores.

Da Proposição 3.1 sabemos que num espaço linear W sobre $\mathbb{K}$ se tem $\alpha \cdot 0 = 0$ para todo o escalar α, e que $\alpha \cdot v \neq 0$ para $\alpha \neq 0 \in \mathbb{K}$ e $v \neq 0 \in W$. Por conseguinte, da definição de independência linear segue:

(i) Um conjunto que contenha o vector zero é linearmente dependente.

(ii) Um conjunto com um único vector não nulo é linearmente independente.

Consideremos alguns exemplos dos conceitos acabados de definir.

Exemplo 3.7. **(a)** Os vectores $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ e $\mathbf{u}_2 = \begin{bmatrix} 5 \\ 10 \end{bmatrix}$ são linearmente dependentes, visto que $\mathbf{u}_2 = 5\mathbf{u}_1$, ou seja

$$5\mathbf{u}_1 - \mathbf{u}_2 = \mathbf{0} \Longleftrightarrow 5 \begin{bmatrix} 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 5 \\ 10 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Assim, o vector $\mathbf{0} \in \mathbb{R}^2$ escreve-se como uma combinação linear de $\mathbf{u}_1$ e $\mathbf{u}_2$ com coeficientes não nulos (os coeficientes são 5 e -1).

Realcemos que dois vectores são linearmente dependentes se e só se um vector é múltiplo do outro (isto é, são colineares).

(b) Os vectores

$$\mathbf{u}_1 = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}$$

são linearmente dependentes já que, $\mathbf{u}_3 = \mathbf{u}_1 + \mathbf{u}_2$, ou seja

$$\mathbf{u}_1 + \mathbf{u}_2 - \mathbf{u}_3 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

A igualdade anterior diz-nos que $\mathbf{0} \in \mathbb{R}^3$ é uma combinação linear de $\mathbf{u}_1, \mathbf{u}_2$ e $\mathbf{u}_3$, com coeficientes $1, 1$ e -1. ♦

Exemplo 3.8. Como vimos, o conjunto das matrizes reais 2×2 com as operações usuais de adição de matrizes e multiplicação de matrizes por escalares reais, é um espaço linear real. O vector zero deste espaço é a matriz nula 2×2. Verifiquemos se o subconjunto $\{A_1, A_2, A_3\}$ deste espaço linear é ou não linearmente independente, onde

$$A_1 = \begin{bmatrix} 1 & 2 \\ 5 & 1 \end{bmatrix}, \qquad A_2 = \begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix}, \qquad A_3 = \begin{bmatrix} 3 & 2 \\ 9 & -1 \end{bmatrix}.$$

Usando a definição de vectores linearmente independentes, tem-se

$$\alpha_1 A_1 + \alpha_2 A_2 + \alpha_3 A_3 = 0 \Longleftrightarrow \begin{bmatrix} \alpha_1 & 2\alpha_1 \\ 5\alpha_1 & \alpha_1 \end{bmatrix} + \begin{bmatrix} \alpha_2 & 0 \\ 2\alpha_2 & -\alpha_2 \end{bmatrix} + \begin{bmatrix} 3\alpha_3 & 2\alpha_3 \\ 9\alpha_3 & -\alpha_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\Longleftrightarrow \begin{bmatrix} \alpha_1 + \alpha_2 + 3\alpha_3 & 2\alpha_1 + 2\alpha_3 \\ 5\alpha_1 + 2\alpha_2 + 9\alpha_3 & \alpha_1 - \alpha_2 - \alpha_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\Longleftrightarrow \begin{cases} \alpha_1 + \alpha_2 + 3\alpha_3 = 0 \\ 2\alpha_1 + 2\alpha_3 = 0 \\ 5\alpha_1 + 2\alpha_2 + 9\alpha_3 = 0 \\ \alpha_1 - \alpha_2 - \alpha_3 = 0. \end{cases}$$

Este sistema homogéneo tem soluções não nulas, nomeadamente a solução $(\alpha_1, \alpha_2, \alpha_3) = (1, 2, -1)$, pelo que o conjunto dado é linearmente dependente. De facto, $A_1 + 2A_2 - A_3$ é igual à matriz nula.

♦

Um subconjunto $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ de $\mathbb{R}^n$ é linearmente independente se a equação

$$\underbrace{\alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \cdots + \alpha_p \mathbf{u}_p}_{\text{combinação linear de } \mathbf{u}_1, \ldots, \mathbf{u}_p} = \mathbf{0},$$

só admite a solução $\alpha_1 = \alpha_2 = \cdots = \alpha_p = 0$. Como o produto de uma matriz A por um vector $\mathbf{x}$ é uma combinação linear das colunas de A (ver Definição 1.11, pág. 34) a equação anterior é equivalente a

$$\alpha_1 \mathbf{u}_1 + \cdots + \alpha_p \mathbf{u}_p = \mathbf{0} \Longleftrightarrow A\mathbf{x} = \mathbf{0},$$

onde A é a matriz cujas colunas são os vectores $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p$ e $\mathbf{x} = (\alpha_1, \alpha_2, \ldots, \alpha_p)$. Assim, o subconjunto $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ de $\mathbb{R}^n$ é linearmente independente se e só se a única solução do sistema homogéneo $A\mathbf{x} = \mathbf{0}$ é $\mathbf{x} = \mathbf{0}$. Podemos então enunciar o seguinte resultado.

Proposição 3.5. Os vectores coluna de uma matriz A são linearmente independentes se e só se a única solução do sistema homogéneo $A\mathbf{x} = \mathbf{0}$ é a solução nula. Equivalentemente, se e só se $N(A) = \{\mathbf{0}\}$.

Exemplo 3.9. 1. Se A é uma matriz quadrada do tipo $n \times n$, então os vectores coluna de A formam um conjunto linearmente independente se e só se a característica de A é igual a n, uma vez que isto é equivalente ao sistema homogéneo só ter a solução nula (ver proposições 1.10 e 1.12 nas páginas 59 e 62).

2. Se A é uma matriz com mais colunas do que linhas, o sistema $A\mathbf{x} = \mathbf{0}$ tem mais incógnitas que equações. Neste caso os vectores coluna de A formam um conjunto linearmente dependente uma vez que o sistema $A\mathbf{x} = \mathbf{0}$ é indeterminado (a característica de A é no máximo igual ao número de equações do sistema, portanto existem incógnitas livres). Assim, qualquer subconjunto de $\mathbb{R}^n$ com cardinalidade superior a n é linearmente dependente, visto que se colocarmos os vectores deste conjunto numa matriz, por colunas, a matriz terá mais colunas que linhas.

♦

Sublinhe-se que enquanto a independência linear de vectores de $\mathbb{R}^n$ se traduz no estudo de um sistema homogéneo de equações lineares, noutros espaços lineares a equação (3.6) não é necessariamente um sistema. Ou seja, enquanto que para estudarmos a independência linear de vectores de $\mathbb{R}^n$ podemos considerar a matriz A cujas colunas são estes vectores e estudar as soluções do sistema $A\mathbf{x} = \mathbf{0}$, o mesmo não acontece noutros espaços lineares onde teremos de recorrer à definição (ver Exemplo 3.8) ou a resultados que obteremos a seguir.

Proposição 3.6. Um conjunto $S = \{v_1, v_2, \ldots, v_p\}$ com $p > 1$ vectores distintos, é linearmente dependente se e só se pelo menos um dos vectores de S é combinação linear dos outros $(p-1)$ vectores de S.

Demonstração.($\Longrightarrow$): O conjunto $S = \{v_1, v_2, \ldots, v_p\}$ é linearmente dependente se e só se a equação $\alpha_1 v_1 + \cdots + \alpha_p v_p = 0$ é satisfeita com pelo menos um dos α's não nulo. Suponhamos que $\alpha_i \neq 0$. São válidas as equivalências:

$$\begin{aligned}
&\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_p v_p = 0 \\
&\quad \Longleftrightarrow \alpha_i v_i = -\alpha_1 v_1 - \cdots - \alpha_{i-1} v_{i-1} - \alpha_{i+1} v_{i+1} - \cdots - \alpha_p v_p \\
&\quad \Longleftrightarrow v_i = \underbrace{-\frac{\alpha_1}{\alpha_i} v_1 - \cdots - \frac{\alpha_{i-1}}{\alpha_i} v_{i-1} - \frac{\alpha_{i+1}}{\alpha_i} v_{i+1} - \cdots - \frac{\alpha_p}{\alpha_i} v_p}_{\text{combinação linear de } v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_p}.
\end{aligned}$$

Isto é, v_i é uma combinação linear de vectores de S, distintos de v_i.

($\Longleftarrow$): Se um dos vectores de S, digamos v_i, é uma combinação linear dos outros vectores de S, então tem-se

$$v_i = \alpha_1 v_1 + \cdots + \alpha_{i-1}v_{i-1} + \alpha_{i+1}v_{i+1} + \cdots + \alpha_p v_p \Longleftrightarrow$$
$$\alpha_1 v_1 + \cdots - v_i + \cdots + \alpha_p v_p = 0.$$

Os coeficientes da combinação linear da equação anterior não são todos nulos (o coeficiente de v_i é igual a (-1)), e portanto o conjunto S é linearmente dependente.

$\square$

A Proposição 3.6 garante que um conjunto é linearmente independente se nenhum dos vectores do conjunto é combinação linear de outros vectores do conjunto.

Exemplo 3.10. As Figuras 3.2 ilustram exemplos de conjuntos linearmente independentes e dependentes.

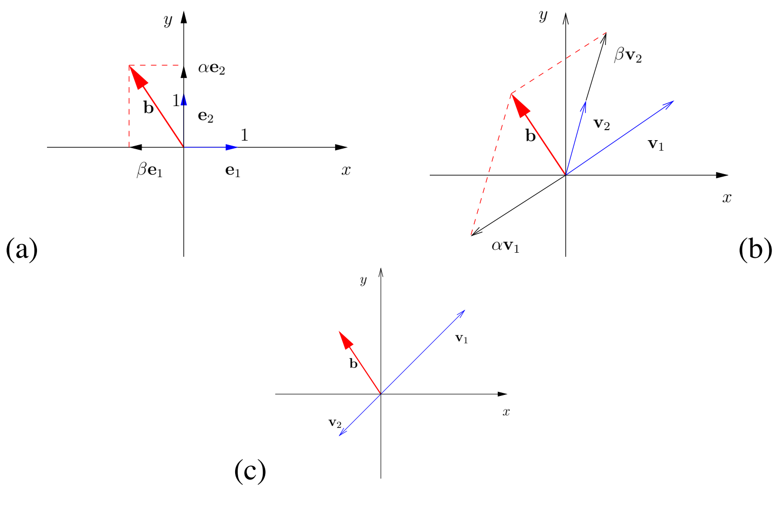


Figura 3.2: (a) Qualquer vector $\mathbf{b} \in \mathbb{R}^2$ é combinação linear de $\mathbf{e}_1 = (1,0)$ e $\mathbf{e}_2 = (0,1)$; (b) O vector $\mathbf{b}$ é combinação linear de $\mathbf{v}_1$ e $\mathbf{v}_2$; (c) $\mathbf{b}$ não é combinação linear de $\mathbf{v}_1$ e $\mathbf{v}_2$.

Fig. 3.2-(a): Conjuntos linearmente independentes: $\{\mathbf{e}_1\}$, $\{\mathbf{e}_2\}$, $\{\beta\mathbf{e}_1\}$, $\{\alpha\mathbf{e}_2\}$, $\{\mathbf{b}\}$, $\{\mathbf{e}_1,\mathbf{e}_2\}$, $\{\mathbf{e}_1,\alpha\mathbf{e}_2\}$, $\{\beta\mathbf{e}_1,\mathbf{e}_2\}$, $\{\mathbf{e}_1,\mathbf{b}\}$, $\{\mathbf{e}_2,\mathbf{b}\}$, $\{\beta\mathbf{e}_1,\mathbf{b}\}$, $\{\alpha\mathbf{e}_2,\mathbf{b}\}$, $\{\alpha\mathbf{e}_2,\beta\mathbf{e}_1\}$, etc.

Conjunto linearmente dependente: $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{b}\}$.

Fig. 3.2-(b): Conjuntos linearmente independentes: $\{\mathbf{v}_1\}$, $\{\mathbf{v}_2\}$, $\{\mathbf{b}\}$, $\{\mathbf{v}_1, \mathbf{v}_2\}$, $\{\mathbf{v}_1, \mathbf{b}\}$ e $\{\mathbf{v}_2, \mathbf{b}\}$.

Conjunto linearmente dependente: $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{b}\}$.

Fig. 3.2-(c): Conjuntos linearmente independentes: $\{\mathbf{v}_1\}$, $\{\mathbf{v}_2\}$, $\{\mathbf{v}_1, \mathbf{b}\}$ e $\{\mathbf{v}_2, \mathbf{b}\}$.

Conjunto linearmente dependente: $\{\mathbf{v}_1, \mathbf{v}_2\}$.

♦

Vimos, na página 122, que ao removermos de um subconjunto S de $\mathbb{R}^n$ vectores que sejam combinações lineares de outros vectores do conjunto, o conjunto resultante ainda gera o mesmo espaço linear que o conjunto S. Este resultado é válido em qualquer espaço linear. Ou seja, removendo de um dado conjunto linearmente dependente, vectores que sejam combinações lineares de outros vectores do conjunto, o conjunto resultante e o conjunto dado geram o mesmo subespaço linear, conforme se enuncia na proposição a seguir.

Proposição 3.7. Seja $S = \{v_1, \ldots, v_p\}$ um subconjunto linearmente dependente de um espaço linear W e v_k um vector de S que é combinação linear de vectores de S, distintos de v_k.

O conjunto S e o conjunto S' obtido de S removendo v_k, isto é, $S' = S \backslash \{v_k\}$, geram o mesmo subespaço linear. Isto é,

$$\operatorname{Span} S' = \operatorname{Span} S.$$

Demonstração. A demonstração é inteiramente análoga à prova apresentada na página 122 para subconjuntos de $\mathbb{R}^n$. □

Exemplo 3.11. Sejam v_1, v_2, v_3 e v_4 vectores não nulos de um espaço linear e $V = \operatorname{Span}\{v_1, v_2, v_3, v_4\}$. Suponhamos que os vectores v_i verificam as relações:

(a) $v_1 + 2v_2 = 0$ (b) $v_1 - v_2 + v_3 = 0$ (c) $v_4 \in \operatorname{Span}\{v_1, v_2, v_3\}$.

Determinemos um subconjunto de $\{v_1, v_2, v_3, v_4\}$, com o menor número possível de vectores que gere V.

Da última proposição sabemos que se removermos de um conjunto um vector que seja combinação linear de outros vectores do conjunto, o novo conjunto ainda gera o mesmo espaço linear. Assim, da alínea (c) sabemos que v_4 é uma combinação linear de v_1, v_2, v_3, logo $V = \mathrm{Span}\{v_1, v_2, v_3, v_4\} = \mathrm{Span}\{v_1, v_2, v_3\}$. De igual modo, a alínea (b) diz-nos que $v_3 = v_2 - v_1$, ou seja, v_3 é uma combinação linear de v_2 e v_1, e portanto $V = \mathrm{Span}\{v_1, v_2, v_3\} = \mathrm{Span}\{v_1, v_2\}$. Finalmente, por (a), o vector v_2 é combinação linear de v_1 e portanto

$$V = \mathrm{Span}\{v_1, v_2\} = \mathrm{Span}\{v_1\}.$$

Claramente o subespaço V não pode ser gerado por menos do que um vector. ♦

Apresentamos a seguir a noção de base.

Definição 3.6. Base
Um subconjunto (finito) B de um espaço linear $W \neq \{0\}$ diz-se uma *base* de W se são verificadas as duas condições seguintes.

1. B é linearmente independente.

2. B gera W, isto é, $\mathrm{Span}(B) = W$.

Nota 18. *Uma base nunca pode conter o vector zero do espaço já que, qualquer conjunto que contenha este vector é linearmente dependente (ver quadro na página 126).*

Exemplo 3.12. Para os vectores representados na Figura 3.2-(a)-(c) (pág. 129) podemos formar as seguintes bases de $\mathbb{R}^2$.

(a) $\{\mathbf{e}_1, \mathbf{e}_2\}, \{\mathbf{e}_1, \alpha\mathbf{e}_2\}, \{\beta\mathbf{e}_1, \mathbf{e}_2\}, \{\mathbf{e}_1, \mathbf{b}\}, \{\mathbf{e}_2, \mathbf{b}\}, \{\beta\mathbf{e}_1, \mathbf{b}\}, \{\alpha\mathbf{e}_2, \mathbf{b}\}, \{\alpha\mathbf{e}_2, \beta\mathbf{e}_1\}$, etc.

(b) $\{\mathbf{v}_1, \mathbf{v}_2\}, \{\mathbf{v}_1, \mathbf{b}\}$ e $\{\mathbf{v}_2, \mathbf{b}\}$.

(c) $\{\mathbf{v}_1, \mathbf{b}\}$ e $\{\mathbf{v}_2, \mathbf{b}\}$.

♦

Exemplo 3.13. Consideremos os vectores de $\mathbb{R}^3$ representados na Figura 3.3- (a).

Os três vectores $\mathbf{e}_1 = (1, 0, 0)$, $\mathbf{e}_2 = (0, 1, 0)$ e $\mathbf{e}_3 = (0, 0, 1)$ formam o que se designa por *base canónica* de $\mathbb{R}^3$. De facto, qualquer vector de $\mathbb{R}^3$ é uma combinação linear destes três vectores, já que

$$\mathbf{x} = (x_1, x_2, x_3) = x_1(1, 0, 0) + x_2(0, 1, 0) + x_3(0, 0, 1) = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3.$$

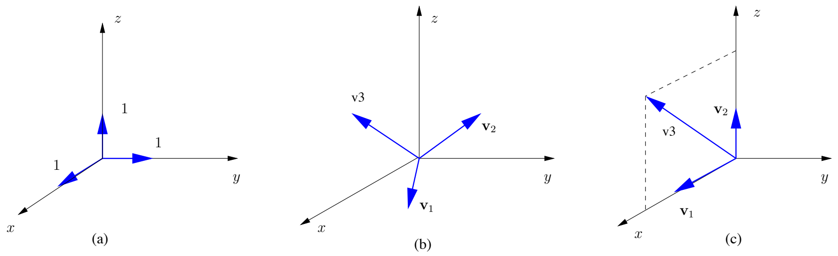


Figura 3.3: (a) Base canónica de $\mathbb{R}^3$: $\{(1,0,0),(0,1,0),(0,0,1)\}$; (b) $\{\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3\}$ é uma base de $\mathbb{R}^3$; (c) $\{\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3\}$ não é uma base de $\mathbb{R}^3$.

Além disso, $\{\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3\}$ é linearmente independente, uma vez que

$$\begin{aligned}\alpha\mathbf{e}_1+\beta\mathbf{e}_2+\gamma\mathbf{e}_3=\mathbf{0} &\iff \alpha(1,0,0)+\beta(0,1,0)+\gamma(0,0,1)=(0,0,0)\\ &\iff (\alpha,\beta,\gamma)=(0,0,0).\end{aligned}$$

Por conseguinte, o conjunto $\{\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3\}$ é linearmente independente e gera $\mathbb{R}^3$.

Na Figura 3.3-(b) os vectores $\mathbf{v}_1,\mathbf{v}_2$ e $\mathbf{v}_3$ formam uma base de $\mathbb{R}^3$. Na Figura 3.3-(c) os vectores $\mathbf{v}_1,\mathbf{v}_2$ e $\mathbf{v}_3$ não formam uma base de $\mathbb{R}^3$ já que estes três vectores geram o plano xz e portanto não geram $\mathbb{R}^3$, tem-se

$$\operatorname{Span}\{\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3\}=\operatorname{Span}\{\mathbf{v}_1,\mathbf{v}_2\}.$$

♦

Exemplo 3.14.

1. O conjunto $B=\{(1,2),(0,1),(2,5)\}$ não é uma base de $\mathbb{R}^2$, apesar de $\operatorname{Span}(B)=\mathbb{R}^2$. De facto, B não é linearmente independente já que $(2,5)=2(1,2)+(0,1)$.

2. O conjunto $B=\{(1,2)\}$ é linearmente independente (ver quadro na pág. 126) e não é uma base de $\mathbb{R}^2$, já que $\operatorname{Span}(B)\neq\mathbb{R}^2$. De facto, $\operatorname{Span}(B)$ é uma recta que passa pela origem e tem a direcção do vector $(1,2)$.

3. O conjunto $B=\{(1,2),(1,1)\}$ gera $\mathbb{R}^2$ uma vez que os vectores de B não são colineares. Além disso, B é linearmente independente já que o sistema

$$\alpha_1\begin{bmatrix}1\\2\end{bmatrix}+\alpha_2\begin{bmatrix}1\\1\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix}\iff\begin{bmatrix}1&1\\2&1\end{bmatrix}\begin{bmatrix}\alpha_1\\\alpha_2\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix},$$

só admite a solução trivial $(\alpha_1, \alpha_2) = (0, 0)$, visto que $\begin{vmatrix} 1 & 1 \\ 2 & 1 \end{vmatrix} = -1$. Por conseguinte, B é uma base de $\mathbb{R}^2$.

♦

Exemplo 3.15. Qualquer vector $(x_1, x_2, \ldots, x_n)$ de $\mathbb{R}^n$ é combinação linear dos vectores do conjunto

$$BC = \{(1, 0, \ldots, 0), (0, 1, \ldots, 0), \ldots, (0, 0, \ldots, 1)\} \subset \mathbb{R}^n,$$

uma vez que

$$(x_1, x_2, \ldots, x_n) = x_1(1, 0, \ldots, 0) + x_2(0, 1, \ldots, 0) + \cdots + x_n(0, 0, \ldots, 1).$$

Os vectores de BC são linearmente independentes já que,

$$\begin{aligned} \mathbf{0} = \alpha_1(1, 0, \ldots, 0) + \cdots + \alpha_n(0, 0, \ldots, 1) &\iff (0, \ldots, 0) = (\alpha_1, \ldots, \alpha_n) \\ &\iff \alpha_1 = \cdots = \alpha_n = 0. \end{aligned}$$

Logo, o conjunto BC é uma base de $\mathbb{R}^n$. A base BC é designada por *base canónica* de $\mathbb{R}^n$. ♦

Como consequência da Proposição 3.7 podemos obter o resultado que se segue.

Proposição 3.8. Seja $W \neq \{0\}$ um espaço linear e S um subconjunto (finito) de W que gera W. Existe um subconjunto de S que é uma base de W.

Demonstração. Se S é linearmente independente, então S já é uma base de W. Se S não é linearmente independente, pela Proposição 3.7, podemos ir removendo vectores de S por forma a obter um subconjunto $S' \subset S$ que seja linearmente independente. Se o conjunto S' tem um único vector ele é necessariamente não nulo pois, por hipótese, $W \neq \{0\}$, e portanto S' é conjunto linearmente independente. Como $W = \operatorname{Span} S = \operatorname{Span} S'$ (cf. Proposição 3.7), então S' é uma base de W. □

O teorema a seguir é um resultado fundamental na teoria dos espaços lineares porquanto, desde que se fixe uma base ordenada no espaço linear, ele permitirá estabelecer uma correspondência biunívoca entre vectores de um espaço linear real (resp. complexo) e vectores de $\mathbb{R}^n$ (resp. $\mathbb{C}^n$).

Teorema 3.1. Seja $B = \{v_1, v_2, \dots, v_n\}$ uma base de um espaço linear W.

Qualquer vector x de W escreve-se, de forma única, como combinação linear dos vectores da base B. Isto é, existem escalares α_i únicos tais que

$$x = \alpha_1 v_1 + \cdots + \alpha_n v_n.$$

Demonstração. Obviamente que qualquer $x \in W$ é uma combinação linear dos vectores de B já que, por definição de base, o conjunto gerado por B é W. Para mostrar que esta combinação linear é única vamos supor, por redução ao absurdo, que existe um vector $x \in W$ que se escreve como combinação linear dos vectores de B de duas formas distintas.

$$x = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n \tag{3.7}$$
$$x = \beta_1 v_1 + \beta_2 v_2 + \cdots + \beta_n v_n \tag{3.8}$$

com $\alpha_i \neq \beta_i$ para algum $i = 1, 2, \dots, n$. Subtraindo (3.7) a (3.8) obtemos:

$$0 = (\alpha_1 - \beta_1)v_1 + (\alpha_2 - \beta_2)v_2 + \cdots + (\alpha_n - \beta_n)v_n.$$

Como $B = \{v_1, v_2, \dots, v_n\}$ é uma base, este conjunto é linearmente independente, e portanto a última igualdade só se verifica se $(\alpha_i - \beta_i) = 0$ para todo $i = 1, 2, \dots, n$. Isto é, quando $\alpha_i = \beta_i$, o que é uma contradição. $\square$

Dada uma base $B = \{v_1, v_2, \dots, v_n\}$ podemos ordenar os seus vectores. Para designar que uma base está *ordenada* usamos parênteses curvos em vez de chavetas. Por exemplo, o conjunto $B = \{(1, 2), (1, 1)\}$ é uma base de $\mathbb{R}^2$ que pode ser ordenado de duas formas distintas. Nomeadamente, podemos considerar as bases ordenadas $B_1 = ((1, 2), (1, 1))$ e $B_2 = ((1, 1), (1, 2))$.

Definição 3.7. Vector das coordenadas

Seja $B = (v_1, v_2, \dots, v_n)$ uma base ordenada de um espaço linear W. As coordenadas de um vector $x \in W$ na base B são os coeficientes $\alpha_1, \alpha_2, \dots, \alpha_n$ da combinação linear

$$x = \alpha_1 v_1 + \alpha_2 v_2 + \cdots \alpha_n v_n.$$

O vector $\mathbf{x}_B = (\alpha_1, \alpha_2, \dots, \alpha_n)$ é designado por *vector das coordenadas* de x na base (ordenada) B.

Nota 19. *1. Se W é um espaço linear real, então o vector das coordenadas $\mathbf{x}_B$, na definição anterior, é um vector de $\mathbb{R}^n$ (mesmo que o espaço linear não seja $\mathbb{R}^n$) pois neste caso os escalares α_i são números reais.*

2. *Quando em $\mathbb{R}^n$ e $\mathbb{C}^n$, não indicamos o índice B no vector das coordenadas supomos que a base B que estamos a considerar é a base canónica do espaço linear em questão. Esta notação está de acordo com a notação usada na definição de $\mathbb{R}^n$ e de $\mathbb{C}^n$. Ver ainda Exemplo 3.15, na página 133.*

3. *O Teorema 3.1 garante a unicidade do vector das coordenadas numa base fixada, para todo o vector x do espaço linear.*

Exemplo 3.16. No Exemplo 3.14-3 (pág. 132) vimos que $B = \{(1, 2), (1, 1)\}$ é uma base de $\mathbb{R}^2$. Esta base pode ser ordenada de duas formas distintas. Sejam $B_1 = ((1, 2), (1, 1))$ e $B_2 = ((1, 1), (1, 2))$. Determinemos as coordenadas de $\mathbf{x} = (10, 4)$ nas bases ordenadas B_1 e B_2. Como

$$(10, 4) = -6(1, 2) + 16(1, 1),$$

temos $\mathbf{x}_{B_1} = (-6, 16)$ e $\mathbf{x}_{B_2} = (16, -6)$. ♦

Exemplo 3.17. Consideremos a base ordenada de $\mathbb{R}^2$, $B = (\mathbf{b}, \mathbf{c}) = ((2, 2), (-2, 2))$ e o vector $\mathbf{v} = (10, 6)$. As coordenadas de $\mathbf{v}$ na base B são $\mathbf{v}_B = (4, -1)$, já que

$$(10, 6) = \alpha(2, 2) + \beta(-2, 2) = (2\alpha - 2\beta, 2\alpha + 2\beta) \iff \begin{cases} 2\alpha - 2\beta = 10 \\ 2\alpha + 2\beta = 6 \end{cases}$$
$$\iff \alpha = 4 \text{ e } \beta = -1.$$

Na Figura 3.4 ilustramos geometricamente este exemplo. ♦

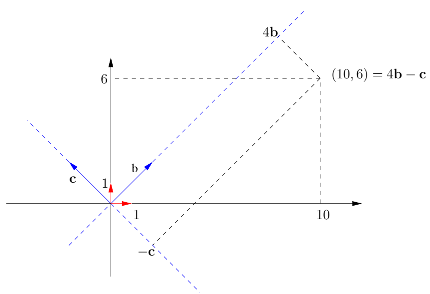


Figura 3.4: O vector das coordenadas de $\mathbf{v} = (10, 6)$ na base $B = (\mathbf{b}, \mathbf{c})$ é $\mathbf{v}_B = (4, -1)$. Isto é, $\mathbf{v} = 4\mathbf{b} - \mathbf{c}$.

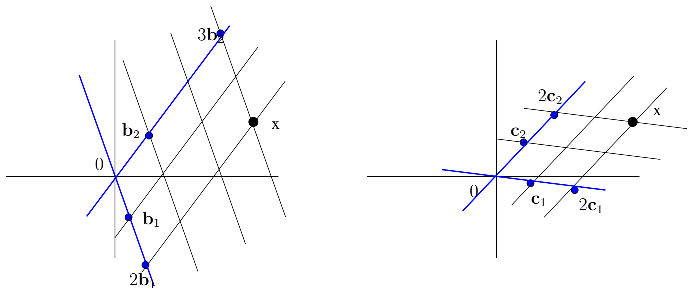


Figura 3.5: Duas bases distintas $B = (\mathbf{b}_1, \mathbf{b}_2)$ e $B' = (\mathbf{c}_1, \mathbf{c}_2)$ do mesmo espaço linear, e as coordenadas do vector $\mathbf{x}$ nessas bases: $\mathbf{x}_B = (2, 3)$ e $\mathbf{x}_{B'} = (2, 2)$.

Exemplo 3.18. Na Figura 3.5 ilustramos a representação do vector $\mathbf{x}$ em duas bases ordenadas distintas, nomeadamente nas bases $B = (\mathbf{b}_1, \mathbf{b}_2)$ e $B' = (\mathbf{c}_1, \mathbf{c}_2)$. O vector das coordenadas de $\mathbf{x}$ nestas bases é

$$\mathbf{x}_B = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \quad \text{e} \quad \mathbf{x}_{B'} = \begin{bmatrix} 2 \\ 2 \end{bmatrix},$$

uma vez que

$$\mathbf{x} = 2\mathbf{b}_1 + 3\mathbf{b}_2 \qquad \text{e} \qquad \mathbf{x} = 2\mathbf{c}_1 + 2\mathbf{c}_2.$$

◆

Exercício 3.2. Considere a base ordenada $B = (u, v)$ de um certo espaço linear. Sabendo que $\mathbf{x}_B = (1, 2)$, determine o vector das coordenadas de x nas bases ordenadas seguintes.

a) $B_1 = (u, 2v)$.
b) $B_2 = (-u, u + v)$.
c) $B_3 = (u - v, u + v)$.
d) $B_4 = (4u, 6v)$.

Solução: $\mathbf{x}_{B_1} = (1, 1)$, $\mathbf{x}_{B_2} = (1, 2)$, $\mathbf{x}_{B_3} = (-1/2, 3/2)$, $\mathbf{x}_{B_4} = (1/4, 1/3)$.

▲

É fácil mostrar que fixada uma base ordenada $B = (v_1, \ldots, v_n)$ num espaço linear W, um subconjunto $\{u_1, \ldots, u_p\}$ de W é linearmente independente se e só se o conjunto dos vectores das coordenadas $\{(\mathbf{u}_1)_B, \ldots, (\mathbf{u}_p)_B\}$ é linearmente independente. Para tal, considerem-se as combinações lineares

$$\begin{aligned} u_1 &= \beta_{11}v_1 + \cdots + \beta_{1n}v_n \\ u_2 &= \beta_{21}v_1 + \cdots + \beta_{2n}v_n \\ &\vdots \\ u_p &= \beta_{p1}v_1 + \cdots + \beta_{pn}v_n \end{aligned} \tag{3.9}$$

Por definição de vector das coordenadas numa certa base, as igualdades anteriores significam que os vectores das coordenadas de $u_1, \ldots, u_p$ na base B são:

$$\begin{aligned} (\mathbf{u}_1)_B &= (\beta_{11}, \ldots, \beta_{1n}) \\ &\vdots \\ (\mathbf{u}_p)_B &= (\beta_{p1}, \ldots, \beta_{pn}). \end{aligned}$$

De (3.9) tem-se que $\alpha_1 u_1 + \cdots + \alpha_p u_p = 0$ é equivalente a

$$(\alpha_1\beta_{11} + \cdots + \alpha_p\beta_{p1})v_1 + \cdots + (\alpha_1\beta_{1n} + \cdots + \alpha_p\beta_{pn})v_n = 0.$$

Como os vectores $v_1, \ldots, v_n$ são linearmente independentes, da igualdade anterior segue que $\alpha_1 u_1 + \cdots + \alpha_p u_p = 0$ se e só se

$$\begin{cases} \alpha_1\beta_{11} + \cdots + \alpha_p\beta_{p1} = 0 \\ \qquad\vdots \\ \alpha_1\beta_{1n} + \cdots + \alpha_p\beta_{pn} = 0 \end{cases} \iff \alpha_1 \begin{bmatrix} \beta_{11} \\ \vdots \\ \beta_{1n} \end{bmatrix} + \cdots + \alpha_p \begin{bmatrix} \beta_{p1} \\ \vdots \\ \beta_{pn} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\iff \alpha_1(\mathbf{u}_1)_B + \cdots + \alpha_p(\mathbf{u}_p)_B = \mathbf{0}.$$

Por conseguinte, $\{u_1, \ldots, u_p\}$ é linearmente independente se e só se $\{(\mathbf{u}_1)_B, \ldots, \mathbf{u}_p)_B\}$ é linearmente independente. Para referência futura enunciamos a seguir o que acabámos de mostrar.

Proposição 3.9. Fixada uma base ordenada B no espaço linear W, um subconjunto $\{u_1, \ldots, u_p\}$ de W é linearmente independente se e só se o conjunto dos vectores das coordenadas $\{(\mathbf{u}_1)_B, \ldots, (\mathbf{u}_p)_B\}$ é linearmente independente.

O teorema que enunciamos a seguir mostra que todas as bases de um espaço linear têm o mesmo cardinal.

Teorema 3.2. Todas as bases de um espaço linear W têm o mesmo número de elementos.

Demonstração. Vamos usar um raciocínio por contradição. Sejam $B = \{v_1, v_2, \ldots, v_n\}$ e $B' = \{u_1, u_2, \ldots, u_p\}$ duas bases de W, e assuma-se que $p > n$.

Consideremos os vectores das coordenadas $(\mathbf{u}_1)_B, (\mathbf{u}_2)_B, \ldots, (\mathbf{u}_p)_B$. Colocando estes vectores numa matriz por colunas obtém-se a seguinte matriz A do tipo $n \times p$.

$$A = \begin{bmatrix} | & | & & | \\ (\mathbf{u}_1)_B & (\mathbf{u}_2)_B & \cdots & (\mathbf{u}_p)_B \\ | & | & & | \end{bmatrix}.$$

Como $p > n$, o sistema homogéneo $A\mathbf{x} = \mathbf{0}$ é indeterminado (tem mais incógnitas que equações) e portanto os vectores coluna de A são linearmente dependentes. Assim, da Proposição 3.9 tem-se que os vectores $u_1, \ldots, u_p$ são linearmente dependentes, e portanto B' não é uma base, o que é uma contradição.

Para o caso $p < n$ a prova é análoga bastando para tal considerar os vectores das coordenadas dos vectores de B na base B'. □

A dimensão de um espaço linear define-se como o número de elementos de uma (qualquer) base do espaço.

Definição 3.8. Dimensão

Chama-se dimensão de um espaço linear W ao número de vectores de uma (qualquer) base do espaço. Designa-se a dimensão de W por $\dim W$.

Se o espaço linear W apenas possui o vector zero, isto é, se $W = \{0\}$, convenciona-se que a sua dimensão é igual a zero.

Decorre do Teorema 3.1 que, uma vez fixada uma base ordenada B num espaço linear real (resp. complexo) W de dimensão n, existe uma correspondência biunívoca entre W e $\mathbb{R}^n$ (resp. $\mathbb{C}^n$). Esta correspondência associa a cada vector do espaço linear o seu vector das coordenadas na base fixada. Esquematicamente,

$$\begin{array}{ccc} (W, B) \text{ espaço linear real de dimensão } n & \longleftrightarrow & \mathbb{R}^n \\ & & \\ x \in W & \longleftrightarrow & \mathbf{x}_B \in \mathbb{R}^n. \end{array} \tag{3.10}$$

Nota 20. *1. Refira-se de novo que neste livro apenas se tratam espaços lineares de dimensão finita.*

Com base nos resultados obtidos, enunciamos a seguir um teorema de grande utilidade prática. Este teorema aplica-se com frequência quando se pretende decidir se um dado subconjunto de um espaço linear é ou não uma base desse espaço linear.

Teorema 3.3. Seja W é um espaço linear de dimensão $n \geq 1$, e S é um subconjunto de W com p vectores. São verdadeiras as afirmações:

1. Se $p > n$, então S é linearmente dependente;
2. Se $p < n$, então S não gera W;
3. Se S é linearmente independente e $p < n$, então pode acrescentar-se $(n-p)$ vectores a S por forma a obter uma base de W;
4. Se $p = n$ e S é linearmente independente, então S é uma base de W;
5. Se $p = n$ e S gera W, então S é uma base de W.

Demonstração.

1. Este resultado segue da demonstração do Teorema 3.2.
2. Suponha-se, por absurdo, que S gera W. Se S gera W, pela Proposição 3.8 existe um subconjunto B de S que é uma base de W. Porém a base B tem no máximo $p < n$ vectores. Logo, B não pode ser uma base já que qualquer base de W tem n vectores (o que é uma contradição).
3. Seja $S = \{v_1, v_2, \ldots, v_p\}$ linearmente independente com $p < n$. Pelo item 2, S não gera W, e portanto existe pelo menos um vector u de W que não pertence ao conjunto gerado por S. Ora o conjunto $S' = \{v_1, v_2, \ldots, v_p, u\}$ também é linearmente independente já que, se na igualdade

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_p v_p + \alpha_{p+1} u = 0,$$

α_{p+1} não fosse igual a zero, então $u = \frac{1}{\alpha_{p+1}} (\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_p v_p)$ era uma combinação linear de elementos de S, o que contraria a hipótese de u não pertencer ao conjunto gerado por S.

Se S' não tiver n elementos, então podemos repetir o processo, aumentando S' com outro vector não pertencente a $\operatorname{Span} S'$, e obter um novo conjunto linearmente independente. Quando chegamos a um conjunto com n elementos este conjunto tem necessariamente que gerar W, caso contrário poderíamos acrescentar um novo vector ao conjunto e obter um conjunto com $n + 1$ elementos o qual, pelo item 1, seria linearmente dependente. Como um conjunto linearmente independente que gera W é uma base de W, o conjunto obtido é uma base de W.

4. Como se viu na prova do item anterior, se S é linearmente independente e tem n elementos, então gera W. Logo, S é uma base.

5. Como $W \neq \{0\}$ (já que por hipótese a dimensão de W é pelo menos 1) e S gera W, a Proposição 3.8 garante que existe um subconjunto $B \subseteq S$, que é uma base de W. Como $\dim W = n$, o número de elementos de B é n. Logo, $B = S$.

$\square$

Do teorema anterior conclui-se que um conjunto gerador de um espaço linear de dimensão n não pode ter menos do que n vectores. Também se conclui que a dimensão de um espaço linear W é o número máximo de vectores linearmente independentes existentes num conjunto gerador de W.

Exemplo 3.19. A dimensão de $\mathbb{R}^n$ é n já que, como se viu no Exemplo 3.15, uma base de $\mathbb{R}^n$ é a base canónica

$$BC = \{(1,0,\ldots,0),(0,1,\ldots,0),\ldots,(0,0,\ldots,1)\} = \{\mathbf{e}_1,\mathbf{e}_2,\ldots,\mathbf{e}_n\}.$$

♦

Exemplo 3.20. Considerem-se os vectores não nulos $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ e $\mathbf{v}_4$ de $\mathbb{R}^3$. Do Teorema 3.3 temos:

- $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ é linearmente dependente uma vez que a dimensão de $\mathbb{R}^3$ é 3 (cf. Teorema 3.3-1)).

- $\{\mathbf{v}_1, \mathbf{v}_2\}$ não é uma base de $\mathbb{R}^3$ uma vez que não gera $\mathbb{R}^3$ (cf. Teorema 3.3-2)).

- Se $S = \{\mathbf{v}_3, \mathbf{v}_4\}$ é linearmente independente e $\mathbf{v}_2 \notin \operatorname{Span} S$, então $\{\mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ é uma base de $\mathbb{R}^3$ (cf. Teorema 3.3-3)).

♦

A finalizar esta secção vamos referir uma importante relação entre a dimensão da soma de subespaços e a dimensão da sua intersecção. Recorde-se, do Exercício 3.1 na página 119, que a soma $U + V$, do subespaço U com o subespaço V, é o conjunto de todas as somas de vectores de U com vectores de V. Saliente-se ainda que o espaço $U + V$ não é em geral a união $U \cup V$, mas contém esta união, como se pode depreender analisando a Figura 3.6.

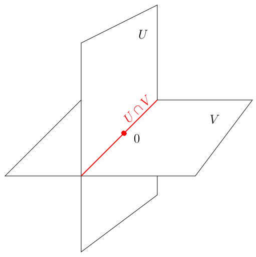


Figura 3.6: A união e a intersecção de subespaços.

Proposição 3.10. Sejam U e V subespaços de um espaço linear W. Verifica-se a igualdade:

$$\dim(U+V) = \dim U + \dim V - \dim(U \cap V). \tag{3.11}$$

Demonstração. Do Exercício 3.1 (pág. 119), sabemos que a intersecção $U \cap V$ é um subespaço de W. Seja $S = \{z_1, z_2, \ldots, z_k\}$ uma base de $U \cap V$. Como $S \subseteq U$ e $S \subseteq V$, o Teorema 3.3 garante que podemos completar S por forma a obter bases para U e para V. Seja $B_U = \{z_1, \ldots, z_k, x_1, \ldots x_p\}$ uma base de U e $B_V = \{z_1, \ldots, z_k, y_1, \ldots y_n\}$ uma base de V. Verifiquemos agora que $B = B_U \cup B_V$ é uma base de $U+V$. É óbvio, a partir da definição de $U+V$, que B é um conjunto gerador para $U+V$, uma vez que vectores $w \in U+V$ são da forma $w = u+v$ com $u \in U$ e $v \in V$. Mostremos agora que B é linearmente independente.

Se

$$\sum_{i=1}^{k} \alpha_i z_i + \sum_{j=1}^{p} \beta_j x_j + \sum_{r=1}^{n} \gamma_r y_r = 0, \tag{3.12}$$

então

$$\sum_{r=1}^{n} \gamma_r y_r = -\left(\sum_{i=1}^{k} \alpha_i z_i + \sum_{j=1}^{p} \beta_j x_j \right) \in U.$$

Como $\sum_{r=1}^{n} \gamma_r y_r \in V$, segue da igualdade anterior que $\sum_{r=1}^{n} \gamma_r y_r \in U \cap V$, e portanto têm de existir escalares δ_i tais que

$$\sum_{r=1}^{n} \gamma_r y_r = \sum_{i=1}^{k} \delta_i z_i \iff \sum_{r=1}^{n} \gamma_r y_r - \sum_{i=1}^{k} \delta_i z_i = 0.$$

Como B_V é linearmente independente, segue da equivalência anterior que todos os γ_i e δ_i são nulos. Logo, a equação (3.12) reduz-se a $\sum_{i=1}^{k} \alpha_i z_i + \sum_{j=1}^{p} \beta_j x_j = 0$. Porém, B_U é linearmente independente e portanto a última equação só tem a solução $\alpha_i = \beta_j = 0$, para todo i, j. Assim, a igualdade (3.12) só é satisfeita quando os α's, γ's e δ's são todos nulos. Isto é, B é linearmente independente. Assim, $B = B_U \cup B_V$ é uma base de $U + V$. Logo,

$$\dim(U+V) = k+p+n = (k+p)+(k+n)-k = \dim U + \dim V - \dim(U \cap V).$$

□

3.3.1 Bases e dimensão dos quatro subespaços fundamentais

Na Secção 3.2.2 chamámos dimensão do espaço das linhas de uma matriz A e do núcleo de A, respectivamente, à característica de A e ao grau de indeterminação do sistema $A\mathbf{x} = \mathbf{0}$. Nesta secção mostramos que essas designações correspondem de facto às dimensões desses espaços lineares. Aproveitamos ainda para mostrar que a dimensão do espaço das colunas é igual à dimensão do espaço das linhas.

Teorema 3.4. Seja R uma matriz em escada obtida da matriz A por eliminação de Gauss, e $\mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_k$ as linhas não nulas de R. São verdadeiras as afirmações seguintes.

(a) Uma base B para o espaço das linhas de A é constituída pelas linhas não nulas de R. Isto é, $B = \{\mathbf{l}_1, \ldots, \mathbf{l}_k\}$ é uma base de $EL(A)$.

(b) Uma base para o espaço das colunas de A é constituída pelas colunas de A correspondentes às colunas de R que contêm pivôs.

(c) $\dim EL(A) = \dim EC(A) = k$.

(d) $\operatorname{car}(A) = \operatorname{car}(A^T) = k$.

Demonstração. (a): Da Proposição 3.4 sabemos que as linhas não nulas de R geram o espaço das linhas de A. Para mostrar que as linhas não nulas de R formam uma base de $EL(A)$ basta provar que são linearmente independentes. Para tal considere-se a matriz P constituída pelas linhas não nulas de R, respeitando a ordem pela qual aparecem em R. Aplicando o método de eliminação de Gauss à matriz transposta de P obtém-se uma matriz em escada com um pivô em todas as colunas. Assim, a única solução do sistema $P^T\mathbf{x} = \mathbf{0}$ é a solução nula e portanto as colunas de P^T (isto é, as linhas não nulas de R) são linearmente independentes (cf. Proposição 3.5, pág. 127).

(b): Aplicando operações elementares à matriz transposta de R é possível anular as linhas de R^T que correspondem às colunas de R sem pivôs. Isto significa que as colunas de uma matriz em escada que não contêm pivôs são combinações lineares das colunas com pivô à sua esquerda. Assim, como a matriz A e R têm a mesma característica, existem no máximo $k = \operatorname{car}(A)$ colunas de A que são linearmente independentes. Seja R' a matriz que se obtém de R suprimindo as colunas sem pivô, e A' a matriz que se obtém de A suprimindo as colunas correspondentes às colunas de R sem pivô. Os sistemas $R'\mathbf{x} = \mathbf{0}$ e $A'\mathbf{x} = \mathbf{0}$ têm as mesmas soluções. Como todas as colunas de R' têm pivô, o sistema $R'\mathbf{x} = \mathbf{0}$ só possui a solução nula. Logo, a única solução de $A'\mathbf{x} = \mathbf{0}$ é a solução nula, e portanto as colunas de A' são linearmente independentes. Conclui-se ainda que

$$\dim EC(A) = \operatorname{car}(A) = k.$$

(c): A igualdade é consequência imediata de (a) e de (b) e do facto da característica de A ser igual a k.

(d): É consequência imediata de (c), visto que

$$\operatorname{car}(A) = k = \dim EC(A) = \dim EL(A) = \dim EC(A^T) = \operatorname{car}(A^T).$$

□

Verificamos de seguida que a dimensão do núcleo de uma matriz A é igual ao grau de indeterminação do sistema $A\mathbf{x} = \mathbf{0}$.

Teorema 3.5. A dimensão do núcleo de uma matriz A é igual ao grau de indeterminação do sistema $A\mathbf{x} = \mathbf{0}$.

Demonstração. Seja A uma matriz $p \times n$. O núcleo de A é o conjunto das soluções de $A\mathbf{x} = \mathbf{0}$. Suponha-se que a característica de A é igual a k, pelo que o grau de indeterminação do sistema é $(n - k)$. Isto é, o sistema $A\mathbf{x} = \mathbf{0}$ tem $(n - k)$ incógnitas livres. Reordenando (se necessário) as colunas da matriz, o que só afecta a ordenação das incógnitas, podemos supor que as incógnitas livres são as primeiras $(n-k)$ incógnitas. Assim, as soluções do sistema $A\mathbf{x} = \mathbf{0}$ são

$$\begin{aligned}
\begin{bmatrix} x_1 \\ \vdots \\ x_{n-k} \\ x_{n-k+1} \\ \vdots \\ x_n \end{bmatrix} &= \begin{bmatrix} x_1 \\ \vdots \\ x_{n-k} \\ \alpha_{11}x_1 + \cdots \alpha_{1(n-k)}x_{n-k} \\ \vdots \\ \alpha_{n1}x_1 + \cdots \alpha_{n(n-k)}x_{n-k} \end{bmatrix} \\
&= x_1 \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ \alpha_{11} \\ \vdots \\ \alpha_{n1} \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \\ \alpha_{12} \\ \vdots \\ \alpha_{n2} \end{bmatrix} + \cdots + x_{n-k} \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \alpha_{1(n-k)} \\ \vdots \\ \alpha_{n(n-k)} \end{bmatrix} \\
&= x_1\mathbf{u}_1 + x_2\mathbf{u}_2 + \cdots + x_{n-k}\mathbf{u}_{n-k}.
\end{aligned}$$

Da igualdade anterior resulta que $N(A) = \mathrm{Span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{n-k}\}$. Para mostrar que a dimensão do núcleo é $(n-k)$, basta provar que o conjunto $S = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{n-k}\}$ é linearmente independente, ou seja, que S forma uma base para $N(A)$. Este conjunto é obviamente linearmente independente já que, o sistema $P\mathbf{x} = 0$, onde P é a matriz cujas colunas são $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{n-k}$ (por esta ordem), só tem a solução nula (ver Proposição 3.5). Por conseguinte, S é uma base de $N(A)$, e a dimensão de $N(A)$ coincide com o grau de indeterminação de $A\mathbf{x} = \mathbf{0}$, que é igual a $(n-k)$.

$\square$

Como consequência imediata deste teorema e do facto da soma da característica de uma matriz A com o grau de indeterminação do sistema $A\mathbf{x} = \mathbf{0}$ ser igual ao número de colunas de A, tem-se o importante teorema sobre as dimensões dos quatro subespaços fundamentais associados a uma matriz.

Teorema 3.6. Teorema da dimensão para matrizes
Seja A uma matriz $p \times n$. Verificam-se as igualdades

a) $\dim EL(A) + \dim N(A) = n =$ número de colunas de A.
b) $\dim EC(A) + \dim N(A^T) = p =$ número de linhas de A.

Demonstração. A alínea a) é exactamente a igualdade (3.3) onde se substituiu $\mathrm{car}(A)$ por $\dim EL(A)$.

Como a igualdade da alínea a) é válida para qualquer matriz, em particular é válida para A^T. Ou seja,

$$\dim EL(A^T) + \dim N(A^T) = p,$$

uma vez que p é igual ao número de colunas de A^T. Assim, o resultado b) segue da igualdade anterior e do facto de $EL(A^T) = EC(A)$. $\square$

Para finalizar esta secção, vamos relacionar as dimensões de $U+V$ e de $U\cap V$ com as dimensões do espaço das colunas e do núcleo de uma matriz, no caso em que U e V são subespaços de $\mathbb{R}^n$.

Sejam B_U e B_V bases, respectivamente, dos subespaços $U \subset \mathbb{R}^n$ e $V \subset \mathbb{R}^n$ dadas por

$$B_U = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\} \qquad B_V = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}.$$

Relembremos que

$$\begin{aligned} U + V &= \{\mathbf{x} = \mathbf{u} + \mathbf{v} : \ \mathbf{u} \in U \text{ e } \mathbf{v} \in V\} \subset \mathbb{R}^n \\ U \cap V &= \{\mathbf{y} : \ \mathbf{y} \in U \text{ e } \mathbf{y} \in V\} \subset \mathbb{R}^n. \end{aligned}$$

Assim, como B_U e B_V são bases, respectivamente, de U e de V, qualquer vector $\mathbf{u} \in U$ e $\mathbf{v} \in V$ é uma combinação linear dos vectores da base do espaço linear respectivo. Logo,

$$\begin{aligned} \mathbf{x} \in U + V &\Longleftrightarrow \mathbf{x} = \alpha_1\mathbf{u}_1 + \cdots + \alpha_k\mathbf{u}_k + \beta_1\mathbf{v}_1 + \cdots + \beta_p\mathbf{v}_p \\ \mathbf{y} \in U \cap V &\Longleftrightarrow \mathbf{y} = \alpha_1\mathbf{u}_1 + \cdots + \alpha_k\mathbf{u}_k = \beta_1\mathbf{v}_1 + \cdots + \beta_p\mathbf{v}_p, \end{aligned}$$

para alguns $\alpha_i, \beta_i \in \mathbb{R}$. Destas equivalências podemos concluir:

$$\begin{aligned} U + V &= \mathrm{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{v}_1, \ldots, \mathbf{v}_p\} \\ \mathbf{y} \in U \cap V &\Longrightarrow \alpha_1\mathbf{u}_1 + \cdots + \alpha_k\mathbf{u}_k - (\beta_1\mathbf{v}_1 + \cdots + \beta_p\mathbf{v}_p) = \mathbf{0}. \end{aligned} \tag{3.13}$$

Consideremos agora uma matriz A cujas colunas são os vectores da base de U juntamente com os da base de V:

$$A = \begin{bmatrix} | & & | & | & & | \\ \mathbf{u}_1 & \cdots & \mathbf{u}_k & \mathbf{v}_1 & \cdots & \mathbf{v}_p \\ | & & | & | & & | \end{bmatrix}.$$

É óbvio que $EC(A) = \text{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{v}_1, \ldots, \mathbf{v}_p\}$, e portanto uma base para $U + V$ é uma base para o espaço das colunas de A. Logo, uma base para $U + V$ é constituída pelas colunas de A correspondentes às colunas com pivô de uma matriz em escada por linhas obtida de A por eliminação de Gauss. Assim,

$$\dim(U + V) = \dim EC(A) = \text{car}(A).$$

Além disso, por construção da matriz A, o número de colunas de A satisfaz as igualdades

$$\text{Número de colunas de } A = k + p = \dim U + \dim V.$$

Do Teorema da dimensão para matrizes (Teorema 3.6) e do Teorema 3.4, tem-se

$$\text{Número de colunas de } A = \dim EC(A) + \dim N(A). \tag{3.14}$$

Da Proposição 3.10 sabemos que

$$\dim U + \dim V = \dim(U + V) + \dim(U \cap V).$$

Tendo em conta que $\dim(U+V) = \dim EC(A)$ e que o número de colunas de A é igual a $\dim U + \dim V$, da igualdade (3.14) obtém-se

$$\underbrace{\dim U + \dim V}_{\text{número de colunas de } A} = \underbrace{\dim(U + V)}_{\dim EC(A)} + \dim(U \cap V) \iff \dim(U \cap V) = \dim N(A).$$

Nota 21. *Apesar de $U \cap V$ e $N(A)$ terem a mesma dimensão, estes subespaços não são em geral iguais. Note-se que $U \cap V \subset \mathbb{R}^n$ e $N(A) \subset \mathbb{R}^{k+p}$. Para determinar $U \cap V$ podemos usar a equação* (3.13) *como se exemplifica a seguir.*

Exemplo 3.21. Determinemos bases para $U + V$ e $U \cap V$, onde U e V são os subespaços

$$U = \text{Span}\{(0, 0, 1), (1, -1, 2)\} \qquad V = \left\{(x, y, z) \in \mathbb{R}^3 : \, x = y - z\right\}.$$

Como se verifica facilmente, os vectores $\mathbf{u}_1 = (0, 0, 1)$ e $\mathbf{u}_2 = (1, -1, 2)$ são linearmente independentes e portanto formam uma base para U. Como,

$$\begin{aligned} V &= \left\{(x, y, z) \in \mathbb{R}^3 : \, x = y - z\right\} = \{(y - z, y, z) : \, y, z \in \mathbb{R}\} \\ &= \{y(1, 1, 0) + z(-1, 0, 1) : \, y, z \in \mathbb{R}\} = \text{Span}\{(1, 1, 0), (-1, 0, 1)\}, \end{aligned}$$

e os vectores $\mathbf{v}_1 = (1, 1, 0)$ e $\mathbf{v}_2 = (-1, 0, 1)$ são linearmente independentes, estes vectores formam uma base de V.

Construa-se a matriz A cujas colunas são os vectores $\mathbf{v}_1, \mathbf{v}_2, \mathbf{u}_1$ e $\mathbf{u}_2$, isto é, as colunas de A são vectores de uma base de V e de uma base de U. Aplicando o método de eliminação de Gauss a esta matriz, obtém-se

$$A = \begin{bmatrix} 1 & -1 & 0 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 2 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & -1 & 0 & 1 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 4 \end{bmatrix} = R.$$

Assim, uma base de $EC(A)$, e portanto do subespaço $U + V$, é $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{u}_1\}$. Logo, $\dim(U + V) = 3$. Além disso, podemos já concluir qual é a dimensão de $U \cap V$, uma vez que esta dimensão é igual à dimensão do núcleo de A. Do Teorema da dimensão para matrizes, tem-se $\dim N(A) = 4 - \operatorname{car} A = 4 - 3 = 1$. Determinemos agora uma base para $U \cap V$. Os vectores $\mathbf{y} \in U \cap V$ são vectores da forma

$$\begin{aligned} &\mathbf{y} = \alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 = \beta_1 \mathbf{v}_1 + \beta_2 \mathbf{v}_2 \\ &\quad \Longleftrightarrow \beta_1 \mathbf{v}_1 + \beta_2 \mathbf{v}_2 - \alpha_1 \mathbf{u}_1 - \alpha_2 \mathbf{u}_2 = \mathbf{0} \\ &\quad \Longleftrightarrow A\mathbf{w} = \mathbf{0} \Longleftrightarrow R\mathbf{w} = \mathbf{0} \quad \text{com } \mathbf{w} = (\beta_1, \beta_2, -\alpha_1, -\alpha_2). \end{aligned}$$

Resolvendo o sistema $R\mathbf{w} = \mathbf{0}$, obtém-se

$$\alpha_1 = -4\alpha_2, \quad \beta_2 = -2\alpha_2, \quad \beta_1 = \beta_2 + \alpha_2 = -\alpha_2.$$

Consequentemente, os vectores $\mathbf{y} \in U \cap V$ satisfazem:

$$\begin{aligned} &\mathbf{y} = \alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 = \alpha_2(\mathbf{u}_2 - 4\mathbf{u}_1) = \alpha_2(1, -1, -2) \\ &\text{e} \\ &\mathbf{y} = \beta_1 \mathbf{v}_1 + \beta_2 \mathbf{v}_2 = -\alpha_2(\mathbf{v}_1 + 2\mathbf{v}_2) = \alpha_2(1, -1, -2). \end{aligned}$$

Logo, uma base para $U \cap V$ é $\{(1, -1, -2)\}$, confirmando-se assim que $\dim U \cap V = 1$. ♦

3.3.2 Característica e sistemas de equações lineares

Nesta secção reformulamos alguns resultados obtidos para espaços lineares em termos de sistemas de equações lineares. Neste contexto, a noção de característica de uma matriz é fundamental, sendo por isso importante entender este conceito sob diferentes pontos de vista. Historicamente, a caracterização da característica de uma matriz apresentada na Proposição 3.12 abaixo, foi a definição adoptada por G. Frobenius[2] em 1879. Esta proposição estabelece que a característica de uma matriz pode ser determinada com recurso ao cálculo de determinantes de submatrizes.

Começamos por estabelecer um lema cujo resultado será usado nas demonstrações das proposições 3.11 e 3.12.

Lema 3.1. Se A é uma matriz $p \times n$, e P e Q são matrizes invertíveis de ordem p e n respectivamente, satisfazendo a igualdade

$$PAQ = B,$$

então $\operatorname{car}(A) = \operatorname{car}(B)$.

Demonstração. Comecemos por mostrar que dada uma matriz invertível P, as matrizes PA e A têm a mesma característica.

Do Teorema 1.4 (pág. 57) sabemos que qualquer matriz invertível é um produto de matrizes elementares, logo P é um produto de matrizes elementares. Como a multiplicação (à esquerda) de uma matriz por matrizes elementares não altera a sua característica, resulta que

$$\operatorname{car}(PA) = \operatorname{car}(A), \quad \text{para qualquer } P \text{ invertível.} \tag{3.15}$$

Uma vez que a transposição de matrizes não altera a característica (cf. Teorema 3.4-(d)), tem-se

$$\operatorname{car}(AQ) = \operatorname{car}(AQ)^T = \operatorname{car}(Q^T A^T) = \operatorname{car}(A^T) = \operatorname{car}(A), \tag{3.16}$$

onde na penúltima igualdade aplicámos (3.15). Por conseguinte, de (3.15) e (3.16), resulta

$$\operatorname{car}(B) = \operatorname{car}(PAQ) = \operatorname{car}(P(AQ)) = \operatorname{car}(AQ) = \operatorname{car}(A).$$

□

[2] Ferdinand Georg Frobenius (1849 – 1917), matemático alemão com inúmeras contribuições na área de equações diferenciais. No âmbito da álgebra linear, refira-se que Frobenius apresentou a primeira demonstração do Teorema de Cayley-Hamilton que abordaremos no próximo capítulo.

Proposição 3.11. Forma normal de uma matriz de característica k
Seja A uma matriz de característica k. Existem matrizes invertíveis P e Q tais que

$$PAQ = \begin{bmatrix} M_{k\times k} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

onde $M_{k\times k}$ é uma submatriz de A de ordem k, invertível, e $\mathbf{0}$ designa matrizes nulas.

Demonstração. Seja A uma matriz do tipo $p \times n$. Sendo $\operatorname{car}(A) = k$, a matriz A tem k linhas e k colunas linearmente independentes. Podemos pois efectuar sucessivas trocas de linhas em A e colocar k linhas linearmente independentes nas primeiras k linhas de uma matriz. Isto corresponde a multiplicar A, à esquerda, por uma matriz invertível (igual ao produto de matrizes de permutação correspondentes às trocas de linhas efectuadas). Em seguida podemos usar operações elementares sobre a matriz construída para reduzir a linhas nulas todas as linhas abaixo da linha k. Designemos por P a matriz (invertível) igual ao produto de matrizes elementares correspondentes às operações elementares efectuadas (troca de linhas e redução de linhas a linhas nulas). Assim,

$$PA = \begin{bmatrix} U_{k\times n} \\ \mathbf{0} \end{bmatrix} = R, \tag{3.17}$$

onde a matriz $U_{k\times n}$, do tipo $k \times n$, possui k colunas linearmente independentes, uma vez que, pelo Lema 3.1, se tem $\operatorname{car}(R) = \operatorname{car}(PA) = \operatorname{car}(A) = k$.

A matriz $R^T = \begin{bmatrix} U^T_{k\times n} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} W_{n\times k} & \mathbf{0} \end{bmatrix}$ tem portanto k linhas linearmente independentes. Assim, usando o mesmo argumento que anteriormente, podemos colocar k linhas linearmente independentes de W nas primeiras k linhas e depois reduzir a linhas nulas as linhas abaixo dessas primeiras k linhas. Tal corresponde à existência de uma matriz invertível N, verificando

$$NR^T = \begin{bmatrix} V_{k\times k} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

Logo, como N é invertível, a igualdade anterior é equivalente a

$$R^T = N^{-1} \begin{bmatrix} V_{k\times k} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \Longleftrightarrow R = \begin{bmatrix} V^T_{k\times k} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} (N^{-1})^T.$$

Da expressão (3.17), tem-se

$$PA = R = \begin{bmatrix} V^T_{k\times k} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} (N^{-1})^T \Longleftrightarrow PAN^T = \begin{bmatrix} V^T_{k\times k} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

onde na equivalência anterior usámos o facto $(N^{-1})^T = (N^T)^{-1}$ (ver Exercício 2.6, pág. 103). A matriz $V^T_{k\times k}$ é uma submatriz de A uma vez que foi obtida de A suprimindo linhas e colunas (linearmente dependentes). Além disso, pelo Lema 3.1, tem-se $\operatorname{car}(V^T_{k\times k}) = \operatorname{car}(A) = k$, pelo que $V^T_{k\times k}$ é invertível (cf. Proposição 1.12, pág. 62). Tomando $Q = N^T$ e $M_{k\times k} = V^T_{k\times k}$, obtemos o resultado enunciado.

□

Nota 22. *Uma submatriz M como a do enunciado da proposição anterior pode obter-se do seguinte modo: (i) reduzir A a uma matriz em escada por linhas T; (ii) suprimir em T as linhas nulas e as colunas sem pivô, obtendo-se assim uma matriz quadrada; (iii) Tomar M igual à matriz que se obtém suprimindo em A as linhas e as colunas correspondentes às linhas e colunas que foram suprimidas em T.*

A proposição que enunciamos a seguir estabelece que a característica de uma matriz pode obter-se calculando determinantes de submatrizes.

Proposição 3.12. A característica de uma matriz A, do tipo $p \times n$, é igual à ordem da maior submatriz de A invertível.

Isto é, se $\operatorname{car}(A) = k$, então existe pelo menos uma submatriz de A invertível de ordem k e não existem outras submatrizes invertíveis de ordem superior.

Demonstração. Se $\operatorname{car}(A) = k$, pela Proposição 3.11 existem matrizes invertíveis P e Q tais que $PAQ = \begin{bmatrix} M_{k\times k} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$, onde $M_{k\times k}$ é uma submatriz de A (de ordem k) invertível.

Falta mostrar que não existe nenhuma submatriz (quadrada) de A de ordem superior a k que seja invertível. Para tal, suponha-se por absurdo que W é uma submatriz de A invertível de ordem superior a k. Usando troca de linhas e de colunas, podemos determinar duas matrizes invertíveis P e Q, tais que o produto PAQ tem a forma

$$PAQ = \begin{bmatrix} W & X \\ Y & Z \end{bmatrix}.$$

Considere-se as matrizes em blocos $E = \begin{bmatrix} I & \mathbf{0} \\ -YW^{-1} & I \end{bmatrix}$, $F = \begin{bmatrix} I & -W^{-1}X \\ \mathbf{0} & I \end{bmatrix}$ e $S = Z - YW^{-1}X$, onde I designa a matriz identidade. Calculando o produto $EPAQF$ de matrizes em blocos como se indica na Secção 1.6, tem-se

$$EPAQF = \begin{bmatrix} W & \mathbf{0} \\ \mathbf{0} & S \end{bmatrix}.$$

Uma vez que as matrizes E, P, Q e F são invertíveis, EP e QF também o são. Logo, usando o Lema 3.1 resulta

$$\operatorname{car}(EPAQF) = \operatorname{car}(A) = \operatorname{car}\left(\begin{bmatrix} W & \mathbf{0} \\ \mathbf{0} & S \end{bmatrix}\right) \geq \operatorname{car}(W) > k.$$

Tem-se portanto a contradição $\operatorname{car}(A) > k$. □

Exercício 3.3. Verifique que as matrizes E, F, Q e P na demonstração anterior possuem os tamanhos adequados para efectuar o produto $EPAQF$. Para tal, comece por fixar o tamanho de W. ▲

Exemplo 3.22. Usemos a proposição anterior para determinar a característica de

$$A = \begin{bmatrix} 1 & 3 & 2 \\ 2 & 6 & 4 \\ 1 & 1 & 1 \end{bmatrix}.$$

Como $\det(A) = 0$, a característica de A não é 3. Consideremos, por exemplo, a submatriz $M = \begin{bmatrix} 1 & 3 \\ 1 & 1 \end{bmatrix}$ que se obtém de A suprimindo a linha 2 e a coluna 3. A matriz M tem determinante igual a -2, e portanto é invertível. Logo, a característica de A é 2.

Na demonstração do Teorema 3.4 a submatriz M é obtida de A suprimindo colunas e linhas dependentes. Além disso, M tem linhas e colunas linearmente independentes (M é invertível). Podemos portanto concluir que os conjuntos $\{\mathbf{l}_1, \mathbf{l}_3\}$ e $\{\mathbf{c}_1, \mathbf{c}_2\}$, onde $\mathbf{l}_i$ e $\mathbf{c}_i$ designam respectivamente a linha i e a coluna i de A, são linearmente independentes (como facilmente se confirma uma vez que são conjuntos de dois vectores não colineares). ♦

Sistemas de equações lineares

Reformulamos agora em termos de sistemas de equações lineares alguns dos resultados obtidos para espaços lineares.

Como é sabido (Proposição 1.3), tem-se a equivalência

$$A\mathbf{x} = \mathbf{b} \text{ é possível} \iff \mathbf{b} \text{ é combinação linear das colunas de } A.$$

Ou seja, um sistema $A\mathbf{x} = \mathbf{b}$ é possível se e só se $\mathbf{b} \in EC(A)$.

Por outro lado, a Proposição 1.1 diz-nos que $\operatorname{car}(A) = \operatorname{car}([A \mid \mathbf{b}])$ se e só se o sistema $A\mathbf{x} = \mathbf{b}$ é possível. Podemos portanto enunciar o seguinte resultado.

Proposição 3.13. Para um sistema linear $A\mathbf{x} = \mathbf{b}$ são equivalentes as afirmações:

(a) $A\mathbf{x} = \mathbf{b}$ é possível.

(b) $\mathbf{b}$ pertence ao espaço das colunas de A, isto é, $\mathbf{b} \in EC(A)$.

(c) A característica de A e da matriz aumentada $[A \,|\, \mathbf{b}]$ são iguais.

Proposição 3.14. Seja A uma matriz $p \times n$. As afirmações seguintes são equivalentes.

(a) $A\mathbf{x} = \mathbf{0}$ só admite a solução $\mathbf{x} = \mathbf{0}$.

(b) $N(A) = \{\mathbf{0}\}$.

(c) As colunas de A são vectores linearmente independentes.

(d) Qualquer que seja $\mathbf{b} \in \mathbb{C}^p$, o sistema $A\mathbf{x} = \mathbf{b}$ admite no máximo uma solução (pode não admitir nenhuma).

Demonstração. As equivalências $(a) \iff (b) \iff (c)$ são o enunciado da Proposição 3.5.

Mostremos agora a equivalência $(a) \iff (d)$:

$(a) \Rightarrow (d)$: O sistema $A\mathbf{x} = \mathbf{b}$ é possível ou impossível. Se é possível, pode ter uma ou mais soluções. Sejam $\mathbf{u}$ e $\mathbf{v}$ soluções distintas do sistema. Temos de mostrar que se $A\mathbf{u} = \mathbf{b}$ e $A\mathbf{v} = \mathbf{b}$, então $\mathbf{u} = \mathbf{v}$. Mas

$$A\mathbf{u} - A\mathbf{v} = \mathbf{0} \iff A(\mathbf{u} - \mathbf{v}) = \mathbf{0},$$

ou seja, $(\mathbf{u} - \mathbf{v})$ é solução de $A\mathbf{x} = \mathbf{0}$. Como, por hipótese, a única solução do sistema $A\mathbf{x} = \mathbf{0}$ é a solução nula, resulta $\mathbf{u} - \mathbf{v} = \mathbf{0}$. Assim, o sistema $A\mathbf{x} = \mathbf{b}$ admite no máximo uma solução.

$(d) \Rightarrow (a)$: Se $A\mathbf{x} = \mathbf{b}$ admite no máximo uma solução, qualquer que seja $\mathbf{b} \in \mathbb{R}^p$, em particular $A\mathbf{x} = \mathbf{0}$ admite no máximo uma solução. Como $\mathbf{x} = \mathbf{0}$ é sempre solução de $A\mathbf{x} = \mathbf{0}$, temos (a).

$\square$

A proposição a seguir é uma consequência imediata de vários resultados anteriores, ficando por isso a sua demonstração como exercício.

Proposição 3.15. Seja A uma matriz $n \times n$. São equivalentes as afirmações seguintes.

(a) A é invertível.

(b) $\det(A) \neq 0$.

(c) $N(A) = \{\mathbf{0}\}$.

(d) As colunas de A formam uma base de $\mathbb{C}^n$.

(e) As linhas de A formam uma base de $\mathbb{C}^n$.

(f) Todo o $\mathbf{b} \in \mathbb{C}^n$ pertence a $EC(A)$.

Finalizamos esta secção com um resultado elementar mas de grande utilidade prática na resolução de equações lineares.

Teorema 3.7. A solução geral de um sistema não homogéneo, $A\mathbf{x} = \mathbf{b}$, é a soma da solução geral do sistema homogéneo associado, $A\mathbf{x} = \mathbf{0}$, com uma solução particular de $A\mathbf{x} = \mathbf{b}$. Ou seja, qualquer solução de $A\mathbf{x} = \mathbf{b}$ é da forma

$$\mathbf{x} = \mathbf{x}_p + \mathbf{x}_h,$$

onde $\mathbf{x}_h$ é uma solução de $A\mathbf{x} = \mathbf{0}$ e $\mathbf{x}_p$ é uma solução de $A\mathbf{x} = \mathbf{b}$.

Demonstração. Seja $\mathbf{x}_p$ um vector tal que $A\mathbf{x}_p = \mathbf{b}$ Então, $\mathbf{x}$ é solução do sistema não homogéneo se e só se $(\mathbf{x} - \mathbf{x}_p)$ pertence ao núcleo de A, uma vez que $A(\mathbf{x} - \mathbf{x}_p) = A\mathbf{x} - A\mathbf{x}_p = \mathbf{b} - \mathbf{b} = \mathbf{0}$. Assim, as soluções do sistema não homogéneo são obtidas adicionando a $\mathbf{x}_p$ todas as soluções de $A\mathbf{x} = \mathbf{0}$. $\square$

No exemplo que apresentamos a seguir ilustramos como a aplicação do método de eliminação de Gauss ao sistema $A\mathbf{x} = \mathbf{b}$ nos fornece a solução geral deste sistema na forma $\mathbf{x} = \mathbf{x}_p + \mathbf{x}_h$, onde $\mathbf{x}_p$ e $\mathbf{x}_h$ designam, respectivamente, uma solução particular de $A\mathbf{x} = \mathbf{b}$ e a solução geral do sistema homogéneo associado.

Exemplo 3.23. Considere-se o sistema

$$\begin{cases} x - \ \ y + 2z = 1 \\ x + 2y - \ \ z = 2. \end{cases}$$

Geometricamente, temos dois planos em $\mathbb{R}^3$ que se intersectam segundo uma recta. Assim, uma solução particular do sistema é um ponto desta recta. Pretendemos escrever a solução geral do sistema como a soma de uma solução particular com a solução geral do sistema homogéneo associado.

Apliquemos o método de eliminação de Gauss para determinar a solução geral do sistema na forma $\mathbf{x} = \mathbf{x}_p + \mathbf{x}_h$.

$$\left[\begin{array}{ccc|c} 1 & -1 & 2 & 1 \\ 1 & 2 & -1 & 2 \end{array}\right] \xrightarrow{L_2 - L_1} \left[\begin{array}{ccc|c} 1 & -1 & 2 & 1 \\ 0 & 3 & -3 & 1 \end{array}\right] = R$$

A matriz em escada por linhas R diz-nos que o sistema tem uma variável livre que vamos escolher como sendo a variável associada à coluna que não possui pivô, ou seja, tomamos z como variável livre. A solução geral do sistema é então dada por

$$\mathbf{x} = (x, y, z) = \left(\frac{4}{3} - z, \frac{1}{3} + z, z\right) = \left(\frac{4}{3}, \frac{1}{3}, 0\right) + z(-1, 1, 1). \tag{3.18}$$

É claro que uma solução particular de $A\mathbf{x} = \mathbf{b}$ pode ser obtida da solução geral deste sistema fazendo, por exemplo, as variáveis livres iguais a zero. Ou seja, de (3.18), podemos tomar

$$\mathbf{x}_p = (4/3, 1/3, 0).$$

Além disso, o sistema homogéneo associado, $A\mathbf{x} = \mathbf{0}$, tem solução geral igual à solução geral do sistema $R\mathbf{x} = \mathbf{0}$ que facilmente se verifica ser

$$\mathbf{x}_h = z(-1, 1, 1).$$

Assim, (3.18) exprime a solução geral de $A\mathbf{x} = \mathbf{b}$ na forma pretendida.

O núcleo da matriz dos coeficientes do sistema (ou seja, a solução geral $\mathbf{x}_h$) é uma recta que passa pela origem e a solução geral do sistema não homogéneo (isto é, $\mathbf{x}_p + \mathbf{x}_h$) é uma recta que não passa pela origem. A recta correspondente à solução geral do sistema não homogéneo obtém-se adicionando o vector $\mathbf{x}_p$ à recta correspondente ao núcleo da matriz dos coeficientes. Uma ilustração deste resultado é apresentada na Figura 3.7. Saliente-se ainda que, enquanto o conjunto solução geral de um sistema homogéneo (isto é, o núcleo de uma matriz) é um subespaço linear de $\mathbb{C}^n$, o mesmo não acontece com a solução geral de um sistema não homogéneo (o vector nulo não pertence ao conjunto das soluções). ♦

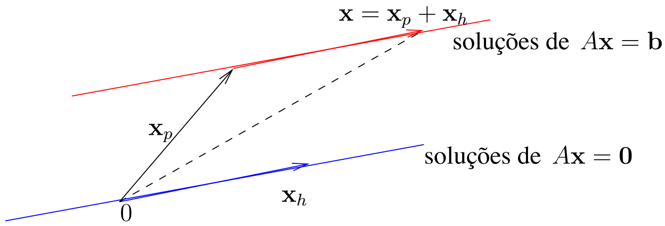


Figura 3.7: A solução geral de um sistema não homogéneo.

3.4 Matriz de mudança de base

Como vimos (pág. 138 e Teorema 3.1), existe uma correspondência biunívoca entre vectores de um espaço linear W e os respectivos vectores das coordenadas numa base ordenada fixada em W. Se fixarmos duas bases ordenadas distintas B e B' num espaço linear W, então qualquer vector $x \in W$ é representado por dois vectores distintos $\mathbf{x}_B$ e $\mathbf{x}_{B'}$. A questão que naturalmente se coloca é a de saber qual a relação entre os vectores das coordenadas $\mathbf{x}_B$ e $\mathbf{x}_{B'}$ que representam o mesmo vector em bases distintas. Trata-se pois de um problema de mudança de coordenadas, porventura já familiar ao leitor noutros contextos.

Como veremos, fixadas duas bases ordenadas num espaço linear, existe uma matriz invertível que relaciona as coordenadas de um qualquer vector numa das bases com as suas coordenadas na outra base. Ou seja, para todo o vector $x \in W$, existe uma matriz invertível M tal que $\mathbf{x}_{B'} = M\mathbf{x}_B$. A matriz M recebe a designação de matriz de mudança de base.

Sejam $B = (v_1, v_2, \ldots, v_n)$ e B' duas bases ordenadas do espaço linear W e x um qualquer vector de W cujos vectores das coordenadas nas bases B e B' são $\mathbf{x}_B$ e $\mathbf{x}_{B'}$. Comecemos por mostrar que, se existe uma matriz invertível M que verifica

$$\mathbf{x}_{B'} = M\mathbf{x}_B, \quad \text{para todo } x \in W, \tag{3.19}$$

então esta matriz é única. Para tal considere-se matrizes invertíveis M e S que verificam (3.19), isto é, tais que $M\mathbf{x}_B = \mathbf{x}_{B'} = S\mathbf{x}_B$ para todo $x \in W$. Em particular, para os vectores da base B, tem-se

$$S^{-1}M(\mathbf{v}_1)_B = (\mathbf{v}_1)_B,\ \ S^{-1}M(\mathbf{v}_2)_B = (\mathbf{v}_2)_B,\ \ldots,\ S^{-1}M(\mathbf{v}_n)_B = (\mathbf{v}_n)_B. \tag{3.20}$$

Os vectores das coordenadas de $v_1, v_2, \dots v_n$ na base B são:

$$\begin{array}{lcl} v_1 = 1 \times v_1 + 0 \times v_2 + \cdots + 0 \times v_n & \rightsquigarrow & (\mathbf{v}_1)_B = (1,0,0,\dots,0) = \mathbf{e}_1 \\ v_2 = 0 \times v_1 + 1 \times v_2 + \cdots + 0 \times v_n & \rightsquigarrow & (\mathbf{v}_2)_B = (0,1,0,\dots,0) = \mathbf{e}_2 \\ & & \\ v_n = 0 \times v_1 + 0 \times v_2 + \cdots + 1 \times v_n & \rightsquigarrow & (\mathbf{v}_n)_B = (0,0,0,\dots,1) = \mathbf{e}_n. \end{array}$$

Logo, as relações (3.20) são equivalentes a

$$S^{-1}M\mathbf{e}_1 = \mathbf{e}_1, \quad S^{-1}M\mathbf{e}_2 = \mathbf{e}_2, \quad \dots \quad , S^{-1}M\mathbf{e}_n = \mathbf{e}_n.$$

Destas igualdades, e da Definição 1.12 (pág. 39) de produto de matrizes, resulta

$$S^{-1}MI_n = I_n,$$

onde I_n é a matriz identidade de ordem n. Como $S^{-1}M = I_n$, multiplicando por S ambos os membros desta igualdade obtém-se $S = M$ conforme pretendido.

Determinemos agora a matriz M que satisfaz (3.19). Se M verifica (3.19) para todo $x \in W$, em particular satisfaz esta relação para os vectores que formam a base B, ou seja,

$$\begin{aligned} (\mathbf{v}_1)_{B'} &= M(\mathbf{v}_1)_B = M\mathbf{e}_1 = \text{coluna 1 de } M \\ (\mathbf{v}_2)_{B'} &= M(\mathbf{v}_2)_B = M\mathbf{e}_2 = \text{coluna 2 de } M \\ &\vdots \\ (\mathbf{v}_n)_{B'} &= M(\mathbf{v}_n)_B = M\mathbf{e}_n = \text{coluna } n \text{ de } M. \end{aligned}$$

Logo, M é a matriz cujas colunas são os vectores das coordenadas na base B' dos vectores da base B. Isto é,

$$M = \begin{bmatrix} | & | & & | \\ (\mathbf{v}_1)_{B'} & (\mathbf{v}_2)_{B'} & \cdots & (\mathbf{v}_n)_{B'} \\ | & | & & | \end{bmatrix}. \tag{3.21}$$

Verifiquemos agora que M é invertível. A matriz M tem colunas linearmente independentes uma vez que, sendo os vectores $v_1, \dots, v_n$ linearmente independentes (são vectores de uma base) os vectores $(\mathbf{v}_1)_{B'}, (\mathbf{v}_2)_{B'}, \dots, (\mathbf{v}_n)_{B'}$ também o são (cf. Proposição 3.9). Por conseguinte, a Proposição 3.15 garante que M é invertível.

Definição 3.9. Matriz de mudança de base
Sejam B e B' duas bases ordenadas de um espaço linear W. Chama-se *matriz de mudança da base* B para a base B', à matriz quadrada invertível M, ou $M_{B' \leftarrow B}$, que verifica:

$$\mathbf{x}_{B'} = M\mathbf{x}_B \quad \text{(ou} \quad \mathbf{x}_{B'} = M_{B' \leftarrow B}\mathbf{x}_B), \quad \text{para todo } x \in W, \tag{3.22}$$

onde $\mathbf{x}_B$ e $\mathbf{x}_{B'}$ são os vectores das coordenadas de x, respectivamente, na base B e na base B'.

Como a matriz M, de mudança da base B para a base B', é invertível, multiplicando a igualdade $\mathbf{x}_{B'} = M\mathbf{x}_B$ por M^{-1} obtém-se $\mathbf{x}_B = M^{-1}\mathbf{x}_{B'}$. Logo, da definição de matriz de mudança de base, tem-se que M^{-1} é a matriz de mudança da base B' para a base B.

Enunciamos no teorema seguinte o que se provou relativamente à forma da matriz de mudança de base.

Teorema 3.8. Sejam $B = (v_1, v_2, \ldots, v_n)$ e B' duas bases ordenadas de um espaço linear W. A matriz que efectua a mudança da base B para a base B' é a matriz cujas colunas são os vectores das coordenadas na base B' dos vectores que formam a base B (respeitando a ordem de B). Isto é,

$$M_{B' \leftarrow B} = \begin{bmatrix} | & | & & | \\ (\mathbf{v}_1)_{B'} & (\mathbf{v}_2)_{B'} & \cdots & (\mathbf{v}_n)_{B'} \\ | & | & & | \end{bmatrix}.$$

A matriz $M_{B' \leftarrow B}$ é invertível e a sua inversa é a matriz que efectua a mudança da base B' para a base B. Ou seja, para todo o vector $x \in W$ verifica-se

$$\mathbf{x}_{B'} = M_{B' \leftarrow B}\mathbf{x}_B \qquad \text{e} \qquad \mathbf{x}_B = M_{B' \leftarrow B}^{-1}\mathbf{x}_{B'} = M_{B \leftarrow B'}\mathbf{x}_{B'}.$$

Exemplo 3.24. Considerando em $\mathbb{R}^2$ as base ordenadas $B = ((1,2),(1,7))$ e $B' = ((0,2),(1,1))$, determinemos a matriz M de mudança da base B para a base B'.

Designemos os vectores da base B por $\mathbf{u} = (1,2)$ e $\mathbf{v} = (1,7)$. Calculemos os vectores das coordenadas na base B' dos vectores $\mathbf{u}$ e $\mathbf{v}$.

$$\begin{aligned}(1,2) &= \alpha(0,2)+\beta(1,1) = (\beta, 2\alpha+\beta) \iff \beta = 1 \text{ e } 2\alpha+\beta = 2\\ &\iff \beta = 1 \text{ e } \alpha = 1/2 \iff \mathbf{u}_{B'} = (1/2, 1).\\ (1,7) &= \alpha(0,2)+\beta(1,1) = (\beta, 2\alpha+\beta) \iff \beta = 1 \text{ e } 2\alpha+\beta = 7\\ &\iff \beta = 1 \text{ e } \alpha = 3 \iff \mathbf{v}_{B'} = (3, 1).\end{aligned}$$

Assim,

$$M = \begin{bmatrix} 1/2 & 3 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} | & | \\ \mathbf{u}_{B'} & \mathbf{v}_{B'} \\ | & | \end{bmatrix}.$$

Usemos agora a matriz M para calcular as coordenadas do vector $\mathbf{x}$ sabendo que o vector das coordenadas de $\mathbf{x}$ na base B é $\mathbf{x}_B = (2/5, 3/5)$.

O vector das coordenadas de $\mathbf{x}$ na base B' é:

$$\mathbf{x}_{B'} = M\mathbf{x}_B = \begin{bmatrix} 1/2 & 3 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 2/5 \\ 3/5 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

Logo,

$$\mathbf{x} = 2(0,2) + (1,1) = (1,5).$$

♦

Exemplo 3.25. Determinemos a matriz M de mudança da base $B = (u_1, u_2, u_3)$ para a base $B' = (v_1, v_2, v_3)$ sabendo que

$$\begin{aligned} v_1 &= u_1 - u_2 + u_3 \\ v_2 &= u_2 - u_3 \\ v_3 &= u_1 + u_3. \end{aligned} \tag{3.23}$$

A matriz M é a matriz que tem nas colunas os vectores das coordenadas $(\mathbf{u}_1)_{B'}, (\mathbf{u}_2)_{B'}$, e $(\mathbf{u}_3)_{B'}$. Calculemos estes vectores. Somando as duas primeiras equações de (3.23) obtemos $u_1 = v_1 + v_2$. Substituindo esta expressão de u_1 na última equação de (3.23) tem-se $u_3 = -v_1 - v_2 + v_3$. Finalmente, substituindo u_3 na segunda equação obtém-se $u_2 = v_3 - v_1$. Assim,

$$\begin{aligned} u_1 &= v_1 + v_2 & &\rightsquigarrow & (\mathbf{u}_1)_{B'} &= (1,1,0) \\ u_2 &= -v_1 + v_3 & &\rightsquigarrow & (\mathbf{u}_2)_{B'} &= (-1,0,1) \\ u_3 &= -v_1 - v_2 + v_3 & &\rightsquigarrow & (\mathbf{u}_3)_{B'} &= (-1,-1,1). \end{aligned}$$

Assim,

$$M_{B' \leftarrow B} = \begin{bmatrix} 1 & -1 & -1 \\ 1 & 0 & -1 \\ 0 & 1 & 1 \end{bmatrix}.$$

Refira-se que das expressões (3.23) é imediato determinar a matriz inversa de $M_{B' \leftarrow B}$, uma vez que $M^{-1}_{B' \leftarrow B}$ é a matriz que efectua a mudança da base B' para a base B. Nomeadamente,

$$M_{B \leftarrow B'} = \begin{bmatrix} 1 & 0 & 1 \\ -1 & 1 & 0 \\ 1 & -1 & 1 \end{bmatrix} = \begin{bmatrix} | & | & | \\ (\mathbf{v}_1)_B & (\mathbf{v}_2)_B & (\mathbf{v}_3)_B \\ | & | & | \end{bmatrix}.$$

♦

3.5 Exemplos de espaços lineares

Finalizamos este capítulo capresentando alguns exemplos de conjuntos munidos de operações de adição e multiplicação por escalares que possuem a estrutura de espaço linear e outros que não possuem essa estrutura para as operações aí definidas. Aproveitamos estes exemplos para propor algumas resoluções de exercícios de aplicação dos conceitos estudados neste capítulo.

1. Como vimos na página 113 e no Exemplo 3.15 (pág. 133), o conjunto $\mathbb{R}^n = \{(x_1, x_2, \ldots, x_n) : x_i \in \mathbb{R}, \text{para } i = 1, \ldots, n\}$ munido das operações usuais de adição de vectores e multiplicação por escalares reais *é um espaço linear real* de dimensão n. A base canónica de $\mathbb{R}^n$ é dada por $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\} \subset \mathbb{R}^n$, com

$$\begin{aligned} \mathbf{e}_1 &= (1, 0, \ldots, 0, 0) \\ \mathbf{e}_2 &= (0, 1, \ldots, 0, 0) \\ &\vdots \\ \mathbf{e}_n &= (0, 0, \ldots, 0, 1). \end{aligned}$$

Considere-se o subconjunto $S = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$ de $\mathbb{R}^4$, com

$$\mathbf{u}_1 = (1, -1, 0, 1),\ \ \mathbf{u}_2 = (1, 1, 1, -2),\ \ \mathbf{u}_3 = (3, 1, 2, -3),\ \ \mathbf{u}_4 = (5, 1, 3, -4).$$

Pretendemos descrever o subespaço gerado por S através de equações. Para tal, consideremos a matriz A cujas colunas são os vectores de S. Ora, o subespaço $V = \operatorname{Span} S$ é o conjunto de todas as combinações lineares de elementos de S, isto é, o conjunto de todos os vectores $\mathbf{x} = (x, y, z, w) \in \mathbb{R}^4$ para os quais o sistema $A\mathbf{u} = \mathbf{x}$ é possível. Equivalentemente, V é o conjunto de todos os vectores $\mathbf{x} = (x, y, z, w) \in \mathbb{R}^4$ para os quais se tem $\mathrm{car}[A|\mathbf{x}] = \mathrm{car}[A]$. Verifiquemos agora que a igualdade $\mathrm{car}[A|\mathbf{x}] = \mathrm{car}[A]$ nos fornece equações (cartesianas) para $V = \mathrm{Span}\, S$. Aplicando o método de eliminação de Gauss à matriz aumentada $[A|\mathbf{x}]$ tem-se

$$\left[\begin{array}{cccc|c} 1 & 1 & 3 & 5 & x \\ -1 & 1 & 1 & 1 & y \\ 0 & 1 & 2 & 3 & z \\ 1 & -2 & -3 & -4 & w \end{array}\right] \underset{L_2+L_1}{\overset{L_4-L_1}{\longrightarrow}} \left[\begin{array}{cccc|c} 1 & 1 & 3 & 5 & x \\ 0 & 2 & 4 & 6 & x+y \\ 0 & 1 & 2 & 3 & z \\ 0 & -3 & -6 & -9 & w-x \end{array}\right]$$

$$\underset{L_3-1/2L_2}{\overset{L_4+3/2L_2}{\longrightarrow}} \left[\begin{array}{cccc|c} 1 & 1 & 3 & 5 & x \\ 0 & 2 & 4 & 6 & x+y \\ 0 & 0 & 0 & 0 & z-\frac{x}{2}-\frac{y}{2} \\ 0 & 0 & 0 & 0 & w-\frac{x}{2}-\frac{3y}{2} \end{array}\right].$$

Assim, $\mathrm{car}[A|\mathbf{x}] = \mathrm{car}[A]$ se e só se

$$\begin{cases} z - \frac{x}{2} - \frac{y}{2} = 0 \\ w - \frac{x}{2} - \frac{3y}{2} = 0 \end{cases} \iff \begin{cases} x + y - 2z = 0 \\ x - 3y - 2w = 0. \end{cases}$$

Podemos concluir que o subespaço $V = \mathrm{Span}\, S$ tem dimensão dois (o espaço das colunas de A tem dimensão 2) e é definido pelas duas equações (cartesianas) acima.

2. **Conjunto $\mathbb{C}^n$:**

   Na página 113 definimos $\mathbb{C}^n$ como o conjunto dos n-uplos de números complexos e munimos este conjunto de operações de adição e multiplicação por escalares definidas em (3.2). Vimos também que $\mathbb{C}^n$ munido destas operações é um espaço linear complexo. No entanto, $\mathbb{C}^n$ pode ser igualmente encarado como espaço linear real, como veremos a seguir para o caso $n = 2$ (o caso geral é inteiramente análogo).

   Consideremos
   $$\mathbb{C}^2 = \{(z_1, z_2) : z_1, z_2 \in \mathbb{C}\},$$
   com as operações
   $$(z_1, z_2) + (w_1, w_2) = (z_1 + w_1, z_2 + w_2) \qquad \alpha \cdot (z_1, z_2) = (\alpha z_1, \alpha z_2),$$

para $(z_1, z_2), (w_1, w_2) \in \mathbb{C}^2$ e $\alpha \in \mathbb{C}$.

$\mathbb{C}^2$ é fechado para a multiplicação por escalares complexos (ao contrário do que acontece com $\mathbb{R}^2$).

Seja $(z_1, z_2) = (a_1 + ib_1, a_2 + ib_2)$ um qualquer vector de $\mathbb{C}^2$. O vector $(z_1, z_2) \in \mathbb{C}^2$ pode ser escrito nas formas:

$$(z_1, z_2) = (a_1 + ib_1, a_2 + ib_2) = (a_1 + ib_1)(1, 0) + (a_2 + ib_2)(0, 1) \tag{3.24}$$

ou

$$\begin{aligned} (z_1, z_2) &= (a_1 + ib_1, a_2 + ib_2) \\ &= a_1(1, 0) + b_1(i, 0) + a_2(0, 1) + b_2(0, i). \end{aligned} \tag{3.25}$$

No primeiro caso, (z_1, z_2) está escrito como combinação linear de $(1, 0)$ e $(0, 1)$ com coeficientes complexos, e no segundo caso está escrito como combinação linear com coeficientes reais de $(1, 0), (i, 0), (0, 1)$ e $(0, i)$. As igualdades (3.24) e (3.25) dizem-nos que o conjunto $\{(1, 0), (0, 1)\}$ gera $\mathbb{C}^2$ enquanto espaço linear complexo, e o conjunto

$$\{(1, 0), (i, 0), (0, 1), (0, i)\},$$

gera $\mathbb{C}^2$ como espaço linear real. Expressamos este facto por:

$$\begin{aligned} \mathbb{C}^2 &= \operatorname{Span}_{\mathbb{R}} \{(1, 0), (i, 0), (0, 1), (0, i)\} \\ \mathbb{C}^2 &= \operatorname{Span}_{\mathbb{C}} \{(1, 0), (0, 1)\}. \end{aligned}$$

Para mostrar que estes conjuntos formam uma base, resta verificar que são linearmente independentes. Usando as igualdades (3.24) e (3.25) com o membro do lado esquerdo igual a $\mathbf{0} = (0 + i0, 0 + i0)$ temos, respectivamente,

$$\begin{aligned} (0 + i0, 0 + i0) &= (a_1 + ib_1)(1, 0) + (a_2 + ib_2)(0, 1) = (a_1 + ib_1, a_2 + ib_2) \\ &\iff a_1 + ib_1 = 0 + i0 \text{ e } a_2 + ib_2 = 0 + i0, \end{aligned}$$

e

$$\begin{aligned} (0 + i0, 0 + i0) &= a_1(1, 0) + b_1(i, 0) + a_2(0, 1) + b_2(0, i) = (a_1 + ib_1, a_2 + ib_2) \\ &\iff a_1 = b_1 = a_2 = b_2 = 0. \end{aligned}$$

Logo, os conjuntos $\{(1, 0), (0, 1)\}$ e $\{(1, 0), (i, 0), (0, 1), (0, i)\}$ são linearmente independentes no contexto correspondente. Assim,

- Uma base do espaço linear *complexo* $\mathbb{C}^2$ é $\{(1,0),(0,1)\}$, e a dimensão de $\mathbb{C}^2$ como espaço linear complexo é 2.
- Uma base do espaço linear *real* $\mathbb{C}^2$ é $\{(1,0),(i,0),(0,1),(0,i)\}$, e a dimensão de $\mathbb{C}^2$ como espaço linear real é 4.

Analogamente, considerando $\mathbb{C}^n$ um espaço linear real, a sua dimensão é $2n$ e considerando $\mathbb{C}^n$ um espaço linear complexo a sua dimensão é igual a n.

Consideremos agora o seguinte subconjunto do espaço linear complexo $\mathbb{C}^2$

$$S = \{(1+i, 2i), (2, 2+2i)\}.$$

Uma vez que a dimensão de $\mathbb{C}^2$ é igual a 2, o subconjunto S será uma base se os vectores de S forem linearmente independentes (ver Teorema 3.3). Verifiquemos se S é ou não linearmente independente. Usando a definição de independência linear tem-se que $\alpha_1(1+i,2i)+\alpha_2(2,2+2i) = (0+0i, 0+0i)$, para escalares $\alpha_1 = a_1+ib_1$ e $\alpha_2 = a_2+ib_2$, é equivalente a

$$(a_1+ib_1)(1+i,2i)+(a_2+ib_2)(2,2+2i) = (0+0i,0+0i) \iff$$

$$\begin{aligned}((a_1-b_1+2a_2)+i(a_1+b_1+2b_2)\,,\,(-2b_1+2a_2-2b_2)+i(2a_1+2b_2+2a_2))\\ =(0+0i,0+0i)\end{aligned}$$

$$\iff \begin{cases} a_1-b_1+2a_2 & =0\\ a_1+b_1+2b_2 & =0\\ -2b_1+2a_2-2b_2 & =0\\ 2a_1+2b_2+2a_2 & =0.\end{cases}$$

Nas equivalências anteriores usámos as operações de adição e multiplicação de números complexos e as suas propriedades, bem como o facto de dois complexos serem iguais se as respectivas partes reais e imaginárias forem iguais. Se necessitar de rever estas operações pode consultar o Apêndice A. O sistema anterior admite a solução $(a_1,b_1,a_2,b_2) = (1,-1,-1,0)$, ou seja, a solução $\alpha_1 = 1-i$ e $\alpha_2 = -1$, pelo que S é linearmente dependente. Logo, S não é uma base do espaço linear complexo $\mathbb{C}^2$.

3. **Conjunto das matrizes reais do tipo** $p \times n$

   Seja $\mathcal{M}_{p\times n}$ o conjunto de todas as matrizes reais $p\times n$ munido das operações usuais de adição de matrizes e multiplicação de matrizes por um número real definidas pelas expressões (1.8) e (1.9) (pág. 38). Como vimos no final da Secção 3.1, as propriedades destas operações garantem que $\mathcal{M}_{p\times n}$ é um espaço linear real. Relembremos que o vector zero de $\mathcal{M}_{p\times n}$ é a matriz nula.

Consideremos o espaço $\mathcal{M}_{2\times 2}$. Qualquer matriz $\begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathcal{M}_{2\times 2}$ pode escrever-se como combinação linear de 4 matrizes do tipo 2×2:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = a \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + b \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + c \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + d \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Além disso, o conjunto ordenado $B = \left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right)$ é linearmente independente uma vez que

$$\alpha_1 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \alpha_2 \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + \alpha_3 \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + \alpha_4 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},$$

é equivalente a

$$\begin{bmatrix} \alpha_1 & \alpha_2 \\ \alpha_3 & \alpha_4 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \Longleftrightarrow \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0.$$

Logo, B é uma base de $\mathcal{M}_{2\times 2}$ e portanto $\dim \mathcal{M}_{2\times 2} = 4$. A base B é designada por base canónica de $\mathcal{M}_{2\times 2}$.

Como B é uma base, qualquer matriz A de $\mathcal{M}_{2\times 2}$ escreve-se de forma única como combinação linear dos vectores de B. Por exemplo, o vector das coordenadas da matriz $A = \begin{bmatrix} 1 & 5 \\ 4 & 3 \end{bmatrix}$ na base ordenada B é $\mathbf{A}_B = (1, 5, 4, 3) \in \mathbb{R}^4$.

4. **Conjunto das matrizes simétricas**

   Consideremos agora o conjunto $Sym(2)$ das matrizes reais e simétricas 2×2, isto é, matrizes A tais que $A^T = A$. Este conjunto é um subespaço de $\mathcal{M}_{2\times 2}$ já que a soma de duas matrizes simétricas ainda é uma matriz simétrica, e a multiplicação de uma matriz simétrica por um escalar real também é simétrica. Ou seja, $Sym(2)$ é fechado para a adição e multiplicação por escalares reais. Logo, $Sym(2)$ é um subespaço de $\mathcal{M}_{2\times 2}$ e portanto também é um espaço linear (ver Definição 3.3 e Nota 16).

Qualquer matriz simétrica, do tipo 2×2, pode escrever-se como uma combinação linear de três matrizes simétricas:

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix} = a\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + b\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + c\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = aS_1 + bS_2 + cS_3.$$

As matrizes simétricas S_1, S_2, S_3 são linearmente independentes como se verifica facilmente usando a definição de independência linear. Portanto, as matrizes S_1, S_2 e S_3 formam uma base do espaço linear das matrizes reais e simétricas do tipo 2×2. Assim, $Sym(2)$ é um espaço linear de dimensão 3.

Considerando a base ordenada $B = (S_1, S_2, S_3)$, o vector das coordenadas de $A = \begin{bmatrix} 2 & 5 \\ 5 & 3 \end{bmatrix} \in Sym(2)$ na base B é o vector de $\mathbb{R}^3$:

$$\mathbf{A}_B = (2, 5, 3).$$

5. **Conjunto das matrizes anti-simétricas**

Considere-se o conjunto das matrizes reais anti-simétricas do tipo 2×2, ou seja, matrizes que satisfazem a igualdade $A = -A^T$. Este conjunto é um subespaço de $\mathcal{M}_{2\times2}$ uma vez que é fechado para a adição e multiplicação por escalares reais. As matrizes reais anti-simétricas de segunda ordem são matrizes da forma

$$A = \begin{bmatrix} 0 & a \\ -a & 0 \end{bmatrix}, \quad a \in \mathbb{R}.$$

Estas matrizes podem escrever-se na forma

$$A = \begin{bmatrix} 0 & a \\ -a & 0 \end{bmatrix} = a\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix},$$

pelo que a matriz $C = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ gera o espaço linear das matrizes anti-simétricas de segunda ordem. Como um conjunto com um único vector não nulo é necessariamente linearmente independente, o conjunto $\{C\}$ é uma base do espaço linear das matrizes anti-simétricas de segunda ordem. Portanto, o espaço linear das matrizes anti-simétricas de ordem 2 tem dimensão é igual a 1.

6. Como vimos (Proposição 3.2), o núcleo de uma matriz (ou seja o conjunto das soluções de um sistema homogéneo) é um espaço linear. Contudo, o conjunto S das soluções de um sistema não homogéneo $A\mathbf{x} = \mathbf{b}$, com $\mathbf{b} \neq \mathbf{0}$, não é um espaço linear. De facto, se $\mathbf{u}, \mathbf{v}$ são soluções do sistema $A\mathbf{x} = \mathbf{b}$, tem-se $A\mathbf{u} + A\mathbf{v} = A(\mathbf{u} + \mathbf{v}) = 2\mathbf{b}$, e portanto $\mathbf{u} + \mathbf{v}$ não é solução do sistema. Assim, S não é fechado para a adição (nem para a multiplicação por escalares).

7. Seja P o conjunto dos polinómios de grau $n\,(n \geq 1)$ de coeficientes reais, munido das operações usuais de adição de polinómios e multiplicação de um polinómio por um número real.

   Este conjunto não é um espaço linear visto não ser fechado para a adição. Por exemplo, para $p_1(t) = t^n \in P$ e $p_2(t) = -t^n \in P$, tem-se $p_1 + p_2 = 0 \notin P$ (0 denota o polinómio constante igual a zero, que não tem grau ≥ 1).

8. **Conjunto dos polinómios de grau menor ou igual a n**

   O conjunto P_n dos polinómios de coeficientes reais de grau menor ou igual a n, munido das operações usuais de adição de polinómios e de multiplicação de um polinómio por um número real, é um espaço linear real.

   É óbvio que qualquer polinómio $p(t) = a_0 + a_1 t + \cdots a_n t^n$ de P_n se escreve como combinação linear dos monómios do conjunto $\{1, t, t^2, \ldots, t^n\}$. Este conjunto é linearmente independente já que, a equação

   $$\alpha_0 + \alpha_1 t + \cdots \alpha_n t^n = 0 \times 1 + 0 \times t + \cdots + 0 \times t^n, \quad \text{para todo } t \in \mathbb{R},$$

   é equivalente a $\alpha_0 = \cdots = \alpha_n = 0$. A base $\{1, t, t^2, \ldots, t^n\}$ é designada por *base canónica* de P_n. Assim, $\dim P_n = n + 1$.

   Consideremos o polinómio $p(t) = 3t^2 - t^3$ do espaço linear P_3 (dos polinónios de grau menor ou igual a 3). O vector das coordenadas de p na base ordenada $BC = (1, t, t^2, t^3)$ é $\mathbf{p}_{BC} = (0, 0, 3, -1)$, já que

   $$p(t) = 0 \times 1 + 0 \times t + 3t^2 - t^3.$$

   Consideremos agora o seguinte subconjunto S de P_3

   $$S = \left\{1 + t^2 + 4t^3, -1 + t + 2t^3, t + t^2 + 6t^3, 2 + 3t + 2t^2 + t^3\right\}.$$

   Pretendemos saber se este conjunto de quatro polinómios é ou não uma base de P_3. Para tal, vamos fixar a base canónica em P_3 e usar a correspondência biunívoca, estabelecida na página 138, entre vectores de P_3 e vectores de $\mathbb{R}^4$. Os vectores das coordenadas, na base (canónica) BC de P_3, dos vectores de S são:

$$\begin{array}{lcl} p_1(t) = 1 + t^2 + 4t^3 & \rightsquigarrow & (\mathbf{p_1})_{BC} = (1, 0, 1, 4) \\ p_2(t) = -1 + t + 2t^3 & \rightsquigarrow & (\mathbf{p_2})_{BC} = (-1, 1, 0, 2) \\ p_3(t) = t + t^2 + 6t^3 & \rightsquigarrow & (\mathbf{p_3})_{BC} = (0, 1, 1, 6) \\ p_4(t) = 2 + 3t + 2t^2 + t^3 & \rightsquigarrow & (\mathbf{p_4})_{BC} = (2, 3, 2, 1). \end{array}$$

Da Proposição 3.9 sabemos que os vectores de S são linearmente independentes se e só se os respectivos vectores das coordenadas são linearmente independentes. Colocando os vectores das coordenadas numa matriz, por exemplo, por linhas, temos

$$A = \begin{bmatrix} 1 & 0 & 1 & 4 \\ -1 & 1 & 0 & 2 \\ 0 & 1 & 1 & 6 \\ 2 & 3 & 2 & 1 \end{bmatrix}. \tag{3.26}$$

Como a terceira linha de A é a soma das duas primeiras, as linhas de A são linearmente dependentes, e por conseguinte o conjunto S é linearmente dependente. Assim, S não é uma base de P_3.

Determinamos agora um subconjunto de S, com o maior número possível de elementos, que seja linearmente independente. Conforme segue da demonstração da Proposição 3.12, basta determinar uma submatriz P de A, invertível, e de ordem igual à característica de A. As linhas de A correspondentes às linhas de P são linearmente independentes (ver Nota 22, pág. 150).

Considere-se a submatriz P que se obtém de A suprimindo a primeira linha e a primeira coluna de A. A matriz P tem determinante diferente de zero ($\det P = -13$) e portanto a característica de A é 3, conforme decorre da Proposição 3.12. Além disso, a segunda, terceira e quarta linhas de A são linearmente independentes, e portanto o subconjunto $S' = \{p_2, p_3, p_4\} \subset S$ é linearmente independente.

Alternativamente poderíamos ter usado o método de eliminação de Gauss para determinar uma base para o espaço das linhas da matriz A, e os polinómios correspondentes aos vectores dessa base formariam o subconjunto S' pretendido.

9. Seja S o conjunto dos polinómios p de grau menor ou igual a n tais que $p(0) = 1$. Ou seja,

$$S = \{p \in P_n : \; p(0) = 1\} = \{1 + a_1 t + a_2 t^2 + \ldots + a_n t^n\}.$$

S não é um subespaço de P_n uma vez que o vector zero de P_n não pertence a S, isto é, o polinómio constante igual a zero não pertence a S.

10. **Conjunto das funções reais de variável real**

O conjunto $\mathcal{F}$ de todas as funções reais de variável real munido das operações

$$(f+g)(x) = f(x) + g(x), \quad (\alpha \cdot f)(x) = \alpha f(x), \quad \alpha \in \mathbb{R}, f, g \in \mathcal{F},$$

é um espaço linear real. É fácil verificar que as operações acima satisfazem as propriedades da definição de espaço linear, e que o elemento neutro da adição é a função (constante) que toma o valor zero em todos os pontos de $\mathbb{R}$.

Considere-se em $\mathcal{F}$ o subconjunto $S = \{\cos^2 x, \operatorname{sen}^2 x, \cos(2x)\}$. O conjunto S não é linearmente independente, uma vez que $\cos(2x)$ é uma combinação linear de $\cos^2 x$ e $\operatorname{sen}^2 x$ (relembre que $\cos(2x) = \cos^2 x - \operatorname{sen}^2 x$).

Exercícios

1. Determine um conjunto gerador para o núcleo das matrizes:

a) $A = \begin{bmatrix} 2 & 1 & -2 & 0 \\ 4 & 2 & -4 & 0 \end{bmatrix}$,

b) $B = \begin{bmatrix} 2 & 6 & 0 & 2 \\ 0 & 1 & 2 & 0 \\ 1 & 3 & 0 & 1 \end{bmatrix}$,

c) $C = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & -3 \end{bmatrix}$.

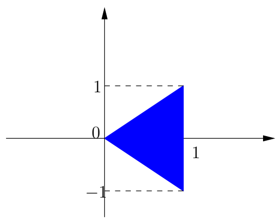
2. Exprima cada um dos seguintes vectores de $\mathbb{R}^3$ como combinação linear de $\mathbf{u} = (2,1,4)$, $\mathbf{v} = (1,-1,3)$ e $\mathbf{w} = (3,2,5)$.

a) $(-9,-7,-15)$
b) $(6,11,6)$
c) $(0,0,0)$

3. Considere $\mathbf{v}_1 = (2,1,0,3)$, $\mathbf{v}_2 = (3,-1,5,2)$ e $\mathbf{v}_3 = (-1,0,2,1)$, vectores de $\mathbb{R}^4$. Quais dos vectores seguintes pertencem ao espaço $\mathrm{Span}(\{\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}\})$?

a) $(2,3,-7,3)$
b) $(0,0,0,0)$
c) $(1,1,1,1)$
d) $(-4,6,-13,4)$

4. Para o triângulo T indicado na figura diga quais das afirmações são correctas.



I. T é um subespaço linear de $\mathbb{R}^2$.
II. A recta que passa pelos vértices $(0,0)$ e $(1,-1)$ de T é um subespaço linear de $\mathbb{R}^2$.
III. A recta que passa pelos vértices $(1,-1)$ e $(1,1)$ de T é um subespaço linear de $\mathbb{R}^2$.
IV. O conjunto formado pelos vectores $(0,0)$ e $(1,1)$ é uma base de $\mathbb{R}^2$.

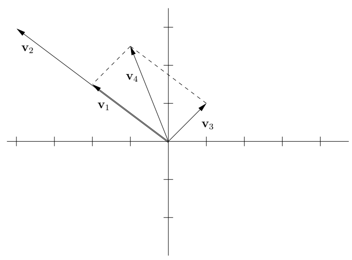
5. Na figura seguinte estão representados os vectores $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ e $\mathbf{v}_4$. Considere ainda os conjuntos

$$A = \{\mathbf{v}_1, \mathbf{v}_2\}, \quad B = \{\mathbf{v}_1, \mathbf{v}_3\}$$

$$C = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_4\}, \quad D = \{\mathbf{v}_2, \mathbf{v}_4\}$$

$$E = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}.$$

Diga quais destes conjuntos geram $\mathbb{R}^2$.



6. Diga, justificando, quais dos seguintes conjuntos são subespaços lineares de $\mathbb{R}^4$.

a) O conjunto de vectores da forma $(a,0,0,1)$ com a real
b) O conjunto de vectores da forma $(a,b,0,0)$ com a,b reais.
c) O conjunto de vectores da forma (a,b,c,d) com $b = a+c-d$ e $c = 2d$, sendo a,b,c,d reais.
d) O conjunto de vectores da forma (a,b,c,d) com a,b,c,d positivos.

e) O conjunto de vectores da forma (a, b, c, d) com $c = a+b+1$ e $d = 2a-b$, sendo a, b, c e d reais.

7. Dê um exemplo de uma matriz A do tipo 2×2, e de um vector $\mathbf{b}$ de $\mathbb{R}^2$ tal que $\mathbf{b}$ não pertença ao conjunto gerado pelas colunas de A.

8. Construa uma matriz 3×3, que não esteja em escada por linhas, cujas colunas não geram $\mathbb{R}^3$.

9. Construa uma matriz simétrica 3×3, de característica 2 e tal que $(4, 2, 6)$ pertença ao espaço das colunas da matriz.

10. Considere os vectores $\mathbf{u} = (2, 1, 3, 4)$ e $\mathbf{v} = (4, 1, 6, 5)$.

Acrescente vectores da base canónica de $\mathbb{R}^4$ ao conjunto $S = \{\mathbf{u}, \mathbf{v}\}$ por forma a obter uma base de $\mathbb{R}^4$.

11. Seja $S = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$ o subconjunto de $\mathbb{R}^4$ formado pelos vectores:

$$\mathbf{u}_1 = (1, 0, 1, -1), \quad \mathbf{u}_2 = (0, 1, 1, -1),$$

$$\mathbf{u}_3 = (-1, -1, -2, 2), \quad \mathbf{u}_4 = (-2, 0, -2, 1).$$

Indique o valor lógico das afirmações seguintes.

a) S é uma base de $\mathbb{R}^4$.

b) S gera $\mathbb{R}^4$.

c) Removendo de S o vector $\mathbf{u}_3$ e acrescentando o vector $(0, 0, 1, 0)$, obtém-se uma base de $\mathbb{R}^4$.

d) Acrescentando a S o vector

$$\mathbf{w} = (0, 0, 0, 1),$$

obtém-se um conjunto que gera $\mathbb{R}^4$.

12. Verifique se $S = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$ é linearmente dependente. Caso o seja, indique um subconjunto linearmente independente com o maior número possível de vectores, e escreva os restantes elementos de S como combinação linear desses vectores.

a) Em $\mathbb{R}^4$, $\mathbf{u}_1 = (1, 1, 1, 1)$, $\mathbf{u}_2 = (1, 2, 2, 2)$ $\mathbf{u}_3 = (1, 2, 3, 3)$ e $\mathbf{u}_4 = (1, 2, 3, 4)$.

b) Em $\mathbb{R}^3$, $\mathbf{u}_1 = (0, 1, 2)$, $\mathbf{u}_2 = (1, 0, 2)$, $\mathbf{u}_3 = (1, 2, 0)$ e $\mathbf{u}_4 = (1, 2, 2)$.

13. Diga, justificando, quais dos seguintes conjuntos são subespaços lineares dos espaços vectoriais indicados.

a) $\{(x, y) \in \mathbb{R}^2 : 2x + y = 0\}$.

b) $\{(x, y) \in \mathbb{R}^2 : xy = 0\}$.

c) $\{(x, y, z) \in \mathbb{R}^3 : x+2y-z = 0$ e $x-2y-z = 0\}$.

d) $\{(x, y, z, w) \in \mathbb{R}^4 : x+y+z+w = 2\}$.

14. Determine a dimensão e uma base para a solução geral de cada um dos sistemas seguintes.

a) $\begin{cases} x + y - z &= 0 \\ -2x - y + 2z &= 0 \\ -x + z &= 0 \end{cases}$

b) $\begin{cases} 3x + y + z + t &= 0 \\ 5x - y + z - t &= 0. \end{cases}$

15. Para cada uma das matrizes seguintes, encontre a dimensão e uma base para: o núcleo da matriz; o espaço gerado pelas linhas da matriz; e para o espaço gerado pelas colunas da matriz.

$$A = \begin{bmatrix} 1 & -3 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$$B = \begin{bmatrix} 2 & 1 & 5 \\ 6 & 3 & -8 \end{bmatrix},$$

$$C = \begin{bmatrix} 2 & -1 & 0 & 1 \\ 3 & 1 & -1 & 0 \\ 5 & 0 & -1 & 1 \end{bmatrix}.$$

16. Utilize a informação da tabela seguinte para determinar em cada um dos casos a dimensão do espaço gerado pelas linhas da matriz A, do espaço gerado pelas colunas de A, do núcleo de A e do núcleo de A^T (matriz transposta de A).

	(a)	(b)	(c)
A	3×3	3×3	3×3
car A	3	2	1

17. Seja A uma matriz $n \times p$, tal que a dimensão do núcleo de A é 2, a dimensão do núcleo de A^T é 1, e a dimensão do espaço das linhas de A é 2. Quais dos valores seguintes são válidos?

a) $n = 4$ e $p = 5$ b) $n = 5$ e $p = 4$.
c) $n = 3$ e $p = 4$. d) $n = 4$ e $p = 3$.

18. Utilize a informação da tabela seguinte para determinar se o correspondente sistema de equações lineares não homogéneo $A\mathbf{x} = \mathbf{b}$ é possível. Em caso afirmativo, indique o número de variáveis livres da solução geral.

	(a)	(b)	(c)	(d)
A	5×9	9×5	4×4	6×2
car A	2	2	0	2
car$(A\,\|\mathbf{b})$	2	3	0	2

19. Sejam $\mathbf{u}, \mathbf{v}$ e $\mathbf{w}$ os vectores coluna de uma matriz A, com $\mathbf{u}$ e $\mathbf{v}$ linearmente independentes e $\mathbf{w} = 2\mathbf{u} - \mathbf{v}$. Indique o valor lógico das afirmações seguintes.

a) $\det(A) \neq 0$
b) Existe uma submatriz de A de ordem 2 invertível.
c) $\mathrm{car}(A) = 1$.
d) Uma base para o espaço das linhas de A é $\{\mathbf{u}, \mathbf{v}\}$.
e) A dimensão do espaço das linhas de A é 2.

20. Seja A uma matriz real 4×4 e $\mathbf{b} \in \mathbb{R}^4$. Responda às questões seguintes.

a) Se o espaço das colunas de A não é $\mathbb{R}^4$, que pode dizer a respeito do núcleo de A?
b) Se o núcleo de A não é o subespaço $\{\mathbf{0}\}$, que pode dizer a respeito do espaço das colunas de A?
c) Se o espaço das colunas de A é $\mathbb{R}^4$, que pode dizer a respeito das soluções do sistema $A\mathbf{x} = \mathbf{b}$?
d) Se o núcleo de A é $\{\mathbf{0}\}$, que pode dizer a respeito das soluções do sistema $A\mathbf{x} = \mathbf{b}$?

21. Sejam A e B duas matrizes 3×3 tais que $\mathrm{car}(A) = 2$ e $\mathrm{car}(B) = 3$. Indique o valor lógico das afirmações seguintes.

a) A e B são invertíveis.
b) O núcleo de B é $\{\mathbf{0}\}$.
c) O sistema $A\mathbf{x} = \mathbf{0}$ tem grau de indeterminação 1.
d) Existem matrizes invertíveis P e Q tais que $A = PBQ$.

22. Determine as coordenadas de $\mathbf{x}$ na base ordenada B para:

a) $\mathbf{x} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}, B = \left(\begin{bmatrix} 5 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right);$

b) $\mathbf{x} = \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}, B = \left(\begin{bmatrix} 1 \\ -1 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} \right).$

23. Para as bases B indicadas, determine o vector que tem por vector das coordenadas o vector $\mathbf{x}_B$ dado.

a) $B = \left(\begin{bmatrix} 5 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right), \mathbf{x}_B = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$

b) $B = \left(\begin{bmatrix} 1 \\ -1 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} \right)$,

$\mathbf{x}_B = \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix}$.

24. Seja $B = (\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$ uma base ordenada de $\mathbb{R}^3$, onde $\mathbf{u}_1 = (2,1,0)$, $\mathbf{u}_2 = (-1,1,1)$, e $\mathbf{u}_3 = (1,1,1)$. Determine a matriz de mudança da base B para a base canónica de $\mathbb{R}^3$.

25. Determine a matriz que efectua a mudança da base canónica de $\mathbb{R}^2$ para a base $B = ((2,-2),(3,4))$. Use essa matriz para obter $\mathbf{w}_B$ onde $\mathbf{w} = (2,2)$.

26. Diga se é verdadeira ou falsa cada uma das proposições seguintes, justificando a sua resposta.

a) O conjunto de todas as soluções de um sistema homogéneo de n equações lineares a k incógnitas (com $k \neq n$) é um subespaço de $\mathbb{R}^n$.

b) As colunas de uma matriz real do tipo $n \times n$ invertível, formam uma base de $\mathbb{R}^n$.

27. Exprima a matriz $\begin{bmatrix} 5 & 9 \\ 0 & 5 \end{bmatrix}$ como combinação linear de $\begin{bmatrix} 2 & 1 \\ 0 & 4 \end{bmatrix}$, $\begin{bmatrix} 1 & -1 \\ 0 & 3 \end{bmatrix}$ e $\begin{bmatrix} 3 & 2 \\ 0 & 5 \end{bmatrix}$. Verifique ainda se estas três últimas matrizes formam ou não uma base para o espaço linear das matrizes 2×2 triangulares superiores.

28. Seja

$$V = \left\{ (x,y,z) \in \mathbb{R}^3 \, : \, 2x + y - z = 0 \right\}.$$

Diga qual das afirmações seguintes é verdadeira.

a) V não é subespaço linear de $\mathbb{R}^3$.

b) $\{(1,2,0),(0,1,1)\}$ é uma base de V.

c) $\{(1,2,0),(0,1,1),(1,3,1)\}$ é uma base de V.

d) $\{(1,0,2),(0,1,1)\}$ é uma base de V.

29. Sejam u, v, w vectores linearmente independentes de um espaço linear real U. Diga qual das afirmações seguintes é necessariamente verdadeira.

a) A dimensão de U é 3.

b) A dimensão de $\operatorname{Span}(\{u,v,w,u+v\})$ é 4.

c) A dimensão de $\operatorname{Span}(\{u,v,w,u+v\})$ é igual à dimensão de $\operatorname{Span}(\{u,v,w\})$.

d) A dimensão de U é inferior a 3.

30. Determine se os vectores seguintes são ou não linearmente independentes. Caso o não sejam, indique um subconjunto linearmente independente com o maior número possível de elementos.

a) No espaço dos polinómios de grau menor ou igual a 3: $p_1(t) = 1, p_2(t) = 1 + t$, $p_3(t) = 1 + t + t^2$ e $p_4(t) = 1 + t + t^2 + t^3$.

b) No espaço $M_{2\times 2}$, das matrizes reais de segunda ordem:

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \text{ e } \begin{bmatrix} 0 & 0 \\ 0 & -5 \end{bmatrix}.$$

31. Considere a base ordenada

$$B = \left(1, 1+t, 2t+t^2\right)$$

do espaço linear dos polinómios de grau menor ou igual a 2. As coordenadas de $p(t) = 3 + t + t^2$ na base B são:

a) $(8,-3,1)$ b) $(5,-1,1)$

c) $(6, -1, 1)$ d) $(4, -1, 1)$

32. Recorde que $M_{2\times 2}$ designa o espaço das matrizes reais de segunda ordem. Diga quais dos seguintes conjuntos são subespaços lineares do espaço linear indicado, e em caso afirmativo diga qual a sua dimensão e indique uma base.

a) $\mathrm{W} = \{(x, y, z, w) \in \mathbb{R}^4 : x-y+z-w = 0,\ -4y + z = 0,\ x - w = 0\}$.
b) $W = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in M_{2\times 2} : a \in \mathbb{Z} \right\}$.
c) Em $M_{2\times 2}$, o conjunto das matrizes invertíveis.
d) No espaço linear dos polinómios reais de variável real de grau menor ou igual a 5, o conjunto dos polinómios $a_0+a_1t+a_2t^2$, tais que $a_0+a_1 = 0$.
e) $\mathrm{Span}(S)$, onde S é o subconjunto do espaço das funções reais de variável real, com as operações usuais, definido por $S = \{\cos^2(t)-\mathrm{sen}^2(t), \cos(2t)+\mathrm{sen}(t), \mathrm{sen}(t)\}$.

33. Seja V o espaço linear dos polinómios reais de variável real de grau menor ou igual a 3, com as operações usuais de adição e multiplicação por um número real.

a) Determine uma base de V e indique qual a dimensão de V. Calcule as coordenadas do polinómio $(1-t)(1+t)$ na base escolhida.
b) Considere o subconjunto $S \subset V$ definido por $S = \{1 - 2t, 1 + t^2, t, 1 + 2t - 3t^2, t^2\}$. Diga, justificando, se S é uma base de V.
c) Diga qual a dimensão do espaço linear $\mathrm{Span}(S)$, e determine uma base para esse espaço.
d) Seja W o subconjunto de todos os polinómios de V que não se anulam em 0. Diga se W é um subespaço linear de V. Em caso afirmativo, indique a sua dimensão e uma base.

34. Considere as bases ordenadas $B_1 = (1+t, t)$ e $B_2 = (t - 1, t + 2)$ do espaço linear real dos polinómios de coeficientes reais de grau menor ou igual a 1. A matriz de mudança da base B_1 para a base B_2 é:

a) $\begin{pmatrix} -1 & \frac{2}{3} \\ 0 & \frac{1}{3} \end{pmatrix}$. b) $\begin{pmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} \end{pmatrix}$.
c) $\begin{pmatrix} -1 & 2 \\ 1 & 1 \end{pmatrix}$. d) $\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$.

35. Para cada par de subespaços U e V encontre a dimensão e uma base para $U + V$ e para $U \cap V$.

a) $U = \mathrm{Span}\{(1, 0, 2, 0)\}$ e
$V = \{(x, y, z, w) \in \mathbb{R}^4 : y + 2z - w = 0,$
$-y + 3w = 0 \text{ e } z = 0\}$.
b) $U = \mathrm{Span}\{(-1, 1, 1, 1), (1, 0, -1, 0)\}$
e
$V = \{(x, y, z, w) \in \mathbb{R}^4 : -x + y - 2w = 0 \text{ e } y - z = 0$

Capítulo 4

Valores e vectores próprios

Os valores e vectores próprios desempenham um papel central em diversas áreas da matemática aplicada, da física, da economia, da engenharia etc.. A diagonalização de matrizes, intimamente relacionada com os conceitos de valores e vectores próprios, desempenha um papel crucial no estudo de sistemas de equações diferenciais, nomeadamente no estudo qualitativo da dinâmica do sistema definido por essas equações.

Na Secção 4.1 estudam-se as propriedades que resultam da definição de valor e vector próprio de uma matriz. A Secção 4.2 é dedicada ao problema da diagonalização de matrizes. Mais tarde, no Capítulo 7, estuda-se a diagonalização (ortogonal) de matrizes simétricas e aplica-se essa diagonalização na identificação de cónicas e superfícies.

A Secção 4.4 apresentamos algumas aplicações dos conceitos anteriormente tratados. Nessa secção, destacamos o papel dos valores e vectores próprios no estudo de sistemas dinâmicos (discretos e contínuos). Em particular, é ilustrado o papel dos valores e vectores próprios no algoritmo PageRank usado pelo motor de busca Google, e na previsão a prazo de sistemas descritos por matrizes de Markov, os quais modelam vários problemas de que a dinâmica de populações é um exemplo. A Secção 4.4.2, é dedicada à resolução de sistemas lineares de equações diferenciais ordinárias descritos por matrizes diagonalizáveis. Aproveitamos ainda para introduzir nessa secção a exponencial de matrizes.

4.1 Valores e vectores próprios de matrizes

Os valores próprios de uma matriz quadrada são escalares que satisfazem uma certa equação matricial. Recorde-se que quando nos referimos a escalares estamos a considerar números reais ou complexos. A cada valor próprio correspondem certos vectores que recebem a designação de vectores próprios. Os números complexos não podem ser evitados quando se lida com valores próprios, uma vez que mesmo uma matriz real pode ter valores próprios complexos. É assim essencial que o leitor possua alguns conhecimentos de números complexos (pelo que deve consultar o Anexo A caso necessite).

Eis a definição de valor e vector próprio de uma matriz.

Definição 4.1. Um escalar λ diz-se um *valor próprio* de uma matriz quadrada A se existe um vector não nulo $\mathbf{x}$, tal que

$$A\mathbf{x} = \lambda\mathbf{x}. \tag{4.1}$$

A um vector não nulo $\mathbf{x}$ que verifica a equação (4.1) chama-se *vector próprio* de A associado ao valor próprio λ.

O par $(\lambda, \mathbf{x})$ diz-se um *par próprio* de A se λ é um valor próprio de A e $\mathbf{x}$ é um vector próprio de A associado a λ.

O conjunto dos valores próprios de uma matriz A designa-se por *espectro* de A e denota-se por $\sigma(A)$.

A terminologia inglesa para valor e vector próprio é respectivamente *"eigenvalue"* e *"eigenvector"*, enquanto que em português do Brasil se usam as designações de autovalor e autovector.

Exemplo 4.1. O vector $\mathbf{x} = (2, 1)$ é um vector próprio da matriz $A = \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}$ já que

$$A\mathbf{x} = \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 10 \\ 5 \end{bmatrix} = 5 \begin{bmatrix} 2 \\ 1 \end{bmatrix} = 5\mathbf{x}.$$

Da igualdade anterior, concluimos que $\mathbf{x}$ é um vector próprio de A associado ao valor próprio $\lambda = 5$. Ou seja, $(5, \mathbf{x})$ é um par próprio de A.

♦

A equação $A\mathbf{x} = \lambda\mathbf{x}$ pode reescrever-se na seguinte forma equivalente

$$A\mathbf{x} = \lambda\mathbf{x} \Longleftrightarrow A\mathbf{x} - \lambda\mathbf{x} = \mathbf{0} \Longleftrightarrow (A - \lambda I)\mathbf{x} = \mathbf{0}, \tag{4.2}$$

onde I é a matriz identidade. Assim, a definição de vector próprio é equivalente à existência de uma solução não nula do sistema homogéneo $(A - \lambda I)\mathbf{x} = \mathbf{0}$. Ora, um sistema homogéneo com matriz dos coeficientes quadrada possui soluções não nulas se e só se o determinante da matriz dos coeficientes é nulo (Proposição 2.2, pág. 95). Podemos portanto enunciar a proposição:

Proposição 4.1. O escalar λ é um valor próprio da matriz quadrada A se e só se satisfaz a equação

$$\det(A - \lambda I) = 0. \tag{4.3}$$

Um vector próprio $\mathbf{x}$ associado ao valor próprio λ é uma solução não nula do sistema homogéneo $(A - \lambda I)\mathbf{x} = \mathbf{0}$.

A equação (4.3) e a solução geral do sistema $(A - \lambda I)\mathbf{x} = \mathbf{0}$, recebem designações que a seguir se especificam.

Definição 4.2. Chama-se *equação característica* da matriz A à equação (na variável λ)

$$\det(A - \lambda I) = 0.$$

O espaço

$$E(\lambda) = \{\mathbf{v} : \ (A - \lambda I)\mathbf{v} = \mathbf{0}\},$$

diz-se o *espaço próprio* do valor próprio λ.

Note-se que o espaço próprio $E(\lambda)$ é a solução geral do sistema $(A - \lambda I)\mathbf{x} = \mathbf{0}$, ou seja, o núcleo de $(A - \lambda I)$. Isto é,

$$\boxed{E(\lambda) = N(A - \lambda I).}$$

É importante observar que resulta imediatamente da igualdade $A\mathbf{x} = \lambda\mathbf{x}$ que, se A é uma matriz real e λ é um valor próprio complexo não real, um vector próprio $\mathbf{x}$ associado a λ é necessariamente um vector complexo. Nesse caso é conveniente encarar A como uma matriz complexa e $E(\lambda) = N(A - \lambda I)$ como subespaço de $\mathbb{C}^n$. Se λ é um valor próprio real da matriz real A, existem vectores próprios em $\mathbb{R}^n$ associados a λ, e nesse caso é usual tomar para $E(\lambda)$ apenas o conjunto dos vectores de $\mathbb{R}^n$ que satisfazem $A\mathbf{x} = \lambda\mathbf{x}$. Por isso, quando uma matriz real A só tem valores próprios reais, consideramos os espaços próprios $E(\lambda) = N(A - \lambda I)$ como subespaços de $\mathbb{R}^n$, caso contrário estes espaços são considerados subespaços de $\mathbb{C}^n$.

Exemplo 4.2. Determinar os valores próprios e os espaços próprios da matriz

$$A = \begin{bmatrix} 1 & -1/2 \\ -1/2 & 1 \end{bmatrix}.$$

Os valores próprios são as soluções da equação característica $\det(A - \lambda I) = 0$, isto é,

$$\det(A - \lambda I) = \begin{vmatrix} 1-\lambda & -1/2 \\ -1/2 & 1-\lambda \end{vmatrix} = (1-\lambda)^2 - \frac{1}{4} = 0 \Longleftrightarrow 1 - \lambda = \pm\frac{1}{2}$$
$$\Longleftrightarrow \lambda = \frac{3}{2} \text{ ou } \lambda = \frac{1}{2}.$$

Assim, os valores próprios de A são $\lambda_1 = \frac{3}{2}$ e $\lambda_2 = \frac{1}{2}$.

Os espaços próprios de $\lambda_1 = \frac{3}{2}$ e $\lambda_2 = \frac{1}{2}$ são, respectivamente, as soluções gerais dos sistemas homogéneos $(A - \lambda_i I)\mathbf{x} = \mathbf{0}$ para $i = 1, 2$. Assim,

$$\left(A - \frac{3}{2}I\right)\mathbf{x} = \mathbf{0} \Longleftrightarrow \begin{bmatrix} -1/2 & -1/2 \\ -1/2 & -1/2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Longleftrightarrow -a - b = 0.$$

Logo,

$$E\left(\frac{3}{2}\right) = \{(-b, b);\ b \in \mathbb{R}\} = \operatorname{Span}\{(-1, 1)\}.$$

Para $\lambda_2 = \frac{1}{2}$, tem-se

$$\left(A - \frac{1}{2}I\right)\mathbf{x} = \mathbf{0} \Longleftrightarrow \begin{bmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Longleftrightarrow a - b = 0.$$

Portanto,

$$E\left(\frac{1}{2}\right) = \{(b, b);\ b \in \mathbb{R}\} = \operatorname{Span}\{(1, 1)\}.$$

♦

A equação característica de uma matriz A de ordem n é uma equação polinomial uma vez que $\det(A - \lambda I)$ é um polinómio de grau n em λ. É fácil verificar que assim é usando a definição de determinante como a soma de produtos elementares de entradas (multiplicados pelo respectivo sinal). De facto, como a matriz $(A - \lambda I)$ difere da matriz A apenas nas entradas da diagonal principal, todos os produtos elementares de entradas de $(A - \lambda I)$ são polinómios em λ de grau menor ou igual a n e o único produto elementar de grau n é $(a_{11} - \lambda) \cdots (a_{nn} - \lambda)$. Assim, $\det(A - \lambda I)$ é a soma de um polinómio de grau n com polinómios de grau inferior, ou seja, $\det(A - \lambda I)$ é um polinómio de grau n.

Definição 4.3. Seja λ um escalar, A uma matriz $n \times n$, e I a matriz identidade de ordem n. O polinómio de grau n em λ, definido por $p(\lambda) = \det(A - \lambda I)$, é denominado *polinómio característico* de A.

A seguir sumarizamos algumas equivalências anteriormente referidas.

Proposição 4.2. Se A é uma matriz quadrada e λ um escalar, são equivalentes as afirmações:

a) λ é um valor próprio de A;

b) O sistema $(A - \lambda I)\mathbf{x} = \mathbf{0}$ tem soluções não nulas;

c) O núcleo de $(A - \lambda I)$ não é trivial, isto é, $N(A - \lambda I) \neq \{\mathbf{0}\}$;

d) Existe um vector $\mathbf{x}$ não nulo tal que $A\mathbf{x} = \lambda\mathbf{x}$;

e) λ é uma raiz do polinómio característico $p(\lambda) = \det(A - \lambda I)$.

O Teorema Fundamental da Álgebra[1] afirma que um polinómio (numa variável), de coeficientes complexos, de grau $n \geq 1$ tem n raízes (contando as raízes repetidas de acordo com a sua multiplicidade). Estas raízes podem ser simples ou múltiplas (com diferentes graus de multiplicidade). As raízes complexas de polinómios de coeficientes reais ocorrem em pares de conjugados. De facto, se p é um polinómio de coeficientes reais e $p(\lambda) = 0$, então $0 = \overline{p(\lambda)} = p(\overline{\lambda})$.

Por conseguinte, o polinómio característico de uma matriz A de ordem n,

$$p(\lambda) = \det(A - \lambda I) = b_0 + b_1\lambda + \cdots + b_{n-1}\lambda^{n-1} + (-1)^n\lambda^n, \tag{4.4}$$

tem n raízes $\lambda_1, \ldots, \lambda_n$, podendo por isso ser escrito como um produto de n factores:

$$p(\lambda) = \det(A - \lambda I) = (\lambda_1 - \lambda)(\lambda_2 - \lambda)\cdots(\lambda_n - \lambda). \tag{4.5}$$

Note-se que na factorização (4.5) pode haver factores repetidos.

De seguida introduzimos alguma terminologia usada para caracterizar valores próprios.

[1] Existe um grande número de provas do denominado Teorema Fundamental da Álgebra, algumas de natureza topológica, outras de natureza algébrica ou ainda de natureza analítica. As provas analíticas são do âmbito da Análise complexa e usam nomeadamente o Teorema do integral de Cauchy, ou o Teorema de Liouville ou ainda o designado princípio do argumento. No site `http://www.cut-the-knot.org/do_you_know/fundamental2.shtml`, poderá encontrar várias provas deste teorema bem como várias referências.

Definição 4.4. Seja λ um valor próprio da matriz A.

- A *multiplicidade algébrica* de λ é número de vezes que a raiz λ aparece repetida no polinómio característico de A. Isto é, $mult\,alg(\lambda_i) = k_i$ se e só se $p(\lambda) = (\lambda_1 - \lambda)^{k_1} \cdots (\lambda_s - \lambda)^{k_s}$, onde o espectro de A é $\sigma(A) = \{\lambda_1, \ldots, \lambda_s\}$, com $\lambda_i \neq \lambda_j$.
- λ diz-se um valor próprio *simples* quando $mult\,alg(\lambda) = 1$.
- A *multiplicidade geométrica* de λ é a dimensão do núcleo de $(A - \lambda I)$, isto é, $\dim E(\lambda)$. Dito de outra forma: $mult\,geom(\lambda)$ é o número máximo de vectores próprios linearmente independentes associados a λ.
- λ diz-se um valor próprio *semi-simples* quando $mult\,alg(\lambda) = mult\,geom(\lambda)$.

Na proposição seguinte apresentamos um resultado de utilidade prática, em particular quando se pretende decidir sobre a existência de valores próprios sem os calcular explicitamente. Para tal, é necessário definir o que se entende por traço de uma matriz quadrada.

Definição 4.5. Chama-se *traço* de uma matriz quadrada à soma das entradas da sua diagonal principal, e designamos por $\operatorname{tr}(A)$ o traço de A .

Proposição 4.3. Seja $A = [a_{ij}]_{i,j=1,\ldots,n}$ e $\lambda_1, \lambda_2, \ldots, \lambda_n$ os valores próprios de A. São satisfeitas as igualdades:

$$\begin{aligned} \det(A) &= \lambda_1\lambda_2\cdots\lambda_n, \\ \text{e} & \\ \operatorname{tr}(A) &= a_{11} + a_{22} + \cdots + a_{nn} = \lambda_1 + \lambda_2 + \cdots + \lambda_n. \end{aligned}$$

Antes de passarmos à demonstração desta proposição, notemos que no caso particular de uma matriz 2×2 a sua demontração é muito simples. O polinómio característico da matriz $A = [a_{ij}]_{i,j=1,2}$ é

$$\begin{aligned} p(\lambda) = \det(A - \lambda I) = \begin{vmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{vmatrix} &= (a_{11} - \lambda)(a_{22} - \lambda) - a_{12}a_{21} \\ &= \lambda^2 - \underbrace{(a_{11} + a_{22})}_{\text{traço de } A}\lambda + \underbrace{a_{11}a_{22} - a_{12}a_{21}}_{\det(A)}. \end{aligned}$$

Por outro lado, se λ_1 e λ_2 são as raízes de $p(\lambda)$, então podemos escrever o polinómio na forma

$$p(\lambda) = (\lambda_1 - \lambda)(\lambda_2 - \lambda) = \lambda^2 - (\lambda_1 + \lambda_2)\lambda + \lambda_1\lambda_2.$$

Comparando as duas expressões obtidas para $p(\lambda)$ segue o resultado enunciado na proposição.

Demonstração. O termo independente do polinómio característico de A, $p(\lambda) = \det(A - \lambda I)$, é $p(0) = \det(A)$. Por outro lado, usando a factorização de p em termos das suas raízes (expressão (4.5)) temos $p(0) = \lambda_1\lambda_2 \cdots \lambda_n$, ficando assim mostrado que o produto dos valores próprios é igual ao determinante da matriz.

Para provar a relação entre o traço da matriz e os seus valores próprios, note-se que usando a factorização (4.5), o coeficiente do termo λ^{n-1} de p é

$$(-1)^{n-1}(\lambda_1 + \lambda_2 + \cdots + \lambda_n).$$

Se mostrarmos que este coeficiente é igual a $(-1)^{n-1}(a_{11} + a_{22} + \cdots + a_{nn})$, provamos que o traço de A é igual à soma dos valores próprios. Para tal, vamos usar indução sobre n. Quando $n = 1$, o resultado é trivialmente satisfeito.

Suponha-se (hipótese de indução) que para qualquer matriz $A_{n-1} = [a_{ij}]$ de ordem $(n-1)$ o coeficiente do termo em λ^{n-2} do seu polinómio característico é $(-1)^{n-2}(a_{11} + a_{22} + \cdots + a_{n-1,n-1})$. Isto é,

$$\det(A_{n-1} - \lambda I) = (-1)^{n-1}\lambda^{n-1} + (-1)^{n-2}(a_{11} + a_{22} + \cdots + a_{n-1,n-1})\lambda^{n-2} + \text{ t.o.i.,} \tag{4.6}$$

onde t.o.i. designa termos de ordem inferior, ou seja, termos que envolvem potências λ^k com $k < n-2$.

Recorrendo ao desenvolvimento de Laplace (ver página 98) segundo a última linha da matriz A (de ordem n), tem-se

$$\begin{aligned}
\det(A - \lambda I) &= \begin{vmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{12} & a_{22} - \lambda & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda \end{vmatrix} \\
&= (a_{nn} - \lambda)\det(A_{n-1} - \lambda I) + a_{n1}C_{n1} + \cdots + a_{n,n-1}C_{n,n-1} \\
&= (a_{nn} - \lambda)\det(A_{n-1} - \lambda I) + a_{n1}q_1(\lambda) + \cdots + a_{n,n-1}q_{n-1}(\lambda),
\end{aligned}$$

onde os q_j's designam polinómios em λ de grau menor ou igual a $(n-2)$.

Usando a hipótese de indução, nomeadamente a expressão (4.6), o primeiro termo da soma anterior satisfaz a igualdade

$$(a_{nn} - \lambda)\det(A_{n-1} - \lambda I) = (-1)^n \lambda^n + \\ + (-1)^{n-1}\lambda^{n-1}(a_{11} + a_{22} + \cdots + a_{n-1,n-1} + a_{nn}) + \text{ t.o.i.}.$$

Finalmente, substituindo a expressão anterior na expressão de $\det(A - \lambda I)$, tem-se

$$\det(A-\lambda I) = (-1)^n\lambda^n + (-1)^{n-1}\lambda^{n-1}(a_{11}+a_{22}+\cdots+a_{n-1,n-1}+a_{nn}) + \text{ t.o.i.},$$

e portanto o enunciado é válido para qualquer n. □

Exemplo 4.3. Consideremos a matriz

$$A = \begin{bmatrix} 3 & 1 & 4 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

Como A é triangular superior, a matriz $(A - \lambda I)$ também o é, e portanto o seu determinante é igual ao produto das entradas da sua diagonal principal. Ou seja, o polinómio característico de A é $p(\lambda) = (3-\lambda)^2(2-\lambda)$. Assim, a matriz A tem:

- um valor próprio igual a 3 de multiplicidade algébrica dois.
- um valor próprio simples que é 2.

Refira-se que o polinómio característico é do terceiro grau e as suas três raízes são contadas considerando a raiz 3 duas vezes e a raiz 2 uma vez.

Confirmando os resultados da proposição anterior, tem-se

$$\det(A) = 3 \times 3 \times 2 = 18 \quad \text{e} \quad \operatorname{tr}(A) = 3 + 3 + 2 = 8,$$

onde a raiz repetida do polinómio característico é considerada de acordo com a sua multiplicidade. ♦

Nota 23. *No que se segue abreviamos por vezes o enunciado da proposição anterior dizendo que o determinante (resp. o traço) de uma matriz é igual ao produto (resp. a soma) dos valores próprios da matriz, subentendendo que os valores próprios múltiplos são considerados de acordo com as suas multiplicidades.*

É consequência imediata da proposição anterior o resultado que a seguir se enuncia.

Corolário 4.1. Uma matriz é invertível se e só se zero não é um valor próprio da matriz.

Exercício 4.1. Mostre que se $(\lambda, \mathbf{x})$ é um par próprio de uma matriz invertível A, então $\left(\frac{1}{\lambda}, \mathbf{x}\right)$ é um par próprio de A^{-1}. ▲

As matrizes reais podem ter valores próprios complexos (ver Exemplo 4.4). Sendo o polinómio característico de uma matriz real um polinónio real, as suas raízes complexas ocorrem em pares de conjugados. Isto significa que se $(a+ib)$ é um valor próprio complexo (não real) de uma matriz real, então o seu conjugado $(a-ib)$ também é valor próprio dessa matriz.

Antes de estabelecermos a relação existente entre vectores próprios correspondentes a valores próprios complexos conjugados definimos a matriz conjugada de uma matriz.

Definição 4.6. A matriz conjugada da matriz C é a matriz $\overline{C}$ cujas entradas são os conjugados das entradas de C.

O conjugado de um vector $\mathbf{u}$ é o vector $\overline{\mathbf{u}}$ cujas componentes são os conjugados das componentes de $\mathbf{u}$.

Refira-se que o conjugado de um número real coincide consigo próprio, e portanto se C é uma matriz real tem-se $C = \overline{C}$.

Proposição 4.4. Se $\lambda \in \mathbb{C}$ é um valor próprio de uma matriz real A, então $\overline{\lambda}$ também é um valor próprio de A. Além disso, se $\mathbf{u}$ é um vector próprio de A associado a $\lambda \in \mathbb{C}$, então $\overline{\mathbf{u}}$ é um vector próprio de A associado a $\overline{\lambda}$.

Demonstração. Seja $(\lambda, \mathbf{u})$ um par próprio de A, isto é, $A\mathbf{u} = \lambda\mathbf{u}$. Tomando o conjugado da igualdade $A\mathbf{u} = \lambda\mathbf{u}$, tem-se

$$\overline{A\mathbf{u}} = \overline{(\lambda\mathbf{u})} \iff \overline{A}\,\overline{\mathbf{u}} = \overline{\lambda}\,\overline{\mathbf{u}} \iff A\overline{\mathbf{u}} = \overline{\lambda}\,\overline{\mathbf{u}},$$

onde aplicámos a igualdade $\overline{A} = A$ uma vez que, por hipótese, A é real.

Por definição de valor e vector próprio de A, a igualdade $A\overline{\mathbf{u}} = \overline{\lambda}\,\overline{\mathbf{u}}$ significa que $\overline{\lambda}$ é um valor próprio de A e $\overline{\mathbf{u}}$ é um vector próprio associado. □

Exemplo 4.4. A matriz $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ possui os seguintes valores próprios:

$$\det(A - \lambda I) = \begin{vmatrix} -\lambda & 1 \\ -1 & -\lambda \end{vmatrix} = \lambda^2 + 1 = 0 \iff \lambda = i \text{ ou } \lambda = -i.$$

O espaço próprio $E(i)$ é o núcleo da matriz $(A - iI)$, ou seja o conjunto dos vectores $\mathbf{x}$ que verificam

$$(A - iI)\mathbf{x} = 0 \iff \begin{bmatrix} -i & -1 \\ 1 & -i \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \iff -ia = b.$$

Logo, $E(i) = \{(a, -ia) : a \in \mathbb{C}\} = \mathrm{Span}\{(1, -i)\}$. Como a valores próprios conjugados correspondem vectores próprios conjugados, tem-se

$$E(-i) = \left\{(\bar{a}, \overline{(-ia)}) : a \in \mathbb{C}\right\} = \{\bar{a}\,(1, i) : a \in \mathbb{C}\} = \mathrm{Span}\{(1, i)\}.$$

♦

4.1.1 Valores próprios e comportamento de $f(\mathbf{x}) = A\mathbf{x}$

Nesta secção analisamos algumas relações entre os valores e vectores próprios de uma matriz real A de ordem n e o comportamento da função $f : \mathbb{R}^n \to \mathbb{R}^n$ que aplica um vector $\mathbf{x} \in \mathbb{R}^n$ no vector $A\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x} \mapsto A\mathbf{x}$.

A função f, definida por $f(\mathbf{x}) = A\mathbf{x}$, é uma função linear, ou seja, uma função que verifica $f(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha f(\mathbf{x}) + \beta f(\mathbf{y})$ para quaisquer $\alpha, \beta \in \mathbb{R}$ e $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Como veremos no Capítulo 6, qualquer função linear f de $\mathbb{R}^n$ em $\mathbb{R}^n$ pode ser escrita na forma $f(\mathbf{x}) = A\mathbf{x}$, com A uma matriz $n \times n$.

Estudaremos aqui com algum detalhe dois casos: 1) A matriz A só tem valores próprios reais; 2) A matriz A é 2×2 e tem um par de valores próprios complexos conjugados. Como veremos ainda neste capítulo, estes dois casos são aqueles que importa estudar se se pretende entender o caso geral de funções f definidas por matrizes A diagonalizáveis.

Caso 1: A matriz A só tem valores próprios reais.

Seja λ um valor próprio real da matriz real A e $f(\mathbf{x}) = A\mathbf{x}$. O espaço próprio $E(\lambda)$ é o espaço gerado pelos vectores próprios associados ao valor próprio λ da matriz A, logo

$$f(\mathbf{x}) = A\mathbf{x} = \lambda\mathbf{x}, \quad \text{para todo} \quad \mathbf{x} \in E(\lambda).$$

A igualdade anterior diz-nos que, se $\mathbf{x} \in E(\lambda)$ então o vector $f(\mathbf{x})$ é um múltiplo de $\mathbf{x}$, e portanto $f(\mathbf{x})$ também pertence a $E(\lambda)$.

Ou seja, se $\lambda \in \mathbb{R}$ qualquer vector do espaço próprio $E(\lambda)$ é aplicado por f num vector do espaço próprio. Na Figura 4.1 ilustramos este facto.

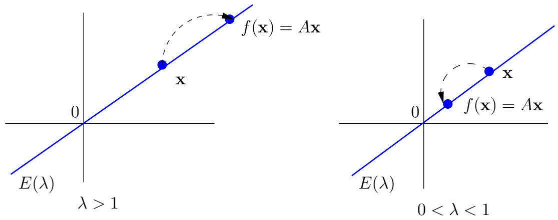


Figura 4.1: Os subespaços próprios de valores próprios reais são invariantes por f.

Um subconjunto S do domínio de uma função g diz-se *invariante* por g se qualquer vector de S é aplicado por g num vector de S .

Se A é uma matriz real de ordem n, os subespaços próprios correspondentes a valores próprios reais são subespaços de $\mathbb{R}^n$ invariantes por $f(\mathbf{x}) = A\mathbf{x}$.

Exemplo 4.5. A matriz A do Exemplo 4.2 (pág. 176) tem espectro $\sigma(A) = \{\frac{3}{2}, \frac{1}{2}\}$. A função $f(\mathbf{x}) = A\mathbf{x}$ é

$$f(\mathbf{x}) = f(x_1, x_2) = \begin{bmatrix} 1 & -1/2 \\ -1/2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 - \frac{x_2}{2} \\ \\ \frac{-x_1}{2} + x_2 \end{bmatrix}.$$

Ou seja, $f(x_1, x_2) = (x_1 - \frac{x_2}{2}, \frac{-x_1}{2} + x_2)$.

Os espaços próprios de A, obtidos no referido exemplo, são

$$E\left(\frac{3}{2}\right) = \text{Span}\{(-1,1)\} \quad \text{e} \quad E\left(\frac{1}{2}\right) = \text{Span}\{(1,1)\}.$$

Geometricamente, estes espaços próprios são rectas que passam pela origem e têm as direcções dos vectores $(-1, 1)$ e $(1, 1)$. Assim, a função f: "contrai" vectores na direcção definida por $E\left(\frac{1}{2}\right)$ já que, para $\mathbf{x} \in E\left(\frac{1}{2}\right)$ a respectiva imagem por f é $f(\mathbf{x}) = \frac{\mathbf{x}}{2}$; e "expande" vectores na direcção definida por $E\left(\frac{3}{2}\right)$, visto que $f(\mathbf{x}) = \frac{3\mathbf{x}}{2}$ para $\mathbf{x} \in E\left(\frac{3}{2}\right)$. Na Figura 4.2 ilustramos este facto.

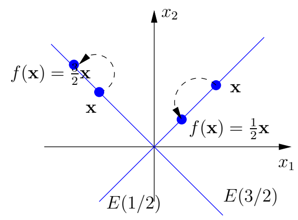


Figura 4.2: A função f "expande" vectores na direcção $E\left(\frac{3}{2}\right)$ e "contrai" na direcção $E\left(\frac{1}{2}\right)$.

Caso 2: A é uma matriz real, 2×2, e tem um par de valores próprios complexos conjugados.

Do caso anterior, sabemos que os subespaços próprios de valores próprios reais, de uma matriz real A, são subespaços invariantes para a função $f : \mathbb{R}^n \to \mathbb{R}^n$, definida por $f(\mathbf{x}) = A\mathbf{x}$. Quando os valores próprios de A são complexos, os respectivos espaços próprios não são subespaços de $\mathbb{R}^n$, apesar da função f estar definida de $\mathbb{R}^n$ em $\mathbb{R}^n$.

Consideremos a matriz A do Exemplo 4.4 (pág. 182) e a função $f(\mathbf{x}) = A\mathbf{x}$. Essa matriz tem valores próprios complexos $\pm i$, e a função $f(\mathbf{x}) = A\mathbf{x}$ é definida por

$$f(\mathbf{x}) = A\mathbf{x} \Longleftrightarrow f(x_1, x_2) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -x_2 \\ x_1 \end{bmatrix}.$$

Geometricamente, f actua no vector $\mathbf{x}$ rodando-o em torno da origem de um ângulo $\dfrac{\pi}{2}$ no sentido directo (ou anti-horário). A Figura 4.3 ilustra este facto. Como se observa neste exemplo, se aplicarmos f, sucessivamente, a um vector $\mathbf{x}$, ao fim de 4 aplicações voltamos a obter o vector $\mathbf{x}$. Na Figura 4.3 denotamos por

$$f^k(\mathbf{x}) = (\underbrace{f \circ f \circ \cdots \circ f}_{\text{k vezes}})(\mathbf{x}),$$

a composição de f consigo própria k vezes (isto é, a transformação obtida por k aplicações sucessivas de f). Note-se que sendo $f(\mathbf{x}) = A\mathbf{x}$ se tem $f^k(\mathbf{x}) = A^k\mathbf{x}$.

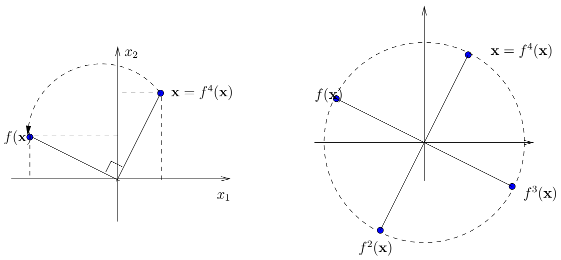


Figura 4.3: A função $f(\mathbf{x}) = A\mathbf{x}$, em que A tem valores próprios $\pm i$ representa uma rotação em $\mathbb{R}^2$.

É óbvio que, exceptuando a origem, nenhum vector de $\mathbb{R}^2$ é aplicado num múltiplo de si próprio. Ou seja, os únicos subespaços de $\mathbb{R}^2$ invariantes por f são $\{(0,0)\}$ e $\mathbb{R}^2$.

Consideremos agora uma matriz A do tipo 2×2 com valores próprios complexos $\lambda = a \pm ib$, com $b \neq 0$,

$$A = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}.$$

Sugere-se como exercício que verifique que esta matriz tem valores próprios $a \pm ib$.

Recorde ainda (Anexo A) que há uma correspondência biunívoca entre pontos do plano de coordenadas (a, b) e números complexos $a + ib$. Usando coordenadas polares, $a = \rho \cos\theta$ e $b = \rho \operatorname{sen}\theta$, o número complexo $\lambda = a + ib$ escreve-se na forma polar: $\lambda = \rho(\cos\theta + i \operatorname{sen}\theta)$. O valor $\rho = |\lambda| = \sqrt{a^2 + b^2}$ é a distância de λ à origem, e θ é o ângulo entre a parte positiva do eixo real e o ponto (a, b) (com $-\pi < \theta \leq \pi$). A Figura A.2 da página 512, ilustra a representação polar de um número complexo $a + ib$.

Assim, a matriz A pode escrever-se como um produto de duas matrizes

$$A = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} = \begin{bmatrix} \rho & 0 \\ 0 & \rho \end{bmatrix} \begin{bmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{bmatrix} = DR.$$

A matriz $R = \begin{bmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{bmatrix}$ é uma *matriz de rotação*, visto que para $\mathbf{x} \in \mathbb{R}^2$ o vector $R\mathbf{x}$ é o vector de $\mathbb{R}^2$ que se obtém rodando $\mathbf{x}$ de um ângulo θ, no sentido directo, em torno da origem (como facilmente se verifica).

A matriz $D = \begin{bmatrix} \rho & 0 \\ 0 & \rho \end{bmatrix}$ representa uma *expansão* se $\rho = |\lambda| > 1$, e uma *contracção* se $\rho = |\lambda| < 1$ já que, $D\mathbf{x} = \rho\mathbf{x}$.

Assim, a função $f(\mathbf{x}) = A\mathbf{x} = DR\mathbf{x}$ actua sobre um vector $\mathbf{x} \in \mathbb{R}^2$ rodando este vector e depois expandindo, contraindo ou mantendo-o, respectivamente nos casos em que $|\lambda| > 1, |\lambda| < 1$ e $|\lambda| = 1$. A Figura 4.4 ilustra esse comportamento de f sobre vectores de $\mathbb{R}^2$.

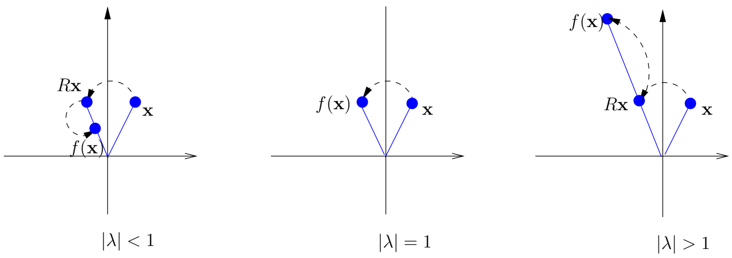


Figura 4.4: A matriz A tem λ como valor próprio complexo e $f(\mathbf{x}) = A\mathbf{x} = DR\mathbf{x}$.

Como veremos na secção seguinte, o comportamento geral de uma função $f : \mathbb{R}^n \to \mathbb{R}^n$ definida por $f(\mathbf{x}) = A\mathbf{x}$, em que A é uma matriz (real) diagonalizável é bem ilustrado pelos dois casos apresentados.

Consideremos agora um exemplo de uma matriz A possuindo valores próprios reais e complexos.

Exemplo 4.6. Consideremos $f(\mathbf{x}) = A\mathbf{x}$ com $A = \begin{bmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} & 0 \\ \frac{1}{2} & \frac{\sqrt{3}}{2} & 0 \\ 0 & 0 & 1.2 \end{bmatrix}$.

A matriz A tem valores próprios $\lambda_1 = \frac{\sqrt{3}}{2} + i\frac{1}{2}$, $\lambda_2 = \frac{\sqrt{3}}{2} - i\frac{1}{2}$ e $\lambda_3 = 1.2$.

O valor próprio λ_3 é real e o seu espaço próprio é gerado pelo vector $(0, 0, 1)$ (isto é, $E(1.2)$ é o eixo dos zz). Logo, vectores do eixo dos zz são aplicados por f em vectores do eixo dos zz por uma expansão de factor 1.2.

Atendendo à forma da matriz A (diagonal por blocos) é fácil verificar que f aplica vectores do plano xy em vectores deste plano. Como o valor próprio λ_1 (e portanto o seu conjugado λ_2) tem módulo $|\lambda_1| = |\lambda_2| = \sqrt{\frac{3}{4} + \frac{1}{4}} = 1$, a função f aplica um vector $\mathbf{u}$ do plano xy num vector que se obtém de $\mathbf{u}$ por rotação (em torno do eixo dos zz). O ângulo desta rotação é $\pi/6$ (note que $\operatorname{sen}\pi/6 = 1/2$). Por exemplo $f(1,1,0) = (\frac{\sqrt{3}-1}{2}, \frac{\sqrt{3}+1}{2}, 0)$. Para qualquer outro vector $\mathbf{v} = (s_1, s_2, s_3)$, a imagem por f deste vector é o vector $f(\mathbf{v})$ que tem terceira coordenada $1.2s_3$ (expansão na direcção do eixo dos zz), e duas primeiras coordenadas respectivamente, $\cos(\pi/6)s_1 - \operatorname{sen}(\pi/6)s_2$ e $\operatorname{sen}(\pi/6)s_1 + \cos(\pi/6)s_2$ (rotação de (s_1, s_2) de $\pi/6$ em torno da origem).

A Figura 4.5 ilustra aplicações sucessivas de f ao vector $\mathbf{u}$ do plano xy, ao vector $\mathbf{w}$ do eixo dos zz, e ao vector $\mathbf{v}$ que não pertence a estes subespaços. Esta figura ilustra ainda o facto das imagens de aplicações sucessivas de f a vectores $\mathbf{v}$ não pertencentes ao eixo dos zz nem ao plano xy, estarem sobre hélices inscritas num cilindro.

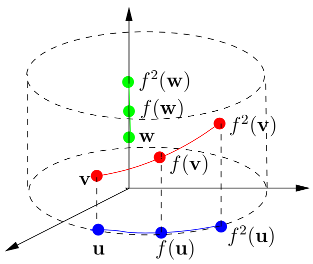


Figura 4.5: Aplicações sucessivas de $f(\mathbf{x}) = A\mathbf{x}$ para a matriz A do Exemplo 4.6.

♦

4.2 Diagonalização de matrizes

O problema central tratado nesta secção é o de saber se dada uma matriz de ordem n existe uma base de $\mathbb{C}^n$ formada por vectores próprios. Quando tal acontece a matriz diz-se diagonalizável, ou ainda, que a matriz é semelhante a uma matriz diagonal. O processo de diagonalização de matrizes desempenha um papel importante em álgebra linear sendo inúmeras as suas aplicações. Por exemplo, a diagonalização de matrizes é utilizada na interpretação da dinâmica de modelos físicos, em computação gráfica, e na identificação de cónicas e de superfícies quadráticas.

Definição 4.7. Duas matrizes quadradas A e B dizem-se *semelhantes* se existe uma matriz invertível P tal que

$$A = PBP^{-1}.$$

Definição 4.8. Uma matriz quadrada A diz-se *diagonalizável* se é semelhante a uma matriz diagonal D. Isto é, se existe uma matriz invertível P tal que

$$A = PDP^{-1}.$$

A uma matriz P tal que, $A = PDP^{-1}$ com D diagonal, chama-se *matriz que diagonaliza* A, ou matriz *diagonalizante* de A.

Comecemos por mostrar que os valores próprios de matrizes semelhantes são iguais.

Proposição 4.5. Matrizes semelhantes têm o mesmo polinómio característico. Em particular, os valores próprios são os mesmos e ocorrem com as mesmas multiplicidades.

Demonstração. Sejam A e B matrizes semelhantes. Isto é, existe uma matriz invertível P tal que $A = PBP^{-1}$. Tem-se

$$\begin{aligned}\det(A - \lambda I) &= \det(PBP^{-1} - \lambda I) = \det\left(P(B - \lambda I)P^{-1}\right)\\ &= \det P \det(B - \lambda I) \det(P^{-1}) = \det(B - \lambda I).\end{aligned}$$

Nas igualdades anteriores aplicámos os seguintes factos: o determinante do produto é igual ao produto dos determinantes; o determinante da inversa é o inverso do determinante. Da última igualdade segue que o polinómio característico de A é igual ao de B e portanto A e B possuem os mesmos valores próprios com as mesmas multiplicidades. □

Da Proposição 4.3 sabemos que o traço e o determinante de uma matriz de ordem n são respectivamente iguais à soma e ao produto dos n valores próprios da matriz, por conseguinte, da proposição anterior, segue o corolário que passamos a enunciar.

Corolário 4.2. Matrizes semelhantes têm o mesmo traço e o mesmo determinante.

Se A é uma matriz diagonalizável, isto é, $A = PDP^{-1}$ com D diagonal, pela Proposição 4.5 os valores próprios de D são os valores próprios de A. Como D é diagonal, os seus valores próprios são as entradas da diagonal principal. Logo, $D = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$, onde $\lambda_1, \lambda_2, \ldots, \lambda_n$ são os valores próprios de A.

No caso de A ser diagonalizável, a questão que agora se coloca é a de saber construir uma matriz P que diagonaliza A. O teorema seguinte mostra como construir uma tal matriz, oferecendo simultaneamente uma condição necessária e suficiente para que uma matriz seja diagonalizável.

Teorema 4.1. Uma matriz A do tipo $n \times n$ é diagonalizável se e só se possui n vectores próprios linearmente independentes. Ou seja, se e só se existe uma base de $\mathbb{C}^n$ constituída por vectores próprios de A.

Além disso, se $A = PDP^{-1}$ com $D = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$, então para todo $i = 1, \ldots, n$, a coluna i de P é um vector próprio de A associado ao valor próprio λ_i.

Demonstração. A igualdade $A = PDP^{-1}$ é equivalente a $AP = PD$. O produto AP é a matriz cujas colunas são o produto de A pelas colunas de P (ver Definição 1.12, pág. 39). Assim, designando as colunas de P por $\mathbf{c}_i$, tem-se

$$AP = A \begin{bmatrix} | & | & & | \\ \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ A\mathbf{c}_1 & A\mathbf{c}_2 & \cdots & A\mathbf{c}_n \\ | & | & & | \end{bmatrix}.$$

Por outro lado,

$$PD = \begin{bmatrix} | & | & & | \\ \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

$$= \begin{bmatrix} | & | & & | \\ \lambda_1\mathbf{c}_1 & \lambda_2\mathbf{c}_2 & \cdots & \lambda_n\mathbf{c}_n \\ | & | & & | \end{bmatrix}.$$

Logo, $AP = PD$ se e só se a i-ésima coluna de P verifica $A\mathbf{c}_i = \lambda_i\mathbf{c_i}$, para todo $i = 1, \ldots, n$. Ou seja, se e só se $\mathbf{c}_i$ é um vector próprio de A associado ao valor próprio λ_i.

A matriz P é invertível se e só se tem n colunas linearmente independentes (cf. Proposição 3.15, pág. 153). Conclui-se portanto que é necessário e suficiente para que A seja diagonalizável que existam n vectores próprios de A linearmente independentes. $\square$

Note-se que nem todas as matrizes são diagonalizáveis como se verifica no exemplo seguinte.

Exemplo 4.7. Considere a matriz $A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$.

Verifiquemos se esta matriz é ou não diagonalizável.

Uma vez que a matriz A é triangular, os valores próprios de A são $\lambda_1 = 2$ e $\lambda_2 = 1$ (as entradas da diagonal principal). O valor próprio 2 é simples e o valor próprio 1 tem multiplicidade algébrica 2, visto que o polinómio característico de A é $p(\lambda) = (2-\lambda)(1-\lambda)^2$.

Para que A seja diagonalizável têm de existir 3 vectores próprios linearmente independentes. Determinemos os espaços próprios.

$$(A - 2I)\mathbf{x} = \mathbf{0} \Longleftrightarrow \begin{bmatrix} 0 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Longleftrightarrow \begin{cases} -b + c = 0 \\ c = 0 \end{cases}$$

Logo,

$$E(2) = \left\{(a,b,c) \in \mathbb{R}^3 : \ b = c = 0\right\} = \{(a,0,0) : \ a \in \mathbb{R}\} = \operatorname{Span}\{(1,0,0)\}.$$

Para $\lambda_2 = 1$:

$$(A - I)\mathbf{x} = \mathbf{0} \Longleftrightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Longleftrightarrow \begin{cases} a = 0 \\ c = 0 \end{cases}$$

$$E(1) = \left\{(a,b,c) \in \mathbb{R}^3 : \ a = c = 0\right\} = \{(0,b,0) : \ b \in \mathbb{R}\} = \operatorname{Span}\{(0,1,0)\}.$$

Como a dimensão de cada espaço próprio é igual a 1, existem no máximo dois vectores próprios linearmente independentes (um vector retirado de cada espaço próprio). Ou seja, não existe um número suficiente de vectores próprios (que seria 3) para a matriz ser diagonalizável. Portanto, a matriz A não é diagonalizável. ♦

Pelo teorema anterior sabemos que é condição necessária e suficiente para uma matriz A, de ordem n, ser diagonalizável que possua n vectores próprios linearmente independentes, ou seja, que exista uma base do espaço linear complexo $\mathbb{C}^n$ formada por vectores próprios de A. Como veremos adiante (Proposição 4.6

e Corolário 4.4), esta condição é equivalente à soma das dimensões dos espaços próprios ser igual a n.

Na proposição seguinte mostramos que vectores próprios associados a valores próprios distintos são necessariamente linearmente independentes.

Proposição 4.6. Seja $\{\lambda_1, \lambda_2, \ldots, \lambda_k\}$ um conjunto de valores próprios distintos da matriz A de ordem n.

1) Se $\{(\lambda_1, \mathbf{x}_1), (\lambda_2, \mathbf{x}_2), \ldots, (\lambda_k, \mathbf{x}_k)\}$ é um conjunto de pares próprios de A, então o conjunto $S = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ é linearmente independente.

2) Se B_i é uma base de $E(\lambda_i)$, então o conjunto $B = B_1 \cup B_2 \cup \cdots \cup B_k$ é linearmente independente.

Demonstração. 1): Suponhamos, por absurdo, que S é linearmente dependente, e que $S' = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p\}$ é um subconjunto de S com o maior número possível de vectores linearmente independentes (isto é, S' é uma base de $\operatorname{Span} S$ cuja existência é assegurada pela Proposição 3.8, pág. 133).

Qualquer vector $\mathbf{x}_j$ de $S \setminus S'$ escreve-se de forma única como combinação linear dos vectores de S' (cf. Teorema 3.1). Ou seja, existem escalares α_i únicos tais que

$$\mathbf{x}_j = \alpha_1 \mathbf{x}_1 + \cdots + \alpha_p \mathbf{x}_p. \tag{4.7}$$

Multiplicando (à esquerda) a igualdade anterior pela matriz $(A - \lambda_j I)$, obtemos

$$\begin{aligned} (A - \lambda_j I)\mathbf{x}_j = \alpha_1 (A - \lambda_j I)\mathbf{x}_1 + \cdots + \alpha_p (A - \lambda_j I)\mathbf{x}_p \quad &\iff \\ \mathbf{0} = \alpha_1 (\lambda_1 - \lambda_j)\mathbf{x}_1 + \cdots + \alpha_p (\lambda_p - \lambda_j)\mathbf{x}_p, & \end{aligned}$$

já que $\mathbf{x}_i$ é um vector próprio de A associado a λ_i. Como S' é linearmente independente, segue da última igualdade que $\alpha_i(\lambda_i - \lambda_j) = 0$, para todo o $i = 1, \ldots, p$. Sendo distintos os valores próprios de A, obtemos de $\alpha_i(\lambda_i - \lambda_j) = 0$ que $\alpha_i = 0$ para todo $i = 1, \ldots, p$. Por conseguinte, de (4.7) resulta que $\mathbf{x}_j$ é o vector zero, o que contraria a hipótese de $\mathbf{x}_j$ ser vector próprio. Logo, S é linearmente independente.

2): Para mostrar que $B = B_1 \cup B_2 \cup \cdots \cup B_k$ é linearmente independente, basta provar que B é uma base de $E(\lambda_1) + \cdots + E(\lambda_k)$ (ver demostração da Proposição 3.10, pág. 141). Ou equivalentemente, mostrar que para $j = 1, \ldots, k$ se tem

$$X_j = E(\lambda_j) \cap (E(\lambda_1) + \cdots + E(\lambda_{j-1}) + E(\lambda_{j+1}) + \cdots + E(\lambda_k)) = \{\mathbf{0}\}\,.$$

Suponhamos, por absurdo, que existe um vector $\mathbf{x} \in X_j$ tal que $\mathbf{x} \neq \mathbf{0}$. Como o vector $\mathbf{x}$ pertence a $E(\lambda_j)$, tem-se

$$A\mathbf{x} = \lambda_j \mathbf{x}. \tag{4.8}$$

Uma vez que $\mathbf{x}$ também pertence a $E(\lambda_1)+\cdots+E(\lambda_{j-1})+E(\lambda_{j+1})+\cdots+E(\lambda_k)$, podemos escrever

$$\mathbf{x} = \mathbf{v}_1 + \cdots + \mathbf{v}_{j-1} + \mathbf{v}_{j+1} + \cdots + \mathbf{v}_k, \tag{4.9}$$

para certos vectores $\mathbf{v}_i \in E(\lambda_i)$. Multiplicando (à esquerda) a igualdade (4.9), respectivamente, por A e por λ_j, obtemos

$$\begin{aligned} A\mathbf{x} &= A\mathbf{v}_1 + \cdots + A\mathbf{v}_{j-1} + A\mathbf{v}_{j+1} + \cdots + A\mathbf{v}_k \\ &= \lambda_1\mathbf{v}_1 + \cdots + \lambda_{j-1}\mathbf{v}_{j-1} + \lambda_{j+1}\mathbf{v}_{j+1} + \cdots + \lambda_k\mathbf{v}_k \\ \lambda_j\mathbf{x} &= \lambda_j\mathbf{v}_1 + \cdots + \lambda_j\mathbf{v}_{j-1} + \lambda_j\mathbf{v}_{j+1} + \cdots + \lambda_j\mathbf{v}_k. \end{aligned}$$

De (4.8) segue que as últimas expressões são iguais, pelo que

$$(\lambda_1 - \lambda_j)\mathbf{v}_1 + + \cdots + (\lambda_{j-1} - \lambda_j)\mathbf{v}_{j-1} + (\lambda_{j+1} - \lambda_j)\mathbf{v}_j + \cdots + (\lambda_k - \lambda_j)\mathbf{v}_k = \mathbf{0}.$$

Como pelo item 1) os vectores $\mathbf{v}_1, \ldots, \mathbf{v}_{j-1}, \mathbf{v}_{j+1}, \ldots, \mathbf{v}_k$ são linearmente independentes, da última igualdade resulta que $(\lambda_i - \lambda_j) = 0$ para todo $i = 1, \ldots, k$ e $i \neq j$, o que contraria a hipótese dos valores próprios serem distintos. Logo, $X_j = \{\mathbf{0}\}$. Como j é qualquer, temos $X_j = \{\mathbf{0}\}$ para todo $j = 1, \ldots, k$, o que prova que B é uma base de $E(\lambda_1) + \cdots + E(\lambda_k)$. □

Como corolário da proposição anterior e do Teorema 4.1 podemos enunciar o seguinte resultado.

Corolário 4.3. Uma matriz $n \times n$ com n valores próprios distintos é diagonalizável.

Exemplo 4.8. Verifiquemos que $A = \begin{bmatrix} 4 & 0 & 1 \\ 2 & 3 & 2 \\ 1 & 0 & 4 \end{bmatrix}$ é diagonalizável, e determinemos uma matriz P que diagonaliza A.

O polinómio característico de A é

$$\det(A-\lambda I)=\begin{vmatrix}4-\lambda & 0 & 1\\ 2 & 3-\lambda & 2\\ 1 & 0 & 4-\lambda\end{vmatrix}=(3-\lambda)\begin{vmatrix}4-\lambda & 1\\ 1 & 4-\lambda\end{vmatrix}$$

$$=(3-\lambda)\left(\lambda^2-8\lambda+15\right)=(3-\lambda)^2(5-\lambda).$$

Assim, a matriz A tem valores próprios $\lambda_1=3$ e $\lambda_2=5$, com multiplicidades algébricas 2 e 1, respectivamente. Saliente-se ainda que no cálculo de $\det(A-\lambda I)$ aplicámos o desenvolvimento de Laplace utilizando a segunda coluna, o que produziu imediatamente uma factorização do polinómio (e portanto uma raiz).

Calculemos bases para os espaços próprios de A.

$$(A-3I)\mathbf{x}=\mathbf{0}\iff\begin{bmatrix}1 & 0 & 1\\ 2 & 0 & 2\\ 1 & 0 & 1\end{bmatrix}\begin{bmatrix}a\\ b\\ c\end{bmatrix}=\begin{bmatrix}0\\ 0\\ 0\end{bmatrix}\iff a+c=0$$

$$\begin{aligned}E(3)=\left\{(a,b,c)\in\mathbb{R}^3:\ a=-c\right\}&=\{(-c,b,c):\ b,c\in\mathbb{R}\}\\ &=\operatorname{Span}\{(-1,0,1),(0,1,0)\}.\end{aligned}$$

$$\begin{aligned}(A-5I)\mathbf{x}=\mathbf{0}\iff\begin{bmatrix}-1 & 0 & 1\\ 2 & -2 & 2\\ 1 & 0 & -1\end{bmatrix}\begin{bmatrix}a\\ b\\ c\end{bmatrix}=\begin{bmatrix}0\\ 0\\ 0\end{bmatrix}&\iff\begin{cases}-a \qquad\ + \ c=0\\ 2a-2b+2c=0\end{cases}\\ &\iff a=c \text{ e } b=2a.\end{aligned}$$

$$\begin{aligned}E(5)=\left\{(a,b,c)\in\mathbb{R}^3:\ a=c \text{ e } b=2a\right\}&=\{(a,2a,a):\ a\in\mathbb{R}\}\\ &=\operatorname{Span}\{(1,2,1)\}.\end{aligned}$$

Como $\{(-1,0,1),(0,1,0)\}$ e $\{(1,2,1)\}$ são bases, respectivamente, dos subespaços $E(3)$ e $E(5)$, tem-se

$$\dim E(3)=2 \quad \text{e} \quad \dim E(5)=1.$$

Conclui-se portanto que a multiplicidade geométrica de $\lambda=3$ é dois, e a de $\lambda=5$ é um. Por conseguinte, a matriz A só tem valores próprios semi-simples.

Existem três vectores próprios de A linearmente independentes, e portanto A é diagonalizável.

Uma matriz P que diagonaliza A (isto é, tal que $A = PDP^{-1}$ com D diagonal), é uma matriz cujas colunas formam uma base constituída por vectores próprios de A. É claro que P depende da forma como se constrói a matriz D. Assim, se escolhermos $D = \operatorname{diag}(3, 5, 3)$, a matriz P possui, na 1ª e 3ª colunas vectores próprios associados ao valor próprio $\lambda_1 = 3$, e na segunda coluna um vector próprio associado a $\lambda_2 = 5$. Para que P seja invertível, temos de escolher vectores próprios linearmente independentes. Por exemplo,

$$P = \begin{bmatrix} -1 & 1 & 0 \\ 0 & 2 & 1 \\ 1 & 1 & 0 \end{bmatrix} \quad \text{e} \quad D = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

Também podemos considerar, por exemplo,

$$P = \begin{bmatrix} -1 & 0 & 1 \\ 0 & 1 & 2 \\ 1 & 0 & 1 \end{bmatrix} \quad \text{e} \quad D = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix},$$

correspondendo a uma outra colocação dos valores próprios de A na diagonal principal de D. ♦

Exemplo 4.9. Seja $A = \begin{bmatrix} 2 & 5 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 5 \end{bmatrix}$.

Usando o desenvolvimento de Laplace ao longo da terceira coluna de $(A-\lambda I)$, temos

$$\det(A-\lambda I) = \begin{vmatrix} 2-\lambda & 5 & 0 \\ -1 & -\lambda & 0 \\ 0 & 0 & 5-\lambda \end{vmatrix} = (5-\lambda)\left[-\lambda(2-\lambda)+5\right] = (5-\lambda)(\lambda^2-2\lambda+5).$$

Logo, 5 e $1 \pm 2i$ são valores próprios de A, já que

$$(\lambda^2 - 2\lambda + 5) = 0 \iff \lambda = \frac{2 \pm \sqrt{-16}}{2} = 1 \pm 2i.$$

A matriz A é diagonalizável uma vez que tem três valores próprios distintos (cf. Corolário 4.3). Para factorizar A na forma $A = PDP^{-1}$ vamos calcular os espaços próprios considerando-os como subespaços de $\mathbb{C}^n$ (a matriz é real mas tem valores próprios complexos).

$$(A-5I)\mathbf{x} = \begin{bmatrix} -3 & 5 & 0 \\ -1 & -5 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \iff \begin{cases} -3a + 5b &= 0 \\ -a - 5b &= 0 \end{cases} \iff a = b = 0.$$

Logo,

$$E(5) = \{(0,0,c) : c \in \mathbb{C}\} = \mathrm{Span}\{(0,0,1)\}.$$

$$(A-(1+2i)I)\mathbf{x} = \begin{bmatrix} 1-2i & 5 & 0 \\ -1 & -1-2i & 0 \\ 0 & 0 & 4-2i \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\iff \begin{cases} -x-(1+2i)y &= 0 \\ (4-2i)z &= 0 \end{cases} \iff x = -(1+2i)y \text{ e } z = 0.$$

Note-se que a matriz $A - (1+2i)I$ deverá ter determinante igual a zero uma vez que $1+2i$ é valor próprio de A. Ou seja, as linhas de $A-(1+2i)I$ são linearmente dependentes. Desta observação podemos concluir (sem verificação adicional) que as duas primeiras linhas da matriz são linearmente dependentes, e consequentemente o sistema $(A-(1+2i)I)\mathbf{x} = \mathbf{0}$ reduz-se às duas equações indicadas acima. Logo,

$$E(1+2i) = \{(-(1+2i)y, y, 0) :\ y \in \mathbb{C}\} = \mathrm{Span}\{(1+2i,-1,0)\}.$$

Como a valores próprios complexos conjugados correspondem vectores próprios conjugados, tem-se

$$E(1-2i) = \{(-(1-2i)\bar{y}, \bar{y}, 0) :\ y \in \mathbb{C}\} = \mathrm{Span}\{(1-2i,-1,0)\}.$$

Assim, uma matriz P que diagonaliza A e a correspondente matriz diagonal D podem ser

$$P = \begin{bmatrix} 1+2i & 1-2i & 0 \\ -1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} 1+2i & 0 & 0 \\ 0 & 1-2i & 0 \\ 0 & 0 & 5 \end{bmatrix}.$$

Sugere-se que confirme a igualdade $A = PDP^{-1}$. ♦

Como vimos, uma matriz A de ordem n é diagonalizável se e só se a soma das dimensões dos espaços próprios de A (ou seja, a soma das multiplicidades geométricas dos valores próprios de A) for igual a n. Já se observou no Exemplo 4.7 que existem matrizes cuja soma das multiplicidades geométricas dos seus valores próprios é inferior à ordem da matriz. Coloca-se naturalmente a questão de saber se essa soma pode ser superior a n. A resposta a esta questão é negativa como se deduz da proposição seguinte.

Proposição 4.7. Se λ é um valor próprio da matriz A, então

$$mult\, geom(\lambda) \leq mult\, alg(\lambda).$$

Demonstração. Seja A uma matriz de ordem n e μ um valor próprio de A com multiplicidade algébrica igual a k e multiplicidade geométrica igual a r.

É óbvio que $r \leq n$, caso contrário $r = \dim N(A - \mu I)$ seria maior do que n, o que é impossível já que a ordem de A é n. De igual modo, $k \leq n$ já que $\det(A - \lambda I)$ é um polinómio de grau n e portanto não tem raízes de multiplicidade superior a n.

Provemos agora que $r \leq k \leq n$. Suponhamos, por absurdo, que $r > k$. Ou seja, que existem r vectores próprios $\mathbf{u}_1, \ldots, \mathbf{u}_r$ linearmente independentes associados ao valor próprio μ. Podemos completar o conjunto $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ por forma a obter uma base de $\mathbb{C}^n$. Seja $B = \{\mathbf{u}_1, \ldots, \mathbf{u}_r, \mathbf{v}_1, \ldots, \mathbf{v}_{n-r}\}$ uma tal base, e P a matriz cujas colunas são os vectores de B, colocados segundo a ordem pela qual aparecem em B. As primeiras r colunas $\mathbf{u}_i$ de P verificam $A\mathbf{u}_i = \mu\mathbf{u}_i$, e portanto PAP^{-1} é uma matriz em blocos da forma

$$PAP^{-1} = T = \begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix},$$

onde T_{11} é uma matriz diagonal do tipo $r \times r$ com todas as entradas na diagonal principal iguais a μ, isto é, $T_{11} = \mu I_r$. A matriz A e a matriz $T = PAP^{-1}$ têm o mesmo polinómio característico (cf. Proposição 4.5), ou seja

$$p(\lambda) = \det(A - \lambda I) = \det(T - \lambda I) = \begin{vmatrix} T_{11} - \lambda I_r & T_{12} \\ 0 & T_{22} - \lambda I_{n-r} \end{vmatrix}.$$

Efectuando r aplicações sucessivas do desenvolvimento de Laplace segundo a primeira coluna, tem-se

$$p(\lambda) = \det(T - \lambda I) = \det(T_{11} - \lambda I_r)\det(T_{22} - \lambda I_{n-r}) = (\mu - \lambda)^r q(\lambda),$$

onde q é um polinómio de grau $n - r$. Da expressão anterior conclui-se que μ é uma raiz de p com multiplicidade algébrica pelo menos $r > k$, o que é uma contradição.

□

A soma das multiplicidades algébricas dos valores próprios de uma matriz de ordem n é exactamente n, consequentemente segue como corolário da proposição anterior, da Proposição 4.6 e do Teorema 4.1, o seguinte:

Corolário 4.4. Uma matriz A é diagonalizável se e só se todo o valor próprio λ de A satisfaz a igualdade

$$mult\, alg(\lambda) = mult\, geom(\lambda).$$

Ou seja, uma matriz é diagonalizável se e só todos os valores próprios são semi-simples.

Terminamos esta secção referindo alguns resultados sobre diagonalização de matrizes reais com valores próprios complexos.

Valores próprios complexos

Recordemos que se λ é um valor próprio complexo não real de uma matriz real A de ordem n, e $\mathbf{x}$ um vector próprio associado a λ, então $\mathbf{x}$ não é um vector de $\mathbb{R}^n$. Assim, se a matriz real A é diagonalizável e tem valores próprios complexos, então a matriz P na factorização $A = PDP^{-1}$ possui entradas complexas (ver Exemplo 4.9). É habitual designar-se este facto dizendo que a matriz A é *diagonalizável em* $\mathbb{C}^n$.

Quando uma matriz diagonalizável A tem valores próprios complexos a factorização $A = PDP^{-1}$ (com D diagonal), não é a factorização mais conveniente para, por exemplo, estudar o comportamento geométrico da função $f : \mathbb{R}^n \to \mathbb{R}^n$, definida por $\mathbf{x} \mapsto A\mathbf{x} = PDP^{-1}\mathbf{x}$, uma vez que $P^{-1}\mathbf{x}$ não pertence a $\mathbb{R}^n$. No sentido de esclarecer esta questão, iremos mostrar que se A é diagonalizável em $\mathbb{C}^n$, então existem matrizes <u>reais</u> M e Σ tais que $A = M\Sigma M^{-1}$, com Σ uma matriz diagonal por blocos, com um bloco diagonal (correspondente aos valores próprios reais), e blocos 2×2 da forma

$$S = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}. \tag{4.10}$$

(correspondentes a cada par de valores próprios complexos $\lambda = a \pm bi$). Ou seja,

$$\Sigma = \begin{bmatrix} D & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & S_1 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & S_k \end{bmatrix}, \tag{4.11}$$

onde as matrizes S_i são da forma (4.10), a matriz D é uma matriz diagonal tendo na diagonal principal os valores próprios reais de A e $\mathbf{0}$ designa matrizes nulas.

Antes de procedermos à demonstração deste teorema estabelecem-se alguns resultados preliminares.

Define-se a parte real e a parte imaginária de um vector $\mathbf{u} \in \mathbb{C}^n$ como sendo os vectores de $\mathbb{R}^n$ cujas componentes são, respectivamente, a parte real e a parte imaginária das componentes correspondentes de $\mathbf{u}$. Por exemplo, se $\mathbf{u} = (5 - i, -2 + 3i)$, então $\operatorname{Re}\mathbf{u} = (5, -2)$ e $\operatorname{Im}\mathbf{u} = (-1, 3)$.

Lema 4.1. Os vectores $\mathbf{u} \in \mathbb{C}^n$ e $\overline{\mathbf{u}} \in \mathbb{C}^n$ são linearmente independentes sobre $\mathbb{C}$ se e só se $\operatorname{Re}\mathbf{u}$ e $\operatorname{Im}\mathbf{u}$ são vectores (de $\mathbb{R}^n$) linearmente independentes.

Demonstração. Qualquer vector $\mathbf{u}$ de $\mathbb{C}^n$ escreve-se na forma $\mathbf{u} = \operatorname{Re}\mathbf{u} + i\operatorname{Im}\mathbf{u}$. Como $\operatorname{Re}\mathbf{u} = \operatorname{Re}\overline{\mathbf{u}}$ e $\operatorname{Im}\overline{\mathbf{u}} = -\operatorname{Im}\mathbf{u}$, tem-se

$$\begin{aligned}(\alpha + i\beta)\mathbf{u} + (\gamma + i\delta)\overline{\mathbf{u}} &= (\alpha + i\beta)(\operatorname{Re}\mathbf{u} + i\operatorname{Im}\mathbf{u}) + (\gamma + i\delta)(\operatorname{Re}\mathbf{u} - i\operatorname{Im}\mathbf{u}) \\ &= [(\alpha + \gamma)\operatorname{Re}\mathbf{u} + (\delta - \beta)\operatorname{Im}\mathbf{u}] + i\,[(\beta + \delta)\operatorname{Re}\mathbf{u} + (\alpha - \gamma)\operatorname{Im}\mathbf{u}] .\end{aligned}$$

Logo, $(\alpha + i\beta)\mathbf{u} + (\gamma + i\delta)\overline{\mathbf{u}} = 0 + 0i$ é equivalente ao sistema

$$\begin{cases} (\alpha + \gamma)\operatorname{Re}\mathbf{u} + (\delta - \beta)\operatorname{Im}\mathbf{u} = 0 \\ (\beta + \delta)\operatorname{Re}\mathbf{u} + (\alpha - \gamma)\operatorname{Im}\mathbf{u} = 0. \end{cases} \tag{4.12}$$

Uma vez que se verifica a equivalência

$$\alpha = \beta = \gamma = \delta = 0 \Longleftrightarrow \begin{cases} \alpha + \gamma = 0 \\ \alpha - \gamma = 0 \\ \delta + \beta = 0 \\ \delta - \beta = 0, \end{cases}$$

os vectores $\mathbf{u}$ e $\overline{\mathbf{u}}$ são linearmente independentes se e só se $\operatorname{Re}\mathbf{u}$ e $\operatorname{Im}\mathbf{u}$ são linearmente independentes. $\square$

Proposição 4.8. Seja A uma matriz real 2×2 com valores próprios $a\pm bi$ ($b \neq 0$) e $\mathbf{v} \in \mathbb{C}^2$ um vector próprio associado a $\lambda = a - bi$. Então,

$$A = MSM^{-1}, \quad \text{com} \quad M = \begin{bmatrix} | & | \\ \operatorname{Re}\mathbf{v} & \operatorname{Im}\mathbf{v} \\ | & | \end{bmatrix} \quad \text{e} \quad S = \begin{bmatrix} a & -b \\ b & a \end{bmatrix},$$

onde $\operatorname{Re}\mathbf{v}$ e $\operatorname{Im}\mathbf{v}$ designam, respectivamente, a parte real e a parte imaginária do vector $\mathbf{v}$.

Demonstração. O Lema 4.1 garante que as colunas de M são linearmente independentes, uma vez que o vectores $\mathbf{v}$ e $\overline{\mathbf{v}}$ são vectores próprios associados a valores próprios distintos, e portanto linearmente independentes (cf. Proposição 4.6). Assim, M é invertível e $A = MSM^{-1}$ é equivalente a $AM = MS$. Necessitamos pois de mostrar que $AM = MS$, com M e S da forma indicada no enunciado. Ora,

$$\begin{aligned} MS &= \begin{bmatrix} | & | \\ \operatorname{Re}\mathbf{v} & \operatorname{Im}\mathbf{v} \\ | & | \end{bmatrix} \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \\ &= \begin{bmatrix} M\begin{bmatrix} a \\ b \end{bmatrix} & M\begin{bmatrix} -b \\ a \end{bmatrix} \end{bmatrix} && \text{(pela Definição 1.12)} \\ &= \begin{bmatrix} | & | \\ a\operatorname{Re}\mathbf{v} + b\operatorname{Im}\mathbf{v} & -b\operatorname{Re}\mathbf{v} + a\operatorname{Im}\mathbf{v} \\ | & | \end{bmatrix} && \text{(pela Definição 1.11)} \end{aligned}$$

e

$$AM = A\begin{bmatrix} | & | \\ \operatorname{Re}\mathbf{v} & \operatorname{Im}\mathbf{v} \\ | & | \end{bmatrix} = \begin{bmatrix} | & | \\ A\operatorname{Re}\mathbf{v} & A\operatorname{Im}\mathbf{v} \\ | & | \end{bmatrix} \qquad \text{(pela Definição 1.12).}$$

Da definição de valor e vector próprio, temos

$$\begin{aligned} A\mathbf{v} = \lambda\mathbf{v} &\iff A(\operatorname{Re}\mathbf{v} + i\operatorname{Im}\mathbf{v}) = (a - ib)(\operatorname{Re}\mathbf{v} + i\operatorname{Im}\mathbf{v}) \\ &\iff A\operatorname{Re}\mathbf{v} + iA\operatorname{Im}\mathbf{v} = (a\operatorname{Re}\mathbf{v} + b\operatorname{Im}\mathbf{v}) + i(-b\operatorname{Re}\mathbf{v} + a\operatorname{Im}\mathbf{v}) \\ &\iff A\operatorname{Re}\mathbf{v} = a\operatorname{Re}\mathbf{v} + b\operatorname{Im}\mathbf{v} \quad \text{e} \quad A\operatorname{Im}\mathbf{v} = -b\operatorname{Re}\mathbf{v} + a\operatorname{Im}\mathbf{v}, \end{aligned}$$

onde na última equivalência aplicámos o facto de dois vectores complexos serem iguais se e só se as respectivas partes reais e imaginárias forem iguais. Por conseguinte, a igualdade $AM = MS$ é satisfeita. □

Enunciemos agora o teorema já referido.

Teorema 4.2. Seja A uma matriz real, $n \times n$, diagonalizável, e com p valores próprios reais e k pares de valores próprios complexos conjugados ($p + 2k = n$). Existem matrizes reais M e Σ tais que $A = M\Sigma M^{-1}$. A matriz Σ é uma matriz diagonal por blocos da forma

$$\Sigma = \begin{bmatrix} D & 0 & \cdots & 0 \\ 0 & S_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & S_k \end{bmatrix}.$$

Os blocos (na diagonal) de Σ têm as seguintes propriedades:

- O bloco D é uma matriz diagonal de ordem p com entradas na diagonal principal iguais aos valores próprios reais de A, repetidos de acordo com as suas multiplicidades.

- Cada bloco S_j é um bloco 2×2 da forma $\begin{bmatrix} a_j & -b_j \\ b_j & a_j \end{bmatrix}$, com $a_j \pm ib_j$ um par de valores próprios complexos conjugados de A.

As colunas de M têm as seguintes propriedades::

- Para $i = 1, \ldots, p$, a coluna i de M é um vector próprio $\mathbf{v}_i$ associados ao valor próprio real λ_i de A.

- As colunas de $p + 1$ a n são, respectivamente, os pares de vectores $\operatorname{Re}\mathbf{v}_j$ e $\operatorname{Im}\mathbf{v}_j$, ($j = 1, \ldots, k$), onde $\mathbf{v}_j$ é um vector próprio associado ao valor próprio (complexo) $\lambda_j = a_j - ib_j$.

Demonstração. A matriz A é diagonalizável e portanto existe uma base de $\mathbb{C}^n$ constituída por n vectores próprios de A. Seja $\{\mathbf{u}_1, \ldots, \mathbf{u}_p, \mathbf{v}_1, \overline{\mathbf{v}}_1, \ldots, \mathbf{v}_k, \overline{\mathbf{v}}_k\}$ uma base de $\mathbb{C}^n$, em que $\mathbf{u}_i$ é um vector próprio associado a um valor próprio real e $\mathbf{v}_j, \overline{\mathbf{v}}_j$ são vectores próprios associados, respectivamente, ao par de valores próprios complexos $a_j - ib_j, a_j + ib_j$. Do Lema 4.1 e da Proposição 4.6, segue que

$$B = (\mathbf{u}_1, \ldots, \mathbf{u}_p, \operatorname{Re}\mathbf{v}_1, \operatorname{Im}\mathbf{v}_1, \ldots, \operatorname{Re}\mathbf{v}_k, \operatorname{Im}\mathbf{v}_k)$$

é uma base ordenada de $\mathbb{R}^n$.

Coloquem-se os vectores da base B numa matriz M por colunas, respeitando a ordem de B. Efectuando os produtos MA e ΣM, obtém-se $MA = \Sigma M$ (cf. Proposição 4.8). Como as colunas de M formam uma base, a matriz M é invertível e portanto $MA = \Sigma M$ é equivalente a $MAM^{-1} = \Sigma$. $\square$

Ilustremos a aplicabilidade deste teorema à matriz do Exemplo 4.9. Nesse exemplo, verificámos que o espectro de A é $\sigma(A) = \{5, 1+2i, 1-2i\}$, e consequentemente A é diagonalizável. Os espaços próprios são

$$E(5) = \mathrm{Span}\{(0,0,1)\}, \quad E(1-2i) = \mathrm{Span}\{(1-2i,-1,0)\}.$$

Seja $\mathbf{v} = (1-2i,-1,0)$ um vector próprio associado a $\lambda = 1-2i$. Os vectores $\mathrm{Re}\,\mathbf{v}$ e $\mathrm{Im}\,\mathbf{v}$ são

$$\mathrm{Re}\,\mathbf{v} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \quad \text{e} \quad \mathrm{Im}\,\mathbf{v} = \begin{bmatrix} -2 \\ 0 \\ 0 \end{bmatrix}.$$

Por conseguinte, o Teorema 4.2 diz-nos que podemos tomar para M e Σ as matrizes

$$M = \begin{bmatrix} 0 & 1 & -2 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad \text{e} \quad \Sigma = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 1 & -2 \\ 0 & 2 & 1 \end{bmatrix}.$$

Ou seja, $A = M\Sigma M^{-1}$ com M e Σ reais. Sugerimos que compare esta factorização com a factorização $A = PDP^{-1}$ obtida no Exemplo 4.9.

Finalizamos esta secção fazendo uma referência breve ao comportamento da função $f : \mathbb{R}^n \to \mathbb{R}^n$, definida por $f(\mathbf{x}) = A\mathbf{x}$, no caso em que A é diagonalizável e possui valores próprios complexos.

Seja A uma matriz real, de ordem n, diagonalizável. Então, $A = M\Sigma M^{-1}$ com M real e Σ uma matriz real, diagonal por blocos, da forma (4.11). A matriz M é invertível, e pelo Teorema 3.8 a matriz M realiza a mudança da base constituída pelos vectores coluna de M (que é uma base de $\mathbb{R}^n$ constituída por vectores próprios de A correspondentes a valores próprios reais, e pelos vectores das partes reais e imaginárias de vectores próprios associados aos valores próprios complexos) para a base canónica de $\mathbb{R}^n$. A matriz M determina uma mudança de variáveis de $\mathbf{x}$ para $\mathbf{y}$, mediante a igualdade $\mathbf{x} = M\mathbf{y}$, ou seja, $f(\mathbf{x}) = A\mathbf{x} = M\Sigma M^{-1}\mathbf{x} = M\Sigma\mathbf{y}$.

A acção de f sobre um vector $\mathbf{x}$ (ou equivalentemente a acção de A sobre $\mathbf{x}$) pode traduzir-se do seguinte modo: (i) fazer a mudança de variáveis de $\mathbf{x}$ para

$\mathbf{y}$; (ii) fazer actuar a matriz Σ em $\mathbf{y}$; (iii) seguidamente, sobre o vector obtido, efectuar a mudança de variáveis (inversa) para a variável inicial. No diagrama seguinte ilustra-se este processo.

$$\begin{array}{ccc} \mathbb{R}^n \ni \mathbf{x} & \xrightarrow{A} & A\mathbf{x} \in \mathbb{R}^n \\ {\scriptstyle M^{-1}}\big\downarrow & & \big\uparrow{\scriptstyle M} \\ \mathbb{R}^n \ni \mathbf{y} & \xrightarrow{\Sigma} & \Sigma\mathbf{y} \in \mathbb{R}^n \end{array}$$

A matriz Σ é constituída por blocos diagonais e por blocos associados a pares de valores próprios complexos conjugados, os quais são matrizes do tipo estudado no Caso 2 da Secção 4.1.1 (ver página 184). Do estudo efectuado nessa secção, sabemos como actuam os blocos de Σ em vectores de $\mathbb{R}^n$.

Exemplo 4.10. Considere-se a matriz $A = \begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{-\sqrt{3}}{2} \\ \frac{1}{2\sqrt{3}} & \frac{\sqrt{3}}{2} \end{bmatrix}$.

Os valores próprios de A são $\lambda = \frac{1}{2}(\sqrt{3} \pm i)$. Podemos verificar que $\mathbf{v} = \begin{bmatrix} -i\sqrt{3} \\ 1 \end{bmatrix}$ é um vector próprio de A associado ao valor próprio $\lambda = \frac{1}{2}(\sqrt{3} - i)$, já que

$$A\mathbf{v} = \begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{-\sqrt{3}}{2} \\ \frac{1}{2\sqrt{3}} & \frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} -i\sqrt{3} \\ 1 \end{bmatrix} = \frac{\sqrt{3} - i}{2} \begin{bmatrix} -i\sqrt{3} \\ 1 \end{bmatrix} = \lambda\mathbf{v}.$$

A Proposição 4.8 garante que a matriz A é da forma $A = MSM^{-1}$, com

$$S = \begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{-1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \quad \text{e} \quad M = \begin{bmatrix} 0 & -\sqrt{3} \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} | & | \\ \operatorname{Re}\mathbf{v} & \operatorname{Im}\mathbf{v} \\ | & | \end{bmatrix}.$$

Atendendo ao estudo realizado na Secção 4.1.1-Caso 2, sabemos que a matriz S é uma matriz de rotação (os valores próprios têm módulo igual a 1), e que S actua em vectores de $\mathbb{R}^2$ rodando-os (no sentido directo) em torno da origem de um ângulo $\frac{\pi}{6}$ (note que $\cos\frac{\pi}{6} = \frac{\sqrt{3}}{2}$ e $\operatorname{sen}\frac{\pi}{6} = \frac{1}{2}$) .

Na Figura 4.6 encontram-se representadas sucessivas aplicações de S ao vector $\mathbf{x}_0 = (1, 3)$ através de pontos a cor azul. Cada um destes pontos é obtido do anterior por uma rotação de $\pi/6$, ou seja, os pontos correspondentes a aplicações sucessivas de S situam-se sobre uma circunferência de centro na origem e raio igual à distância de $\mathbf{x}_0$ à origem, isto é, de raio igual a $\sqrt{10}$.

As imagens de aplicações sucessivas de A ao mesmo ponto $\mathbf{x}_0$ (representadas na Figura 4.6 a vermelho) estão por sua vez sobre uma elipse. Note-se que $\mathbf{x} = M\mathbf{y}$ é da forma

$$\mathbf{x} = M\mathbf{y} \iff \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \begin{bmatrix} -\sqrt{3}y_1 \\ y_0 \end{bmatrix},$$

e portanto se $\mathbf{x} = (x_0, x_1)$ pertence à circunferência de equação $x_0^2 + x_1^2 = 10$, então o ponto $\mathbf{y} = (y_0, y_1)$ pertence à elipse definida por $y_0^2 + 3y_1^2 = 10$.

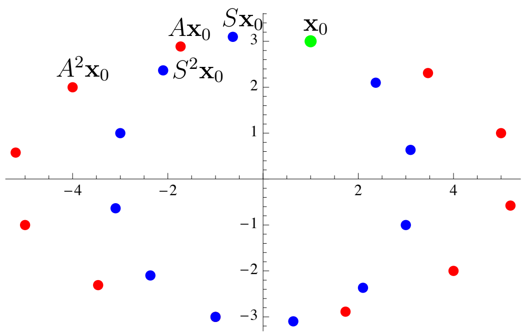


Figura 4.6: Sucessivas aplicações de $f(\mathbf{x}) = A\mathbf{x}$ e de $g(\mathbf{w}) = S\mathbf{w}$ ao ponto $\mathbf{x}_0$, onde $A = MSM^{-1}$. A matriz A tem valores próprios complexos de módulo 1.

4.3 Potências de uma matriz e valores próprios

Certas propriedades dos valores e vectores próprios de potências de uma matriz desempenham um papel fundamental em álgebra linear e nas aplicações. Neste texto apresentam-se alguns exemplos ilustrativos da relevância dos valores e vectores próprios de potências de matrizes, nomeadamente no estudo do comportamento a longo prazo de cadeias de Markov (estudadas na próxima secção), ou na determinação da forma canónica de Jordan de uma matriz (tratada no Capítulo 8).

Comecemos por observar que sendo A uma matriz diagonalizável de ordem n, isto é, $A = PDP^{-1}$ com $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$, qualquer potência positiva de A também é uma matriz diagonalizável, visto que

$$A^k = \underbrace{(PDP^{-1})(PDP^{-1})\cdots(PDP^{-1})}_{k \text{ factores}} = PD^kP^{-1}, \tag{4.13}$$

onde $D^k = \operatorname{diag}(\lambda_1^k, \ldots, \lambda_n^k)$. A igualdade $P^{-1}A^kP = D^k$ implica que os valores próprios de A^k são $\lambda_1^k, \ldots, \lambda_n^k$ (a matriz A^k é semelhante à matriz D^k). Além disso, as colunas de P, que são vectores próprios de A, também são vectores próprios de A^k já que: se $(\lambda, \mathbf{u})$ é um par próprio de A, então $(\lambda^k, \mathbf{u})$ é um par próprio de A^k (ver demonstração da próxima proposição).

Proposição 4.9. Seja $p(\lambda) = (-1)^n\lambda^n + b_{n-1}\lambda^{n-1} + \cdots + b_1\lambda + b_0$ o polinómio característico da matriz A e $\mathbf{u}$ um vector próprio de A. Então

$$p(A)\mathbf{u} = \left[(-1)^n A^n + b_{n-1}A^{n-1} + \cdots + b_1 A + b_0\right]\mathbf{u} = \mathbf{0}.$$

Demonstração. Comecemos por mostrar que se $(\mu, \mathbf{u})$ é um par próprio de A, então $(\mu^k, \mathbf{u})$ é um par próprio de A^k, com k um inteiro positivo. De facto, se $A\mathbf{u} = \mu\mathbf{u}$, resulta

$$\begin{aligned} A^k\mathbf{u} &= A^{k-1}(A\mathbf{u}) = A^{k-1}(\mu\mathbf{u}) = \mu A^{k-1}\mathbf{u} = \mu A^{k-2}(A\mathbf{u}) = \mu^2 A^{k-2}\mathbf{u} \\ &= \cdots = \mu^k\mathbf{u}. \end{aligned}$$

Por definição de valor e vector próprio, a igualdade $A^k\mathbf{u} = \mu^k\mathbf{u}$ significa que μ^k é um valor próprio de A e $\mathbf{u}$ é um vector próprio associado.

Para mostrar que $p(A)\mathbf{u} = \mathbf{0}$ basta mostrar que $p(A)\mathbf{u} = p(\mu)\mathbf{u}$, onde $(\mu, \mathbf{u})$ é um par próprio de A. Aplicando o facto de $(\mu^k, \mathbf{u})$ ser um par próprio de A^k no cálculo de $p(A)\mathbf{u}$, tem-se

$$\begin{aligned} p(A)\mathbf{u} &= (-1)^n A^n\mathbf{u} + b_{n-1}A^{n-1}\mathbf{u} + \cdots + b_1 A\mathbf{u} + b_0\mathbf{u} \\ &= (-1)^n\mu^n\mathbf{u} + b_{n-1}\mu^{n-1}\mathbf{u} + \cdots + b_1\mu\mathbf{u} + b_0\mathbf{u} \\ &= ((-1)^n\mu^n + b_{n-1}\mu^{n-1} + \cdots + b_1\mu + b_0)\mathbf{u} = p(\mu)\mathbf{u} = 0 \times \mathbf{u} = \mathbf{0}, \end{aligned}$$

onde a penúltima igualdade resulta do facto de μ ser valor próprio de A, e portanto raiz de p. □

Corolário 4.5. Seja $p(\lambda) = (-1)^n\lambda^n + b_{n-1}\lambda^{n-1} + \cdots + b_1\lambda + b_0 = 0$ a equação característica da matriz A. Se A tem n vectores próprios linearmente independentes, a matriz A satisfaz a sua equação característica. Isto é,

$$p(A) = (-1)^n A^n + b_{n-1}A^{n-1} + \cdots + b_1 A + b_0 = \mathbf{O},$$

onde $\mathbf{O}$ designa a matriz nula.

Demonstração. Sejam $\mathbf{u}_1, \ldots, \mathbf{u}_n$ vectores próprios linearmente independentes de A e X a matriz cujas colunas são estes vectores. Da proposição anterior tem-se $p(A)\mathbf{u}_i = \mathbf{0}$ para $i = 1, \ldots, n$. Assim, usando a Definição 1.12 de produto de matrizes, obtemos

$$p(A)X = p(A)\begin{bmatrix} | & | & & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \ldots & \mathbf{u}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ p(A)\mathbf{u}_1 & p(A)\mathbf{u}_2 & \ldots & p(A)\mathbf{u}_n \\ | & | & & | \end{bmatrix} = \mathbf{O}.$$

Como as colunas de X são linearmente independentes, a matriz X é invertível (Proposição 3.15). Logo, multiplicando (à direita) a equação matricial $p(A)X = \mathbf{O}$ por X^{-1}, obtém-se $p(A) = \mathbf{O}$. □

O resultado do corolário anterior é igualmente válido no caso da matriz não admitir n vectores próprios linearmente independentes. Esta generalização constitui o famoso Teorema de Cayley-Hamilton[2] que passamos a enunciar.

Teorema 4.3. Cayley-Hamilton
Toda a matriz quadrada verifica a sua equação característica. Ou seja, se $(-1)^n\lambda^n + b_{n-1}\lambda^{n-1} + \cdots + b_1\lambda + b_0$ é o polinómio característico de A, então

$$(-1)^n A^n + b_{n-1}A^{n-1} + \cdots + b_1 A + b_0 I = \mathbf{O}.$$

O exercício guiado seguinte apresenta uma demonstração do Teorema de Cayley-Hamilton.

Exercício 4.2. Mostre o Teorema de Cayley-Hamilton.

Comece por justificar por que razão a matriz adjunta $\operatorname{adj}(A - \lambda I)$ pode ser escrita na forma $\operatorname{adj}(A-\lambda I) = B_{n-1}\lambda^{n-1} + \cdots + B_1\lambda + B_0$, onde B_i são matrizes $n \times n$. Aplique a fórmula (2.14), na página 102, à matriz $(A - \lambda I)$, e conclua que

$$\begin{aligned} -B_{n-1} &= (-1)^n I \\ AB_{n-k} - B_{n-k-1} &= b_{n-k}I \quad \text{para } k = 1, \ldots, n-1 \\ AB_0 &= b_0 I. \end{aligned}$$

Multiplique as igualdades anteriores respectivamente por A^n, A^{n-k} e I, adicione, e obterá $p(A) = \mathbf{O}$.

▲

[2] Arthur Cayley (1821 – 1895), matemático inglês. Sir William Rowan Hamilton (1805 – 1865), físico, astrónomo e matemático irlandês.

4.4 Aplicações: Sistemas dinâmicos

Apresentamos nesta secção alguns exemplos ilustrativos da importância dos valores e vectores próprios de uma matriz no estudo de certos modelos matemáticos. Em primeiro lugar trataremos sistemas dinâmicos discretos e posteriormente equações diferenciais (sistemas dinâmicos contínuos).

Um sistema dinâmico é um modelo matemático que descreve a evolução no tempo do estado de um sistema. O modelo matemático procura descrever o resultado de uma determinada experiência, efectuada repetidas vezes. Nos modelos mais simples, o resultado de cada experiência depende apenas do resultado da experiência anterior. Consideremos o exemplo seguinte.

Exemplo 4.11. Uma universidade tem na totalidade dos seus cursos de licenciatura (com a duração de três anos) um *numerus clausus* de 850 alunos. Em cada ano lectivo, 80% dos estudantes dos cursos de licenciatura transitam de ano (ou terminam, caso estejam no terceiro ano) e 20% ficam retidos no mesmo ano. O número de alunos que frequentam as licenciaturas dessa universidade no ano lectivo k representa-se pelo *vector de estado* $\mathbf{x}_k \in \mathbb{R}^3$ cujas componentes x_{k1}, x_{k2} e x_{k3} são, respectivamente, o número de alunos no primeiro, segundo e terceiro ano das licenciaturas. Suponha-se que número de alunos de licenciatura no ano lectivo 2010/11, era 1600 no primeiro ano, 950 no segundo ano e 1100 no terceiro. Representamos o número de alunos de licenciatura no ano lectivo 2010/11 pelo vector de estado $\mathbf{x}_0 = (1600, 950, 1100)$. O número de alunos de licenciatura no ano lectivo seguinte é representado pelo vector $\mathbf{x}_1$. De acordo com os dados do problema o vector $\mathbf{x}_1$ é

$$\mathbf{x}_1 = \begin{bmatrix} 850 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0.2 & 0 & 0 \\ 0.8 & 0.2 & 0 \\ 0 & 0.8 & 0.2 \end{bmatrix} \mathbf{x}_0 = \mathbf{b} + A\mathbf{x_0}.$$

O número de alunos de licenciatura nos anos lectivos subsequentes é representado pelos vectores de estado $\mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \ldots$. O vector de estado $\mathbf{x}_{k+1}$ é dado por

$$\mathbf{x}_{k+1} = \mathbf{b} + A\mathbf{x}_k \quad k = 0, 1, 2, \ldots \tag{4.14}$$

A equação (4.14) é uma fórmula de recorrência que permite obter o vector de estado de um determinado ano lectivo à custa dos vectores de estado de anos lectivos anteriores. ♦

Estudamos a seguir sistemas dinâmicos discretos lineares do tipo

$$\mathbf{x}_{k+1} = A\mathbf{x}_k, \quad \text{com} \quad \mathbf{x}_k \in \mathbb{R}^n \text{ e } k = 0, 1, 2, 3, \ldots$$

onde A é uma matriz de ordem n. Este tipo de sistemas é também designado por *equação às diferenças*, de 1ª ordem homogénea. Note-se que (4.14) é uma equação às diferenças de 1ª ordem não homogénea.

O análogo contínuo de uma equação às diferenças é uma *equação diferencial*. Uma equação diferencial modela sistemas físicos cujos estados são observados de forma contínua. No final deste capítulo estudaremos este tipo de equações.

4.4.1 Sistemas dinâmicos discretos

Consideremos o sistema dinâmico discreto

$$\mathbf{x}_k = A\mathbf{x}_{k-1}, \quad k = 0, 1, 2, \ldots$$

onde A é uma matriz real do tipo $n \times n$.

Chamamos *órbita* do ponto $\mathbf{x}_0$ ao conjunto de pontos da sucessão $\{\mathbf{x}_s\}$, isto é, $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \ldots$. A órbita do ponto inicial $\mathbf{x}_0$ é determinada pelas potências A^k e por $\mathbf{x}_0$, uma vez que

$$\begin{aligned}
&\mathbf{x}_1 = A\mathbf{x}_0\\
&\mathbf{x}_2 = A\mathbf{x}_1 = A^2\mathbf{x}_0\\
&\vdots\\
&\mathbf{x}_k = A\mathbf{x}_{k-1} = A^k\mathbf{x}_0, \quad k \geq 1.
\end{aligned}$$

Suponhamos que A é diagonalizável e que existe uma base $B = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n)$ de $\mathbb{R}^n$ formada por vectores próprios de A. Nesta base, o ponto inicial $\mathbf{x}_0$ escreve-se como combinação linear (única) dos vectores de B, seja

$$\mathbf{x}_0 = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_n\mathbf{u}_n.$$

Como vimos na demonstração da Proposição 4.9, se $(\lambda_i, \mathbf{u}_i)$ é um par próprio de A, então $(\lambda_i^k, \mathbf{u}_i)$ é um par próprio de A^k. Por conseguinte, o estado do sistema no instante k é dado por

$$\begin{aligned}
\mathbf{x}_k = A^k\mathbf{x}_0 \;&= A^k(c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_n\mathbf{u}_n)\\
&= c_1A^k\mathbf{u}_1 + c_2A^k\mathbf{u}_2 + \cdots + c_nA^k\mathbf{u}_n \qquad (4.15)\\
&= c_1\lambda_1^k\mathbf{u}_1 + c_2\lambda_2^k\mathbf{u}_2 + \cdots + c_n\lambda_n^k\mathbf{u}_n,
\end{aligned}$$

onde $(\lambda_i, \mathbf{u}_i)$ é um par próprio de A, para $i = 1, \ldots, n$.

A expressão (4.15) permite determinar o *comportamento a longo prazo* do sistema. Este comportamento é dado pelo vector $\mathbf{x}_\infty$ definido por

$$\mathbf{x}_\infty = \lim_{k\to\infty} \mathbf{x}_k = \lim_{k\to\infty} \left[c_1\lambda_1^k\mathbf{u}_1 + c_2\lambda_2^k\mathbf{u}_2 + \cdots + c_n\lambda_n^k\mathbf{u}_n\right]. \tag{4.16}$$

Apresentamos a seguir alguns exemplos de sistemas dinâmicos discretos.

Números de Fibonacci e o número de Ouro

É no mínimo surpreendente como, quer na Natureza quer em certas criações artísticas em arquitectura e pintura, se podem encontrar os chamados números de Fibonacci. Um exemplo é a flor do girassol que tem 233 sementes em 144 espirais, números que correspondem aos 12º e 13º termos da sucessão de Fibonacci. Aconselhamos uma visita ao site "goldennumber.net" [3], ou ao site de Ron Knott[4], onde pode encontrar vários exemplos de como as sucessões de Fibonacci aparecem na Arte e na Natureza, bem como outras curiosidades relacionadas com o "número de ouro".

A sucessão de números de Fibonacci foi usada pelo seu criador, Leonardo de Pisa[5] (mais tarde conhecido por Fibonacci), como um modelo matemático simples para descrever o crescimento de uma população de coelhos, nas seguintes condições:

1) Admite-se que os coelhos não morrem;

2) Supõe-se que um casal de coelhos demora dois meses a atingir a maturidade, altura em que se passa a reproduzir mensalmente dando origem a um novo casal de coelhos.

Denotemos por F_k o número de casais de coelhos no mês k. A evolução da população de coelhos pode modelar-se da seguinte forma:

- O processo inicia-se no mês $k = 1$ com um casal de coelhos, isto é, $F_1 = 1$.
- No mês seguinte o número de casais é ainda igual a 1, ou seja, $F_2 = 1$, visto que o casal original está ainda imaturo.

[3] `http://goldennumber.net/`

[4] `http://www.mcs.surrey.ac.uk/Personal/R.Knott/Fibonacci/fibnat.html`

[5] Leonardo de Pisa (1170 — 1250), matemático italiano considerado um dos mais talentosos matemático da Idade Média.

- Decorridos dois meses, tem-se $F_3 = 2$, correspondente ao casal original e a um casal recém-nascido.

A sucessão $\{F_k\}$ obtida pelo processo anterior é designada por *sucessão de Fibonacci*. O termo de ordem k da sucessão de Fibonacci, verifica

$$F_{k+2} = F_{k+1} + F_k \quad \text{com } F_1 = 1 \text{ e } F_2 = 1.$$

Os primeiros termos desta sucessão são

$$1, 1, 2, 3, 5, 8, 13, \ldots$$

A sucessão de Fibonacci pode ser escrita na forma matricial da maneira que indicaremos a seguir. Usaremos os valores próprios e os vectores próprios da matriz que define a sucessão para obter uma expressão explícita (não recursiva) de F_k e provar que o crescimento da sucessão de Fibonacci é do tipo exponencial.

Escrevendo $\mathbf{x}_{k-1} = \begin{bmatrix} F_{k+1} \\ F_k \end{bmatrix}$, a equação $F_{k+2} = F_{k+1} + F_k$ é equivalente a

$$\begin{bmatrix} F_{k+2} \\ F_{k+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} F_{k+1} \\ F_k \end{bmatrix} \Longleftrightarrow \mathbf{x}_k = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \mathbf{x}_{k-1} = A\mathbf{x}_{k-1}, \quad \text{para } k = 1, 2, \ldots \tag{4.17}$$

e $\mathbf{x}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} F_2 \\ F_1 \end{bmatrix}$.

Atendendo à expressão obtida em (4.15), se a matriz A for diagonalizável (e com valores próprios reais), tem-se

$$\mathbf{x}_{k-1} = \begin{bmatrix} F_{k+1} \\ F_k \end{bmatrix} = c_1\lambda_1^{k-1}\mathbf{u}_1 + c_2\lambda_2^{k-1}\mathbf{u}_2, \tag{4.18}$$

onde $(\lambda_1, \mathbf{u}_1)$ e $(\lambda_2, \mathbf{u}_2)$ são pares próprios A, e (c_1, c_2) é o vector das coordenadas de $\mathbf{x}_0$ na base ordenada $(\mathbf{u}_1, \mathbf{u}_2)$.

Comecemos por calcular os valores próprios de A.

$$\det(A - \lambda I) = \begin{vmatrix} 1-\lambda & 1 \\ 1 & -\lambda \end{vmatrix} = \lambda^2 - \lambda - 1.$$

As raízes de $p(\lambda) = \lambda^2 - \lambda - 1$ são

$$\lambda_1 = \frac{1+\sqrt{5}}{2} \approx 1.618\ldots \quad \text{e} \quad \lambda_2 = \frac{1-\sqrt{5}}{2} \approx -0.618\ldots.$$

O valor próprio λ_1 é conhecido por *número de ouro* sendo habitualmente designado pela letra ϕ.

A matriz A tem valores próprios reais e distintos, logo é diagonalizável (cf. Corolário 4.3). Determinemos uma base de $\mathbb{R}^2$ constituída por vectores próprios de A. Para $i = 1, 2$ temos

$$\begin{bmatrix} 1-\lambda_i & 1 \\ 1 & -\lambda_i \end{bmatrix} \mathbf{u}_i = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \iff \begin{cases} (1-\lambda_i)a + \quad b = 0 \\ \qquad a - \lambda_i b = 0 \end{cases} \iff a = \lambda_i b.$$

Logo, podemos tomar $\mathbf{u}_1 = \begin{bmatrix} \lambda_1 \\ 1 \end{bmatrix}$ e $\mathbf{u}_2 = \begin{bmatrix} \lambda_2 \\ 1 \end{bmatrix}$ como vectores próprios associados, respectivamente, a λ_1 e λ_2. As coordenadas de $\mathbf{x}_0$ na base ordenada $(\mathbf{u}_1, \mathbf{u}_2)$ são:

$$\mathbf{x}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} = c_1 \begin{bmatrix} \lambda_1 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} \lambda_2 \\ 1 \end{bmatrix} \iff \begin{cases} c_1\lambda_1 + c_2\lambda_2 = 1 \\ c_1 + \quad c_2 = 1 \end{cases}$$

$$\iff c_1 = 1 - c_2, c_2 = \frac{\lambda_1 - 1}{\lambda_1 - \lambda_2} \iff c_1 = \frac{1}{\sqrt{5}}\phi, c_2 = -\frac{1}{\sqrt{5}}\lambda_2.$$

Assim, a expressão (4.18) toma a forma

$$\mathbf{x}_{k-1} = \begin{bmatrix} F_{k+1} \\ F_k \end{bmatrix} = \frac{1}{\sqrt{5}} \left[\left(\frac{1+\sqrt{5}}{2} \right)^k \begin{bmatrix} \frac{1+\sqrt{5}}{2} \\ 1 \end{bmatrix} - \left(\frac{1-\sqrt{5}}{2} \right)^k \begin{bmatrix} \frac{1-\sqrt{5}}{2} \\ 1 \end{bmatrix} \right].$$

Logo, os termos da sucessão de Fibonacci são dados por

$$F_k = \frac{1}{\sqrt{5}} \left[\left(\frac{1+\sqrt{5}}{2} \right)^k - \left(\frac{1-\sqrt{5}}{2} \right)^k \right], \quad k = 1, 2, \ldots. \tag{4.19}$$

A fórmula (4.19) é conhecida como a *fórmula de Binet* para os números de Fibonacci. Note que nesta fórmula temos que $\left(\frac{1-\sqrt{5}}{2}\right)^k \to 0$ quando $k \to \infty$, já que $0 < \lambda_2 < 1$. Por conseguinte, para k suficientemente grande, o valor de F_k pode aproximar-se por $\frac{1}{\sqrt{5}}\phi^k$, onde $\lambda_1 = \phi > 1$ é o maior valor próprio de A.

Matrizes de Markov

As chamadas cadeias de Markov[6] aparecem naturalmente na modelação matemática de problemas de biologia, química, economia, etc.. Trata-se de sistemas dinâmicos

[6] Andrey Markov (1856-1922), matemático russo.

discretos, $\mathbf{x}_{k+1} = M\mathbf{x}_k$, em que a matriz M é uma matriz cujos vectores coluna são *vectores de probabilidades*, isto é, vectores de componentes não negativas e tais que a soma das componentes é igual a 1.

Por exemplo, suponha-se que o administrador de uma firma de aluguer de viaturas, com três agências localizadas em cidades distintas, pretende estimar o número de carros em cada agência num dado instante. Admita-se que um cliente pode alugar uma viatura numa agência e entregá-la noutra. É claro que o administrador da firma não pode saber de antemão qual o número exacto de viaturas que será entregue numa dada agência, mas pode calcular a percentagem de carros que se encontram numa dada agência. Designe-se por $\mathbf{x}_k = (x_{k1}, x_{k2}, x_{k3})$ o vector de estado do mês k, isto é, o vector em que a componente x_{ki} representa a probabilidade de um carro da frota se encontrar na agência i ($i = 1, 2, 3$) no mês k. Calculando as probabilidades de um carro que está na agência i se encontrar na agência j no mês seguinte, o administrador da firma concluiu que $\mathbf{x}_{k+1} = M\mathbf{x}_k$, onde

$$M = \begin{bmatrix} 1/2 & 0 & 1/10 \\ 1/10 & 7/10 & 1/10 \\ 2/5 & 3/10 & 4/5 \end{bmatrix}.$$

Este é um problema típico que é modelado matematicamente por uma cadeia de Markov com três estados (as agências). A entrada p_{ij} da matriz $M = [p_{ij}]$ designa a probabilidade do sistema se encontrar no estado i quando na observação anterior se encontrava no estado j. A matriz M é designada por *matriz de transição* da cadeia de Markov. Como cada vector coluna de M é um vector de probabilidade, a matriz M goza da propriedade de ter todas as entradas não negativas e da soma das entradas de cada coluna ser constante e igual a 1. Estas matrizes são designadas por *matrizes de Markov* ou *matrizes estocásticas*.

Seja M uma matriz de Markov. Uma *cadeia de Markov* associada a M é uma sucessão de vectores de probabilidade, $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2 \ldots$ satisfazendo

$$\mathbf{x}_{k+1} = M\mathbf{x}_k, \quad k = 0, 1, 2, \ldots \tag{4.20}$$

Como se disse, as matrizes de Markov caracterizam-se por terem entradas não negativas (dizendo-se matrizes não negativas) e a soma das entradas de cada coluna ser igual a 1. Vemos a seguir que estas duas propriedades têm fortes implicações no tipo de valores próprios destas matrizes e no comportamento a longo prazo das cadeias de Markov.

Se A é uma matriz tal que a soma das entradas de cada linha é constante igual a s, então o vector $\mathbf{u} = (1, 1, \ldots, 1)$ satisfaz a igualdade

$$A\mathbf{u} = s\mathbf{u}.$$

A igualdade anterior significa que s é um valor próprio de A e $\mathbf{u}$ é um vector próprio associado. Como o determinante de uma matriz é igual ao determinante da sua transposta, tem-se $\det(A - \lambda I) = \det(A - \lambda I)^T = \det(A^T - \lambda I)$. Ou seja, as matrizes A e A^T têm os mesmos valores próprios. Por conseguinte, se uma matriz tem a soma das entradas de cada coluna constante, esta constante é um valor próprio da matriz.

Proposição 4.10. Se uma matriz A tem a soma das entradas de cada coluna (ou de cada linha) constante, então esta constante é um valor próprio de A.

Em particular, $\lambda = 1$ é um valor próprio de uma matriz de Markov.

Nota 24. *Apesar de uma matriz e a sua transposta terem os mesmos valores próprios, isso não significa que pares próprios de A sejam também pares próprios de A^T. Ou seja, se $\mathbf{v}$ é um vector próprio de A associado ao valor próprio λ, não significa que $\mathbf{v}$ é um vector próprio de A^T associado ao valor próprio λ. Deixamos como exercício encontrar um contra-exemplo.*

Dada uma cadeia de Markov definida pela matriz M, um *vector de equilíbrio*, ou *vector estacionário*, é um vector de probabilidades $\mathbf{q}$ que satisfaz

$$M\mathbf{q} = \mathbf{q}. \tag{4.21}$$

Ou seja, um vector de equilíbrio é um vector próprio de M associado ao valor próprio 1 que também é um vector de probabilidades. Saliente-se que, se $\mathbf{x}_k$ é um vector de equilíbrio, então $\mathbf{x}_j = \mathbf{x}_k$ para $j \geq k$, o que justifica a designação de vector de equilíbrio.

Para uma matriz de Markov, a existência de vectores próprios associados ao valor próprio $\lambda = 1$ que sejam vectores de probabilidades decorre da teoria geral das matrizes não negativas. A teoria das matrizes não negativas (conhecida pela designação de Teoria de Perron-Frobenius[7]) está fora do âmbito deste texto. Utilizamos no entanto alguns resultados fundamentais desta teoria, convidando o leitor interessado a consultar obras especializadas. Para os resultados que aqui utilizamos sugere-se a leitura de Meyer [9].

[7] Oskar Perron (1880 – 1975) e Ferdinand Georg Frobenius (1849 – 1917).

Define-se *raio espectral* de uma matriz como sendo o máximo dos módulos dos seus valores próprios. Isto é,

$$\rho(A) = \max_{i=1,\ldots,n} \{|\lambda_i|\},$$

onde λ_i é um valor próprio da matriz A (de ordem n). Resume-se no quadro seguinte alguns resultados da Teoria de Perron-Frobenius cuja prova pode encontrar em Meyer [9].

A Teoria de Perron-Frobenius para matrizes não negativas garante que uma matriz de Markov admite um vector próprio $\mathbf{u}$, associado ao valor próprio $\lambda = 1$, com todas as componentes positivas. Além disso, o raio espectral de uma matriz de Markov é exactamente 1.

No caso da matriz de Markov M ser positiva (isto é, com todas as entradas positivas), tem-se:

i) O valor próprio $\lambda = 1$ tem multiplicidade algébrica igual a 1 e os restantes valores próprios de M têm módulo inferior a 1.

ii) Se $\mathbf{y}$ é um qualquer vector de probabilidades, então

$$\lim_{k\to\infty} M^k \mathbf{y} = \mathbf{q}, \tag{4.22}$$

onde $\mathbf{q}$ é um vector de probabilidades cujas componentes são todas positivas.

O facto de uma matriz de Markov M possuir um par próprio $(1, \mathbf{u})$ em que o vector próprio $\mathbf{u} = (u_1, \ldots, u_n)$ tem todas as componentes positivas vai implicar que qualquer cadeia de Markov admita pelo menos um vector de equilíbrio. Com efeito, o vector $\mathbf{q} = \frac{1}{\sum_{i=1}^n u_i}\mathbf{u}$ ainda é um vector próprio de M associado ao valor próprio $\lambda = 1$, e além disso é um vector de probabilidades. Logo, $\mathbf{q}$ é um vector de equilíbrio.

Proposição 4.11. Qualquer cadeia de Markov admite pelo menos um vector de equilíbrio.

Saliente-se que o valor próprio $\lambda = 1$ de matrizes de Markov não negativas pode ter multiplicidade geométrica superior a um, e portanto haver mais do que um vector de equilíbrio. No entanto, quando a matriz de Markov é positiva, a unicidade do vector de equilíbrio está garantida, uma vez que o valor próprio $\lambda = 1$ é um valor próprio simples. Além disso, a longo prazo o sistema tende para o vector de equilíbrio independentemente do vector de estado inicial (pelo item (ii) no quadro acima).

No caso da matriz de Markov M ser positiva e diagonalizável, é fácil ver que a expressão (4.16) converge. De facto, como $\lambda_1 = 1$ é um valor próprio de M de multiplicidade algébrica 1, e os restantes valores próprios de M têm módulo inferior a 1, a expressão (4.16) reduz-se a

$$\mathbf{x}_\infty = \lim_{k\to\infty} M^k \mathbf{x}_0 = \lim_{k\to\infty} \left[c_1\lambda_1^k \mathbf{u}_1 + c_2\lambda_2^k \mathbf{u}_2 + \cdots + c_n\lambda_n^k \mathbf{u}_n\right] = c_1\mathbf{u}_1.$$

Como $\mathbf{x}_\infty$ é um vector de probabilidades (de facto é o vector de equilíbrio), a constante c_1 é igual ao inverso da soma das componentes de um vector próprio (positivo) $\mathbf{u}_1$ associado a $\lambda_1 = 1$.

Exemplo 4.12. Suponha-se que anualmente 1.5% da população que vive na área metropolitana de Lisboa (AML) muda-se para outras regiões do país, e 9% da população portuguesa muda-se para AML. Sabendo que no ano de 1970, 18% da população de Portugal vivia na AML, pertende-se determinar qual a distribuição da população portuguesa a longo prazo.

Tomando para vector de estado inicial $\mathbf{x}_0 = (0.18, 0.82)$, correspondendo a ter (em 1970) 18% da população na AML (e portanto 82% fora desta região), a evolução no tempo da percentagem da população portuguesa vivendo na AML é descrita pelo sistema $\mathbf{x}_k = M\mathbf{x}_{k-1}$, $k = 1, 2, \ldots$ onde M é a matriz de Markov

$$M = \begin{bmatrix} 0.985 & 0.09 \\ 0.015 & 0.91 \end{bmatrix}.$$

A matriz M corresponde aos movimentos transcritos na tabela seguinte (onde FAML designa não residentes na AML).

		De	
		AML	FAML
Para	AML	0.985	0.09
	FAML	0.015	0.91

A matriz M é positiva e tem 1 como valor próprio visto que a soma das entradas de cada coluna é igual a 1 (cf. Proposição 4.10). Como a soma dos valores próprios é igual ao traço da matriz (Proposição 4.3), o outro valor próprio é

$\lambda_2 = 0.895$. Uma vez que M é uma matriz positiva já sabíamos que $\lambda_1 = 1$ seria o maior valor próprio e que a sua multiplicidade algébrica seria igual a um. São vectores próprios associados, respectivamente a λ_1 e a λ_2, os vectores $\mathbf{u}_1 = (0.09, 0.015)$ e $\mathbf{u}_2 = (-1, 1)$, como pode confirmar calculando $M\mathbf{u}_1$ e $M\mathbf{u}_2$.

Usando a equação (4.15), no ano k, a percentagem da população portuguesa na AML e fora desta área é dada por

$$\mathbf{x}_k = c_1 \begin{bmatrix} 0.09 \\ 0.015 \end{bmatrix} + c_2(0.895)^k \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \tag{4.23}$$

onde c_1 e c_2 são as coordenadas do vector inicial $\mathbf{x}_0 = (0.18, 0.82)$ na base ordenada $(\mathbf{u}_1, \mathbf{u}_2)$. A longo prazo, a distribuição da população portuguesa tende para

$$\mathbf{x}_\infty = \lim_{k\to\infty} \mathbf{x}_k = c_1 \begin{bmatrix} 0.09 \\ 0.015 \end{bmatrix},$$

uma vez que $(0.895)^k \to 0$ quando $k \to \infty$. Este vector é de facto, o vector de equilíbrio. Uma vez que $\mathbf{x}_\infty$ é um vector de probabilidades tem-se $c_1 = \frac{1}{0.09+0.015} = 0.105$. Pode confirmar-se este resultado calculando as coordenadas c_1 e c_2 de $\mathbf{x}_0$ na base $(\mathbf{u}_1, \mathbf{u}_2)$.

Assim, a longo prazo temos $\mathbf{x}_\infty = \dfrac{1}{0.105}\mathbf{u}_1 \approx (0.86, 0.14)$, ou seja, a longo prazo 86% da população portuguesa viverá na área metropolitana de Lisboa. ♦

Exemplo 4.13. Este exemplo baseia-se no artigo de Kurt Bryan e Tanya Leise intitulado: "The $25000000000[8] eigenvector. The linear Algebra behind Google", publicado em 2006 pela SIAM Review [3].

Nos finais dos anos 90 a empresa fundadora do motor de busca Google[9] apresentou um processo de pesquisa na *net* de palavras chave que listava os resultados segundo a sua relevância. Tal não acontecia com os motores de busca existentes à época, nos quais o utilizador era obrigado a percorrer várias páginas de listagem de sites irrelevantes até encontrar a informação desejada.

Um dos algoritmos usados pelo Google para seriar os sites por ordem decrescente de importância é o denominado algoritmo PageRank.[10]

[8] O valor estimado da empresa Google quando em 2004 se tornou uma empresa pública.

[9] Google é um trocadilho da palavra anglo-saxónica "googol" a qual significa 10^{100}. O termo reflecte o número enorme e sempre crescente de utilizadores da net.

[10] PageRank foi desenvolvido na Universidade de Stanford (USA) por Larry Page e posteriormente por Sergey Brin como parte de um projecto de investigação. Page e Brin fundaram a empresa Google em 1998.

Apresentamos aqui um exemplo muito simples que ilustra a importância da álgebra linear na quantificação da relevância dos sites da net. A relevância de uma dada página é quantificada atribuindo-lhe uma classificação (um número real não negativo). Esta classificação depende do número de links (ligações ou citações) que essa página faz para outras páginas, bem como do número de citações que as outras páginas lhe fazem.

Suponha que o número de páginas (interligadas) numa rede é $n > 1$ e que cada página é designada por um inteiro k. Cada link vai representar-se por uma seta. Uma seta com origem em A e ponto final B, indica um link da página A para a página B. Um exemplo é a rede com cinco páginas representada pelo *grafo direccionado*[11] da Figura 4.7.

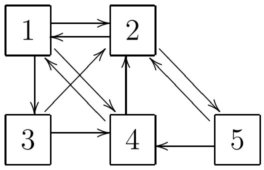


Figura 4.7: Uma rede com cinco páginas. Uma seta de A para B indica um link da página A para a B.

Designemos por x_k o valor da relevância (ou importância) da página k da rede. O valor de x_k é não negativo e $x_j > x_k$ significa que a página j tem mais importância que a página k. Uma forma simples de atribuir a importância a uma dada página k seria considerar x_k igual ao número de setas que entram na página k, ou seja, o número de citações que as outras páginas da rede fazem à página k. Por exemplo, para a rede da Figura 4.7 teríamos $x_1 = 2, x_2 = 4, x_3 = 1, x_4 = 3$ e $x_5 = 1$, significando que a página mais relevante seria a página 2 e as menos importantes a 3 e a 5.

A caracterização anterior é insuficiente visto que uma citação proveniente de uma página pouco importante não deve ter o mesmo valor que uma citação proveniente de outra mais importante, e as autocitações não devem ser consideradas. Uma outra forma de atribuir o valor da importância da página k seria considerar x_k igual à soma do valores das importâncias das páginas que a citam. Isso daria, $x_1 = x_2 + x_4, x_2 = x_1 + x_3 + x_4 + x_5, x_3 = x_1, x_4 = x_1 + x_3 + x_5$ e $x_5 = x_2$.

[11] Um grafo consiste num conjunto de vértices e arestas. Cada aresta liga um par de vértices. Um grafo diz-se direccionado se está atribuído um sentido a cada aresta.

Há contudo uma outra característica a levar em conta neste modelo, nomeadamente o facto de uma página não dever ganhar uma relevância superior só pelo simples facto de fazer muitas citações a outras páginas. Neste sentido, o valor da importância de uma dada página deve ser dividido pelo número de citações que faz a outras páginas da rede, ou seja, se a página j faz um total de s_j links para as outras páginas da rede deve considerar-se que a sua relevância é x_j/s_j (note que s_j é o número de setas que saem do vértice j do grafo). Ou seja, uma página que faz s citações confere a cada página citada o valor $1/s$ da sua importância.

Desta forma, para a rede apresentada na Figura 4.7 teríamos a seguinte modificação nas relações obtidas anteriormente:

$$\begin{array}{ll} x_1 & = \frac{1}{2}x_2 + \frac{1}{2}x_4 \\ x_2 & = \frac{1}{3}x_1 + \frac{1}{2}x_3 + \frac{1}{2}x_4 + \frac{1}{2}x_5 \\ x_3 & = \frac{1}{3}x_1 \\ x_4 & = \frac{1}{3}x_1 + \frac{1}{2}x_3 + \frac{1}{2}x_5 \\ x_5 & = \frac{1}{2}x_2 \end{array} \tag{4.24}$$

As equações anteriores podem reescrever-se na forma matricial $\mathbf{x} = M\mathbf{x}$, onde M é a matriz de Markov

$$M = \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{3} & 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & 0 & 0 & 0 \end{bmatrix} \quad \text{e} \quad \mathbf{x} = (x_1, x_2, x_3, x_4, x_5).$$

Para determinar o valor das relevâncias das cinco páginas web representadas no grafo da Figura 4.7 há que determinar o vector $\mathbf{x}$ que verifica as igualdades (4.24), ou seja, um vector de equilíbrio da cadeia de Markov definida por M. Temos assim de determinar um vector próprio associado ao valor próprio 1 de M que seja um vector de probabilidades. O vector $(15, 18, 5, 12, 9)$ é um vector próprio associado ao valor próprio 1, e portanto um vector de equilíbrio é $\frac{1}{59}(15, 18, 5, 12, 9) \approx (0.25, 0.31, 0.09, 0.20, 0.15)$. Logo, a página mais importante será a página 2, seguida de $1, 4$, 5 e 3.

Evidentemente que no caso concreto da seriação realizada pelo Google a matriz M terá uma grandeza da ordem dos biliões, pelo que no tratamento computacional deste modelo assumem especial relevância os métodos numéricos para cálculo de valores e vectores próprios de matrizes de grandes dimensões.

Convém referir que o vector de equilíbrio pode não ser único, uma vez que, como se observa neste exemplo, a matriz que modela o funcionamento do Google não é necessariamente positiva. Ou seja, o subespaço próprio associado ao valor próprio 1 pode ter dimensão superior a 1. Este caso, é tratado no artigo [3] anteriormente referido, sendo aí apresentado um algoritmo que permite que o Google produza sempre uma listagem de sites ordenados por ordem decrescente de relevância.

Aconselha-se ao leitor interessado em aprofundar os detalhes da implementação do algoritmo PageRank e de outros motores de busca, a leitura de [7]. ♦

4.4.2 Equações diferenciais ordinárias

O análogo contínuo de um sistema dinâmico discreto são os sistemas modelados por equações diferenciais, ou seja, por equações que envolvem uma função e as suas derivadas. Nesta secção abordaremos alguns aspectos da resolução de equações diferenciais ordinárias (EDO[12]) lineares, de primeira ordem. Trataremos equações do tipo $\mathbf{x}'(t) = A\mathbf{x}(t)+\mathbf{b}(t)$, onde A é uma matriz real (constante), $\mathbf{x}(t)$ e $\mathbf{b}(t)$ são vectores $n \times 1$ cujas componentes são funções reais de variável real, e $\mathbf{x}'(t)$ designa a derivada de $\mathbf{x}(t)$, isto é, a função com valor em vectores cujas componentes são as derivadas em ordem a $t \in \mathbb{R}$ das componentes de $\mathbf{x}(t)$. Eis a expressão matricial de $\mathbf{x}$ e $\mathbf{x}'$:

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix}, \qquad \mathbf{x}'(t) = \begin{bmatrix} x_1'(t) \\ \vdots \\ x_n'(t) \end{bmatrix}.$$

Em particular, iremos determinar o conjunto de todas as soluções de um sistema de equações diferenciais (lineares) da forma $\mathbf{x}'(t) = A\mathbf{x}(t)$, em que A é uma matriz diagonalizável. O caso em que A não é diagonalizável é tratado no Capítulo 8. Para além do estudo da solução geral do sistema referido abordaremos outros tópicos relacionados, nomeadamente a exponencial de matrizes e a redução de uma equação diferencial de ordem n, homogénea e de coeficientes constantes, a um sistema do tipo $\mathbf{x}'(t) = A\mathbf{x}(t)$.

Comecemos por precisar a nomenclatura usada na classificação de equações diferenciais.

- Equação diferencial: uma equação que envolve uma função, por exemplo, $\mathbf{x} : \mathbb{R} \to \mathbb{R}^n$, e as suas derivadas.

- Equação diferencial ordinária (EDO): uma equação que envolve uma função de uma única variável real t, e as suas derivadas.

[12] A abreviatura EDO em língua inglesa é ODE, de “ordinary differential equation”.

- Ordem de uma EDO: ordem da derivada de maior ordem que aparece na equação.

Alguns exemplos de equações diferenciais do tipo $\mathbf{x}'(t) = A\mathbf{x}(t) + \mathbf{b}(t)$:

(a) $x' - 2x = 5t$ (onde $\mathbf{x}(t) = [x(t)], \mathbf{b} = [5t]$ e $A = [-2]$)

(b) $\begin{cases} x_1'(t) = 2x_1(t) - x_2(t) \\ x_2'(t) = 4x_1(t) + 10x_2(t) \end{cases} \iff \begin{bmatrix} x_1' \\ x_2' \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ 4 & 10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \iff \mathbf{x}'(t) = A\mathbf{x}(t).$

Quando a matriz A é de ordem superior a 1, como no exemplo (b) anterior, a equação $\mathbf{x}'(t) = A\mathbf{x}(t) + \mathbf{b}(t)$ é um sistema de EDOs lineares de 1ª ordem. Neste texto usaremos indistintamente a designação EDO para uma equação diferencial ou para um sistema de EDOs.

A equação

$$\mathbf{x}'(t) = A\mathbf{x}(t) + \mathbf{b}(t)$$

diz-se homogénea se $\mathbf{b}(t) = \mathbf{0}$. Em geral chama-se a

$$\mathbf{x}'(t) = A\mathbf{x}(t)$$

a equação *homogénea associada* à equação $\mathbf{x}'(t) = A\mathbf{x}(t) + \mathbf{b}(t)$.

Uma *solução* da equação diferencial $\mathbf{x}'(t) = A\mathbf{x}(t) + \mathbf{b}(t)$ é uma função diferenciável $\mathbf{u}$ que verifica a equação, ou seja, tal que $\mathbf{u}'(t) = A\mathbf{u}(t) + \mathbf{b}(t)$. Chama-se *solução geral* de uma equação diferencial ao conjunto de todas as soluções da equação.

Estamos interessados em obter o conjunto de todas as soluções da equação diferencial $\mathbf{x}'(t) = A\mathbf{x}(t)$, onde A é uma matriz (constante) do tipo $n \times n$. Para tal comecemos por abordar o caso mais simples em que $n = 1$, ou seja, de uma equação do tipo $x'(t) = kx(t)$, com k uma constante real.

Exemplo 4.14. Considere-se a equação $x'(t) = 3x(t)$. A solução geral desta equação é o conjunto de todas as funções reais x de variável real cuja derivada é o triplo da função. Esse conjunto solução é constituído por todas as funções da forma $x(t) = ce^{3t}$, onde c designa uma constante real arbitrária. Não existe outro tipo de funções que verifiquem a equação, como pode confirmar usando o Exercício 4.3 adiante. Assim, a solução geral da equação dada, é o subespaço

$$\left\{x(t) = ce^{3t} : \quad c \in \mathbb{R}\right\} = \operatorname{Span}\{e^{3t}\} \tag{4.25}$$

do espaço linear $\mathcal{C}$ das funções reais de variável real, contínuas com derivada contínua. As operações de adição e multiplicação por escalares para as quais $\mathcal{C}$ é um espaço linear são as operações definidas na página 167.

O conjunto (4.25) é uma família de funções parametrizadas por $c \in \mathbb{R}$. Na Figura 4.8 encontram-se representados alguns elementos desta família.

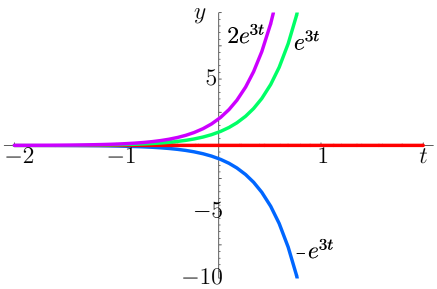


Figura 4.8: A solução geral da equação $x' = 3x$ é $x(t) = ce^{3t}$. A vermelho a função nula, correspondente a $c = 0$; a verde a solução correspondente a $c = x(0) = 1$; a azul a solução correspondente a $c = x(0) = -1$.

Obtivemos uma infinidade de soluções para a equação $x' = 3x$. Porém, se considerarmos o problema de saber quantas soluções da equação tomam um certo valor num dado ponto, a resposta é: uma única solução. Este tipo de problema é designado por *problema de valor inicial* (ou abreviadamente p.v.i.). Por exemplo, o problema de valor inicial

$$\begin{cases} x' = 3x \\ x(0) = 1, \end{cases}$$

apenas possui a solução $x(t) = e^{3t}$, já que impondo a condição inicial $x(0) = 1$ na solução geral da equação resulta

$$x(t) = ce^{3t} \Rightarrow x(0) = 1 \Rightarrow c = 1.$$

♦

Exercício 4.3. Mostre que $u(t)$ é uma solução de $x' = kx$ (com k uma constante real) se e só se o produto $u(t)e^{-kt}$ é uma constante. ▲

Sistemas de EDOs lineares, de primeira ordem, homogéneos

Vimos que no caso de uma EDO homogénea do tipo $x' = kx$ (com $x : \mathbb{R} \to \mathbb{R}$) o conjunto solução geral é o espaço linear gerado por e^{kt}. A solução geral de um sistema de EDOs homogéneo do tipo $\mathbf{x}' = A\mathbf{x}$ é igualmente um subespaço do espaço linear das funções $\mathbf{x} : \mathbb{R} \to \mathbb{R}^n$, contínuas com derivada contínua, com as operações usuais de adição de funções e multiplicação de uma função por um escalar. De facto, é fácil mostrar que:

- A soma de duas soluções de $\mathbf{x}' = A\mathbf{x}$ ainda é uma solução desta equação.

- O produto de uma solução de $\mathbf{x}' = A\mathbf{x}$ por um escalar ainda é uma solução desta equação.

Refira-se que a solução (constante) nula $\mathbf{x}(t) = (0, 0, \ldots, 0)$ é sempre solução do sistema $\mathbf{x}' = A\mathbf{x}$, como não podia deixar de ser uma vez que a solução geral é um subespaço.

Como veremos, a solução geral de $\mathbf{x}' = A\mathbf{x}$, em que A é uma matriz de ordem n, é um (sub)espaço linear de dimensão n. Assim, a determinação da solução geral passa pela obtenção de uma base deste espaço. Uma base para a solução geral recebe a designação de *conjunto fundamental de soluções*, atendendo a que qualquer solução de $\mathbf{x}' = A\mathbf{x}$ se obtém como combinação linear dos elementos do referido conjunto.

Para obter uma base para a solução geral, temos de saber determinar soluções linearmente independentes da equação $\mathbf{x}' = A\mathbf{x}$. A Proposição 4.12 adiante, consequência imediata do *Teorema de existência e unicidade de soluções* de problemas de valor inicial, fornece-nos um teste para a independência linear de soluções da equação $\mathbf{x}' = A\mathbf{x}$. Enunciamos em seguida uma versão do teorema referido cuja prova pode ser encontrada em qualquer livro dedicado ao estudo de equações diferenciais como, por exemplo, Braun [2].

Teorema 4.4. Existência e unicidade de soluções

Seja A uma matriz real de ordem n, $\mathbf{x} : \mathbb{R} \to \mathbb{R}^n$ uma função contínua e $\mathbf{x}'$ a sua derivada.

Existe uma e uma só solução do problema de valor inicial

$$\mathbf{x}' = A\mathbf{x} \quad \text{e} \quad \mathbf{x}(t_0) = (x_1^0, x_2^0, \ldots, x_n^0).$$

Além disso, esta solução existe para todo $t \in \mathbb{R}$.

Deixamos como exercício a demonstração da unicidade de soluções de um problema de valor inicial.

O teorema anterior permite-nos provar que o conjunto solução geral da equação $\mathbf{x}' = A\mathbf{x}$ é um subespaço de dimensão n, onde n é a ordem da matriz A. No exercício a seguir sugere-se uma demonstração.

Exercício 4.4. Mostre que o conjunto solução geral da equação $\mathbf{x}' = A\mathbf{x}$ é um subespaço de dimensão n, onde n é a ordem da matriz A.

Sugestão: Considere o conjunto $B = \{\phi_1(t), \ldots, \phi_n(t)\}$, onde $\phi_j(t)$ é a solução do problema de valor inicial $\mathbf{x}' = A\mathbf{x}$ e $\mathbf{x}(0) = \mathbf{e}_j$, com $\mathbf{e}_j$ o vector da base canónica de $\mathbb{R}^n$, $\mathbf{e}_j = (0, 0, \ldots, 1, \ldots, 0)$. Mostre que B é linearmente independente e que qualquer solução da equação $\mathbf{x}' = A\mathbf{x}$ é uma combinação linear dos elementos de B.

▲

A proposição que enunciamos a seguir fornece um teste de grande utilidade prática na verificação da independência linear de soluções de $\mathbf{x}' = A\mathbf{x}$. Nomeadamente, essa proposição garante que as soluções $\mathbf{u}_1(t), \ldots, \mathbf{u}_k(t)$ de $\mathbf{x}' = A\mathbf{x}$ são linearmente independentes se e só se os vectores de $\mathbb{R}^n$, $\mathbf{u}_1(t_0), \ldots, \mathbf{u}_k(t_0)$ são linearmente independentes, onde t_0 é um valor de t que podemos escolher da forma mais conveniente.

Proposição 4.12. Seja $B = \{\mathbf{u}_1(t), \ldots, \mathbf{u}_k(t)\}$ um conjunto de soluções de $\mathbf{x}' = A\mathbf{x}$ e t_0 um valor fixo de t. O conjunto B é linearmente independente se e só se $\{\mathbf{u}_1(t_0), \ldots, \mathbf{u}_k(t_0)\}$ é linearmente independente em $\mathbb{R}$.

Demonstração. Suponha-se que B é linearmente dependente. Ou seja, existem constantes $c_1, \ldots, c_k$ não todas nulas, tais que $c_1\mathbf{u}_1(t) + \cdots + c_k\mathbf{u}_k(t) = \mathbf{0}$. Calculando esta igualdade em $t = t_0$, tem-se

$$c_1\mathbf{u}_1(t_0) + \cdots + c_k\mathbf{u}_k(t_0) = \mathbf{0},$$

com pelo menos um dos c_i's é não nulo. Logo, $\{\mathbf{u}_1(t_0), \ldots, \mathbf{u}_k(t_0)\}$ é linearmente dependente.

Para a implicação recíproca, suponha-se que existem constantes $c_1, \ldots, c_k$, não todas nulas, tais que $c_1\mathbf{u}_1(t_0) + \cdots + c_k\mathbf{u}_k(t_0) = \mathbf{0}$. Considere-se a função

$$\phi(t) = c_1\mathbf{u}_1(t) + \cdots + c_k\mathbf{u}_k(t).$$

A função ϕ é uma solução da equação diferencial $\mathbf{x}' = A\mathbf{x}$ já que, ϕ é uma combinação linear de soluções. Além disso, $\phi(t_0) = \mathbf{0}$. Pelo Teorema de existência e unicidade de soluções de um problema de valor inicial (Teorema 4.4), o problema de valor inicial $\mathbf{x}' = A\mathbf{x}$ com $\mathbf{x}(t_0) = \mathbf{0}$ só admite a solução nula. Logo, $\phi(t) = \mathbf{0}$ para todo o t, e portanto $c_1\mathbf{u}_1(t) + \cdots + c_k\mathbf{u}_k(t) = \mathbf{0}$ com pelo menos um dos c_i's não nulo. Ou seja, B é linearmente dependente. $\square$

Pretendemos agora determinar a solução geral de $\mathbf{x}' = A\mathbf{x}$ em que $\mathbf{x} : \mathbb{R} \to \mathbb{R}^n$ é uma função contínua e A uma matriz real $n \times n$. Para tal, comecemos por abordar o caso em que A é a matriz diagonal $A = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$. O sistema $\mathbf{x}' = A\mathbf{x}$ reduz-se a n equações do tipo já estudado. Nomeadamente,

$$\mathbf{x}' = A\mathbf{x} \Longleftrightarrow \begin{cases} x_1' = \lambda_1 x_1 \\ x_2' = \lambda_2 x_2 \\ \vdots \\ x_n' = \lambda_n x_n. \end{cases}$$

Como vimos anteriormente, no Exercício 4.3, a solução geral de cada equação $x_i' = \lambda_i x_i$ é dada por $x_i(t) = c_i e^{\lambda_i t}$ onde c_i é uma constante real arbitrária. Logo, a solução geral deste sistema é

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix} = \begin{bmatrix} c_1 e^{\lambda_1 t} \\ c_2 e^{\lambda_2 t} \\ \vdots \\ c_n e^{\lambda_n t}. \end{bmatrix} = c_1 e^{\lambda_1 t}\mathbf{e}_1 + c_2 e^{\lambda_2 t}\mathbf{e}_2 + \cdots + c_n e^{\lambda_n t}\mathbf{e}_n$$

$$= \begin{bmatrix} | & | & & | \\ e^{\lambda_1 t}\mathbf{e}_1 & e^{\lambda_2 t}\mathbf{e}_2 & \cdots & e^{\lambda_n t}\mathbf{e}_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = X(t)\mathbf{c},$$

onde $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ são os vectores da base canónica de $\mathbb{R}^n$. Assim, um conjunto gerador da solução geral é $\left\{e^{\lambda_1 t}\mathbf{e}_1, e^{\lambda_2 t}\mathbf{e}_2, \ldots, e^{\lambda_n t}\mathbf{e}_n\right\}$. Este conjunto é uma base para a solução geral já que, avaliando os elementos deste conjunto em $t_0 = 0$ se obtém a base canónica de $\mathbb{R}^n$, e portanto a Proposição 4.12 garante que $\left\{e^{\lambda_1 t}\mathbf{e}_1, \ldots, e^{\lambda_n t}\mathbf{e}_n\right\}$ é linearmente independente.

Podemos tirar as conclusões que se seguem relativas à solução geral do sistema de EDOs $\mathbf{x}' = A\mathbf{x}$ em que A é uma matriz diagonal do tipo $n \times n$.

- Existem n soluções linearmente independentes da forma $e^{\lambda_i t}\mathbf{e}_i$ para $\mathbf{x}' = A\mathbf{x}$.

- A solução geral do sistema é da forma $\mathbf{x}(t) = X(t)\mathbf{c}$, onde X é uma matriz que tem para colunas os n vectores $e^{\lambda_i t}\mathbf{e}_i$, e $\mathbf{c}$ é um vector coluna constante arbitrário. Como se verifica facilmente, os pares $(\lambda_i, \mathbf{e}_i)$ são pares próprios de A.

- A matriz $X(t)$ é invertível, visto que as suas colunas são linearmente independentes. Consequentemente $\det X(t) \neq 0$ para todo o t.

- A derivada da matriz $X(t)$ satisfaz a equação $\mathbf{X}' = A\mathbf{X}$, onde a derivada de uma matriz é a matriz se que obtém derivando entrada a entrada. É fácil verificar que de facto $X'(t) = AX(t)$.

As conclusões anteriores, válidas para uma matriz diagonal, permanecem válidas para qualquer matriz quadrada, como veremos a seguir.

Definição 4.9. Seja A uma matriz do tipo $n \times n$. Chama-se matriz *solução fundamental* do sistema $\mathbf{x}' = A\mathbf{x}$ a qualquer matriz $X(t)$ cujas colunas sejam n soluções linearmente independentes de $\mathbf{x}' = A\mathbf{x}$.

Visto que a solução geral de um sistema homogéneo de n equações diferenciais lineares de 1ª ordem é um espaço linear de dimensão n, as colunas de uma matriz solução fundamental X formam uma base para este espaço. Assim, a solução geral de $\mathbf{x}' = A\mathbf{x}$ é o conjunto das combinações lineares das colunas de X.

A solução geral de $\mathbf{x}' = A\mathbf{x}$ é dada por

$$\mathbf{x}(t) = X(t)\mathbf{c},$$

onde $X(t)$ é uma matriz solução fundamental do sistema e $\mathbf{c}$ é um vector coluna constante arbitrário.

Note-se que, por definição de produto de uma matriz por um vector, a expressão $X(t)\mathbf{c}$ designa precisamente uma combinação linear (arbitrária) das colunas de X.

Exercício 4.5. Mostre que $X(t)$ é uma matriz solução fundamental do sistema $\mathbf{x}' = A\mathbf{x}$ se e só se $X'(t) = AX(t)$ e $\det(X(0)) \neq 0$.

Relembre que a notação $X'(t)$ designa a matriz cujas entradas são as derivadas das entradas de $X(t)$. ▲

Como determinar n soluções linearmente independentes para $\mathbf{x}' = A\mathbf{x}$? A proposição seguinte responde (parcialmente) a esta questão.

Proposição 4.13. Seja A uma matriz real $n \times n$, e $\mathbf{u}$ um vector constante não nulo.

(1) A função $e^{\lambda t}\mathbf{u}$ é solução de $\mathbf{x}' = A\mathbf{x}$ se e só se λ é valor próprio de A com vector próprio associado $\mathbf{u}$.

(2) Se $\mathbf{x}(t) = \mathbf{x}_1(t) + i\mathbf{x}_2(t)$ é uma solução (complexa) de $\mathbf{x}' = A\mathbf{x}$, então $\operatorname{Re}\mathbf{x}(t) = \mathbf{x}_1(t)$ e $\operatorname{Im}\mathbf{x}(t) = \mathbf{x}_2(t)$ são duas soluções reais de $\mathbf{x}' = A\mathbf{x}$.

(3) Se $\lambda = a + ib$, com $b \neq 0$, é um valor próprio (complexo) de A e $\mathbf{u}$ um vector próprio associado, então $\operatorname{Re}(e^{\lambda t}\mathbf{u})$ e $\operatorname{Im}(e^{\lambda t}\mathbf{u})$ são duas soluções reais linearmente independentes de $\mathbf{x}' = A\mathbf{x}$. Ou seja,

$$e^{at}\left(\cos(bt)\operatorname{Re}\mathbf{u} - \operatorname{sen}(bt)\operatorname{Im}\mathbf{u}\right) \quad \text{e} \quad e^{at}\left(\operatorname{sen}(bt)\operatorname{Re}\mathbf{u} + \cos(bt)\operatorname{Im}\mathbf{u}\right),$$

são duas soluções reais linearmente independentes de $\mathbf{x}' = A\mathbf{x}$.

Demonstração. (1) Como $(e^{\lambda t}\mathbf{u})' = \lambda e^{\lambda t}\mathbf{u}$, tem-se que $\mathbf{z}' = A\mathbf{z}$ para $\mathbf{z}(t) = e^{\lambda t}\mathbf{u}$ se e só se

$$\lambda e^{\lambda t}\mathbf{u} = A(e^{\lambda t}\mathbf{u}) \iff \lambda\mathbf{u} = A\mathbf{u},$$

onde na equivalência anterior se aplicou o facto de $e^{\lambda t}$ nunca se anular.

Ou seja, $e^{\lambda t}\mathbf{u}$ é solução de $\mathbf{x}' = A\mathbf{x}$ se e só se $\mathbf{u}$ é um vector próprio de A associado ao valor próprio λ.

(2) Dizer que $\mathbf{x}(t) = \mathbf{x}_1(t) + i\mathbf{x}_2(t)$ é uma solução complexa de $\mathbf{x}' = A\mathbf{x}$ é equivalente a

$$\begin{aligned}
\mathbf{x}'(t) = A\mathbf{x} &\iff \mathbf{x}_1'(t) + i\mathbf{x}_2'(t) = A\left(\mathbf{x}_1(t) + i\mathbf{x}_2(t)\right) \\
&\iff \mathbf{x}_1'(t) + i\mathbf{x}_2'(t) = A\mathbf{x}_1(t) + iA\mathbf{x}_2(t) \\
&\iff \mathbf{x}_1'(t) = A\mathbf{x}_1(t) \quad \text{e} \quad \mathbf{x}_2'(t) = A\mathbf{x}_2(t).
\end{aligned}$$

Ou seja, $\operatorname{Re}\mathbf{x}(t) = \mathbf{x}_1(t)$ e $\operatorname{Im}\mathbf{x}(t) = \mathbf{x}_2(t)$ são duas soluções reais de $\mathbf{x}' = A\mathbf{x}$.

(3) Do item (1) tem-se que $e^{\lambda t}\mathbf{u}$ é uma solução (complexa) da equação diferencial, e pelo item (2) resulta que $\operatorname{Re}(e^{\lambda t}\mathbf{u})$ e $\operatorname{Im}(e^{\lambda t}\mathbf{u})$ são duas soluções reais da respectiva equação diferencial. A independência linear destas soluções segue do Lema 4.1 (na página 198) e da Proposição 4.12 considerando $t_0 = 0$.

Calculemos a parte real e imaginária de $e^{\lambda t}\mathbf{u}$. Para tal, relembremos que $e^{ibt} = \cos(bt) + i\,\mathrm{sen}(bt)$ (ver (A.1) no Apêndice A).

$$\begin{aligned}
e^{(a+ib)t}(\mathrm{Re}\,\mathbf{u} + i\,\mathrm{Im}\,\mathbf{u}) &= e^{at}e^{ibt}(\mathrm{Re}\,\mathbf{u} + i\,\mathrm{Im}\,\mathbf{u}) \\
&= e^{at}(\cos(bt) + i\,\mathrm{sen}(bt))(\mathrm{Re}\,\mathbf{u} + i\,\mathrm{Im}\,\mathbf{u}) \\
&= e^{at}\left[(\cos(bt)\,\mathrm{Re}\,\mathbf{u} - \mathrm{sen}(bt)\,\mathrm{Im}\,\mathbf{u})\right. \\
&\quad \left. + \; i\,(\mathrm{sen}(bt)\,\mathrm{Re}\,\mathbf{u} + \cos(bt)\,\mathrm{Im}\,\mathbf{u})\right].
\end{aligned}$$

Assim, $\mathrm{Re}(e^{\lambda t}\mathbf{u})$ e $\mathrm{Im}(e^{\lambda t}\mathbf{u})$ são dadas pelas expressões no enunciado.

□

Podemos agora determinar a solução geral de $\mathbf{x}' = A\mathbf{x}$ para matrizes diagonalizáveis A. O caso em que A não é diagonalizável é tratado no Capítulo 8.

Recorde-se que uma matriz A, de ordem n, é diagonalizável se e só se admite n vectores próprios linearmente independentes. Neste caso, pela Proposição 4.13, o sistema $\mathbf{x}' = A\mathbf{x}$ tem n soluções reais linearmente independentes da seguinte forma: (i) $e^{\lambda t}\mathbf{u}$ com λ um valor próprio real de A e $\mathbf{u}$ um vector próprio associado; (ii) para cada par de valores próprios complexos conjugados λ e $\overline{\lambda}$, existem duas soluções reais (linearmente independentes) da forma $\mathrm{Re}(e^{\lambda t}\mathbf{u})$ e $\mathrm{Im}(e^{\lambda t}\mathbf{u})$, onde $\mathbf{u}$ é um vector próprio associado a λ. A solução geral de $\mathbf{x}' = A\mathbf{x}$ é uma combinação linear destas n soluções linearmente independentes.

Apresentamos a seguir um exemplo da determinação da solução geral de um sistema $\mathbf{x}' = A\mathbf{x}$ em que A possui valores próprios reais e complexos.

Exemplo 4.15. Determine-se a solução geral do sistema $\mathbf{x}' = A\mathbf{x}$, em que A é a matriz do Exemplo 4.9 (pág. 194). O sistema correspondente é

$$\begin{cases} x_1' &= 2x_1 + 5x_2 \\ x_2' &= -x_1 \\ x_3' &= 5x_3. \end{cases}$$

A matriz A possui valores próprios 5 e $1 \pm 2i$ e os espaços próprios são:

$$E(5) = \mathrm{Span}\,\{(0,0,1)\}\,, \quad E(1+2i) = \mathrm{Span}\,\{(1+2i,-1,0)\}$$
$$E(1-2i) = \mathrm{Span}\,\{(1-2i,-1,0)\}\,.$$

Usando a Proposição 4.13, são soluções (reais) linearmente independentes do sistema:

$$e^{5t}\begin{bmatrix}0\\0\\1\end{bmatrix}, \quad \mathrm{Re}\left(e^{(1+2i)t}\begin{bmatrix}1+2i\\-1\\0\end{bmatrix}\right) \;\text{ e }\; \mathrm{Im}\left(e^{(1+2i)t}\begin{bmatrix}1+2i\\-1\\0\end{bmatrix}\right).$$

Como $e^{(1+2i)t} = e^t e^{2it} = e^t(\cos(2t) + i\operatorname{sen}(2t))$, resulta

$$\operatorname{Re}\left(e^{(1+2i)t}\begin{bmatrix}1+2i\\-1\\0\end{bmatrix}\right) = e^t\left(\cos(2t)\begin{bmatrix}1\\-1\\0\end{bmatrix} - \operatorname{sen}(2t)\begin{bmatrix}2\\0\\0\end{bmatrix}\right)$$
$$\operatorname{Im}\left(e^{(1+2i)t}\begin{bmatrix}1+2i\\-1\\0\end{bmatrix}\right) = e^t\left(\operatorname{sen}(2t)\begin{bmatrix}1\\-1\\0\end{bmatrix} + \cos(2t)\begin{bmatrix}2\\0\\0\end{bmatrix}\right).$$

Conclui-se que a solução geral do sistema é o conjunto das funções

$$\mathbf{x}(t) = c_1 e^{5t}\begin{bmatrix}0\\0\\1\end{bmatrix} + c_2 e^t\left(\cos(2t)\begin{bmatrix}1\\-1\\0\end{bmatrix} - \operatorname{sen}(2t)\begin{bmatrix}2\\0\\0\end{bmatrix}\right) +$$
$$+ c_3 e^t\left(\operatorname{sen}(2t)\begin{bmatrix}1\\-1\\0\end{bmatrix} + \cos(2t)\begin{bmatrix}2\\0\\0\end{bmatrix}\right),$$

com c_1, c_2 e c_3 constantes reais arbitrárias.

Note-se que uma matriz solução fundamental para este sistema é

$$X(t) = \begin{bmatrix}0 & e^t(\cos(2t) - 2\operatorname{sen}(2t)) & e^t(\operatorname{sen}(2t) + 2\cos(2t))\\ 0 & -e^t\cos(2t) & -e^t\operatorname{sen}(2t)\\ e^{5t} & 0 & 0\end{bmatrix}.$$

♦

Exercício 4.6. Para a matriz A do Exemplo 4.8 (pág. 194), mostre que uma matriz solução fundamental para o sistema $\mathbf{x}' = A\mathbf{x}$ é

$$X(t) = \begin{bmatrix}-e^{3t} & 0 & e^{5t}\\ 0 & e^{3t} & 2e^{5t}\\ e^{3t} & 0 & e^{5t}\end{bmatrix}.$$

Calcule ainda $X(t)X(0)^{-1}$.

▲

Da definição de matriz solução fundamental, vê-se facilmente que não existe uma única matriz solução fundamental. Não é difícil mostrar que quaisquer duas matrizes $X(t), Y(t)$, solução fundamental de um sistema, satisfazem uma relação do tipo $Y(t) = X(t)K$ onde K é uma matriz (constante) invertível.

Exercício 4.7. Mostre que se $X(t)$ e $Y(t)$ são duas quaisquer matrizes solução fundamental do sistema $\mathbf{x}' = A\mathbf{x}$, então existe uma matriz real invertível K tal que $Y(t) = X(t)K$.

Sugestão: Escrever as colunas de Y como combinação linear das colunas de X e usar os factos $\det Y(0) \neq 0$ e $\det X(0) \neq 0$ para mostrar que K é invertível.
▲

A Proposição 4.13 diz-nos como determinar uma base do conjunto das soluções de $\mathbf{x}' = A\mathbf{x}$ no caso em que A é diagonalizável. Conhecida a solução geral da equação homogénea $\mathbf{x}' = A\mathbf{x}$ podemos determinar a solução geral da equação não homogénea $\mathbf{x}' = A\mathbf{x} + \mathbf{b}(t)$ desde que se conheça uma solução desta equação. De facto, à semelhança do que acontece para sistemas de equações lineares (não diferenciais), tem-se:

A solução geral da equação $\mathbf{x}'(t) = A\mathbf{x}(t) + \mathbf{b}(t)$ é a soma de uma solução particular $\mathbf{x}_p$ da equação com a solução geral $\mathbf{x}_h$ da equação homogénea associada, $\mathbf{x}'(t) = A\mathbf{x}(t)$. Ou seja, a solução geral é

$$\{\mathbf{x}(t) : \mathbf{x}(t) = \mathbf{x}_p(t) + \mathbf{x}_h(t)\}.$$

Deixamos como exercício a demonstração deste resultado que se reduz a um mero decalque da demonstração apresentada para o Teorema 3.7 da página 153.

No exemplo que se segue aplicamos o resultado anterior para determinar a solução geral de uma equação diferencial não homogénea.

Exemplo 4.16. Considere-se a seguinte EDO não homogénea $x' - 3x = 2t - 3t^2$.

A função $v(t) = t^2$ é uma solução desta equação como facilmente se verifica. A equação homogénea associada, $x' - 3x = 0$, tem solução geral $x(t) = ce^{3t}$, com $c \in \mathbb{R}$. Logo, a solução geral da equação não homogénea é: $\{ce^{3t} + t^2 : c \in \mathbb{R}\}$.
♦

Exponencial de matrizes

Como se viu, dada uma equação diferencial, $\mathbf{x}' = A\mathbf{x}$, existem várias matrizes solução fundamental dessa equação. A exponencial e^{At} vai ser definida como sendo uma matriz solução fundamental particular do sistema $\mathbf{x}' = A\mathbf{x}$. É possível definir a matriz exponencial e^{At} como uma série de potências da matriz At, semelhante à série de potências que define a função exponencial real. No entanto essa via sai do âmbito deste texto, o leitor interessado poderá consultar, por exemplo, Braun [2].

Relembremos (Exercício 4.5) que uma matriz solução fundamental do sistema $\mathbf{x}' = A\mathbf{x}$ é uma matriz que verifica a equação diferencial para matrizes $X'(t) = AX(t)$.

Definição 4.10. Seja A uma matriz quadrada. A exponencial e^{At} é a matriz solução fundamental do sistema $\mathbf{x}' = A\mathbf{x}$ cujo valor em $t = 0$ é a matriz identidade. Ou seja, e^{At} é a solução do problema de valor inicial

$$\begin{cases} X' = AX \\ X(0) = I, \end{cases}$$

onde I designa a matriz identidade e X uma matriz da mesma ordem de A.

Nota 25. *O problema de valor inicial na definição anterior é um problema de valor inicial para matrizes. Este problema pode ser visto como um conjunto de n problemas de valor inicial para o sistema $\mathbf{x}' = A\mathbf{x}$ (correspondentes a problemas de valor inicial para as colunas de X).*

Tendo em conta a definição anterior e o resultado do Exercício 4.7 (pág. 228), podemos enunciar a proposição seguinte.

Proposição 4.14. Seja A uma matriz real $n \times n$ e $X(t)$ uma qualquer matriz solução fundamental do sistema $\mathbf{x}' = A\mathbf{x}$. A exponencial e^{At} é dada por

$$e^{At} = X(t)X(0)^{-1}.$$

Demonstração. Se e^{At} e $X(t)$ são duas matrizes solução fundamental do sistema $\mathbf{x}' = A\mathbf{x}$, do Exercício 4.7, segue que

$$e^{At} = X(t)K,$$

onde K é uma matriz (constante) invertível. Como por definição $e^{A0} = I$, resulta

$$e^{A0} = I = X(0)K \Longleftrightarrow K = X(0)^{-1}.$$

Por conseguinte, $e^{At} = X(t)X(0)^{-1}$. □

Nota 26. *Uma vez que e^{At} é solução de um problema de valor inicial, segue da unicidade de soluções deste tipo de problemas que a matriz e^{At} é única. Além disso, como e^{At} é uma matriz solução fundamental de $\mathbf{x}' = A\mathbf{x}$, é satisfeita a igualdade $\frac{d}{dt}(e^{At}) = Ae^{At}$ (ver Exercício 4.5).*

Exercício 4.8. a) Mostre que $e^{(A+B)t} = e^{At}e^{Bt}$ se e só se A e B comutam.

b) Use o resultado anterior para mostrar que a matriz inversa de e^{At} é e^{-At}.

Sugestão: Para a alínea a) mostre os resultados:

- Se $AB = BA$, então $X(t) = Be^{At}$ e $Y(t) = e^{At}B$ são matrizes solução fundamental do mesmo p.v.i.. Use a unicidade de soluções de problemas de valor inicial para mostrar que $X(t) = Y(t)$.
- Mostre que $e^{At}e^{Bt}$ e $e^{t(A+B)}$ resolvem o mesmo problema de valor inicial.
- Finalmente, mostre que se $X(t) = Y(t)$, o que necessariamente implica $X'(t) = Y'(t)$, se tem $AB = BA$.

▲

Exemplo 4.17. Calculemos e^{At} para $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$.

No Exemplo 4.4 (pág. 182) calculámos os valores próprios e os espaços próprios desta matriz. Os valores próprios de A são $\lambda = \pm i$ e $E(-i) = \operatorname{Span}\{(1, i)\}$. Logo, pelo item (2) da Proposição 4.13, são soluções (reais) linearmente independentes do sistema $\mathbf{x}' = A\mathbf{x}$, os vectores

$$\operatorname{Re}\left(e^{-it}\begin{bmatrix} 1 \\ i \end{bmatrix}\right) \quad \text{e} \quad \operatorname{Im}\left(e^{-it}\begin{bmatrix} 1 \\ i \end{bmatrix}\right),$$

ou, equivalentemente,

$$\cos(-t)\begin{bmatrix} 1 \\ 0 \end{bmatrix} - \operatorname{sen}(-t)\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} \cos t \\ \operatorname{sen} t \end{bmatrix}, \qquad \operatorname{sen}(-t)\begin{bmatrix} 1 \\ 0 \end{bmatrix} + \cos(-t)\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -\operatorname{sen} t \\ \cos t \end{bmatrix}.$$

Nas igualdades anteriores aplicámos $\cos(-t) = \cos t$ e $\operatorname{sen}(-t) = -\operatorname{sen} t$.

Uma matriz solução fundamental para o sistema $\mathbf{x}' = A\mathbf{x}$ é uma matriz cujas colunas são soluções linearmente independentes do sistema. Neste caso, uma matriz solução fundamental $X(t)$ é, por exemplo,

$$X(t) = \begin{bmatrix} \cos t & -\operatorname{sen} t \\ \operatorname{sen} t & \cos t \end{bmatrix}.$$

Como $X(0)$ é a matriz identidade, temos $e^{At} = X(t)X(0)^{-1} = X(t)$. Verifique ainda que se tivéssemos considerado para $X(t)$ a matriz que se obtém trocando as colunas da matriz acima a expressão $X(t)X(0)^{-1}$ produziria o mesmo resultado (e^{At} é única). ♦

Exercício 4.9. Considere o sistema de equações diferenciais $\mathbf{x}' = A\mathbf{x}$, onde A é uma matriz diagonalizável, isto é, $A = PDP^{-1}$ com D diagonal.

a) Mostre que o sistema dado é equivalente ao sistema $\mathbf{y}' = D\mathbf{y}$, onde $\mathbf{y} = P^{-1}\mathbf{x}$.

b) Use a alínea anterior, para mostrar a igualdade $e^{At} = Pe^{Dt}P^{-1}$.

▲

Equações de ordem n e redução de ordem

Para finalizar esta secção vamos verificar que podemos obter a solução geral de equações diferenciais ordinárias lineares homogéneas, de coeficientes constantes e de ordem superior à primeira, resolvendo um sistema da forma $\mathbf{x}' = A\mathbf{x}$.

Uma *equação diferencial ordinária, linear, homogénea, de coeficientes constantes, e de ordem n*, é uma equação diferencial da forma

$$y^{(n)} + a_{n-1}y^{(n-1)} + \cdots + a_1 y' + a_0 y = 0, \tag{4.26}$$

onde (os coeficientes) $a_i \in \mathbb{R}$ para $i = 1, \ldots, n$, a função $y : \mathbb{R} \to \mathbb{R}$ é real de variável real, e $y^{(k)}$ designa a derivada de ordem k de y.

Introduzindo novas variáveis $x_1 = y, x_2 = y', \ldots, x_n = y^{(n-1)}$, a equação (4.26) é equivalente a um sistema da forma $\mathbf{x}' = A\mathbf{x}$. Nomeadamente

$$\begin{array}{l} x_1 = y \\ x_2 = y' \\ \vdots \\ x_n = y^{(n-1)} \end{array} \quad \Longrightarrow \quad \begin{cases} x_1' = x_2 \\ x_2' = x_3 \\ \vdots \\ x_n' = -a_0x_1 - a_1x_2 - \cdots - a_{n-1}x_n. \end{cases} \tag{4.27}$$

A matriz A do sistema obtido é conhecida por matriz *companheira* da equação (4.26) e tem a forma

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots\cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots\cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots\cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ -a_0 & -a_1 & -a_2 & \cdots & -a_{n-2} & -a_{n-1}. \end{bmatrix}.$$

Exercício 4.10. Mostre que o polinómio característico da matriz companheira da equação (4.26) é $p(\lambda) = (-1)^n[\lambda^n + a_{n-1}\lambda^{(n-1)} + \cdots + a_1\lambda + a_0]$.
Sugestão: Calcule $\det(A - \lambda I)^T$ usando o método de eliminação de Gauss. ▲

Atendendo ao último exercício, o polinómio característico da matriz companheira de uma equação de ordem n pode ser obtido directamente a partir da equação (4.26).

Após reduzir a equação (4.26) a um sistema de EDOs homogéneo de primeira ordem, $\mathbf{x}' = A\mathbf{x}$, a solução geral da equação (4.26) obtém-se da solução geral do sistema (4.27) considerando apenas a primeira componente da solução $\mathbf{x}$ do sistema, visto que fizemos $y = x_1$.

Exemplo 4.18. Determinemos a solução geral da equação diferencial de terceira ordem, $y''' - 6y'' + 11y' - 6y = 0$.

Fazendo $y = x_1, y' = x_2$ e $y'' = x_3$, a equação diferencial reduz-se ao sistema

$$\begin{bmatrix} x_1' \\ x_2' \\ x_3' \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 6 & -11 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

Pelo Exercício 4.10, a equação característica da matriz companheira é $\lambda^3 - 6\lambda^2 + 11\lambda - 6 = 0$. É fácil verificar que esta equação tem raízes $\lambda = 3, \lambda = 2$ e $\lambda = 1$. Como os valores próprios são distintos, a matriz companheira é diagonalizável.

Uma base de vectores próprios é, por exemplo, $\{(1,3,9),(1,2,4),(1,1,1)\}$. Esta base é constituída por vectores próprios associados respectivamente a $3, 2$ e 1. A solução geral do sistema é assim

$$\begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} = c_1e^{3t} \begin{bmatrix} 1 \\ 3 \\ 9 \end{bmatrix} + c_2e^{2t} \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} + c_3e^{t} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Uma vez que fizemos $y = x_1$, a solução geral da equação diferencial de terceira ordem é

$$y(t) = x_1(t) = c_1e^{3t} + c_2e^{2t} + c_3e^{t},$$

onde c_1, c_2 e c_3 são constantes reais arbitrárias. ♦

Exercícios

1. Diga quais dos vectores seguintes são vectores próprios da matriz

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \\ 2 & 0 & 3 \end{bmatrix}.$$

Em caso afirmativo, indique o valor próprio correspondente.

a) $(0,2,2)$ b) $(0,5,0)$ c) $(0,0,0)$
d) $(-1,-1,2)$ e) $(-2,-2,2)$.

2. Verifique se $\lambda = 3$ é ou não um valor próprio de $\begin{bmatrix} 1 & 2 & -3 \\ 0 & 3 & 4 \\ -1 & 1 & 5 \end{bmatrix}$, e em caso afirmativo determine um vector próprio associado.

3. Sejam A e B matrizes de segunda ordem tais que

- $A\mathbf{x}$ é o vector que se obtém de $\mathbf{x}$ projectando-o sobre o eixo dos xx.
- $B\mathbf{x}$ é o vector que se obtém rodando $\mathbf{x}$ de 90º no sentido directo.

Indique o valor lógico das afirmações seguintes:

a) Os valores próprios de A são 1 e -1.
b) O vector $(2,0)$ e $(0,2)$ são vectores próprios de A associados respectivamente aos valores próprios 1 e 0.
c) $\lambda = 1$ é um valor próprio de B.

4. Considere as matrizes

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \text{ e } B = A + 5I.$$

a) Determine os valores próprios e os espaços próprios de A e de B.
b) Compare os resultados obtidos na alínea anterior e estabeleça a relação existente entre os valores próprios e os vectores próprios de uma matriz C e os da matriz $C + kI$.

5. Sem efectuar quaisquer cálculos, determine os valores próprios e dois vectores próprios linearmente independentes para a matriz

$$\begin{bmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 2 \end{bmatrix}.$$

Caso exista, indique uma base de $\mathbb{R}^3$ constituída por vectores próprios.

6. Suponha que $\mathbf{v}$ é um vector próprio de uma matriz invertível M. Mostre que $\mathbf{v}$ também é um vector próprio de M^{-1} e determine o valor próprio de M^{-1} que lhe está associado.

7. Seja A uma matriz 3×3 com valores próprios $0, 2$ e 3.

a) Determine a característica de A.
b) Calcule o determinante de $A^T A$.
c) Calcule os valores próprios de $(A + I)^{-1}$.

8. Para cada uma das matrizes seguintes determine os valores próprios e uma base para cada espaço próprio.

a) $\begin{bmatrix} 3 & 0 \\ 8 & -1 \end{bmatrix}$ b) $\begin{bmatrix} 0 & 3 \\ 4 & 0 \end{bmatrix}$

c) $\begin{bmatrix} -2 & -7 \\ 1 & 2 \end{bmatrix}$ d) $\begin{bmatrix} -2 & 0 & 1 \\ -6 & -2 & 0 \\ 19 & 5 & -4 \end{bmatrix}$

e) $\begin{bmatrix} 5 & 0 & 1 \\ 1 & 1 & 0 \\ -7 & 1 & 0 \end{bmatrix}$

f) $\begin{bmatrix} 0 & 0 & 2 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

9. Determine quais das seguintes matrizes são diagonalizáveis. Em caso afirmativo encontre uma matriz P que diagonaliza a matriz dada A (ou seja, tal que $P^{-1}AP$ é diagonal), indicando a respectiva matriz diagonal.

a) $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ b) $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

c) $\begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 1 & 2 \end{bmatrix}$ d) $\begin{bmatrix} 2 & 0 & -2 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix}$

e) $\begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 2 \end{bmatrix}$

10. Considere a matriz

$$A = \begin{bmatrix} -1 & -2 & -2 \\ 1 & 2 & 1 \\ -1 & -1 & 0 \end{bmatrix}$$

a) Verifique que A é diagonalizável e determine uma matriz diagonal D e uma matriz P tal que $A = PDP^{-1}$.
b) Use a alínea anterior para: (i) calcular A^{25}; (ii) encontrar os valores próprios de A^{25}; (iii) determinar uma base para cada espaço próprio.

11. Para uma matriz real A do tipo $n \times n$, diga se são verdadeiras ou falsas as afirmações seguintes.

a) Se existe uma base de $\mathbb{R}^n$ formada por vectores próprios de A, então A é diagonalizável.
b) A é diagonalizável se tem n vectores próprios.
c) A é diagonalizável se e só se tem n valores próprios contando as suas multiplicidades.
d) Se A é diagonalizável, então tem n valores próprios distintos.
e) Se A tem n valores próprios distintos, então é diagonalizável.
f) Se A é diagonalizável, então é invertível.
g) Se $n = 3$ e A tem dois valores próprios tal que a dimensão de cada espaço próprio é 1, então A é diagonalizável.
h) Se A é uma matriz 5×5 com dois espaços próprios de dimensões 2 e 3, então a matriz A é diagonalizável.

12. Uma matriz L diz-se nilpotente de índice $k \geq 1$ se $L^k = \mathbf{O}$ e $L^{k-1} \neq \mathbf{O}$ (onde $\mathbf{O}$ designa a matriz nula). Considere uma matriz A de ordem n com um único valor próprio μ. Mostre que

a) Se $\dim N(A-\mu I) = n$, então a matriz $(A-\mu I)$ é uma matriz nilpotente de índice 1.
b) Se $\dim N(A-\mu I) < n$, então a matriz $(A-\mu I)$ é uma matriz nilpotente de índice $1 < k \leq n$. (Use o Teorema de Cayley-Hamilton).

13. Seja A uma matriz 3×3 tal que $\dim N(A-3I) = 1$ e

$$N(A-2I) = \{\mathbf{0}\}, \quad N(A) \neq \{\mathbf{0}\}.$$

Diga quais das afirmações seguintes são verdadeiras.

a) 2 não é valor próprio de A.
b) A não é invertível.
c) Zero não é um valor próprio de A.

d) 3 é um valor próprio de A com multiplicidade algébrica menor que a multiplicidade geométrica.
e) Se $\dim N(A) = 2$, a matriz A é semelhante à matriz diagonal $\mathrm{diag}(0, 0, 3)$.
f) A matriz $A - 2I$ é invertível.

14. Sejam $\mathbf{u}, \mathbf{v}$ e $\mathbf{w}$ vectores não nulos de $\mathbb{R}^3$ e A uma matriz 3×3 tal que $A\mathbf{u} = 2\mathbf{u}$, $A\mathbf{v} = \mathbf{0}$ e $A\mathbf{w} = \mathbf{w}$. Diga quais das afirmações seguintes são verdadeiras.

a) $B = \{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ é uma base de $\mathbb{R}^3$.
b) A matriz A é invertível.
c) A matriz A não é diagonalizável.
d) $\lambda = 1$ é um valor próprio de A.

15. Seja $A = [a_{ij}]_{i,j=1,2}$ uma matriz real tal que $a_{11} + a_{22} = 0$ e $\det(A) = 1$. Diga qual das afirmações seguintes é verdadeira:

a) Zero é um valor próprio de A.
b) O polinómio característico de A é $p(\lambda) = \lambda^2 + 1$
c) A matriz A não tem valores próprios complexos.
d) 1 e -1 são valores próprios de A.

16. Para a matriz $A = \begin{bmatrix} 3 & 0 & 0 \\ 0 & \sqrt{2} & 1 \\ 0 & -1 & 0 \end{bmatrix}$ e $f(\mathbf{x}) = A\mathbf{x}$, diga quais das afirmações seguintes são verdadeiras.

a) O eixo dos xx é um subespaço invariante para f.
b) O plano yz é um subespaço invariante para f.
c) O eixo dos zz é um subespaço invariante para f.
d) Se $\mathbf{u}$ é um vector do plano yz, então $\mathbf{w} = f(\mathbf{u})$ é o vector que se obtém de $\mathbf{u}$ por rotação em torno do eixo dos xx de $\pi/2$.
e) Se $\mathbf{u}$ é um vector que não pertence ao plano yz nem ao eixo dos xx, então a sucessão $\mathbf{u}, f(\mathbf{u}), f^2(\mathbf{u}), \ldots$ percorre uma hélice num cilindro cujo eixo é o eixo dos xx.

17. Considere a matriz $A = \begin{bmatrix} 0.97 & 0.05 \\ 0.03 & 0.95 \end{bmatrix}$ cuja soma por colunas é constante, e o vector $\mathbf{x}_0 = \begin{bmatrix} 0.3 \\ 0.7 \end{bmatrix}$.

a) Justifique por que razão $\lambda = 1$ é um valor próprio de A. Sem calcular explicitamente, mostre ainda que $\lambda = 0.92$ é também um valor próprio de A.
b) Mostre que $\mathbf{v} = \begin{bmatrix} 5 \\ 3 \end{bmatrix}$ e $\mathbf{u} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ são vectores próprios, indicando os valores próprios correspondentes.
c) Mostre que pode escrever a sucessão $\mathbf{x}_{k+1} = A\mathbf{x}_k$ $(k = 0, 1, 2, \ldots)$ na forma $\mathbf{x}_{k+1} = c_1 \begin{bmatrix} 5 \\ 3 \end{bmatrix} + c_2 (0.92)^k \begin{bmatrix} 1 \\ -1 \end{bmatrix}$, onde c_1 e c_2 são as componentes de $\mathbf{x}_0$ na base $\{\mathbf{u}, \mathbf{v}\}$.
d) Use os resultados das alíneas anteriores para determinar: $\lim_{k\to\infty} \mathbf{x}_{k+1}$.

18. Seja $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ uma matriz de Markov. Mostre que se $\lambda \neq 1$ é um valor próprio de A, então $\lambda < 1$. Sugestão: use o traço de A.

19. Anualmente, numa certa população de elefantes 2% dos elefantes ficam velhos, 3% dos elefantes velhos morrem e não nascem elefantes. Designe-se por N, V, e M respectivamente o número de elefantes novos, velhos e dos que morrem. Determine o vector estacionário do sistema correspondente:

$$\mathbf{x}_{k+1} = \begin{bmatrix} 0.98 & 0 & 0 \\ 0.2 & 0.97 & 0 \\ 0 & 0.03 & 1 \end{bmatrix} \mathbf{x}_k$$

com $\mathbf{x} = (N, V, M)$.

20. Verifique quais das matrizes seguintes são de Markov e caso o sejam determine um vector de equilíbrio.

a) $\begin{bmatrix} 1/2 & -1/2 \\ 1/2 & -1/2 \end{bmatrix}$ b) $\begin{bmatrix} 1/3 & 1/2 & 1/4 \\ 1/3 & 0 & 1/2 \\ 1/3 & 1/2 & 1/4 \end{bmatrix}$

c) $\begin{bmatrix} 0 & 1/2 & 0 \\ 1 & 1/2 & 1/3 \\ 0 & 0 & 2/3 \end{bmatrix}$

21. Para o problema do Exemplo 4.13 (pág. 216) obteve-se a matriz:

$$M = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 1/2 \\ 1/3 & 1/2 & 0 & 0 \end{bmatrix}.$$

a) Determine o grafo direccionado correspondente.
b) Determine a relevância das páginas da rede.

22. Supondo que a população $P(t)$ de um certo planeta cresce à taxa (instantânea) de 5% ao ano, isto é $P' = 0,05P$, determine ao fim de quanto tempo duplica a população.

23. A equação diferencial $y' = \frac{1}{10}y$ modela o crescimento de uma certa cultura de bactérias. Se a cultura de bactérias tem 100 indivíduos no instante $t = 0$ determine ao fim de quanto tempo atinge 3000.

24. Seja $T(t)$ a temperatura no instante t de uma sala aquecida. A temperatura da sala decresce a razão de 0,008 vezes a diferença entre T e a temperatura exterior fixa a 15 graus. Este problema é modelado pela equação diferencial

$$T' = -0,008(T - 15).$$

a) Faça a mudança de variável $T - 15 = y$ e escreva a correspondente equação diferenciável para y. Determine ainda a respectiva solução geral.
b) Determine ao fim de quanto tempo a temperatura da sala atinge os 18 graus, sabendo que no instante inicial $t = 0$ a temperatura da sala era de 25 graus.

25. Determine a solução geral dos seguintes sistemas de equações diferenciais.

a) $\begin{cases} y_1' &= y_1 + y_2 \\ y_2' &= 5y_1 + y_2 \end{cases}$

b) $\begin{cases} y_1' &= 3y_1 + 2y_2 \\ y_2' &= y_1 + y_2 \end{cases}$

c) $\begin{cases} y_1' &= 3y_1 - 2y_2 \\ y_2' &= y_1 + y_2 \\ y_3' &= y_2 - y_3 \end{cases}$

26. Para cada um dos sistemas do problema anterior determine a solução que satisfaz as condições

a) $y_1(0) = 0$ e $y_2(0) = 0$.
b) $y_1(0) = 2$ e $y_2(0) = 1$.
c) $y_1(0) = -1, y_2(0) = 1$ e $y_3(0) = 0$.

27. Para cada uma das matrizes A seguintes, determine e^{At}.

a) $A = \begin{bmatrix} 2 & -2 \\ 1 & 0 \end{bmatrix}$. b) $A = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

c) $A = \begin{bmatrix} 2 & 0 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

28. Escreva as seguintes equações diferenciais na forma de um sistema de equações diferenciais de 1ª ordem e determine a sua solução geral.

a) $y'' - y' - 6y = 0$.
b) $y''' - 3y'' + 2y' - 1y = 0$.

29. Mostre que t^2 é uma solução da equação diferencial (não homogénea) $y'' - y' - 6y = 2 - 2t - 6t^2$. Use ainda a solução geral da equação da alínea a) do problema anterior, para determinar a solução geral desta equação.

Capítulo 5

Produto interno e ortogonalidade

O conceito de produto interno em $\mathbb{R}^2$ é generalizado neste capítulo a um espaço linear qualquer. Munindo um espaço linear de um produto interno, ficamos habilitados a calcular distâncias entre vectores do espaço bem como a concretizar noções fundamentais como ortogonalidade, projecção ortogonal, bases ortogonais e ortonormadas. Em particular, nas secções 5.3 e 5.5 são discutidos alguns tópicos fundamentais, como sejam a decomposição ortogonal de um espaço linear em subespaços, o Teorema da melhor aproximação e o processo de ortogonalização de Gram-Schmidt. Na Secção 5.4 estudam-se as propriedades de ortogonalidade dos subespaços fundamentais associados a uma matriz.

O produto interno usual em $\mathbb{R}^3$ é aplicado para determinar equações de rectas e planos nesse espaço e calcular distâncias de pontos a planos. Na Secção 5.7.2 o Teorema da melhor aproximação e as propriedades de ortogonalidade dos subespaços fundamentais associados a uma matriz, aplicam-se na determinação de soluções de mínimos quadrados de sistemas lineares.

Finalizamos o capítulo tratando, na Secção 5.7.3, a aproximação de funções reais por polinómios trigonométricos. Neste tipo de aproximação, é crucial a propriedade de ortogonalidade (em relação ao produto interno usual em $\mathbb{C}^n$) da base formada pelas colunas da matriz de Fourier. Relaciona-se ainda o denominado algoritmo da transformada de Fourier rápida com uma factorização da matriz de Fourier.

5.1 Produto interno num espaço linear

Muitos dos resultados de álgebra linear são de natureza geométrica, visto terem surgido da necessidade de generalizar propriedades geométricas básicas de $\mathbb{R}^2$ e $\mathbb{R}^3$ a espaços lineares de dimensão superior a 3. Nestes espaços são bem conhecidos os conceitos de comprimento e perpendicularidade. Procura-se nesta secção generalizar estas noções, não só a $\mathbb{R}^n$ mas também a qualquer outro espaço linear. Como veremos, uma vez definido o conceito de produto interno num espaço linear, é fácil efectuar a generalização das noções de comprimento, distância e ortogonalidade.

Comecemos por relembrar que pontos do plano (ou seja, vectores de $\mathbb{R}^2$) podem escrever-se usando coordenadas polares. De acordo com a Figura 5.1, dados os vectores $\mathbf{u} = (u_1, u_2)$ e $\mathbf{v} = (v_1, v_2)$ podemos escrever

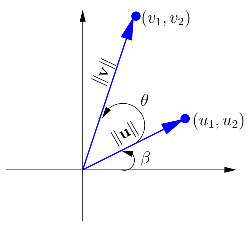


Figura 5.1: Coordenadas polares.

$$\mathbf{u} = (u_1, u_2) \rightsquigarrow \begin{cases} u_1 = \|\mathbf{u}\| \cos\beta \\ u_2 = \|\mathbf{u}\| \operatorname{sen}\beta \end{cases} \qquad \mathbf{v} = (v_1, v_2) \rightsquigarrow \begin{cases} v_1 = \|\mathbf{v}\| \cos(\beta+\theta) \\ v_2 = \|\mathbf{v}\| \operatorname{sen}(\beta+\theta) \end{cases},$$

onde $\|\mathbf{u}\|$ designa o comprimento do vector $\mathbf{u}$, e β e θ os ângulos indicados nessa figura. Calculemos a quantidade $u_1v_1 + u_2v_2$:

$$\begin{aligned} u_1v_1 + u_2v_2 &= \|\mathbf{u}\|\|\mathbf{v}\| \cos\beta\cos(\beta+\theta) + \|\mathbf{u}\|\|\mathbf{v}\| \operatorname{sen}\beta\operatorname{sen}(\beta+\theta) \\ &= \|\mathbf{u}\|\|\mathbf{v}\| \left[\cos\beta\left(\cos\beta\cos\theta - \operatorname{sen}\beta\operatorname{sen}\theta\right) + \operatorname{sen}\beta\left(\operatorname{sen}\beta\cos\theta + \operatorname{sen}\theta\cos\beta\right)\right] \\ &= \|\mathbf{u}\|\|\mathbf{v}\| \left(\cos^2\beta\cos\theta + \operatorname{sen}^2\beta\cos\theta\right) = \|\mathbf{u}\|\|\mathbf{v}\|\cos\theta. \end{aligned}$$

O produto interno de dois vectores $\mathbf{u} = (u_1, u_2)$ e $\mathbf{v} = (v_1, v_2)$ de $\mathbb{R}^2$ define-se como

$$\langle \mathbf{u}, \mathbf{v} \rangle = \langle (u_1, u_2), (v_1, v_2) \rangle = u_1v_1 + u_2v_2 = \|\mathbf{u}\|\|\mathbf{v}\|\cos\theta. \tag{5.1}$$

onde $\|\mathbf{u}\|$ e $\|\mathbf{v}\|$ designam respectivamente os comprimentos de $\mathbf{u}$ e de $\mathbf{v}$, e θ é o ângulo que $\mathbf{u}$ faz com $\mathbf{v}$. A fórmula (5.1) ainda é válida em $\mathbb{R}^3$.

É pois natural tomar como definição de produto interno em $\mathbb{R}^n$ uma generalização da expressão obtida em (5.1). Isto é, para $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, podemos definir um produto interno como

$$\langle \mathbf{u}, \mathbf{v} \rangle = \langle (u_1, u_2, \ldots, u_n), (v_1, v_2, \ldots, v_n) \rangle = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n. \quad (5.2)$$

Se considerarmos $\mathbf{u}$ e $\mathbf{v}$ vectores coluna de $\mathbb{R}^n$, a expressão matricial do produto interno em (5.2) tem a forma

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^T \mathbf{v} = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.$$

É fácil verificar que o produto interno definido em (5.2) goza das seguintes propriedades:

i) *Simetria*: $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$.

ii) *Aditividade*: $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$.

iii) *Homogeneidade*: $\langle \mathbf{u}, k\mathbf{v} \rangle = k\langle \mathbf{u}, \mathbf{v} \rangle$, com $k \in \mathbb{R}$.

iv) *Positividade*: $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$, com $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ se e só se $\mathbf{u} = 0$.

Relembre-se que o comprimento do vector $\mathbf{u} = (u_1, u_2) \in \mathbb{R}^2$ é dado por $\|\mathbf{u}\|^2 = u_1^2 + u_2^2$, ou equivalentemente, $\|\mathbf{u}\|^2 = \langle \mathbf{u}, \mathbf{u} \rangle$. Assim, a propriedade iv) acima é essencial se se pretende definir comprimento à custa de um produto interno.

Coloca-se naturalmente a questão de saber se podemos aplicar a expressão (5.2) para definir um produto interno em $\mathbb{C}^n$, de modo que a propriedade iv) de positividade se mantenha. A resposta é negativa já que, por exemplo, para $(z_1, z_2) \in \mathbb{C}^2$, a expressão $\langle (z_1, z_2), (z_1, z_2) \rangle = z_1^2 + z_2^2$ não é em geral um número real não negativo. Vamos portanto adaptar essa definição de modo a que a propriedade iv) se verifique. Considerando, para $(z_1, z_2), (w_1, w_2) \in \mathbb{C}^2$, a expressão $\langle (z_1, z_2), (w_1, w_2) \rangle = \overline{z}_1 w_1 + \overline{z}_2 w_2$, onde $\overline{z}$ designa o conjugado de z, verificamos que $\langle (z_1, z_2), (z_1, z_2) \rangle = \overline{z}_1 z_1 + \overline{z}_2 z_2 = |z_1|^2 + |z_2|^2$ é um número não negativo.

Assim, para $\mathbf{z}, \mathbf{w} \in \mathbb{C}^n$ define-se um produto interno em $\mathbb{C}^n$ do seguinte modo:

$$\begin{aligned} \langle \mathbf{z}, \mathbf{w} \rangle &= \langle (z_1, z_2, \dots, z_n), (w_1, w_2, \dots, w_n) \rangle \\ &= \overline{z}_1 w_1 + \overline{z}_2 w_2 + \cdots + \overline{z}_n w_n \\ &= \begin{bmatrix} \overline{z}_1 & \overline{z}_2 & \cdots & \overline{z}_n \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} = \mathbf{z}^H \mathbf{w}, \end{aligned} \tag{5.3}$$

onde $\mathbf{z}^H = \overline{\mathbf{z}}^T$ é o *transconjugado* de $\mathbf{z}$, isto é, o vector linha cujas entradas são as conjugadas de $\mathbf{z}$.

Notação:

Para $\mathbf{z} = (a_1 + ib_1, a_2 + ib_2, \dots, a_n + ib_n) \in \mathbb{C}^n$, o *transconjugado* do vector $\mathbf{z}$ é o vector linha:

$$\mathbf{z}^{\mathrm{H}} = \overline{\mathbf{z}}^T = \begin{bmatrix} a_1 - ib_1 & a_2 - ib_2 & \cdots & a_n - ib_n \end{bmatrix}.$$

Resumindo:

Produto interno usual em $\mathbb{R}^n$ e $\mathbb{C}^n$

Define-se o produto interno usual, respectivamente, em $\mathbb{R}^n$ e em $\mathbb{C}^n$, como

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^T \mathbf{v} = \sum_{i=1}^{n} u_i v_i \in \mathbb{R} \quad \text{e} \quad \langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^H \mathbf{v} = \sum_{i=1}^{n} \overline{u}_i v_i \in \mathbb{C}, \tag{5.4}$$

onde $\mathbf{u} = (u_1, \dots, u_n)$ e $\mathbf{v} = (v_1, \dots, v_n)$. A *norma* usual de um vector $\mathbf{u}$ é definida como

$$\|\mathbf{u}\|^2 = \langle \mathbf{u}, \mathbf{u} \rangle,$$

onde $\langle \, , \, \rangle$ é o produto interno usual.

É frequente designar-se o produto interno usual (em $\mathbb{R}^n$ ou $\mathbb{C}^n$) por *produto interno euclidiano*.

Note-se que a norma usual de um vector $\mathbf{u} = (x_1 + iy_1, x_2 + iy_2, \dots, x_n + iy_n)$ de $\mathbb{C}^n$ coincide com a norma usual do vector $(x_1, y_1, x_2, y_2, \dots, x_n, y_n)$ de $\mathbb{R}^{2n}$.

Verifica-se facilmente que exceptuando a propriedade de simetria, a definição (5.3) de produto interno usual em $\mathbb{C}^n$ satisfaz todas as propriedades acima enunciadas para o produto interno usual em $\mathbb{R}^n$. Em vez da simetria temos $\langle \mathbf{u}, \mathbf{v} \rangle =$

$\overline{\langle \mathbf{v}, \mathbf{u} \rangle}$, uma vez que para $\mathbf{u}, \mathbf{v} \in \mathbb{C}^n$, se tem

$$\begin{aligned}\langle \mathbf{u}, \mathbf{v} \rangle &= \overline{u}_1 v_1 + \overline{u}_2 v_2 + \cdots + \overline{u}_n v_n \\ \langle \mathbf{v}, \mathbf{u} \rangle &= \overline{v}_1 u_1 + \overline{v}_2 u_2 + \cdots + \overline{v}_n u_n,\end{aligned}$$

onde $\mathbf{u} = (u_1, \ldots, u_n) \in \mathbb{C}^n$ e $\mathbf{v} = (v_1, \ldots, v_n) \in \mathbb{C}^n$. Refira-se ainda que o produto interno usual em $\mathbb{C}^n$ se restringe ao produto interno em $\mathbb{R}^n$, já que o conjugado de um vector de $\mathbb{R}^n$ é o próprio vector.

Com base nas propriedades do produto interno usual em $\mathbb{R}^n$ e $\mathbb{C}^n$, definimos seguidamente o que se entende por produto interno num qualquer espaço linear real ou complexo.

Definição 5.1. Seja W um espaço linear real (resp. complexo).

Um produto interno em W é uma função que aplica cada par de vectores $u, v \in W$ num escalar real (resp. complexo) $\langle u, v \rangle$, satisfazendo as propriedades:

P1. $\langle u, v \rangle = \langle v, u \rangle$ (resp. $\langle u, v \rangle = \overline{\langle v, u \rangle}$).

P2. $\langle u, v + w \rangle = \langle u, v \rangle + \langle u, w \rangle$.

P3. $\langle u, kv \rangle = k\langle u, v \rangle$ para $k \in \mathbb{R}$ (resp. para $k \in \mathbb{C}$).

P4. $\langle u, u \rangle \geq 0$, com $\langle u, u \rangle = 0$ se e só se $u = 0$.

Passamos a designar um espaço linear (real ou complexo) munido de um produto interno por *espaço euclidiano*[1].

Nota 27. *1. As propriedades P2 e P3 da definição anterior são equivalentes a dizer que, fixado o vector u, a função $f(v) = \langle u, v \rangle$ é linear (ver Definição 6.1, pág. 325). Dito de outra forma: o produto interno é uma função linear na segunda variável. Se o espaço linear é real, pela propriedade de simetria, resulta*

$$\begin{aligned}\langle v + w, u \rangle = \langle u, v + w \rangle &= \langle u, v \rangle + \langle u, w \rangle = \langle v, u \rangle + \langle w, u \rangle \\ \langle ku, v \rangle = \langle v, ku \rangle &= k\langle v, u \rangle = k\langle u, v \rangle,\end{aligned}$$

[1] Na literatura encontram-se várias designações para um espaço linear munido de um produto interno. São frequentes as designações de *espaço euclidiano* se o espaço linear é real, enquanto que espaços lineares complexos munidos de um produto interno são chamados de *espaços hermitianos*. A designação *espaço de Hilbert* é igualmente usada embora se requeira ainda que o espaço seja *completo* (qualquer sucessão de Cauchy de elementos do espaço converge para um elemento do espaço na norma induzida pelo produto interno).

o que significa que o produto interno é também linear na primeira variável. Assim, num espaço linear real o produto interno é uma função linear em ambas as variáveis. Neste caso dizemos que o produto interno é uma função bilinear ou uma forma bilinear.

Quando o espaço linear é complexo, o produto interno é apenas linear na segunda variável (ou seja, verifica P2 e P3), enquanto que na primeira variável se tem

$$\langle v+w,u\rangle \overset{P1}{=} \overline{\langle u,v+w\rangle} \overset{P2}{=} \overline{\langle u,v\rangle} + \overline{\langle u,w\rangle} \overset{P1}{=} \langle v,u\rangle + \langle w,u\rangle,$$
$$\langle ku,v\rangle \overset{P1}{=} \overline{\langle v,ku\rangle} \overset{P3}{=} \overline{k\langle v,u\rangle} = \overline{k}\langle u,v\rangle.$$

Por isso, num espaço linear complexo o produto interno diz-se uma forma sesquilinear.

2. *Noutros textos matemáticos é habitual definir um produto interno num espaço linear complexo como sendo uma função linear na primeira variável (contrariamente ao que definimos). Neste caso a definição de produto interno em $\mathbb{C}^n$ seria $\langle \mathbf{z},\mathbf{w}\rangle = \mathbf{z}^T\overline{\mathbf{w}}$ em vez da nossa definição $\langle \mathbf{z},\mathbf{w}\rangle = \overline{\mathbf{z}}^T\mathbf{w} = \mathbf{z}^H\mathbf{w}$. A definição por nós adoptada é mais conveniente em termos práticos pois facilita a tradução para $\mathbb{C}^n$ de resultados válidos em $\mathbb{R}^n$ (basta substituir o "T" de transposto pelo "H" de Hermite).*

3. *Outras notações comuns para o produto interno $\langle u,v\rangle$ são: $\langle u|v\rangle$, $u|v$ e $u\cdot v$.*

Exemplo 5.1. a) Em $\mathbb{R}^2$, a expressão

$$\langle (x_1,x_2),(y_1,y_2)\rangle = 2x_1y_1 + x_2y_2$$

define um produto interno, já que são satisfeitas todas as propriedades da definição de produto interno.

b) Em $\mathbb{R}^2$, a expressão

$$\langle (x_1,x_2),(y_1,y_2)\rangle = 2x_1y_1 - x_2y_2$$

não define um produto interno, visto que a propriedade P4 da definição não é verificada em geral. Por exemplo, $\langle (0,1),(0,1)\rangle = -1$.

c) Em $\mathbb{R}^2$, a expressão

$$\langle (x_1,x_2),(y_1,y_2)\rangle = x_1y_1 + x_2y_2 - x_1y_2$$

não define um produto interno, uma vez que a propriedade de simetria não se verifica:

$$\langle (y_1, y_2), (x_1, x_2) \rangle = y_1 x_1 + y_2 x_2 - y_1 x_2 \neq \langle (x_1, x_2), (y_1, y_2) \rangle.$$

♦

Exercício 5.1. Mostre que no espaço linear real das matrizes reais $n \times n$, a expressão seguinte define um produto interno:

$$\langle A, B \rangle = \mathrm{tr}(AB^T),$$

onde $\mathrm{tr}(A)$ designa o traço da matriz A (soma das entradas da diagonal principal). Sugestão: Para mostrar que $\langle A, B \rangle = \langle B, A \rangle$ use o facto do traço de uma matriz ser igual ao da sua transposta.

▲

Matriz de Gram

Comecemos por verificar que qualquer produto interno num espaço linear fica completamente determinado se conhecermos os produtos internos dos vectores de uma base do espaço. Como consequência deste resultado, qualquer produto interno é determinado por uma matriz chamada *matriz de Gram*[2].

Considere-se, por exemplo, um espaço linear real W de dimensão 2, munido de um produto interno, e $B = (v_1, v_2)$ uma base ordenada de W. As propriedades de linearidade de um produto interno permitem escrever um produto interno relativamente à base B usando uma matriz. De facto, se x e y são vectores de W tais que $x = x_1 v_1 + x_2 v_2$ e $y = y_1 v_1 + y_2 v_2$, das propriedades P2 e P3 da definição de produto interno resulta

$$\begin{aligned}
\langle x, y \rangle &= \langle x_1 v_1 + x_2 v_2 \,,\, y_1 v_1 + y_2 v_2 \rangle \\
&= x_1 y_1 \langle v_1, v_1 \rangle + x_1 y_2 \langle v_1, v_2 \rangle + x_2 y_1 \langle v_2, v_1 \rangle + x_2 y_2 \langle v_2, v_2 \rangle \\
&= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} \langle v_1, v_1 \rangle & \langle v_1, v_2 \rangle \\ \langle v_2, v_1 \rangle & \langle v_2, v_2 \rangle \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \mathbf{x}_B^T \mathbf{G}\, \mathbf{y}_B,
\end{aligned}$$

onde $\mathbf{x}_B$ e $\mathbf{y}_B$ designam os vectores das coordenadas, respectivamente, de $\mathbf{x}$ e de $\mathbf{y}$ na base B.

[2] Jørgen Pedersen Gram (1850 – 1916), actuário e matemático dinamarquês.

De modo inteiramente análogo, se W é um espaço linear real de dimensão n munido de um produto interno, e $B = (v_1, \ldots, v_n)$ uma base de W, então $\langle x, y\rangle = \mathbf{x}_B^T \mathbf{G}\, \mathbf{y}_B$ onde a matriz $\mathbf{G} = [g_{ij}]_{i,j=1,\ldots n}$ tem entradas $g_{ij} = \langle v_i, v_j\rangle$, e $\mathbf{x}_B$ é o vector das coordenadas de x na base B.

Das propriedades do produto interno podemos concluir:

- Da simetria, temos

$$\langle v_i, v_j\rangle = \langle v_j, v_i\rangle \Longleftrightarrow g_{ij} = g_{ji},$$

  e portanto a matriz $\mathbf{G} = [g_{ij}] = [\langle v_i, v_j\rangle]$ é simétrica ($\mathbf{G} = \mathbf{G}^T$).

- Da positividade ($\langle x, x\rangle > 0$ para todo $x \neq 0$) a matriz $\mathbf{G}$ terá de verificar

$$\mathbf{x}_B^T \mathbf{G}\, \mathbf{x}_B > 0, \quad \text{para todo } \mathbf{x}_B \neq 0. \tag{5.5}$$

  Uma matriz que satisfaz a desigualdade (5.5) diz-se *definida positiva*. Note-se que, sendo a desigualdade (5.5) válida para qualquer vector não nulo, ela é satisfeita em particular por qualquer vector próprio de $\mathbf{G}$. Por conseguinte, se $\mathbf{x}_B$ é um vector próprio de $\mathbf{G}$ associado ao valor próprio λ, tem-se

$$\mathbf{x}_B^T \mathbf{G}\, \mathbf{x}_B = \mathbf{x}_B^T \lambda \mathbf{x}_B = \lambda \|\mathbf{x}_B\|^2 > 0 \Longrightarrow \lambda > 0.$$

  Ou seja, todos os valores próprios de $\mathbf{G}$ são positivos. No Capítulo 7 estabelecem-se critérios que permitem identificar matrizes definidas positivas, em particular estabelece-se que é condição necessária e suficiente para uma matriz ser definida positiva que todos os seus valores próprios sejam positivos.

Exercício 5.2. Repita o procedimento utilizado no caso real para mostrar que num espaço linear complexo munido de um produto interno, a matriz que define o produto interno, relativamente a uma base B, é uma matriz $\mathbf{G}$ tal que:

1. $\mathbf{G} = \mathbf{G}^H$, onde $\mathbf{G}^H = \overline{\mathbf{G}}^T$ designa a matriz transposta da conjugada (também designada por *transconjugada*). Quando $\mathbf{G} = \mathbf{G}^H$ a matriz $\mathbf{G}$ diz-se *hermitiana* (ver Capítulo 7).

2. $\mathbf{x}_B^H \mathbf{G}\, \mathbf{x}_B > 0$, para todo $x \neq 0$ (ou seja, $\mathbf{G}$ é definida positiva).

▲

Nota 28. *A restrição de uma matriz hermitiana aos reais é uma matriz simétrica, uma vez que se* $\mathbf{G}$ *é uma matriz real a igualdade* $\mathbf{G} = \mathbf{G}^H = \overline{\mathbf{G}}^T$ *reduz-se à igualdade* $\mathbf{G} = \mathbf{G}^T$*.*

Nota 29. *Outras notações frequentes para a trasconjugada* A^H *de uma matriz* A *são:* $A^\dagger$ *e* A^**.*

Matriz de Gram

Seja W um espaço linear real (resp. complexo) munido de um produto interno e $B = (v_1, v_2, \ldots, v_n)$ uma base ordenada de W. A matriz $\mathbf{G} = [\langle v_i, v_j\rangle]_{i,j=1,\ldots,n}$, dos produtos internos dos vectores da base B, é designada por *matriz de Gram* ou *matriz da métrica*, relativa à base B. A matriz $\mathbf{G}$ verifica:

1. $\mathbf{G}$ é simétrica (resp. hermitiana);
2. $\mathbf{G}$ é definida positiva, isto é, $\mathbf{x}_B^T \mathbf{G}\, \mathbf{x}_B > 0$ para todo $x \neq 0$ (resp. $\mathbf{x}_B^H \mathbf{G}\, \mathbf{x}_B > 0$ para todo $x \neq 0$) .

Em relação à base B, o produto interno em W escreve-se na forma

$$\langle x, y\rangle = \mathbf{x}_B^T \mathbf{G}\, \mathbf{y}_B \qquad \left(\text{resp. } \langle x, y\rangle = \mathbf{x}_B^H \mathbf{G}\, \mathbf{y}_B\right),$$

onde $\mathbf{x}_B$ e $\mathbf{y}_B$ são, respectivamente, os vectores das coordenadas de x e de y na base B.

Podemos assim concluir:

Num espaço linear de dimensão n, um produto interno fica completamente determinado se forem conhecidos os produtos internos entre os vectores de uma base do espaço, isto é, se for conhecida a respectiva matriz de Gram $\mathbf{G}$.

Exemplo 5.2.

1. A matriz de Gram relativa à base canónica de $\mathbb{R}^n$ para o produto interno usual em $\mathbb{R}^n$ é a matriz identidade.
2. Se a expressão do Exemplo 5.1-b) definisse um produto interno, então na base canónica de $\mathbb{R}^2$ a matriz de Gram respectiva, $G = [g_{ij}]$, teria de ser simétrica e definida positiva. Ora, calculando $\langle \mathbf{e}_i, \mathbf{e}_j\rangle = g_{ij}$ obtemos a matriz:

$$\begin{bmatrix} 2 & 0 \\ 0 & -1 \end{bmatrix}.$$

Esta matriz, apesar de ser simétrica, não é definida positiva pois tem um valor próprio negativo. Logo, a expressão do exemplo referido não define um produto interno.

3. Calculando a expressão do Exemplo 5.1-c) nos vectores da base canónica de $\mathbb{R}^2$, obtém-se a matriz

$$\mathbf{G} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}.$$

A matriz $\mathbf{G}$ não é simétrica, pelo que a referida expressão não define um produto interno.

♦

Exercício 5.3. Sejam B e B' duas bases ordenadas de um espaço linear W munido de um produto interno e M a matriz de mudança da base B' para a base B. Mostre que para quaisquer $x, y \in W$, se tem

$$\langle x, y\rangle = \mathbf{x}_B^H \mathbf{G} \mathbf{y}_B = \mathbf{x}_{B'}^H \mathbf{S} \mathbf{y}_{B'},$$

com $\mathbf{S} = M^H \mathbf{G} M$. ▲

Norma, distância e ângulo.

A propriedade P4, enunciada na definição de produto interno, permite-nos definir uma norma e uma distância entre vectores de um qualquer espaço linear, desde que este espaço esteja munido de um produto interno.

Definição 5.2. Num espaço linear W munido do produto interno $\langle\, , \rangle$ definimos *norma* (ou *comprimento*) de um vector $u \in W$ como sendo

$$\|u\| = \sqrt{\langle u, u\rangle}.$$

A *distância* entre os vectores $u, v \in W$ é definida por

$$\operatorname{dist}(u, v) = \|u - v\|.$$

Em $\mathbb{R}^2$, munido do produto interno usual, tem-se

$$\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2} = \sqrt{\langle \mathbf{x}, \mathbf{x}\rangle}, \quad \text{com } \mathbf{x} = (x_1, x_2).$$

A norma generaliza pois o conceito de comprimento a um espaço euclidiano.

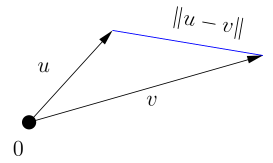


Figura 5.2: Distância entre o vector u e v.

Exercício 5.4. Verifique que num espaço linear W, real ou complexo, se tem

$$\|ku\| = |k|\,\|u\|,$$

para qualquer escalar k e qualquer $u \in W$. ▲

Exemplo 5.3. Considere os vectores $\mathbf{x} = (1, 0, -2, 1)$ e $\mathbf{y} = (i, 1, -2i, 0)$, de $\mathbb{R}^4$ e $\mathbb{C}^4$ respectivamente. Para o produto interno usual nestes espaços lineares, a norma destes vectores é

$$\|\mathbf{x}\| = \sqrt{1 + 4 + 1} = \sqrt{6},$$

e

$$\|\mathbf{y}\|^2 = \mathbf{y}^H\mathbf{y} = \begin{bmatrix} -i & 1 & 2i & 0 \end{bmatrix} \begin{bmatrix} i \\ 1 \\ -2i \\ 0 \end{bmatrix} = [-i^2 + 1 - 4i^2] = 6,$$

e portanto $\|\mathbf{y}\| = \sqrt{\mathbf{y}^H\mathbf{y}} = \sqrt{6}$. ♦

Dado um vector não nulo u, podemos sempre obter um vector v proporcional a u que tenha norma unitária. Para tal, basta dividir o vector u pela sua norma:

$$u \rightsquigarrow v = \frac{u}{\|u\|}.$$

Um vector com norma igual a 1 recebe a designação de *versor*.

Exemplo 5.4. Considere-se $\mathbb{R}^3$ munido do produto interno usual e o subespaço $V = \operatorname{Span}\{(2, 5, 0), (1, 0, 2)\}$. Os vectores $\mathbf{u} = (2, 5, 0)$ e $\mathbf{v} = (1, 0, 2)$ formam um base de V. Determinemos uma base de V constituída por versores

$$\mathbf{u} = (2,5,0) \Rightarrow \|\mathbf{u}\| = \sqrt{29} \Rightarrow \tilde{\mathbf{u}} = \frac{1}{\sqrt{29}}(2,5,0),$$
$$\mathbf{v} = (1,0,2) \Rightarrow \|\mathbf{v}\| = \sqrt{5} \Rightarrow \tilde{\mathbf{v}} = \frac{1}{\sqrt{5}}(1,0,2).$$

Assim, $\{\tilde{\mathbf{u}}, \tilde{\mathbf{v}}\}$ é uma base V constituída exclusivamente por vectores de norma unitária. ♦

Exercício 5.5. Mostre as igualdades seguintes.

a) Num espaço linear real:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \frac{1}{2}\left(\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - \|\mathbf{x}-\mathbf{y}\|^2\right). \tag{5.6}$$

b) Num espaço linear complexo:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \frac{1}{4}\left(\|\mathbf{x}+\mathbf{y}\|^2 - \|\mathbf{x}-\mathbf{y}\|^2 - i\left(\|\mathbf{x}+i\mathbf{y}\|^2 - \|\mathbf{x}-i\mathbf{y}\|^2\right)\right). \tag{5.7}$$

▲

Todas as propriedades que a seguir deduzimos são válidas para qualquer espaço linear munido de um produto interno (quer esse espaço seja real ou complexo). Contudo, algumas das provas só serão realizadas para o caso real sendo o caso complexo deixado como exercício.

Desigualdade de Cauchy-Schwarz

Num espaço linear W munido de um produto interno $\langle\,,\,\rangle$, é válida a desigualdade

$$|\langle u, v \rangle| \leq \|u\|\|v\|, \quad \text{para todo } u, v \in W. \tag{5.8}$$

Ocorre a igualdade na expressão anterior se e só se os vectores u e v são linearmente dependentes.

A desigualdade (5.8) é conhecida por *desigualdade de Cauchy-Schwarz*.

Demonstração. Se um dos vectores é o vector nulo a desigualdade é trivialmente satisfeita. Suponhamos que $v \neq 0$ e considere-se o escalar $t = \frac{\langle v,u \rangle}{\|v\|^2}$. Calculando $\|tv - u\|^2$, tem-se

$$
\begin{aligned}
0 \leq \langle tv-u, tv-u\rangle &\overset{P2,P3}{=} t\langle tv-u, v\rangle - \langle tv-u, u\rangle \\
&\overset{P1}{=} t\,\overline{\langle v, tv-u\rangle} - \overline{\langle u, tv-u\rangle} \\
&\overset{P2,P3}{=} t\,\bar{t}\,\langle v, v\rangle - t\,\overline{\langle v, u\rangle} - \bar{t}\,\overline{\langle u, v\rangle} + \langle u, u\rangle, \qquad (5.9)
\end{aligned}
$$

onde na última igualdade aplicámos o facto do conjugado de um número real ser o próprio número (por exemplo, $\overline{\langle u, u\rangle} = \langle u, u\rangle$).

Como o produto de um número complexo pelo seu conjugado é igual ao quadrado do módulo do complexo, tem-se

$$
\begin{aligned}
t\,\bar{t} &= |t|^2 = \frac{|\langle v, u\rangle|^2}{\|v\|^4} = \frac{|\langle u, v\rangle|^2}{\|v\|^4} \\
t\,\overline{\langle v, u\rangle} &\overset{P1}{=} \frac{\langle v, u\rangle\overline{\langle v, u\rangle}}{\|v\|^2} = \frac{|\langle v, u\rangle|^2}{\|v\|^2} = \frac{|\langle u, v\rangle|^2}{\|v\|^2} \\
\bar{t}\,\overline{\langle u, v\rangle} &= \frac{\overline{\langle v, u\rangle}\,\overline{\langle u, v\rangle}}{\|v\|^2} \overset{P1}{=} \frac{\langle u, v\rangle\overline{\langle u, v\rangle}}{\|v\|^2} = \frac{|\langle u, v\rangle|^2}{\|v\|^2},
\end{aligned}
$$

onde se levou em conta que o módulo de um complexo é igual ao módulo do seu conjugado.

Substituindo as expressões anteriores em (5.9), vem

$$
\begin{aligned}
0 \leq \|tv-u\|^2 &= \frac{|\langle u, v\rangle|^2}{\|v\|^2} - 2\frac{|\langle u, v\rangle|^2}{\|v\|^2} + \|u\|^2 \\
&= \frac{\|u\|^2\|v\|^2 - |\langle u, v\rangle|^2}{\|v\|^2} \iff \|u\|^2\|v\|^2 - |\langle u, v\rangle|^2 \geq 0. \qquad (5.10)
\end{aligned}
$$

Usando o facto de $\langle u, u\rangle = \|u\|^2$ e $\langle v, v\rangle = \|v\|^2$ serem quantidades não negativas, da desigualdade anterior obtém-se a desigualdade de Cauchy-Schwarz.

Verifica-se a igualdade em (5.10) se e só se $\|tv-u\|^2 = 0$, ou seja, se e só se $tv = u$. Como t é um escalar, a igualdade $tv = u$ significa que u e v são linearmente dependentes. □

Se na desigualdade de Cauchy-Schwarz[3] a quantidade $\langle u, v\rangle$ for real (isto acontece necessariamente se o espaço linear é real), para vectores não nulos u

[3]Augustin-Louis Cauchy (1789 – 23 May 1857), matemático francês, e um dos pioneiros da análise matemática; Karl Hermann Amandus Schwarz (1843 – 1921), matemático alemão com inúmeras contribuições em análise complexa.

e v, resulta

$$\frac{|\langle u, v\rangle|}{\|u\|\|v\|} \leq 1 \iff -1 \leq \frac{\langle u, v\rangle}{\|u\|\|v\|} \leq 1.$$

Se θ é um ângulo cuja medida em radianos varia de 0 a π, então $\cos\theta$ toma valores entre -1 e 1 (uma única vez). Podemos assim definir ângulo entre dois vectores da forma que se segue.

Definição 5.3. Sendo u e v vectores de um espaço linear real munido de um produto interno $\langle\, ,\, \rangle$, definimos *ângulo* entre u e v como sendo o único ângulo $0 \leq \theta \leq \pi$ que satisfaz a igualdade

$$\cos\theta = \frac{\langle u, v\rangle}{\|u\|\|v\|}.$$

Exemplo 5.5. Considere em $\mathbb{R}^4$ o produto interno usual. O ângulo entre os vectores $\mathbf{u} = (\sqrt{2}, 0, 0, 1)$ e $\mathbf{v} = (0, 1, 0, \sqrt{3})$ é dado por:

$$\|\mathbf{u}\| = \sqrt{2+1} = \sqrt{3}, \quad \|\mathbf{v}\| = \sqrt{1+3} = 2, \quad \text{e} \quad \langle \mathbf{u}, \mathbf{v}\rangle = \sqrt{3}.$$

Logo,

$$\cos\theta = \frac{\sqrt{3}}{2\sqrt{3}} = \frac{1}{2} \Longrightarrow \theta = \frac{\pi}{3}.$$

♦

Uma das consequências importantes da desigualdade de Cauchy-Schwarz é a chamada desigualdade triangular da norma.

Desigualdade Triangular

Se u e v são vectores de um espaço euclidiano, então

$$\|u+v\| \leq \|u\| + \|v\|.$$

Demonstração. Calcule-se $\|u+v\|^2$ usando as propriedades da definição de produto interno num espaço linear real (o caso complexo é deixado como exercício).

$$\begin{aligned}
\|u+v\|^2 &= \langle u+v, u+v\rangle \overset{P2}{=} \langle u+v, u\rangle + \langle u+v, v\rangle \\
&\overset{P1}{=} \langle u, u+v\rangle + \langle v, u+v\rangle \\
&\overset{P1,P2}{=} \langle u, u\rangle + 2\langle v, u\rangle + \langle v, v\rangle \\
&= \|u\|^2 + \|v\|^2 + 2\langle u, v\rangle \\
&\leq \|u\|^2 + \|v\|^2 + 2\,|\langle u, v\rangle| && \text{(Propriedade do módulo)} \\
&\leq \|u\|^2 + \|v\|^2 + 2\|u\|\|v\|. && \text{(Des. Cauchy-Schwarz)}
\end{aligned}$$

Ou seja,

$$\|u+v\|^2 \leq (\|u\| + \|v\|)^2\,.$$

Como as quantidades na desigualdade anterior são todas não negativas, extraindo a raiz quadrada resulta a desigualdade pretendida. □

Exercício 5.6. Faça a demonstração da desigualdade triangular para o caso do espaço linear ser complexo. Para tal necessita de usar alguns resultados sobre números complexos enunciados no Exercício A.1 (pág. 510). ▲

Vectores ortogonais

Em $\mathbb{R}^2$ dois vectores são ortogonais (ou perpendiculares) se o ângulo entre eles é recto. Ou seja, se o produto interno usual $\langle u, v\rangle = \|u\|\|v\|\cos\theta$ é igual a zero. É portanto natural generalizar esta noção a um espaço linear qualquer, do seguinte modo:

Definição 5.4. Dois vectores u e v de um espaço euclidiano dizem-se *ortogonais* ou *perpendiculares*), se $\langle u, v\rangle = 0$.

Quando u e v são vectores ortogonais escrevemos $u \perp v$.

Sendo u e v dois vectores ortogonais, tem-se $\|u-v\| = \|u+v\|$ (ver Figura 5.3). De facto, se u e v são dois vectores ortogonais, então

$$\begin{aligned}
\|u-v\|^2 &= \langle u-v, u-v\rangle = \langle u, u\rangle - \langle u, v\rangle - \overline{\langle u, v\rangle} + \langle v, v\rangle \\
&= \|u\|^2 + \|v\|^2,
\end{aligned}$$

e

$$\begin{aligned}
\|u+v\|^2 &= \langle u+v, u+v\rangle = \langle u, u\rangle + \langle u, v\rangle + \overline{\langle u, v\rangle} + \langle v, v\rangle \\
&= \|u\|^2 + \|v\|^2.
\end{aligned}$$

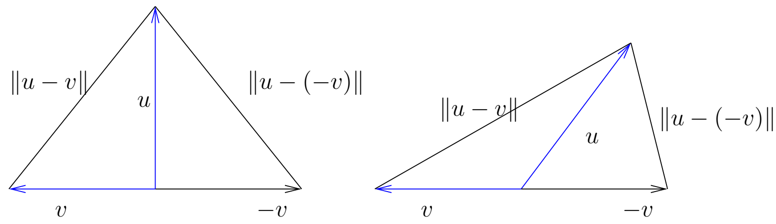


Figura 5.3: Vectores u e v respectivamente ortogonais e não ortogonais.

Acabámos de mostrar que o Teorema de Pitágoras é válido em qualquer espaço linear munido de um produto interno.

Teorema 5.1. Teorema de Pitágoras
Se u e v são vectores ortogonais de um espaço euclidiano, então

$$\|u-v\|^2 = \|u\|^2 + \|v\|^2.$$

Além disso, também se verifica $\|u-v\|^2 = \|u+v\|^2$.

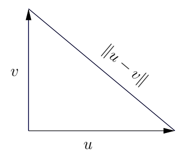


Figura 5.4: Teorema de Pitágoras num espaço euclidiano.

Exercício 5.7. a) Mostre que num espaço linear real a relação de ortogonalidade $u \perp v$ é equivalente a $\|u+v\|^2 = \|u\|^2 + \|v\|^2$.
b) Mostre, através de um exemplo, que num espaço linear complexo é falsa uma das implicações da equivalência da alínea anterior.

▲

Exercício 5.8. Mostre que se u é ortogonal aos vectores v e w, então o vector u é ortogonal a qualquer combinação linear de v e w. Isto é,

$$u \perp v \text{ e } u \perp w \Longrightarrow u \perp (\alpha v + \beta w), \text{ com } \alpha, \beta \text{ escalares.}$$

▲

Projecção ortogonal de um vector sobre outro

Em $\mathbb{R}^2$, a projecção ortogonal do vector $\mathbf{u}$ sobre o vector $\mathbf{v}$ é um vector $\mathbf{p}$ com a direcção de $\mathbf{v}$ e tal que $\mathbf{u}-\mathbf{p}$ é ortogonal a $\mathbf{v}$. Assim, o vector projecção ortogonal de $\mathbf{u}$ sobre $\mathbf{v}$ é um múltiplo de $\mathbf{v}$, isto é, $\mathbf{p} = k\mathbf{v}$. Observe a Figura 5.5.

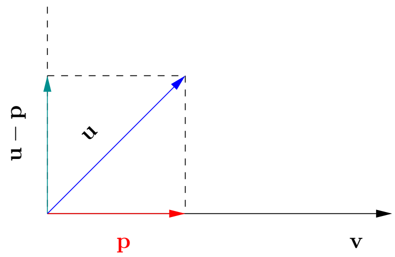


Figura 5.5: O vector $\mathbf{p}$ é a projecção ortogonal de $\mathbf{u}$ sobre $\mathbf{v}$. O vector $\mathbf{u}-\mathbf{p}$ é ortogonal a $\mathbf{v}$.

Num espaço linear real ou complexo, definimos projecção ortogonal de um vector u sobre o vector não nulo v como sendo um vector $p = kv$ (k escalar do espaço), tal que $(u-p) \perp v$. Usando a relação de ortogonalidade $(u-p) \perp v$, podemos determinar o escalar k, e consequentemente o vector p. Ou seja, para $v \neq 0$,

$$\begin{aligned} 0 = \langle v, u-p \rangle = \langle v, u-kv \rangle &= \langle v, u \rangle - k\langle v, v \rangle \\ &= \langle v, u \rangle - k\|v\|^2 \iff k = \frac{\langle v, u \rangle}{\|v\|^2}. \end{aligned}$$

Logo, $p = kv = \dfrac{\langle v, u \rangle}{\|v\|^2} v$. Passamos a designar o vector p por $\operatorname{proj}_v u$.

Definição 5.5. Num espaço linear W munido de um produto interno $\langle\, , \,\rangle$ a *projecção ortogonal* do vector $u \in W$ sobre o vector não nulo $v \in W$, é definida por

$$\operatorname{proj}_v u = \frac{\langle v, u \rangle}{\|v\|^2} v. \tag{5.11}$$

Nota 30. *Num espaço linear complexo não é indiferente usar* $\langle v, u \rangle$ *ou* $\langle u, v \rangle$ *na definição* (5.11)*, de projecção ortogonal. Se tivéssemos definido (como acontece* *em alguns textos), o produto interno como sendo linear na primeira variável, na Definição 5.5 apareceria $\langle u, v\rangle$ em vez de $\langle v, u\rangle$.*

Exemplo 5.6. Consideremos $\mathbb{C}^2$ munido do produto interno usual, e determinemos a projecção ortogonal do vector $\mathbf{x} = (2i, 1)$ sobre $\mathbf{y} = (i, 0)$. Como

$$\langle (i,0)\,,\,(2i,1)\rangle = -2i^2 = 2 \quad \text{e} \quad \|(i,0)\|^2 = -i^2 = 1,$$

tem-se

$$\operatorname{proj}_{\mathbf{y}} \mathbf{x} = \frac{\langle (i,0)\,,\,(2i,1)\rangle}{\|(i,0)\|^2}(i,0) = 2(i,0).$$

♦

Exemplo 5.7. Usemos a expressão (5.11) para confirmar que $(u - \operatorname{proj}_v u\rangle) \perp v$.

$$\begin{aligned}
\langle v\,,\,u - \operatorname{proj}_v u\rangle &= \langle v\,,\,u - \frac{\langle v, u\rangle}{\|v\|^2} v\rangle \\
&= \langle v, u\rangle - \underbrace{\frac{\langle v, u\rangle}{\|v\|^2}}_{\text{escalar}} \|v\|^2 \\
&= \langle v, u\rangle - \langle v, u\rangle = 0.
\end{aligned}$$

♦

Exercício 5.9. Considere o subespaço $S = \operatorname{Span}\{v_1, v_2, \ldots, v_n\}$ de um espaço linear W tal que $\langle v_i, v_j\rangle = 0$ para todo $i \neq j$ e $i, j = 1, \ldots, n$. Seja $x \in W$.

a) Justifique que o vector $w = \operatorname{proj}_{v_1} x + \operatorname{proj}_{v_2} x + \cdots + \operatorname{proj}_{v_n} x$ pertence a S.

b) Sendo w o vector da alínea anterior, mostre que o vector $x - w$ é ortogonal a w.

▲

Consideremos $\mathbb{R}^n$ e $\mathbb{C}^n$ munidos do produto interno usual, e sejam $\mathbf{x}$ e $\mathbf{y}$ vectores destes espaços. Calculemos $\operatorname{proj}_{\mathbf{y}} \mathbf{x}$ usando a expressão matricial do produto interno. Nos cálculos que apresentamos a seguir identificamos matrizes 1×1 com escalares.

a) Para $\mathbf{x}$ e $\mathbf{y} \neq \mathbf{0}$ pertencentes a $\mathbb{R}^n$, tem-se

$$\operatorname{proj}_{\mathbf{y}} \mathbf{x} = \frac{\langle \mathbf{y}, \mathbf{x} \rangle}{\|\mathbf{y}\|^2} \mathbf{y} = \underbrace{\frac{\mathbf{y}^T \mathbf{x}}{\|\mathbf{y}\|^2}}_{\text{escalar}} \mathbf{y} = \mathbf{y} \frac{\mathbf{y}^T \mathbf{x}}{\|\mathbf{y}\|^2} = \frac{1}{\|\mathbf{y}\|^2} \underbrace{(\mathbf{y}\mathbf{y}^T)}_{n \times n} \mathbf{x},$$

onde na última igualdade se aplicou a associatividade do produto de matrizes. Assim, a projecção ortogonal de um vector sobre outro pode obter-se por multiplicação por uma matriz. Isto é,

$$\operatorname{proj}_{\mathbf{y}} \mathbf{x} = P_{\mathbf{y}} \mathbf{x}, \text{ onde } P_{\mathbf{y}} \text{ é a matriz, de ordem } n, \; P_{\mathbf{y}} = \frac{1}{\|\mathbf{y}\|^2} \mathbf{y}\mathbf{y}^T.$$

b) Para $\mathbf{x}$ e $\mathbf{y} \neq \mathbf{0}$ pertencentes a $\mathbb{C}^n$, tem-se

$$\operatorname{proj}_{\mathbf{y}} \mathbf{x} = \frac{\langle \mathbf{y}, \mathbf{x} \rangle}{\|\mathbf{y}\|^2} \mathbf{y} = \underbrace{\frac{\mathbf{y}^H \mathbf{x}}{\|\mathbf{y}\|^2}}_{\text{escalar}} \mathbf{y} = \mathbf{y} \frac{\mathbf{y}^H \mathbf{x}}{\|\mathbf{y}\|^2} = \frac{1}{\|\mathbf{y}\|^2} \underbrace{(\mathbf{y}\mathbf{y}^H)}_{n \times n} \mathbf{x}.$$

Portanto,

$$\operatorname{proj}_{\mathbf{y}} \mathbf{x} = P_{\mathbf{y}} \mathbf{x}, \text{ onde } P_{\mathbf{y}} \text{ é a matriz de ordem } n, \; P_{\mathbf{y}} = \frac{1}{\|\mathbf{y}\|^2} \mathbf{y}\mathbf{y}^H.$$

Do Exemplo 5.7 sabemos que $(x - \operatorname{proj}_y x)$ é ortogonal a y. Logo, para o produto interno usual em $\mathbb{R}^n$ (ou $\mathbb{C}^n$), tem-se que

$$\mathbf{x} - \operatorname{proj}_{\mathbf{y}} \mathbf{x} = \mathbf{x} - P_{\mathbf{y}} \mathbf{x} = (I - P_{\mathbf{y}}) \mathbf{x},$$

é um vector ortogonal a $\mathbf{y}$.

A discussão anterior pode resumir-se da seguinte forma:

Se $\mathbb{R}^n$ (resp. $\mathbb{C}^n$) está munido do produto interno usual, a projecção ortogonal do vector $\mathbf{x}$ sobre $\mathbf{y} \neq \mathbf{0}$ é dada por

$$\operatorname{proj}_{\mathbf{y}} \mathbf{x} = P_{\mathbf{y}} \mathbf{x},$$

onde $P_{\mathbf{y}}$ é a seguinte matriz $n \times n$:

$$P_{\mathbf{y}} = \frac{1}{\|\mathbf{y}\|^2} \mathbf{y}\mathbf{y}^T \qquad \left(\text{resp. } P_{\mathbf{y}} = \frac{1}{\|\mathbf{y}\|^2} \mathbf{y}\mathbf{y}^H\right). \tag{5.12}$$

O vector $(I - P_{\mathbf{y}})\mathbf{x}$ é ortogonal a $\mathbf{y}$.

A matriz $P_{\mathbf{y}}$ é designada por *matriz de projecção ortogonal* (sobre $\mathbf{y}$).

5.2 Bases ortogonais

As bases ortogonais permitem uma simplificação computacional considerável. Como veremos adiante, dada uma base ortogonal de um espaço linear, qualquer vector do espaço é igual à soma das projecções ortogonais do vector sobre os vectores da base. Trata-se de uma generalização de um facto bem conhecido em $\mathbb{R}^2$ e $\mathbb{R}^3$ onde as coordenadas de um vector relativas à base canónica (que é uma base ortonormada) se obtêm das projecções ortogonais do vector sobre vectores com direcções dos eixos coordenados.

Definição 5.6. Conjuntos ortogonais e ortonormados

Um subconjunto $U = \{u_1, u_2, \ldots, u_k\}$ de um espaço euclidiano diz-se um *conjunto ortogonal*, se todos os vectores de U são ortogonais dois a dois. Isto é,

$$\langle u_i, u_j \rangle = 0, \quad \text{para todo } \; i \neq j \text{ e } i, j = 1, \ldots, k.$$

O conjunto U diz-se *ortonormado* se é ortogonal e todos os seus vectores têm norma igual a 1. Ou seja,

$$\langle u_i, u_j \rangle = 0 \quad \text{e} \quad \|u_i\| = 1, \quad \text{para todo } \; i \neq j \text{ e } i, j = 1, \ldots, k.$$

Um conjunto ortogonal (resp. ortonormado) que seja uma base diz-se uma *base ortogonal* (resp. *base ortonormada*).

Exemplo 5.8. Considere-se $\mathbb{R}^2$ munido do produto interno usual.

a) A base $\{(1, 2), (-2, 1)\}$ de $\mathbb{R}^2$ é uma base ortogonal mas não é ortonormada.
b) A base $\{\frac{1}{\sqrt{5}}(1, 2), \frac{1}{\sqrt{5}}(-2, 1)\}$ é uma base ortonormada de $\mathbb{R}^2$.

♦

Como vimos, a partir de um qualquer vector não nulo podemos obter um vector de norma unitária. Assim, dado um conjunto ortogonal que não contenha o vector nulo podemos sempre obter um conjunto ortonormado, bastando para tal multiplicar cada vector do conjunto pelo inverso da respectiva norma. Passamos a designar este procedimento por *normalização*.

$$\begin{array}{ccc} U \text{ ortogonal} & & \tilde{U} \text{ ortonormado} \\ & \xrightarrow{\text{normalização}} & \\ U = \{u_1, \ldots, u_k\} & & \tilde{U} = \left\{ \frac{u_1}{\|u_1\|}, \ldots, \frac{u_k}{\|u_k\|} \right\}. \end{array} \tag{5.13}$$

É óbvio que os conjuntos U e $\tilde{U}$ geram o mesmo espaço linear.

A proposição que se segue garante-nos a independência linear de um conjunto ortogonal.

Proposição 5.1. Qualquer conjunto ortogonal, que não contenha o vector nulo, é linearmente independente.

Demonstração. Seja $U = \{u_1, \ldots, u_k\}$ um conjunto ortogonal (isto é, tal que $\langle u_i, u_j \rangle = 0$ para todo $i, j = 1, \ldots, k$ e $i \neq j$) que não contém o vector zero.

Verifiquemos que a única solução da equação $\alpha_1 u_1 + \cdots + \alpha_k u_k = 0$ é $\alpha_1 = \cdots = \alpha_k = 0$. Seja u_i um qualquer vector de U. São válidas as igualdades

$$\begin{aligned} 0 = \langle u_i, \alpha_1 u_1 + \cdots + \alpha_k u_k \rangle &= \alpha_1 \langle u_i, u_1 \rangle + \cdots + \alpha_i \langle u_i, u_i \rangle + \cdots + \alpha_k \langle u_i, u_k \rangle \\ &= \alpha_i \langle u_i, u_i \rangle \quad \text{(os vectores } u_i \text{ são ortogonais dois a dois)} \\ &= \alpha_i \|u_i\|^2 \iff \alpha_i = 0 \text{ (já que } u_i \neq 0). \end{aligned}$$

Logo, como i é qualquer, obtém-se $\alpha_1 = \cdots = \alpha_k = 0$. $\square$

Consideremos $\mathbb{R}^n$ munido do produto interno usual e o subconjunto $U = \{\mathbf{u}_1, \ldots, \mathbf{u}_k\} \subset \mathbb{R}^n$. Seja A a matriz que cujas colunas são os vectores $\mathbf{u}_1, \ldots, \mathbf{u}_k$. Isto é,

$$A = \begin{bmatrix} | & | & & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \\ | & | & & | \end{bmatrix}.$$

As linhas de A^T são os vectores $\mathbf{u}_1^T, \mathbf{u}_2^T, \ldots, \mathbf{u}_k^T$. O produto $A^T A$ é a matriz (simétrica) de ordem k, dada por

$$\begin{aligned} A^T A &= \begin{bmatrix} — & \mathbf{u}_1^T & — \\ — & \mathbf{u}_2^T & — \\ & \vdots & \\ — & \mathbf{u}_k^T & — \end{bmatrix} \begin{bmatrix} | & | & & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \\ | & | & & | \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{u}_1^T\mathbf{u}_1 & \mathbf{u}_1^T\mathbf{u}_2 & \cdots & \mathbf{u}_1^T\mathbf{u}_k \\ \mathbf{u}_2^T\mathbf{u}_1 & \mathbf{u}_2^T\mathbf{u}_2 & \cdots & \mathbf{u}_2^T\mathbf{u}_k \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{u}_k^T\mathbf{u}_1 & \mathbf{u}_k^T\mathbf{u}_2 & \cdots & \mathbf{u}_k^T\mathbf{u}_k \end{bmatrix} = \begin{bmatrix} \|\mathbf{u}_1\|^2 & \langle \mathbf{u}_1, \mathbf{u}_2 \rangle & \cdots & \langle \mathbf{u}_1, \mathbf{u}_k \rangle \\ \langle \mathbf{u}_1, \mathbf{u}_2 \rangle & \|\mathbf{u}_2\|^2 & \cdots & \langle \mathbf{u}_2, \mathbf{u}_k \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{u}_1, \mathbf{u}_k \rangle & \langle \mathbf{u}_2, \mathbf{u}_k \rangle & \cdots & \|\mathbf{u}_k\|^2 \end{bmatrix}. \end{aligned}$$

Podemos concluir:

Proposição 5.2. Seja A uma matriz real.

1. As colunas de A formam um conjunto ortogonal (em relação ao produto interno usual em $\mathbb{R}^n$) se e só se a matriz A^TA é diagonal.

2. As colunas de A formam um conjunto ortonormado se e só se a matriz A^TA é igual à matriz identidade (isto é, $A^TA = I$).

Exercício 5.10. Considere o produto interno usual em $\mathbb{C}^n$, e A uma matriz complexa. Mostre que:

- As colunas de A formam um conjunto ortogonal se e só se a matriz A^HA é diagonal.

- As colunas de A formam um conjunto ortonormado se e só se a matriz A^HA é igual à matriz identidade (isto é, $A^HA = I$)[4].

▲

O Teorema 5.2 a seguir estabelece que as coordenadas de um vector relativas a uma base ortogonal podem exprimir-se em termos das projecções ortogonais do vector sobre os vectores da base.

Teorema 5.2. Seja $B = (u_1, u_2, \ldots, u_n)$ uma base (ordenada) ortogonal de um espaço linear W.

O vector das coordenadas de $x \in W$ na base B, $\mathbf{x}_B = (c_1, c_2, \ldots, c_n)$, tem componentes

$$c_i = \frac{\langle u_i, x\rangle}{\|u_i\|^2}.$$

Em particular,

$$x = \operatorname{proj}_{u_1} x + \cdots + \operatorname{proj}_{u_n} x.$$

Se B é uma base ortonormada, então $c_i = \langle u_i, x\rangle$.

Demonstração. Sendo B uma base de W, o vector x escreve-se de forma única como combinação linear dos vectores da base. Seja $x = c_1u_1 + \cdots + c_nu_n$, onde os c_i são as coordenadas de x na base B. Logo,

$$\begin{aligned}\langle u_i, x\rangle &= \langle u_i, c_1u_1 + \cdots + c_nu_n\rangle = c_1\langle u_i, u_1\rangle + \cdots + c_n\langle u_i, u_n\rangle \\ &= c_i\langle u_i, u_i\rangle \qquad \text{(a base é ortogonal e portanto } \langle u_i, u_j\rangle = 0 \text{ para } i \neq j).\end{aligned}$$

[4]Uma matriz real A que verifica $A^TA = I$ diz-se uma matriz *ortogonal*, enquanto que se A for complexa e verifica $A^HA = I$, diz-se que A é uma matriz unitária. No Capítulo 7 estudam-se propriedades destas matrizes.

Os vectores u_i são não nulos uma vez que são vectores de uma base, por conseguinte

$$c_i = \frac{\langle u_i, x\rangle}{\langle u_i, u_i\rangle} = \frac{\langle u_i, x\rangle}{\|u_i\|^2}.$$

Da definição de projecção ortogonal de um vector sobre outro (ver (5.11)) tem-se $c_i u_i = \operatorname{proj}_{u_i} x$, e portanto $x = \operatorname{proj}_{u_1} x + \cdots + \operatorname{proj}_{u_n} x$.

No caso da base ser ortonormada, todos os vectores u_i satisfazem $\|u_i\| = 1$, pelo que a afirmação final segue da expressão obtida para os c_i's. □

5.3 Complemento ortogonal de um subespaço

Dois subespaços U e V de um espaço linear W dizem-se subespaços complementares se qualquer vector w de W se escreve na forma $w = u + v$, com $u \in U$ e $v \in V$, e, além disso, a intersecção dos subespaços é $U \cap V = \{0\}$. Ou seja, U e V são *subespaços complementares* em W se $W = U + V$ e $U \cap V = \{0\}$. Quando isto acontece escrevemos $W = U \oplus V$, que se lê "W é a soma directa de U com V".

Subespaços complementares num dado espaço linear W gozam da propriedade da soma das suas dimensões ser igual à dimensão de W. De facto, usando a Proposição 3.10 na página 141, a dimensão do subespaço $U + V$, é dada por

$$\dim(U + V) = \dim U + \dim V - \dim(U \cap V), \tag{5.14}$$

e portanto se U e V são subespaços complementares em W, tem-se $\dim(U \cap V) = 0$, pelo que $\dim W = \dim U + \dim V$.

Em $\mathbb{R}^2$ são exemplos de subespaços complementares duas quaisquer rectas distintas concorrentes na origem. Em $\mathbb{R}^3$, qualquer plano que contenha a origem é um subespaço complementar de uma recta que passa pela origem e não esteja contida no plano.

Estamos interessados em estudar subespaços de um espaço euclidiano que para além de serem complementares sejam ortogonais entre si, no sentido em que qualquer vector de um subespaço é ortogonal a qualquer vector do outro subespaço. Quando isto acontece, um subespaço diz-se o *complemento ortogonal* do outro subespaço.

Definição 5.7. Seja W um espaço linear munido de um produto interno $\langle\, ,\, \rangle$ e S um subespaço de W.

O *complemento ortogonal* de S é o conjunto de todos os vectores de W que são ortogonais a qualquer vector de S. Designamos o complemento ortogonal do subespaço S por $S^{\perp}$. Ou seja,

$$S^{\perp} = \{x \in W : \langle u, x\rangle = 0, \forall u \in S\}.$$

Vamos verificar que a definição apresentada significa que S e $S^{\perp}$ são subespaços complementares em W. Comecemos por mostrar que $S^{\perp}$ é um subespaço.

Proposição 5.3. O complemento ortogonal $S^{\perp}$ do subespaço S é um subespaço.

Demonstração. Para mostrar que $S^{\perp}$ é um subespaço, basta mostrar que é fechado para a adição e para a multiplicação por escalares. Como,

$$\begin{aligned} x \in S^{\perp} &\iff \langle u, x\rangle = 0 \quad \text{para todo} \quad u \in S \\ y \in S^{\perp} &\iff \langle u, y\rangle = 0 \quad \text{para todo} \quad u \in S, \end{aligned}$$

resulta que

$$\begin{aligned} \langle u, x\rangle + \langle u, y\rangle = 0 &\iff \langle u, x+y\rangle = 0 \quad \text{para todo } u \in S \\ k\langle u, x\rangle = 0 &\iff \langle u, kx\rangle = 0 \quad \text{para todo } u \in S \text{ e } k \text{ escalar.} \end{aligned}$$

As equivalências anteriores dizem que os vectores $x + y$ e kx pertencem a $S^{\perp}$, para quaisquer $x, y \in S$ e k escalar. Ou seja, $S^{\perp}$ é subespaço. □

Mostremos agora que um subespaço S e o seu complemento ortogonal $S^{\perp}$ se intersectam apenas no vector zero.

Proposição 5.4. Seja S um subespaço e $S^{\perp}$ o seu complemento ortogonal. Verifica-se:

$$S \cap S^{\perp} = \{0\}.$$

Demonstração. Sendo S e $S^{\perp}$ subespaços, o zero pertence a ambos os conjuntos, isto é, $0 \in S \cap S^{\perp}$. Tome-se $x \in S \cap S^{\perp}$.

Como todos os vectores de $S^{\perp}$ são ortogonais a qualquer vector de S, temos que x é ortogonal a si próprio, ou seja,

$$0 = \langle x, x\rangle = \|x\|^2 \iff x = 0, \qquad \text{(propriedade P4 da Definição 5.1).}$$

□

Exemplo 5.9. Verifiquemos se são ou não complementos ortogonais os seguintes conjuntos de $\mathbb{R}^3$ (munido do produto interno usual).

a) $U = \text{Span}\{(1,2,0),(1,0,1)\}$ e $V = \text{Span}\{(1,0,0)\}$. O subespaço V não é o complemento ortogonal de U já que $\langle(1,0,0),(1,0,1)\rangle = 1$. Ou seja, há vectores de U que não são ortogonais a vectores de V.

b) $U = \{(x,y,z) \in \mathbb{R}^3 : x+y-z=2\}$ e $V = \text{Span}\{(1,1,-1)\}$ não são complementos ortogonais, visto que U não é um subespaço (U é um plano que não contém a origem).

c) $U = \{(x,y,z) \in \mathbb{R}^3 : x+y-z=0\}$ e $V = \text{Span}\{(1,1,-1)\}$.
Os conjuntos U e V são subespaços e uma base para U é $\{(-1,1,0),(1,0,1)\}$. Como $\langle(-1,1,0),(1,1,-1)\rangle = 0$ e $\langle(1,0,1),(1,1,-1)\rangle = 0$, tem-se que $U = V^\perp$. Note-se que dados dois subespaços com bases B e B', para mostrar que os subespaços são ortogonais basta provar que cada vector da base B é ortogonal a todos os vectores da base B' (ver Exercício 5.8).

♦

Teorema da decomposição ortogonal

Já mostrámos que o complemento ortogonal $S^\perp$ do subespaço S é um subespaço e que $S^\perp \cap S = \{0\}$. Para mostrar que S e $S^\perp$ são subespaços complementares em W falta mostrar que qualquer vector de W se escreve como a soma de um vector de S com um vector de $S^\perp$. Este é o conteúdo do Teorema da decomposição ortogonal. Antes de enunciarmos esse teorema mostremos o lema seguinte.

Lema 5.1. Seja $\{u_1,\ldots,u_k\}$ uma base ortogonal do subespaço S de W, e x um vector de W. O vector

$$x_S = \text{proj}_S\, x = \text{proj}_{u_1} x + \cdots + \text{proj}_{u_k} x = \frac{\langle u_1, x\rangle}{\|u_1\|^2}u_1 + \cdots + \frac{\langle u_k, x\rangle}{\|u_k\|^2}u_k \tag{5.15}$$

pertence a S, e o vector

$$\begin{aligned} x - x_S &= x - \text{proj}_S\, x \\ &= x - \left(\frac{\langle u_1,x\rangle}{\|u_1\|^2}u_1 + \cdots + \frac{\langle u_k,x\rangle}{\|u_k\|^2}u_k\right) \end{aligned} \tag{5.16}$$

é ortogonal a S.

Demonstração. O vector x_S pertence obviamente ao subespaço S, uma vez que é a soma dos vectores $\operatorname{proj}_{u_i} x$, que são vectores de S. Vejamos agora que $x - x_S$ é ortogonal a S. Para tal, basta verificar que $x - x_S$ é ortogonal a qualquer vector de uma base de S. Assim,

$$\begin{aligned}
\langle u_i, x - x_S\rangle &= \langle u_i, x - \frac{\langle u_1, x\rangle}{\|u_1\|^2}u_1 - \cdots - \frac{\langle u_k, x\rangle}{\|u_k\|^2}u_k\rangle \\
&= \langle u_i, x\rangle - \frac{\langle u_1, x\rangle}{\|u_1\|^2}\langle u_i, u_1\rangle - \cdots - \frac{\langle u_k, x\rangle}{\|u_k\|^2}\langle u_i, u_k\rangle \\
&= \langle u_i, x\rangle - \frac{\langle u_i, x\rangle}{\|u_i\|^2}\langle u_i, u_i\rangle \quad \text{(já que a base é ortogonal)} \\
&= \langle u_i, x\rangle - \frac{\langle u_i, x\rangle}{\|u_i\|^2}\|u_i\|^2 = 0.
\end{aligned}$$

Logo, $x - x_S$ é ortogonal a S, isto é, $x - x_S$ pertence a $S^\perp$. □

Teorema 5.3. Teorema da decomposição ortogonal

Seja W um espaço euclidiano e S um subespaço de W. Qualquer vector $x \in W$ escreve-se de forma única como a soma de um vector x_S de S com um vector $x_{S^\perp}$ do complemento ortogonal de S. Isto é,

$$x = x_S + x_{S^\perp} \qquad \text{com} \qquad x_S \in S \text{ e } x_{S^\perp} \in S^\perp.$$

Define-se a projecção ortogonal de x sobre S como $\operatorname{proj}_S x = x_S$, e a projecção ortogonal de x sobre o subespaço $S^\perp$ como $\operatorname{proj}_{S^\perp} x = x_{S^\perp}$.

A demonstração deste teorema será realizada supondo que o subespaço S admite uma base ortogonal. A existência de uma base ortogonal é assegurada pelo processo de ortogonalização de Gram-Schmidt que será descrito na Secção 5.5.

Demonstração. Suponha-se que S admite uma base ortogonal e seja x um qualquer vector de W. O vector $x_S = \operatorname{proj}_S x$ definido no Lema 5.1 pertence a S, e o vector $x - x_S$ pertence a $S^\perp$. Logo, qualquer vector x de W é igual a $x = x_S + (x - x_S) = x_S + x_{S^\perp}$. Falta mostrar que esta decomposição é única. Para tal, considere-se $x = x_S + x_{S^\perp}$ e $x = \tilde{x}_S + \tilde{x}_{S^\perp}$ com $x_S, \tilde{x}_S \in S$ e $x_{S^\perp}, \tilde{x}_{S^\perp} \in S^\perp$. Subtraindo estas duas expressões de x, tem-se

$$0 = (x_S - \tilde{x}_S) + (x_{S^\perp} - \tilde{x}_{S^\perp}). \tag{5.17}$$

Os vectores $(x_S - \tilde{x}_S)$ e $(x_{S^\perp} - \tilde{x}_{S^\perp})$ pertencem, respectivamente, a S e a $S^\perp$, já que S e $S^\perp$ são subespaços. Mas a igualdade (5.17) diz que os vectores $(x_S - \tilde{x}_S)$ e $(x_{S^\perp} - \tilde{x}_{S^\perp})$ são simétricos, logo pertencem ambos a $S \cap S^\perp$. Isto significa que $(x_S - \tilde{x}_S) = 0 = (x_{S^\perp} - \tilde{x}_{S^\perp})$, já que $S \cap S^\perp = \{0\}$ (cf. Proposição 5.4). □

Na Figura 5.6 ilustramos o Teorema da decomposição ortogonal.

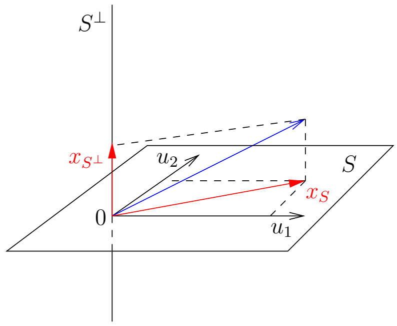


Figura 5.6: Teorema da decomposição ortogonal, onde $\{u_1, u_2\}$ é uma base ortogonal de S.

As proposições 5.3, 5.4, e o Teorema da decomposição ortogonal permitem concluir que dado um subespaço S de um espaço euclidiano W, se tem $W = S \oplus S^\perp$. Consequentemente,

$$\dim W = \dim(S + S^\perp) = \dim S + \dim S^\perp - \dim(S \cap S^\perp) = \dim S + \dim S^\perp.$$

Nota 31. 1. *A projecção ortogonal sobre um subespaço S, de um espaço euclidiano W, pode ser vista como uma função (ou operador) $P_S = \mathrm{proj_S} : W \to S$ tal que*

$$P_S(x) = \begin{cases} x & \textit{se } x \in S \\ 0 & \textit{se } x \in S^\perp. \end{cases}$$

Consequentemente

$$P_S^2(x) = P_S(P_S(x)) = P_S(x) \textit{ para todo } x \in W.$$

Um operador P que satisfaça a condição $P^2 = P$ diz-se idempotente, ou operador de projecção.

Exemplo 5.10. 1) Consideremos um plano em $\mathbb{R}^3$ que contém a origem. Este subespaço é definido por uma equação do tipo $ax + by + cz = 0$, em que a, b e c não são simultaneamente nulos.

Qualquer vector (x, y, z) deste plano é ortogonal ao vector (a, b, c), visto que

$$\langle (a, b, c), (x, y, z) \rangle = ax + by + cz = 0.$$

Uma vez que o plano é um subespaço de $\mathbb{R}^3$ de dimensão dois, o seu complemento ortogonal é um subespaço de $\mathbb{R}^3$ de dimensão um. Logo, a recta que passa pela origem e tem a direcção do vector (a, b, c) é o complemento ortogonal ao plano.

2) Considere-se $\mathbb{R}^2$ munido do produto interno usual, e S a recta de equação $y = x$. Esta recta passa pela origem e tem a direcção do vector $(1, 1)$. Assim, $\{1, 1)\}$ é uma base de S. O complemento ortogonal $S^\perp$ é o conjunto de todos os vectores $(x, y) \in \mathbb{R}^2$ ortogonais a $(1, 1)$, isto é

$$\langle (x, y), (1, 1) \rangle = x + y = 0 \Longleftrightarrow x = -y.$$

Logo

$$S^\perp = \left\{(x, y) \in \mathbb{R}^2 : \ x = -y\right\} = \{(-y, y) : \ y \in \mathbb{R}\} = \operatorname{Span}\{(-1, 1)\}.$$

Por conseguinte, $S^\perp$ é a recta perpendicular à recta S que passa pela origem.

♦

A expressão (5.15), da projecção ortogonal sobre um subespaço S, pressupõe o conhecimento de uma base ortogonal de S. No caso particular em que a dimensão do subespaço S de W é igual a $\dim S = \dim W - 1$, podemos usar o teorema da decomposição ortogonal para determinar $\operatorname{proj}_S x$ sem termos de determinar uma base ortogonal de S. Neste caso, como a dimensão de $S^\perp$ é $\dim S^\perp = 1$, do Teorema da decomposição ortogonal resulta

$$\operatorname{proj}_S x = x - \operatorname{proj}_{S^\perp} x = x - \operatorname{proj}_v x,$$

onde v é um qualquer vector não nulo de $S^\perp$.

Num espaço linear de dimensão n, chama-se *hiperplano* a um subespaço de dimensão $(n - 1)$.

Vejamos agora como se traduz a projecção ortogonal sobre hiperplanos de $\mathbb{R}^n$ para o produto interno usual. Relembremos que, para este produto interno, a projecção ortogonal de um vector $\mathbf{x} \in \mathbb{R}^n$ sobre o vector não nulo $\mathbf{u} \in \mathbb{R}^n$ tem a expressão matricial $P_\mathbf{u}\mathbf{x}$, onde $P_\mathbf{u}$ é definida em (5.12).

Considere-se o subespaço de $\mathbb{R}^n$ gerado pelo vector não nulo $\mathbf{u}$ e designe-se por $\mathbf{u}^\perp$ o seu complemento ortogonal (usamos a notação $\mathbf{u}^\perp$ em vez da notação habitual $(\mathrm{Span}\{\mathbf{u}\})^\perp$). O subespaço $\mathbf{u}^\perp$ é um hiperplano uma vez que $\dim \mathbf{u}^\perp = n-1$. Para qualquer $\mathbf{x} \in \mathbb{R}^n$, o Teorema da decomposição ortogonal garante que

$$\mathbf{x} = \mathrm{proj}_\mathbf{u}\,\mathbf{x} + \mathrm{proj}_{\mathbf{u}^\perp}\,\mathbf{x} = P_\mathbf{u}\mathbf{x} + P_{\mathbf{u}^\perp}\mathbf{x} \iff (I - P_\mathbf{u})\mathbf{x} = P_{\mathbf{u}^\perp}\mathbf{x},$$

significando que a projecção ortogonal sobre o complemento ortogonal ao subespaço gerado por $\mathbf{u}$ é obtida multiplicando pela matriz $(I - P_\mathbf{u})$.

Verifiquemos agora algumas propriedades das matrizes $P_\mathbf{u}$ e $(I - P_\mathbf{u})$:

- Como $P_\mathbf{u} = \frac{1}{\|\mathbf{u}\|^2}\mathbf{u}\mathbf{u}^T$, as matrizes $P_\mathbf{u}$ e $(I - P_\mathbf{u})$ são matrizes simétricas.
- $P_\mathbf{u}^2 = P_\mathbf{u}$ e $(I - P_\mathbf{u})^2 = (I - P_\mathbf{u})$, uma vez que

$$P_\mathbf{u}^2 = P_\mathbf{u}\left(\frac{1}{\|\mathbf{u}\|^2}\mathbf{u}\mathbf{u}^T\right) = \frac{1}{\|\mathbf{u}\|^4}\mathbf{u}\underbrace{\mathbf{u}^T\mathbf{u}}_{\|\mathbf{u}\|^2}\mathbf{u}^T = \frac{1}{\|\mathbf{u}\|^2}\mathbf{u}\mathbf{u}^T = P_\mathbf{u}.$$

$$(I - P_\mathbf{u})^2 = (I - P_\mathbf{u})(I - P_\mathbf{u}) = I - P_\mathbf{u} - P_\mathbf{u} + P_\mathbf{u}^2 = I - P_\mathbf{u}.$$

Exemplo 5.11. Determinar $\mathrm{proj}_\mathbf{u}\,\mathbf{x}$ e $\mathrm{proj}_{\mathbf{u}^\perp}\,\mathbf{x}$ para os seguintes vectores

$$\mathbf{x} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \quad \text{e} \quad \mathbf{u} = \begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix}.$$

Assim, usando a definição temos

$$\mathrm{proj}_\mathbf{u}\,\mathbf{x} = \frac{\langle \mathbf{u}, \mathbf{x}\rangle}{\|\mathbf{u}\|^2}\mathbf{u} = \frac{3+4}{9+1+4}\begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix}.$$

Podemos confirmar este resultado usando a matriz $P_\mathbf{u}$.

$$\mathrm{proj}_\mathbf{u}\,\mathbf{x} = P_\mathbf{u}\mathbf{x} = \frac{1}{\|\mathbf{u}\|^2}\mathbf{u}\mathbf{u}^T\mathbf{x} = \frac{1}{14}\begin{bmatrix} 9 & -3 & 6 \\ -3 & 1 & -2 \\ 6 & -2 & 4 \end{bmatrix}\mathbf{x} = \frac{1}{2}\begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix}.$$

A projecção ortogonal de $\mathbf{x}$ sobre o complemento ortogonal ao espaço gerado por $\mathbf{u}$ é

$$\operatorname{proj}_{\mathbf{u}^\perp}\mathbf{x} = \mathbf{x} - \operatorname{proj}_{\mathbf{u}}\mathbf{x} = \begin{bmatrix}1\\0\\2\end{bmatrix} - \frac{1}{2}\begin{bmatrix}3\\-1\\2\end{bmatrix} = \frac{1}{2}\begin{bmatrix}-1\\1\\2\end{bmatrix} = \mathbf{x} - P_{\mathbf{u}}\mathbf{x} = (I - P_{\mathbf{u}})\mathbf{x}.$$

Note-se que neste caso $\mathbf{u}^\perp$ é um plano perpendicular à recta definida por $\mathbf{u}$. Se para determinar a projecção ortogonal sobre este plano tivéssemos usado a expressão (5.15) necessitaríamos de determinar previamente uma base ortogonal do plano. Sugerimos que confirme o resultado obtido calculando a projecção ortogonal sobre o plano dessa forma. ♦

Teorema da Melhor Aproximação

O Teorema da melhor aproximação está na base de vários resultados usados nas aplicações, com destaque para o problema de mínimos quadrados que abordaremos neste capítulo. Esse teorema afirma que dado um subespaço S de W e um vector $x \in W$, o vector de S mais próximo de x é $\operatorname{proj}_S x = x_S$. Ou seja, que a distância de x a qualquer outro vector u de S é maior ou igual à distância de x a $\operatorname{proj}_S x$:

$$\text{dist}(x, x_S) \leq \text{dist}(x, u) \quad \text{qualquer que seja } u \in S.$$

Usando a definição de distância entre dois vectores (ver Definição 5.2), a desigualdade anterior é equivalente a

$$\|x - x_S\| \leq \|x - u\| \quad \text{para qualquer } u \in S.$$

Teorema 5.4. Teorema da melhor aproximação

Seja W um espaço linear munido de um produto interno, S um subespaço de W e x um vector de W. Então,

$$\|x - x_S\| \leq \|x - u\| \quad \text{para qualquer } u \in S,$$

onde $x_S = \operatorname{proj}_S x$ é a projecção ortogonal de x sobre S.

Demonstração. Pelo Teorema da decomposição ortogonal podemos escrever $x = x_S + x_{S^\perp}$. Seja u um qualquer vector de S. Uma vez que os vectores $x - x_S = x_{S^\perp}$ e $x_S - u \in S$ são ortogonais, pelo Teorema de Pitágoras (pág. 254), temos

$$\|x - u\|^2 = \|\underbrace{x - x_S}_{\in S^\perp} + \underbrace{x_S - u}_{\in S}\|^2 = \|x - x_S\|^2 + \|x_S - u\|^2.$$

Da igualdade anterior obtém-se

$$\|x-u\|^2 \geq \|x-x_S\|^2,$$

equivalente à desigualdade do enunciado. $\square$

O Teorema da melhor aproximação permite determinar a distância de vectores de um espaço linear W a um subespaço S de W. A distância de $x \in W$ a um subespaço S, é a menor das distâncias de x aos vectores de S. Ou seja, a distância dist (x, S), de x ao subespaço S, é

$$\operatorname{dist}(x, S) = \min_{u \in S} \operatorname{dist}(x, u) = \min_{u \in S} \|x-u\| = \|x - x_S\| = \|\operatorname{proj}_{S^\perp} x\|.$$

Distância a um subespaço

Seja W um espaço linear, S um subespaço de W e x um vector de W. A distância de x a S é:

$$\operatorname{dist}(x, S) = \|\operatorname{proj}_{S^\perp} x\|.$$

5.4 Ortogonalidade dos quatro subespaços fundamentais

Nesta secção analisamos as relações de ortogonalidade entre os quatro subespaços fundamentais de $\mathbb{R}^n$ associados a uma matriz real A. Relembremos (ver Secção 3.2, pág. 114) que estes subespaços são: o núcleo de A; o espaço das linhas de A; o espaço das colunas de A; e o núcleo de A^T.

Consideremos $\mathbb{R}^n$ munido do produto interno usual $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T\mathbf{y}$. Se A é uma matriz real $p \times n$,

$$A = \begin{bmatrix} — & \mathbf{l}_1 & — \\ — & \mathbf{l}_2 & — \\ & \vdots & \\ — & \mathbf{l}_p & — \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \\ | & | & & | \end{bmatrix},$$

então:

- As linhas $\mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_p$ de A são vectores de $\mathbb{R}^n$.
- As colunas $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n$ de A são vectores de $\mathbb{R}^p$.

Como estamos a considerar o produto interno usual em $\mathbb{R}^n$, a multiplicação da matriz A por um vector $\mathbf{x} \in \mathbb{R}^n$ é igual a

$$A\mathbf{x} = \begin{bmatrix} — & \mathbf{l}_1 & — \\ — & \mathbf{l}_2 & — \\ & \vdots & \\ — & \mathbf{l}_p & — \end{bmatrix} \begin{bmatrix} | \\ \mathbf{x} \\ | \end{bmatrix} = \begin{bmatrix} \langle \mathbf{l}_1, \mathbf{x} \rangle \\ \langle \mathbf{l}_2, \mathbf{x} \rangle \\ \vdots \\ \langle \mathbf{l}_p, \mathbf{x} \rangle \end{bmatrix}.$$

Logo, o conjunto dos vectores $\mathbf{x} \in \mathbb{R}^n$ que são ortogonais a todas as linhas de A é o conjunto de todos os vectores $\mathbf{x}$ que verificam $A\mathbf{x} = \mathbf{0}$, ou seja, o núcleo de A.

Se $\mathbf{x}$ verifica $\langle \mathbf{x}, \mathbf{l}_i \rangle = 0$ para $i = 1, \ldots, p$, então $\mathbf{x}$ é ortogonal a qualquer combinação linear das linhas $\mathbf{l}_i$, uma vez que

$$\langle \mathbf{x}\,,\, \alpha_1 \mathbf{l}_1 + \ldots + \alpha_p \mathbf{l}_p \rangle = \alpha_1 \langle \mathbf{x}\,,\, \mathbf{l}_1 \rangle + \cdots + \alpha_p \langle \mathbf{x}\,,\, \mathbf{l}_p \rangle = 0.$$

Assim, qualquer vector do espaço das linhas de A é ortogonal ao núcleo de A, isto é,

$$(EL(A))^\perp = N(A).$$

Como a relação anterior é válida para qualquer matriz, em particular é satisfeita pela matriz transposta A^T. Ou seja, $\left(EL(A^T)\right)^\perp = N(A^T)$. Logo,

$$(EC(A))^\perp = N(A^T).$$

Resumindo:

$$\boxed{(EL(A))^\perp = N(A) \quad \text{e} \quad (EC(A))^\perp = N(A^T) \qquad (5.18)}$$

Os espaços $EL(A)$ e $N(A)$ de uma matriz real $p \times n$ são subespaços de $\mathbb{R}^n$, enquanto que $EC(A)$ e $N(A^T)$ são subespaços de $\mathbb{R}^p$. Por conseguinte, de (5.18), resulta

$$\mathbb{R}^n = EL(A) \oplus N(A) \quad \text{e} \quad \mathbb{R}^p = EC(A) \oplus N(A^T).$$

Do Teorema da dimensão (Teorema 3.6, pág. 144), tem-se que $\dim EC(A) = \dim EL(A) = \operatorname{car}(A) = k$ enquanto que a dimensão do núcleo de uma matriz é a diferença entre o número de colunas da matriz e a sua característica. Na Figura 5.7 ilustramos estes factos usando um diagrama apresentado por Strang em [11].

A expressão (5.15) para a projecção ortogonal de um vector sobre um subespaço, pressupõe o conhecimento de uma base ortogonal do subespaço. Os resultados do próximo exercício permitem calcular a projecção ortogonal sobre um subespaço de $\mathbb{R}^n$ sem determinar uma base ortogonal desse subespaço.

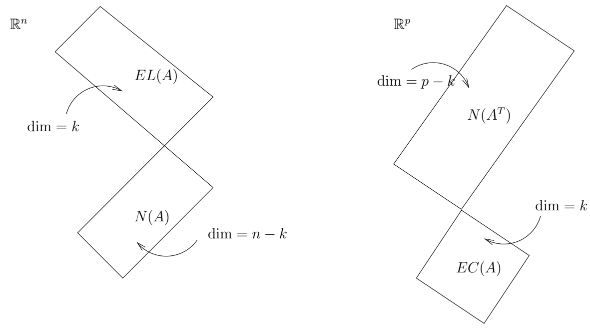


Figura 5.7: Ortogonalidade dos quatro subespaços fundamentais para uma matriz real A do tipo $p \times n$, de característica k.

Exercício 5.11. Considere uma matriz real A do tipo $p \times n$ tal que $A^T A$ é invertível, e a matriz quadrada $P = A(A^T A)^{-1} A^T$. Mostre que:

(a) $P\mathbf{b} = \mathbf{b}$ se $\mathbf{b} \in EC(A)$.

(b) $P\mathbf{b} = \mathbf{0}$ se $\mathbf{b} \in (EC(A))^{\perp}$.

(c) $P^2 = P$.

(d) Justifique que as propriedades das alíneas (a)-(c) equivalem a afirmar que a matriz P fornece a projecção ortogonal sobre o espaço das colunas de A.

(e) Verifique que se A tem uma única coluna não nula $\mathbf{a}$, então $\operatorname{proj}_{\mathbf{a}} \mathbf{x} = P\mathbf{x}$ para qualquer $\mathbf{x} \in \mathbb{R}^p$.

(f) Use a matriz P para determinar a projecção ortogonal de $(1, 2, 0)$ sobre $S = \operatorname{Span}\{(1, 1, 1), (-2, 0, 1)\}$.

▲

Nota 32. *O exercício anterior mostra que a matriz P dá a projecção ortogonal sobre o espaço das colunas de A (desde que $A^T A$ seja invertível). Assim, pode calcular-se a projecção ortogonal sobre um subespaço de $\mathbb{R}^n$ (munido do produto interno usual) tomando para A a matriz que tenha por colunas uma base (não necessariamente ortogonal) do subespaço linear em questão.*

No Teorema 5.8, pág. 301, estabelecem-se condições para invertibilidade da matriz $A^T A$.

Exercício 5.12. Considere o produto interno usual em $\mathbb{C}^n$, S um subespaço de $\mathbb{C}^n$ e $B = (\mathbf{u}_1, \dots, \mathbf{u}_k)$ uma base ortonormada de S. Mostre que a projecção ortogonal sobre S é dada por

$$\operatorname{proj}_S \mathbf{x} = QQ^H\mathbf{x} \quad \text{para todo } \mathbf{x} \in \mathbb{C}^n, \tag{5.19}$$

onde Q é a matriz que tem por colunas os vectores da base B (pela mesma ordem). Sugestão: Use o Exercício 5.11 com as adaptações convenientes ao caso de matrizes complexas (substitua nesse exercício o símbolo 'T' por 'H'). ▲

5.5 Ortogonalização de Gram-Schmidt e decomposição QR

Como se observa na expressão (5.15), a projecção ortogonal sobre um subespaço S tem uma expressão simples, desde que se conheça uma base ortogonal de S. A questão que naturalmente se coloca é a da existência de uma base ortogonal para qualquer subespaço. A resposta a esta questão é fornecida pelo denominado *processo de ortogonalização de Gram-Schmidt*[5] o qual permite obter uma base ortogonal a partir de uma base dada. Par tal considere-se um conjunto de vectores linearmente independentes $V = \{v_1, v_2, \dots, v_k\}$ de um espaço euclidiano. O processo de ortogonalização de Gram-Schmidt constrói a partir de V, de forma recursiva, um conjunto ortogonal $U = \{u_1, u_2, \dots, u_k\}$. Os passos deste algoritmo são detalhados a seguir.

Passo 1: Tome-se para u_1 um dos vectores de V:

$$u_1 = v_1.$$

Passo 2: Considere-se outro vector v_2 de V, e construa-se a projecção ortogonal de v_2 sobre u_1. Do Exemplo 5.7, sabemos que o vector $v_2 - \operatorname{proj}_{u_1} v_2$ é ortogonal a u_1 (observe a ilustração na Figura 5.5 da página 255). Tome-se para u_2 o vector $v_2 - \operatorname{proj}_{u_1} v_2$, isto é,

$$u_2 = v_2 - \operatorname{proj}_{u_1} v_2 = v_2 - \frac{\langle u_1, v_2 \rangle}{\|u_1\|^2} u_1.$$

[5] Este processo recebeu os nomes dos matemáticos Jørgen Pedersen Gram (1850 – 1916) e Erhard Schmidt (1876 – 1959), embora possa ser encontrado em trabalhos anteriores de Laplace e Cauchy.

Passo 3: Considere-se o subespaço S gerado por $\{u_1, u_2\}$. Este subespaço tem dimensão igual a dois já que os vectores u_1, u_2 são ortogonais, e portanto linearmente independentes (cf. Proposição 5.1, pág. 259). Além disso, como u_2 é combinação linear de $u_1 = v_1$ e de v_2, tem-se $\mathrm{Span}\{u_1, u_2\} \subseteq \mathrm{Span}\{v_1, v_2\}$. Como, por hipótese, $\{v_1, v_2\}$ é linearmente independente, tem-se $S = \mathrm{Span}\{u_1, u_2\} = \mathrm{Span}\{v_1, v_2\}$.

Considere-se v_3 um outro vector de V. O Lema 5.1 garante que o vector $v_3 - \mathrm{proj}_S\, v_3$ é um vector ortogonal a u_1 e u_2. Tome-se para u_3 este vector

$$u_3 = v_3 - \mathrm{proj}_S\, v_3 = v_3 - \mathrm{proj}_{u_1} v_3 - \mathrm{proj}_{u_2} v_3,$$

onde na segunda igualdade se aplicou o facto de $\{u_1, u_2\}$ ser uma base ortogonal de S.

Passo i+1: Repetir o passo anterior considerando para S o subespaço gerado pelos vectores u_i construídos nos passos anteriores, considerar um vector $v_{i+1} \in V$ ainda não utilizado, e tomar para $u_{i+1} = v_{i+1} - \mathrm{proj}_S\, v_{i+1}$. Repetir este processo até esgotar os vectores de V.

Acabámos de ilustrar o denominado processo de ortogonalização de Gram-Schmidt.

Teorema 5.5. Ortogonalização de Gram-Schmidt

Seja $V = \{v_1, v_2, \ldots, v_k\}$, com $k > 1$, um conjunto linearmente independente de um espaço linear munido de um produto interno. O conjunto $U = \{u_1, u_2, \ldots, u_k\}$, formado pelos vectores

$$\begin{aligned}
u_1 &= v_1 \\
u_2 &= v_2 - \frac{\langle u_1, v_2\rangle}{\|u_1\|^2} u_1 \\
u_3 &= v_3 - \frac{\langle u_1, v_3\rangle}{\|u_1\|^2} u_1 - \frac{\langle u_2, v_3\rangle}{\|u_2\|^2} u_2 \\
&\vdots \\
u_k &= v_k - \mathrm{proj}_{u_1} v_k - \mathrm{proj}_{u_2} v_k - \cdots - \mathrm{proj}_{u_{k-1}} v_k,
\end{aligned}$$

é ortogonal.

Os conjuntos U e V geram o mesmo espaço.

Demonstração. Para mostrar que o conjunto U é ortogonal vamos usar indução sobre o número j de vectores de U. Para $j = 2$, como o vector $v_2 - \operatorname{proj}_{u_1} v_2$ é ortogonal a u_1 (ver Exemplo 5.7), os vectores u_1 e u_2 são ortogonais.

Suponha-se que o conjunto $\{u_1, \ldots, u_p\}$ é ortogonal (hipótese de indução). Mostremos que o conjunto $\{u_1, \ldots, u_p, u_{p+1}\}$ também é ortogonal. Calculando $\langle u_i, u_{p+1} \rangle$, para qualquer $i = 1, \ldots, p$, tem-se

$$\begin{aligned}
\langle u_i, u_{p+1} \rangle &= \left\langle u_i, v_{p+1} - \sum_{j=1}^{p} \frac{\langle u_j, v_{p+1} \rangle}{\|u_j\|^2} u_j \right\rangle \\
&= \langle u_i, v_{p+1} \rangle - \sum_{j=1}^{p} \frac{\langle u_j, v_{p+1} \rangle}{\|u_j\|^2} \langle u_i, u_j \rangle \\
&= \langle u_i, v_{p+1} \rangle - \frac{\langle u_i, v_{p+1} \rangle}{\|u_i\|^2} \langle u_i, u_i \rangle \\
&= \langle u_i, v_{p+1} \rangle - \langle u_i, v_{p+1} \rangle = 0,
\end{aligned}$$

onde na terceira igualdade aplicámos a hipótese de indução. Logo, U é ortogonal.

Falta mostrar que U e V geram o mesmo espaço. Os vectores de U, construídos a partir dos vectores de V, são da forma

$$u_i = v_i - \operatorname{proj}_{u_1} v_i - \operatorname{proj}_{u_2} v_i - \cdots - \operatorname{proj}_{u_{i-1}} v_i,$$

isto é, os u_i's são combinações lineares dos vectores de V, e portanto $\operatorname{Span} U \subseteq \operatorname{Span} V$. O conjunto U é ortogonal, e portanto U é linearmente independente pela Proposição 5.1 (pág. 259). Como as dimensões de $\operatorname{Span} U$ e $\operatorname{Span} V$ são iguais e $\operatorname{Span} U \subseteq \operatorname{Span} V$, tem-se necessariamente $\operatorname{Span} U = \operatorname{Span} V$. $\square$

Exercício 5.13. Determinemos uma base ortogonal para o subespaço

$$S = \operatorname{Span}\{(1,1,0), (0,1,0), (0,1,1)\}.$$

É fácil verificar que os vectores $\mathbf{v}_1 = (1,1,0)$, $\mathbf{v}_2 = (0,1,0)$ e $\mathbf{v}_3 = (0,1,1)$ formam uma base para S. Esta base não é ortogonal já que, por exemplo, $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = 1$.

Apliquemos o processo de ortogonalização de Gram-Schmidt para obter uma base $\{\mathbf{u}_1, \mathbf{u_2}, \mathbf{u}_3\}$ de S que seja ortogonal.

$$\begin{aligned}
\mathbf{u}_1 &= \mathbf{v}_1 = (1,1,0)\\
\mathbf{u}_2 &= \mathbf{v}_2 - \operatorname{proj}_{\mathbf{u}_1}\mathbf{v}_2 = (0,1,0) - \frac{1}{2}(1,1,0) = \frac{1}{2}(-1,1,0)\\
\mathbf{u}_3 &= \mathbf{v}_3 - \operatorname{proj}_{\mathbf{u}_1}\mathbf{v}_3 - \operatorname{proj}_{\mathbf{u}_2}\mathbf{v}_3\\
&= (0,1,1) - \frac{1}{2}(1,1,0) - \frac{1/2}{1/2}\left(\frac{-1}{2},\frac{1}{2},0\right) = (0,0,1).
\end{aligned}$$

▲

Decomposição QR

A factorização de uma matriz A na forma $A = QR$, onde Q é uma matriz cujas colunas formam um conjunto ortonormado e R é uma matriz triangular superior invertível, é conhecida como *decomposição* QR de A. Esta factorização tem variadas aplicações numéricas, nomeadamente no que respeita a algoritmos para determinação de valores próprios de matrizes de grandes dimensões, sendo também frequentemente utilizada na resolução de problemas de mínimos quadrados (ver Secção 5.7.2, pág. 295). A decomposição QR surge naturalmente como consequência do processo de ortogonalização de Gram-Schmidt.

Considere-se uma matriz A cujos vectores coluna $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n$ são linearmente independentes

$$A = \begin{bmatrix} | & | & & | \\ \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \\ | & | & & | \end{bmatrix}.$$

Podemos aplicar o processo de ortogonalização de Gram-Schmidt ao conjunto $V = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n\}$ e obter um conjunto ortogonal U. Se a este conjunto aplicarmos a normalização referida em (5.13), obtemos um conjunto ortonormado de vectores:

$$\begin{array}{ccccc} V = \{\mathbf{c}_1, \ldots, \mathbf{c}_n\} & \xrightarrow{\text{Gram-Schmidt}} & U & \xrightarrow{\text{Normalização}} & \{\mathbf{q}_1, \ldots, \mathbf{q}_n\}. \\ & & \downarrow & & \downarrow \\ & & \text{Ortogonal} & & \text{Ortonormado} \end{array}$$

O conjunto V, o conjunto U e o conjunto $\{\mathbf{q}_1, \ldots, \mathbf{q}_n\}$ geram o mesmo subespaço. Assim, cada vector $\mathbf{c}_i \in V$ pode escrever-se como combinação linear dos vectores

$\mathbf{q}_1, \ldots, \mathbf{q}_n$. Como o conjunto formado pelos $\mathbf{q}_i$'s é uma base ortonormada para o subespaço gerado por V, do Teorema 5.2 resulta

$$\mathbf{c}_i = \operatorname{proj}_{\mathbf{q}_1} \mathbf{c}_1 + \cdots + \operatorname{proj}_{\mathbf{q}_n} \mathbf{c}_n.$$

Equivalentemente,

$$\begin{aligned}
\mathbf{c}_1 &= \langle \mathbf{q}_1, \mathbf{c}_1 \rangle \mathbf{q}_1 + \langle \mathbf{q}_2, \mathbf{c}_1 \rangle \mathbf{q}_2 + \cdots + \langle \mathbf{q}_n, \mathbf{c}_1 \rangle \mathbf{q}_n \\
\mathbf{c}_2 &= \langle \mathbf{q}_1, \mathbf{c}_2, \rangle \mathbf{q}_1 + \langle \mathbf{q}_2, \mathbf{c}_2 \rangle \mathbf{q}_2 + \cdots + \langle \mathbf{q}_n, \mathbf{c}_2 \rangle \mathbf{q}_n \\
&\ \vdots \\
\mathbf{c}_n &= \langle \mathbf{q}_1, \mathbf{c}_n \rangle \mathbf{q}_1 + \langle \mathbf{q}_2, \mathbf{c}_n \rangle \mathbf{q}_2 + \cdots + \langle \mathbf{q}_n, \mathbf{c}_n \rangle \mathbf{q}_n.
\end{aligned}$$

Suponha-se que os vectores de $U = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ respeitam a ordem pela qual foram construídos através do processo de ortogonalização de Gram-Schmidt. Ou seja, suponha-se que cada vector $\mathbf{q}_i = \frac{\mathbf{u}_i}{\|\mathbf{u}_i\|}$ é ortogonal a cada um dos vectores $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_{i-1}$ para $1 < i \leq n$. Como $\langle \mathbf{c}_j, \mathbf{q}_i \rangle = 0$ para $j = 1, \ldots, i-1$, as igualdades anteriores reduzem-se a

$$\begin{aligned}
\mathbf{c}_1 &= \langle \mathbf{q}_1, \mathbf{c}_1 \rangle \mathbf{q}_1 \\
\mathbf{c}_2 &= \langle \mathbf{q}_1, \mathbf{c}_2 \rangle \mathbf{q}_1 + \langle \mathbf{q}_2, \mathbf{c}_2 \rangle \mathbf{q}_2 \\
&\ \vdots \\
\mathbf{c}_n &= \langle \mathbf{q}_1, \mathbf{c}_n \rangle \mathbf{q}_1 + \langle \mathbf{q}_2, \mathbf{c}_n \rangle \mathbf{q}_2 + \cdots + \langle \mathbf{q}_n, \mathbf{c}_n \rangle \mathbf{q}_n.
\end{aligned}$$

Exprimindo estas igualdades na forma matricial, e relembrando que a coluna j do produto QR de duas matrizes é a matriz Q multiplicada pela coluna j de R (reveja as definições 1.12 e 1.11, pág. 34), tem-se

$$\underbrace{\begin{bmatrix} | & | & & | \\ \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_n \\ | & | & & | \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} | & | & & | \\ \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \\ | & | & & | \end{bmatrix}}_{Q} \underbrace{\begin{bmatrix} \langle \mathbf{q}_1, \mathbf{c}_1 \rangle & \langle \mathbf{q}_1, \mathbf{c}_2 \rangle & \cdots & \langle \mathbf{q}_1, \mathbf{c}_n \rangle \\ 0 & \langle \mathbf{q}_2, \mathbf{c}_2 \rangle & \cdots & \langle \mathbf{q}_2, \mathbf{c}_n \rangle \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \langle \mathbf{q}_n, \mathbf{c}_n \rangle \end{bmatrix}}_{R}.$$

Cada vector $\mathbf{q}_i$ é o vector $\mathbf{u}_i$ normalizado, onde $\mathbf{u}_i$ é obtido pelo processo de ortogonalização de Gram-Schmidt à custa de $\{\mathbf{c}_1, \ldots, \mathbf{c}_{i-1}\}$. Assim,

$$\mathbf{q}_i = \begin{cases} \frac{\mathbf{c}_1}{\|\mathbf{c}_1\|} & \text{para } i = 1 \\ \frac{1}{\mu_i}\left(\mathbf{c}_i - \sum_{j=1}^{i-1} \frac{\langle \mathbf{u}_j, \mathbf{c}_i \rangle}{\|\mathbf{u}_j\|^2} \mathbf{u}_j\right) & \text{para } 1 < i \leq n, \end{cases}$$

com $\mu_i = \left\| \mathbf{c}_i - \sum_{j=1}^{i-1} \frac{\langle \mathbf{u}_j, \mathbf{c}_i \rangle}{\|\mathbf{u}_j\|^2} \mathbf{u}_j \right\|$. Ou seja, para $i > 1$ podemos escrever

$$\mathbf{c}_i = \mu_i \mathbf{q}_i + \sum_{j=1}^{i-1} \frac{\langle \mathbf{u}_j, \mathbf{c}_i \rangle}{\|\mathbf{u}_j\|^2} \mathbf{u}_j.$$

Logo, as entradas da diagonal principal da matriz R são dadas por

$$\begin{aligned}
\langle \mathbf{q}_1, \mathbf{c}_1 \rangle &= \langle \frac{\mathbf{c}_1}{\|\mathbf{c}_1\|}, \mathbf{c}_1 \rangle = \|\mathbf{c}_1\|, \\
\langle \mathbf{q}_i, \mathbf{c}_i \rangle &= \mu_i \langle \mathbf{q}_i, \mathbf{q}_i \rangle + \sum_{j=1}^{i-1} \frac{\langle \mathbf{u}_j, \mathbf{c}_i \rangle}{\|\mathbf{u}_j\|^2} \langle \mathbf{q}_i, \mathbf{u}_j \rangle \\
&= \mu_i \|\mathbf{q}_i\|^2 = \mu_i,
\end{aligned}$$

onde na penúltima igualdade aplicámos o facto de cada vector $\mathbf{u}_i$ (e portanto $\mathbf{q}_i$) ser ortogonal aos vectores $\mathbf{u}_1, \ldots, \mathbf{u}_{i-1}$. Ou seja, as entradas da diagonal principal da matriz R são positivas. Enunciamos no teorema seguinte o que acabamos de provar.

Teorema 5.6. Decomposição QR

Se A é uma matriz do tipo $p \times n$ cujas colunas formam um conjunto linearmente independente, então existem matrizes Q e R tais que

$$A = QR,$$

sendo

(i) Q uma matriz cujas colunas formam um conjunto ortonormado, ou seja, $Q^T Q = I$ no caso de A ser real, e $Q^H Q = I$ no caso de A ser complexa.

(ii) R uma matriz triangular superior cujas entradas na diagonal principal são todas positivas.

Além disso, a factorização $A = QR$ é única.

Demonstração. Consideramos o caso em que A é uma matriz real deixando o caso complexo como exercício. Falta mostrar que a factorização $A = QR$ é única. Comecemos por notar que se provarmos que a matriz R na factorização $A = QR$ é única, segue de imediato que a matriz Q é única, uma vez que sendo R invertível, se tem

$$Q_1 R = Q_2 R \Longleftrightarrow Q_1 = Q_2.$$

Provemos agora que a matriz R é única. Suponha-se que existem matrizes R_1 e R_2 triangulares superiores com todas as entradas da diagonal principal positivas, e matrizes Q_1 e Q_2 cujas colunas formam conjuntos ortonormados, tais que $A = Q_1R_1 = Q_2R_2$. Como o conjunto das colunas de Q_1 e Q_2 é ortonormado (isto é, $Q_1^TQ_1 = I = Q_2^TQ_2$), e as matrizes R_1 e R_2 são invertíveis, verifica-se

$$R_1^TQ_1^TQ_1R_1 = A^TA = R_2^TQ_2^TQ_2R_2 \Longleftrightarrow R_1^TR_1 = R_2^TR_2 \Rightarrow R_1R_2^{-1} = (R_1^T)^{-1}R_2^T.$$

O produto de matrizes triangulares superiores (resp. inferiores) é uma matriz triangular superior (resp. inferior), a inversa de uma matriz triangular superior (resp. inferior) é uma matriz triangular superior (resp. inferior), e a transposta de uma matriz triangular superior é uma matriz triangular inferior. Assim, da igualdade anterior, obtemos

$$\underbrace{R_1R_2^{-1}}_{\text{triangular superior}} = \underbrace{(R_1^T)^{-1}R_2^T}_{\text{triangular inferior}} \Longrightarrow R_1R_2^{-1}\text{é diagonal.}$$

Qualquer matriz triangular superior invertível R pode escrever-se como o produto de uma matriz diagonal cujas entradas na diagonal são as entradas da diagonal principal de R por uma matriz triangular superior com 1's na diagonal principal, ou seja,

$$R_1 = D_1U_1 \quad \text{e} \quad R_2 = D_2U_2,$$

com D_1 e D_2 matrizes diagonais invertíveis, e U_1 e U_2 matrizes triangulares superiores com 1's na diagonal principal. Logo, para a matriz diagonal $R_1R_2^{-1} = D$, tem-se

$$R_1R_2^{-1} = D \Longleftrightarrow D_1U_1U_2^{-1}D_2^{-1} = D \Longrightarrow U_1U_2^{-1} = D_1^{-1}DD_2.$$

Como a matriz $U_1U_2^{-1}$ é triangular superior com 1's na diagonal, e $D_1^{-1}DD_2$ é uma matriz diagonal, a igualdade $U_1U_2^{-1} = D_1^{-1}DD_2$ implica $U_1U_2^{-1} = I$, isto é,

$$U_1U_2^{-1} = I \Longleftrightarrow U_1 = U_2.$$

Além disso, se $U_1 = U_2$, então $R_1^TR_1 = A^TA = R_2^TR_2$ é equivalente a $U_1^TD_1^2U_1 = U_1^TD_2^2U_1$. Como U_1 é invertível, resulta desta igualdade que $D_1^2 = D_2^2$. Como as entradas na diagonal principal de D_1 e de D_2 são positivas por hipótese, e D_1 e D_2 são matrizes diagonais, a igualdade $D_1^2 = D_2^2$ é equivalente a $D_1 = D_2$. Logo, $R_1 = R_2$. □

Exemplo 5.12. Considere-se a seguinte matriz

$$A = \begin{bmatrix} 1 & 0 \\ 2 & 1 \\ 0 & 1 \end{bmatrix}.$$

As colunas $\mathbf{c_1}$ e $\mathbf{c_2}$ de A são linearmente independentes, uma vez que a matriz A tem característica 2. Apliquemos o processo de Gram-Schmidt ao conjunto dos vectores coluna de A, e normalizemos os vectores obtidos.

$$\begin{aligned} &\mathbf{u}_1 = \mathbf{c_1} = (1,2,0) && \rightsquigarrow \quad \mathbf{q_1} = \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|} = \frac{1}{\sqrt{5}}(1,2,0) \\ &\mathbf{u}_2 = \mathbf{c_2} - \frac{\langle \mathbf{u}_1, \mathbf{c_2} \rangle}{\|\mathbf{u}_1\|^2}\mathbf{u}_1 = \frac{1}{5}(-2,1,5) && \rightsquigarrow \quad \mathbf{q_2} = \frac{\mathbf{u}_2}{\|\mathbf{u}_2\|} = \frac{1}{\sqrt{30}}(-2,1,5). \end{aligned}$$

Como $\langle \mathbf{q_1}, \mathbf{c_1} \rangle = \frac{5}{\sqrt{5}}$, $\langle \mathbf{q_1}, \mathbf{c_2} \rangle = \frac{2}{\sqrt{5}}$ e $\langle \mathbf{q_2}, \mathbf{c_2} \rangle = \frac{6}{\sqrt{30}}$, a factorização $A = QR$ é

$$Q = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & -2/\sqrt{6} \\ 2 & 1/\sqrt{6} \\ 0 & 5/\sqrt{6} \end{bmatrix}, \qquad R = \frac{1}{\sqrt{5}} \begin{bmatrix} 5 & 2 \\ 0 & \sqrt{6} \end{bmatrix}.$$

♦

5.6 Determinantes, produto vectorial e produto misto em $\mathbb{R}^3$

Nesta secção fazemos uma interpretação geométrica do determinante de uma matriz 3×3, e aproveitamos para introduzir alguns conceitos usados com frequência em Cálculo, como seja o produto vectorial e o produto misto.

O produto vectorial de dois vectores de $\mathbb{R}^3$ será definido como sendo um vector de $\mathbb{R}^3$ que goza da propriedade de ser ortogonal ao plano gerado pelos vectores dados (quando esses vectores gerarem um plano).

Definição 5.8. Produto vectorial

Sejam $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3$ com $\mathbf{x} = (x_1, x_2, x_3)$ e $\mathbf{y} = (y_1, y_2, y_3)$. Definimos *produto vectorial*, ou *produto externo*, de $\mathbf{x}$ por $\mathbf{y}$, como sendo o seguinte vector de $\mathbb{R}^3$:

$$\begin{aligned}
\mathbf{x} \times \mathbf{y} &= (x_2y_3 - x_3y_2, x_3y_1 - x_1y_3, x_1y_2 - x_2y_1) \\
&= \left(\begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix}, - \begin{vmatrix} x_1 & x_3 \\ y_1 & y_3 \end{vmatrix}, \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} \right) \\
&= \begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix} \mathbf{e}_1 - \begin{vmatrix} x_1 & x_3 \\ y_1 & y_3 \end{vmatrix} \mathbf{e}_2 + \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} \mathbf{e}_3,
\end{aligned} \tag{5.20}$$

onde $\mathbf{e}_1, \mathbf{e}_2$ e $\mathbf{e}_3$ designam os três vectores da base canónica de $\mathbb{R}^3$.

Por analogia com o desenvolvimento de Laplace para o cálculo do determinante ao longo da primeira linha, é frequente escrever a igualdade (5.20) na forma:

$$\mathbf{x} \times \mathbf{y} = \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix}. \tag{5.21}$$

A expressão (5.21) formalmente não tem sentido (não faz sentido calcular um determinante de uma matriz cujas entradas da primeira linha são vectores e as outras entradas números) mas funciona como uma mnemónica para a expressão que define o produto vectorial.

Relembrando que o determinante de uma matriz muda de sinal quando se trocam duas linhas da matriz, quer usando (5.21) ou (5.20), temos: $\mathbf{x} \times \mathbf{y} = -(\mathbf{y} \times \mathbf{x})$. Ou seja, o produto vectorial é *anticomutativo*.

Definição 5.9. Produto misto

Sendo $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^3$ com $\mathbf{x} = (x_1, x_2, x_3)$, $\mathbf{y} = (y_1, y_2, y_3)$ e $\mathbf{z} = (z_1, z_2, z_3)$, define-se *produto misto*, ou produto *triplo*, de $\mathbf{x}, \mathbf{y}$ e $\mathbf{z}$ (por esta ordem) por:

$$\begin{aligned}
\mathbf{x} \cdot (\mathbf{y} \times \mathbf{z}) &= \langle \mathbf{x} \,, \mathbf{y} \times \mathbf{z} \rangle \\
&= x_1(y_2z_3 - y_3z_2) + x_2(y_3z_1 - y_1z_3) + x_3(y_1z_2 - y_2z_1).
\end{aligned}$$

Como consequência imediata do desenvolvimento de Laplace (ao longo da primeira linha) para o cálculo de um determinante, e da definição de produto vectorial, o produto misto de três vectores é exactamente igual ao determinante de uma matriz cujas linhas são os vectores $\mathbf{x}, \mathbf{y}$ e $\mathbf{z}$. Isto é,

$$\mathbf{x} \cdot (\mathbf{y} \times \mathbf{z}) = \begin{vmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{vmatrix}. \tag{5.22}$$

Das propriedades do determinante de uma matriz, resultam imediatas as seguintes propriedades do produto misto:

$$\begin{aligned} \mathbf{x} \cdot (\mathbf{y} \times \mathbf{z}) &= \mathbf{y} \cdot (\mathbf{z} \times \mathbf{x}) = \mathbf{z} \cdot (\mathbf{x} \times \mathbf{y}) \\ &= -\mathbf{y} \cdot (\mathbf{x} \times \mathbf{z}) = -\mathbf{x} \cdot (\mathbf{z} \times \mathbf{y}) = -\mathbf{z} \cdot (\mathbf{y} \times \mathbf{x}). \end{aligned}$$

Uma propriedade importante do produto vectorial é a seguinte:

O vector $\mathbf{x} \times \mathbf{y}$ é ortogonal a $\mathbf{x}$ e a $\mathbf{y}$. Se $\mathbf{x}$ e $\mathbf{y}$ são linearmente independentes, o vector $\mathbf{x} \times \mathbf{y}$ é perpendicular ao plano gerado por $\mathbf{x}$ e $\mathbf{y}$.

Para mostrar a afirmação anterior, basta notar que $\langle \mathbf{x}, \mathbf{x} \times \mathbf{y} \rangle = \mathbf{x} \cdot (\mathbf{x} \times \mathbf{y})$ corresponde ao determinante de uma matriz com duas linhas iguais, portanto igual a zero. Da mesma forma $\langle \mathbf{y}, \mathbf{x} \times \mathbf{y} \rangle = 0$.

Exercício 5.14. Use as propriedades do determinante de uma matriz para mostrar:

a) $(\mathbf{y} + \mathbf{w}) \times \mathbf{x} = (\mathbf{y} \times \mathbf{x}) + (\mathbf{w} \times \mathbf{x})$.
b) $k(\mathbf{x} \times \mathbf{y}) = (k\mathbf{x}) \times \mathbf{y} = \mathbf{x} \times (k\mathbf{y})$, com k escalar.
c) $\mathbf{x} \times \mathbf{x} = \mathbf{0}$.

▲

Usemos as propriedades do Exercício 5.14 para mostrar a igualdade

$$\mathbf{x} \times \mathbf{y} = (\mathbf{x} - \operatorname{proj}_{\mathbf{y}} \mathbf{x}) \times \mathbf{y}. \tag{5.23}$$

Da definição de projecção ortogonal, $\operatorname{proj}_{\mathbf{y}} \mathbf{x} = \dfrac{\langle \mathbf{y}, \mathbf{x} \rangle}{\|\mathbf{y}\|^2} \mathbf{y}$, e das propriedades a)-c) do exercício anterior, tem-se

$$\begin{aligned} (\mathbf{x} - \operatorname{proj}_{\mathbf{y}} \mathbf{x}) \times \mathbf{y} &\overset{\text{a), b)}}{=} \mathbf{x} \times \mathbf{y} - (\operatorname{proj}_{\mathbf{y}} \mathbf{x}) \times \mathbf{y} \\ &\overset{\text{b)}}{=} \mathbf{x} \times \mathbf{y} - \frac{\langle \mathbf{y}, \mathbf{x} \rangle}{\|\mathbf{y}\|^2} (\mathbf{y} \times \mathbf{y}) \\ &\overset{\text{c)}}{=} \mathbf{x} \times \mathbf{y}. \end{aligned}$$

Exercício 5.15. Mostre que para os vectores $\mathbf{e}_1, \mathbf{e}_2$ e $\mathbf{e}_3$ da base canónica de $\mathbb{R}^3$ se tem:

$$\mathbf{e}_1 \times \mathbf{e}_2 = \mathbf{e}_3, \quad \mathbf{e}_2 \times \mathbf{e}_3 = \mathbf{e}_1, \quad \mathbf{e}_3 \times \mathbf{e}_1 = \mathbf{e}_2.$$

▲

Exercício 5.16. Mostre as propriedades seguintes e conclua que o produto vectorial não é associativo.

a) $\mathbf{x} \times (\mathbf{y} \times \mathbf{z}) = \langle \mathbf{x}\,, \mathbf{z}\rangle \mathbf{y} - \langle \mathbf{x}\,, \mathbf{y}\rangle \mathbf{z}$.
b) $(\mathbf{x} \times \mathbf{y}) \times \mathbf{z} = \langle \mathbf{x}\,, \mathbf{z}\rangle \mathbf{y} - \langle \mathbf{y}\,, \mathbf{z}\rangle \mathbf{x}$.

▲

Para além das propriedades que enunciámos nos exercícios anteriores, referimos em seguida uma propriedade essencial na interpretação geométrica do produto vectorial.

Proposição 5.5. Para quaisquer $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3$ verifica-se a *identidade de Lagrange*:

$$\|\mathbf{x} \times \mathbf{y}\|^2 = \|\mathbf{x}\|^2 \|\mathbf{y}\|^2 - (\langle \mathbf{x}\,,\, \mathbf{y}\rangle)^2\,.$$

Além disso, a igualdade anterior é equivalente a

$$\|\mathbf{x} \times \mathbf{y}\|^2 = \|\mathbf{x}\|^2 \|\mathbf{y}\|^2 \operatorname{sen}^2 \theta,$$

onde θ é o ângulo entre $\mathbf{x}$ e $\mathbf{y}$.

Demonstração. Para mostrar a identidade de Lagrange[6] basta calcular $\|\mathbf{x} \times \mathbf{y}\|^2$ e $\|\mathbf{x}\|^2 \|\mathbf{y}\|^2 - (\langle \mathbf{x}\,,\, \mathbf{y}\rangle)^2$. Para tal, considere-se $\mathbf{x} = (x_1, x_2, x_3)$ e $\mathbf{y} = (y_1, y_2, y_3)$.

$$\|\mathbf{x} \times \mathbf{y}\|^2 = (x_2 y_3 - x_3 y_2)^2 + (x_3 y_1 - x_1 y_3)^2 + (x_1 y_2 - x_2 y_1)^2$$

e

$$\|\mathbf{x}\|^2 \|\mathbf{y}\|^2 - (\langle \mathbf{x}\,,\, \mathbf{y}\rangle)^2 = (x_1^2 + x_2^2 + x_3^2)(y_1^2 + y_2^2 + y_3^2) - (x_1 y_1 + x_2 y_2 + x_3 y_3)^2.$$

Desenvolvendo os membros direitos das duas expressões anteriores e comparando os resultados, confirma-se a igualdade dessas expressões.

[6]Joseph Louis Lagrange (1736 — 1813), matemático francês.

Da identidade de Lagrange e da definição de ângulo entre os vectores $\mathbf{x}$ e $\mathbf{y}$, resulta

$$\begin{aligned}\|\mathbf{x}\|^2\|\mathbf{y}\|^2 - (\langle \mathbf{x}\,,\,\mathbf{y}\rangle)^2 &= \|\mathbf{x}\|^2\|\mathbf{y}\|^2 - \|\mathbf{x}\|^2\|\mathbf{y}\|^2\cos^2\theta = \|\mathbf{x}\|^2\|\mathbf{y}\|^2(1-\cos^2\theta)\\ &= \|\mathbf{x}\|^2\|\mathbf{y}\|^2\operatorname{sen}^2\theta.\end{aligned}$$

□

Nota 33. *A igualdade $\|\mathbf{x}\times\mathbf{y}\|^2 = \|\mathbf{x}\|^2\|\mathbf{y}\|^2\operatorname{sen}^2\theta$ diz-nos que se $\mathbf{x}$ e $\mathbf{y}$ são vectores não nulos, o vector $\mathbf{x}\times\mathbf{y}$ é nulo se e só se $\theta = 0$ ou $\theta = \pi$. Ou seja, para $\mathbf{x}\neq\mathbf{0}$ e $\mathbf{y}\neq\mathbf{0}$, o vector $\mathbf{x}\times\mathbf{y} = \mathbf{0}$ se e só se $\mathbf{x}$ e $\mathbf{y}$ têm a mesma direcção (isto é, se e só se $\mathbf{x}$ e $\mathbf{y}$ são colineares).*

Exercício 5.17. Mostre que o determinante de uma matriz A cujas linhas são os vectores $\mathbf{u}, \mathbf{v}$ e $\mathbf{u}\times\mathbf{v}$, por esta ordem, é sempre positivo, nomeadamente $\det(A) = \|\mathbf{u}\times\mathbf{v}\|^2$. Deduza ainda que se as linhas de A são os vectores $\mathbf{u}, \mathbf{v}$ e $-(\mathbf{u}\times\mathbf{v})$, por esta ordem, então $\det(A)$ é negativo.
Sugestão: Use a expressão (5.22). ▲

Vimos que $\mathbf{u}\times\mathbf{v}$ é ortogonal ao plano gerado por $\mathbf{u}$ e $\mathbf{v}$, e que o seu comprimento é dado pela identidade de Lagrange. No caso de $\mathbf{u}$ e $\mathbf{v}$ serem linearmente independentes, os três vectores $\mathbf{u}, \mathbf{v}$ e $\mathbf{u}\times\mathbf{v}$ são linearmente independentes e podemos usar estes vectores para definir um referencial em $\mathbb{R}^3$. O sentido de $\mathbf{u}\times\mathbf{v}$ depende da orientação desse referencial.

A chamada regra da mão direita[7] é um processo prático para decidir a orientação positiva de um referencial em $\mathbb{R}^3$. Um terno ordenado $(\mathbf{u}, \mathbf{v}, \mathbf{w})$ de vectores linearmente independentes de $\mathbb{R}^3$ diz-se um *triedro directo* , ou uma *base positiva* de $\mathbb{R}^3$, se está de acordo com a regra da mão direita, isto é, se se pode dispor os dedos indicador e médio da mão direita por forma a que apontem, respectivamente, no sentido de $\mathbf{u}$ e de $\mathbf{v}$, então o dedo polegar apontará no sentido de $\mathbf{w}$.

Considerando os vectores da base canónica de $\mathbb{R}^3$, são exemplos de triedros directos:

$$(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)\,, \quad (\mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_1)\,, \quad (\mathbf{e}_3, \mathbf{e}_1, \mathbf{e}_2)\,.$$

É fácil verificar que uma base ordenada $(\mathbf{e}_i, \mathbf{e}_j, \mathbf{e}_k)$, formada por vectores da base canónica de $\mathbb{R}^3$, é um triedro directo se e só se o determinante da matriz cujas colunas são os vectores $\mathbf{e}_i, \mathbf{e}_j, \mathbf{e}_k$, por esta ordem, tiver o valor 1. Fixada uma

[7] A regra da mão direita foi introduzida pelo engenheiro electrónico e físico inglês, Sir John Ambrose Fleming (1849 – 1945).

base ordenada $B = (\mathbf{e}_i, \mathbf{e}_j, \mathbf{e}_k)$ qualquer outra base ordenada $(\mathbf{u}, \mathbf{v}, \mathbf{w})$ de $\mathbb{R}^3$ está relacionada com a base B através da matriz de mudança de base M. É natural considerar-se que uma base ordenada $(\mathbf{u}, \mathbf{v}, \mathbf{w})$ é positiva se $\det M > 0$, e que é negativa se $\det M < 0$.

Definição 5.10. Triedro directo

Uma base ordenada $(\mathbf{u}, \mathbf{v}, \mathbf{w})$ de $\mathbb{R}^3$, diz-se um *triedro directo* ou uma *base positiva*, se o determinante da matriz cujas colunas (ou linhas) são os vectores $\mathbf{u}, \mathbf{v}, \mathbf{w}$, por esta ordem, é positivo.

Se $\mathbf{u}$ e $\mathbf{v}$ são vectores linearmente independentes de $\mathbb{R}^3$, o vector $\mathbf{u} \times \mathbf{v}$ é ortogonal a $\mathbf{u}$ e a $\mathbf{v}$, pelo que $B = (\mathbf{u}, \mathbf{v}, \mathbf{u} \times \mathbf{v})$ é uma base de $\mathbb{R}^3$. Seja A a matriz cujas colunas são os vectores $\mathbf{u}, \mathbf{v}$ e $\mathbf{u} \times \mathbf{v}$, por esta ordem. Da expressão (5.22), tem-se que $\det A = \|\mathbf{u} \times \mathbf{v}\|^2$, e portanto $(\mathbf{u}, \mathbf{v}, \mathbf{u} \times \mathbf{v})$ é uma base positiva.

Interpretação geométrica do determinante de matrizes 3×3

Nesta secção relacionamos os conceitos de produto vectorial e produto misto com os conceitos de área e volume da geometria elementar.

Considere-se um paralelogramo definido por dois vectores $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3$, como se indica na Figura 5.8.

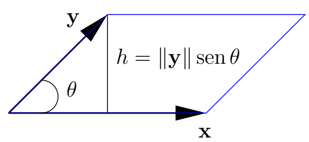


Figura 5.8: Paralelogramo definido por dois vectores $\mathbf{x}$ e $\mathbf{y}$ fazendo um ângulo θ entre si.

Tomando como base do paralelogramo o lado definido por $\mathbf{x}$, a altura h do paralelogramo é igual à norma do vector $(\mathbf{y} - \operatorname{proj}_{\mathbf{x}} \mathbf{y})$. Como exercício pode verificar que

$$h = \|\mathbf{y} - \operatorname{proj}_{\mathbf{x}} \mathbf{y}\| = \|\mathbf{y}\| \operatorname{sen} \theta,$$

onde θ é o ângulo entre $\mathbf{x}$ e $\mathbf{y}$.

A área $\mathcal{A}$ do paralelogramo é igual ao comprimento da base vezes a altura. Isto é,

$$\mathcal{A} = \|\mathbf{x}\| h.$$

Logo,

$$\mathcal{A} = \|\mathbf{x}\|\|\mathbf{y}\| \operatorname{sen}\theta. \tag{5.24}$$

Usando a identidade de Lagrange (Proposição 5.5), concluimos que a norma do produto vectorial de $\mathbf{x}$ por $\mathbf{y}$ é igual à área do paralelogramo definido por $\mathbf{x}$ e $\mathbf{y}$. Note-se que o ângulo θ varia entre 0 e π, e portanto $\operatorname{sen}\theta > 0$.

Exercício 5.18. Mostre que o valor dado pela fórmula (5.24) não depende da aresta que se escolhe para base do paralelogramo. ▲

A área $\mathcal{A}$ de um paralelogramo definido por dois vectores não nulos $\mathbf{x}$ e $\mathbf{y}$ de $\mathbb{R}^3$ é igual à norma do produto vectorial de $\mathbf{x}$ por $\mathbf{y}$:

$$\mathcal{A} = \|\mathbf{x} \times \mathbf{y}\|.$$

A área de um paralelogramo definido por dois vectores $\mathbf{x} = (x_1, x_2, 0)$ e $\mathbf{y} = (y_1, y_2, 0)$, do plano xy, de $\mathbb{R}^3$ é

$$\begin{aligned}\mathcal{A} = \|\mathbf{x} \times \mathbf{y}\| = \|(0, 0, x_1y_2 - x_2y_1)\| = \sqrt{(x_1y_2 - x_2y_1)^2} = |x_1y_2 - x_2y_1| \\ = \left|\det\left(\begin{bmatrix} x_1 & x_2 \\ y_1 & y_2 \end{bmatrix}\right)\right|.\end{aligned}$$

Exercício 5.19. Determine a área do triângulo cujos vértices são $V_1 = (1, -1, 2)$, $V_2 = (0, 4, 3)$ e $V_3 = (1, 0, 1)$.

▲

Consideremos agora o paralelipípedo definido pelos vectores $\mathbf{x}, \mathbf{y}$ e $\mathbf{z}$ (Figura 5.9).

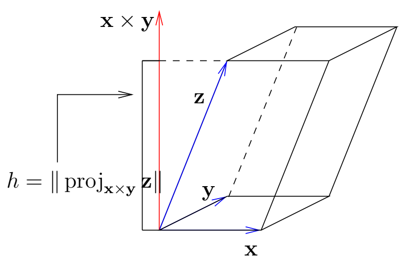


Figura 5.9: Paralelipípedo definido pelos vectores $\mathbf{x}$,$\mathbf{y}$ e $\mathbf{z}$.

O volume deste paralelipípedo é igual à área da base vezes a altura h. Tomando como base o paralelogramo definido por $\mathbf{x}$ e $\mathbf{y}$, a área da base é igual a $\|\mathbf{x}\times\mathbf{y}\|$, enquanto que a altura é igual à norma da projecção ortogonal de $\mathbf{z}$ sobre $\mathbf{x}\times\mathbf{y}$ (o vector $\mathbf{x}\times\mathbf{y}$ é ortogonal ao plano gerado por $\mathbf{x}$ e $\mathbf{y}$). Assim, o volume $\mathcal{V}$ é dado por:

$$\begin{aligned}\mathcal{V} = \|\mathbf{x}\times\mathbf{y}\|h &= \|\mathbf{x}\times\mathbf{y}\|\|\operatorname{proj}_{\mathbf{x}\times\mathbf{y}}\mathbf{z}\| \\ &= \|\mathbf{x}\times\mathbf{y}\|\left\|\frac{\langle\mathbf{x}\times\mathbf{y}\,,\,\mathbf{z}\rangle}{\|\mathbf{x}\times\mathbf{y}\|^2}(\mathbf{x}\times\mathbf{y})\right\| \\ &= \|\mathbf{x}\times\mathbf{y}\|\frac{|\langle\mathbf{x}\times\mathbf{y}\,,\,\mathbf{z}\rangle|}{\|\mathbf{x}\times\mathbf{y}\|^2}\|\mathbf{x}\times\mathbf{y}\| \qquad \text{(ver Exercício 5.4)} \\ &= |\langle\mathbf{x}\times\mathbf{y}\,,\,\mathbf{z}\rangle| = |\mathbf{z}\cdot(\mathbf{x}\times\mathbf{y})|\,.\end{aligned}$$

Deixamos como exercício verificar que o volume de um paralelipípedo é sempre dado pela expressão acima, independentemente de qual das faces se considera como base.

Conclui-se que o volume de um paralelipípedo gerado por três vectores linearmente independentes $\mathbf{x}, \mathbf{y}$ e $\mathbf{z}$ de $\mathbb{R}^3$, é igual ao módulo do determinante da matriz cujas linhas são os vectores $\mathbf{x}, \mathbf{y}$ e $\mathbf{z}$. Ou seja,

$$\mathcal{V} = |\mathbf{z}\cdot(\mathbf{x}\times\mathbf{y})| = \left|\det\begin{bmatrix} — & \mathbf{x} & — \\ — & \mathbf{y} & — \\ — & \mathbf{z} & — \end{bmatrix}\right|.$$

Na igualdade anterior usou-se o facto do produto misto de três vectores ser igual ao determinante da matriz 3×3 cujas linhas são esses vectores. Resumindo:

- A área do paralelogramo definido pelos vectores $\mathbf{x} = (x_1, x_2)$ e $\mathbf{y} = (y_1, y_2)$ é
$$\mathcal{A} = \left|\det\begin{bmatrix} x_1 & x_2 \\ y_1 & y_2\end{bmatrix}\right|.$$
- O volume do paralelipípedo definido pelos vectores $\mathbf{x} = (x_1, x_2, x_3), \mathbf{y} = (y_1, y_2, y_3)$ e $\mathbf{z} = (z_1, z_2, z_3)$ é
$$\mathcal{V} = \left|\det\begin{bmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3\end{bmatrix}\right|.$$

Para terminar esta secção refira-se que as noções de paralelipípedo e de volume em $\mathbb{R}^3$ podem ser generalizadas a $\mathbb{R}^n$. Um paralelipípedo em $\mathbb{R}^3$ definido pelos vectores linearmente independentes $\mathbf{x}, \mathbf{y}$ e $\mathbf{z}$ é o conjunto de todas as combinações lineares destes três vectores com coeficientes variando no intervalo $[0, 1]$. Define-se um paralelipípedo em $\mathbb{R}^n$, gerado pelos vectores linearmente independentes $\mathbf{u_1}, \ldots, \mathbf{u}_n$, como sendo o conjunto:

$$\{\alpha_1 \mathbf{u}_1 + \cdots + \alpha_n \mathbf{u}_n : \quad 0 \le \alpha_1 \le 1, \ldots, 0 \le \alpha_n \le 1\}.$$

A noção de volume de um paralelipípedo é generalizada de forma óbvia sob a designação de *medida*. A medida $\mathcal{M}$ de um paralelipípedo de $\mathbb{R}^n$ é o módulo do determinante da matriz cujas linhas (ou colunas) são n vectores $\mathbf{u_1}, \ldots, \mathbf{u}_n$, geradores do paralelipípedo. Isto é,

$$\mathcal{M} = \left| \det \begin{bmatrix} — & \mathbf{u}_1 & — \\ — & \mathbf{u}_2 & — \\ & \vdots & \\ — & \mathbf{u}_n & — \end{bmatrix} \right|.$$

5.7 Aplicações

Nesta secção abordamos as seguintes aplicações de conceitos estudados neste capítulo: (i) equações de rectas e planos em $\mathbb{R}^3$; (ii) mínimos quadrados; (iii) transformada de Fourier discreta. No cerne destas aplicações estão várias propriedades, anteriormente estudadas, de um espaço linear munido de um produto interno.

5.7.1 Equações de rectas e planos em $\mathbb{R}^3$

Comecemos por relembrar que no Capítulo 1 estabelecemos uma correspondência biunívoca entre pontos e vectores do espaço tridimensional. Esta correspondência associa a um ponto P do espaço o vector de posição $\mathbf{p} = \overrightarrow{OP}$, isto é, o vector aplicado na origem O. As coordenadas do ponto P são as coordenadas do respectivo vector de posição $\mathbf{p}$.

No Capítulo 3, vimos que em $\mathbb{R}^3$ uma recta que passe pela origem, ou um plano que contenha a origem, são subespaços de dimensão, respectivamente, 1 e 2. Isto quer dizer que uma recta que passa pela origem é gerada por um vector $\mathbf{u}$ (não nulo), que determina a direcção da recta, enquanto que um plano que passe pela origem é gerado por dois vectores $\mathbf{u}$ e $\mathbf{v}$ linearmente independentes. O conjunto gerado por um vector é o conjunto $\{t\mathbf{u} : t \in \mathbb{R}\}$, de todos os múltiplos do vector, enquanto o conjunto gerado por dois vectores $\mathbf{u}$ e $\mathbf{v}$ é o conjunto de todas as combinações lineares desses vectores, ou seja $\{t\mathbf{u} + s\mathbf{v} : t, s \in \mathbb{R}\}$. Assim, uma recta que passa pela origem e tem a direcção do vector $\mathbf{u}$ é o conjunto

$$\left\{\mathbf{x} \in \mathbb{R}^3 : \mathbf{x} = t\mathbf{u}, \quad t \in \mathbb{R}\right\},$$

e um plano que contém a origem e os vectores (não colineares) $\mathbf{u}$ e $\mathbf{v}$ é o conjunto

$$\left\{\mathbf{x} \in \mathbb{R}^3 : \mathbf{x} = t\mathbf{u} + s\mathbf{v}, \quad t, s \in \mathbb{R}\right\}.$$

As rectas que não passam pela origem (resp. os planos que não contêm a origem) não são subespaços de $\mathbb{R}^3$, no entanto estas rectas (resp. estes planos) podem ser obtidos por translação de uma recta paralela (resp. um plano paralelo) que passa pela origem. De facto, uma recta que passa por um ponto P_0 e tem a direcção do vector $\mathbf{u}$ é o conjunto dos pontos X tais que os vectores $\overrightarrow{P_0X}$ é colinear com $\mathbf{u}$, isto é, $\overrightarrow{P_0X} = t\mathbf{u}$, com $t \in \mathbb{R}$. Como $\overrightarrow{P_0X} = \overrightarrow{OX} - \overrightarrow{OP_0} = \mathbf{x} - \mathbf{p}_0$ (ver Capítulo 1, pág. 28), tem-se que a recta que passa por P_0 é o conjunto

$$\{\mathbf{x} \in \mathbb{R}^3 : \mathbf{x} = \mathbf{p}_0 + t\mathbf{u},\ t \in \mathbb{R}\} = \mathbf{p}_0 + \mathrm{Span}\{\mathbf{u}\}.$$

Ou seja, a recta que passa por P_0 e tem a direcção de $\mathbf{u}$ é a translação da recta $\mathrm{Span}\{\mathbf{u}\}$ pelo vector $\mathbf{p}_0$.

De igual modo, um plano que contém o ponto P_0 pode ser obtido por translação de um plano paralelo que passa na origem, gerado pelos vectores (linearmente independentes) $\mathbf{u}$ e $\mathbf{v}$. Ou seja, um plano que contém P_0 e é paralelo ao plano gerado pelos vectores $\mathbf{u}$ e $\mathbf{v}$ é o conjunto de pontos

$$\{\mathbf{x} \in \mathbb{R}^3 : \mathbf{x} = \mathbf{p}_0 + t\mathbf{u} + s\mathbf{v},\ t, s \in \mathbb{R}\} = \mathbf{p}_0 + \mathrm{Span}\{\mathbf{u}, \mathbf{v}\}.$$

Na Figura 5.10 ilustram-se as translações de rectas e planos por um vector $\mathbf{p}_0$.

As equações $\mathbf{x} = \mathbf{p}_0 + t\mathbf{u}$ e $\mathbf{x} = \mathbf{p}_0 + t\mathbf{u} + s\mathbf{v}$ que definem, respectivamente, uma recta e um plano que passam por P_0, são designadas por *equações vectoriais*. Quando nestas equações se explicitam as coordenadas dos vectores envolvidos, essas equações recebem a designação de *equações paramétricas*.

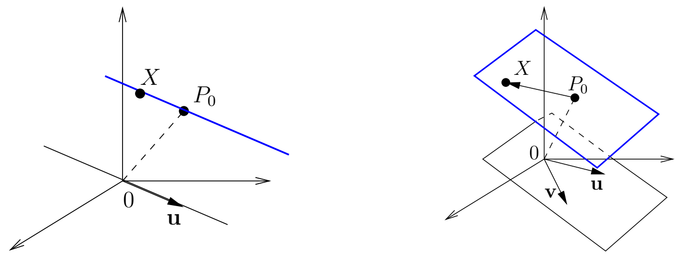


Figura 5.10: Plano e recta que passam por P_0 como translações de um plano e de uma recta que passam pela origem.

- As equações *paramétricas* de uma recta em $\mathbb{R}^3$ que passa por $P_0 = (x_0, y_0, z_0)$ e tem a direcção do vector não nulo $\mathbf{u} = (u_1, u_2, u_3)$ são:

$$\mathbf{x} = \mathbf{p}_0 + t\mathbf{u} \iff \begin{cases} x = x_0 + tu_1 \\ y = y_0 + tu_2 \\ z = z_0 + tu_3 \end{cases} \quad t \in \mathbb{R}.$$

- As equações *paramétricas* de um plano em $\mathbb{R}^3$ que passa por $P_0 = (x_0, y_0, z_0)$ e é paralelo ao plano gerado pelos vectores linearmente independentes $\mathbf{u} = (u_1, u_2, u_3)$ e $\mathbf{v} = (v_1, v_2, v_3)$ são:

$$\mathbf{x} = \mathbf{p}_0 + t\mathbf{u} + s\mathbf{v} \iff \begin{cases} x = x_0 + tu_1 + sv_1 \\ y = y_0 + tu_2 + sv_2 \\ z = z_0 + tu_3 + sv_3 \end{cases} \quad t, s \in \mathbb{R}.$$

Exemplo 5.13.

1. As equações paramétricas de uma recta, com a direcção do vector $\mathbf{u} = (-1, 1, 2)$ e que passa pelo ponto $P_0 = (1, -2, 3)$, são

$$x = 1 - t, \quad y = -2 + t, \quad z = 3 + 2t, \quad t \in \mathbb{R}.$$

2. As equações paramétricas de um plano que passa pelo ponto $P_0 = (1, -2, 3)$ e é paralelo ao plano gerado pelos vectores $\mathbf{u} = (1, 1, 0)$ e $\mathbf{v} = (2, 0, 1)$, são

$$x = 1 + t + 2s, \quad y = -2 + t, \quad z = 3 + s, \quad t, s \in \mathbb{R}.$$

♦

É frequente pretender-se determinar uma equação de um plano conhecendo um ponto $P_0 = (x_0, y_0, z_0)$ do plano e uma direcção normal $\mathbf{n} = (a, b, c) \neq (0, 0, 0)$ (observe a Figura 5.11). Dizer que o vector $\mathbf{n}$ é normal ao plano significa que

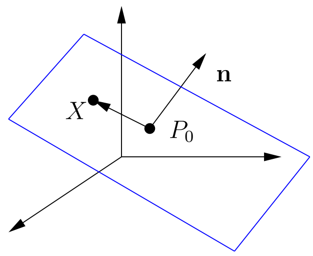


Figura 5.11: O vector $\mathbf{n}$ é normal ao plano que passa por P_0.

qualquer vector do plano é ortogonal a $\mathbf{n}$. A condição do vector genérico $\overrightarrow{P_0X}$ do plano ser ortogonal a $\mathbf{n}$ é traduzida pela equação

$$\langle \overrightarrow{P_0X},\, \mathbf{n} \rangle = 0 \iff \langle \mathbf{x} - \mathbf{p}_0,\, \mathbf{n} \rangle = 0.$$

Em coordenadas, a equação anterior reduz-se a

$$\begin{aligned} \langle (x - x_0, y - y_0, z - z_0)\,,\,(a, b, c) \rangle = 0 &\iff a(x - x_0) + b(y - y_0) + c(z - z_0) = 0 \\ &\iff ax + by + cz = d, \end{aligned}$$

onde $d = ax_0 + by_0 + cz_0$.

Uma *equação cartesiana* de um plano que passa por $P_0 = (x_0, y_0, z_0)$ e é perpendicular a $\mathbf{n} = (a, b, c)$ é

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0,$$

ou, equivalentemente,

$$ax + by + cz = d, \qquad \text{com } \; d = ax_0 + by_0 + cz_0.$$

Exemplo 5.14. No Exemplo 5.13 determinámos equações paramétricas de um plano passando pelo ponto $P_0 = (1, -2, 3)$ e contendo os vectores $\mathbf{u} = (1, 1, 0)$ e $\mathbf{v} = (2, 0, 1)$. As equações paramétricas obtidas foram: $x = 1 + t + 2s$, $y = -2 + t$ e $z = 3 + s$.

Eliminando os parâmetros s, t nestas equações (isto é, determinando uma equação que não envolva t e s) podemos determinar uma equação cartesiana do plano. Assim, substituindo os valores $t = y + 2$ e $s = z - 3$ na igualdade para x, obtemos para equação do plano

$$x = 1 + y + 2 + 2z - 6 \iff x - y - 2z = -3.$$

Esta equação diz-nos que o vector $\mathbf{n} = (1, -1, -2)$ é perpendicular ao plano. Uma vez que são conhecidos dois vectores $\mathbf{u}$ e $\mathbf{v}$ geradores do plano, podemos confirmar que $\mathbf{n}$ é normal ao plano calculando o vector $\mathbf{u} \times \mathbf{v}$ o qual é necessariamente um múltiplo do vector $\mathbf{n}$. Calculando este vector obtemos $\mathbf{u} \times \mathbf{v} = (1, -1, -2)$. ♦

Sabemos que três pontos não colineares, P_0, P_1 e P_2, definem um plano. Para determinar uma equação do plano definido por esses pontos, note-se que os três pontos definem dois vectores (linearmente independentes) do plano, por exemplo, $\mathbf{u} = \overrightarrow{P_0P_1}$ e $\mathbf{v} = \overrightarrow{P_0P_2}$ (veja Figura 5.12). Assim, podemos obter equações paramétricas e cartesianas, para o plano, respectivamente das seguintes equações:

- $\mathbf{x} = \mathbf{p}_0 + t\mathbf{u} + s\mathbf{v} = \mathbf{p}_0 + t(\mathbf{p}_1 - \mathbf{p}_0) + s(\mathbf{p}_2 - \mathbf{p}_0)$.
- $\langle \mathbf{x} - \mathbf{p}_0 \,,\, \mathbf{u} \times \mathbf{v} \rangle = 0$.

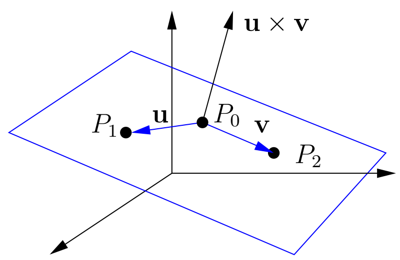


Figura 5.12: o vector $\mathbf{n} = \mathbf{u} \times \mathbf{v}$ é normal ao plano que passa por P_0.

Exemplo 5.15. Determinar uma equação cartesiana do plano Π definido pelos pontos $P_0 = (1, -2, 3)$, $P_1 = (0, 3, 0)$ e $P_2 = (-1, 0, 1)$.

Os vectores $\mathbf{u} = \mathbf{p}_1 - \mathbf{p}_0 = (-1, 5, -3)$ e $\mathbf{v} = \mathbf{p}_2 - \mathbf{p}_0 = (-2, 2, -2)$ definem o plano. O vector $\mathbf{u} \times \mathbf{v} = (-4, 4, 8)$ é ortogonal a Π. Logo, sendo $X = (x, y, z)$

um ponto genérico de Π, uma equação cartesiana é

$$\begin{aligned}\langle \mathbf{x} - \mathbf{p}_0 \, , \, \mathbf{u} \times \mathbf{v} \rangle = 0 &\Longleftrightarrow \langle (x-1, y+2, z-3) \, , \, (-4, 4, 8) \rangle = 0 \\ &\Longleftrightarrow -4x + 4y + 8z = 12.\end{aligned}$$

A equação obtida é equivalente à equação do plano do Exemplo 5.14 (os três pontos P_0, P_1 e P_2 pertencem a esse plano). ♦

Vejamos agora como obter equações cartesianas da recta que passa por $P_0 = (x_0, y_0, z_0)$ e possui a direcção do vector $\mathbf{u}$. Esta recta é a translação de uma recta paralela que passa na origem, ou seja, é a soma $\mathbf{p}_0 + \mathrm{Span}\{\mathbf{u}\} = \mathbf{p}_0 + S$. O complemento ortogonal do subespaço S é um plano, gerado por dois vectores (linearmente independentes) $\mathbf{v}, \mathbf{w}$, e portanto a recta é o conjunto de pontos $X = (x, y, z)$ tais que $\overrightarrow{P_0X}$ é ortogonal a $\mathbf{v}$ e a $\mathbf{w}$. Assim, a recta é o conjunto de pontos X que satisfazem as equações:

$$\begin{cases} \langle \overrightarrow{P_0X} \, , \, \mathbf{v} \rangle & = \langle (x - x_0, y - y_0, z - z_0), \mathbf{v} \rangle = 0 \\ \langle \overrightarrow{P_0X} \, , \, \mathbf{w} \rangle & = \langle (x - x_0, y - y_0, z - z_0), \mathbf{w} \rangle = 0. \end{cases}$$

Calculando os produtos internos nas expressões anteriores, temos

Uma recta que passa por $P_0 = (x_0, y_0, z_0)$ e é perpendicular aos vectores $\mathbf{v} = (v_1, v_2, v_3)$ e $\mathbf{w} = (w_1, w_2, w_3)$ tem *equações cartesianas*:

$$\begin{cases} v_1(x - x_0) + v_2(y - y_0) + v_3(z - z_0) = 0 \\ w_1(x - x_0) + w_2(y - y_0) + w_3(z - z_0) = 0. \end{cases}$$

Estas equações reescrevem-se na forma

$$\begin{cases} v_1x + v_2y + v_3z = d_1 \\ w_1x + w_2y + w_3z = d_2. \end{cases}$$

Exemplo 5.16. Determinemos equações cartesianas da recta que passa por $P_0 = (1, -2, 3)$ e por $P_1 = (2, -3, 1)$. Esta recta tem a direcção do vector $\mathbf{u} = \overrightarrow{P_0P_1} = (1, -1, -2)$, ou seja, a recta é o conjunto de pontos $\mathbf{p}_0 + \mathrm{Span}\{\mathbf{u}\} = \mathbf{p}_0 + S$. O complemento ortogonal $S^\perp$ é o conjunto dos vectores (a, b, c) ortogonais a $\mathbf{u}$, isto é,

$$\langle (a, b, c) \, , \, (1, -1, -2) \rangle = 0 \Longleftrightarrow a - b - 2c = 0 \Longleftrightarrow a = b + 2c.$$

Assim,

$$S^\perp = \{(b + 2c, b, c) : \; b, c \in \mathbb{R}\} = \mathrm{Span}\,\{(1, 1, 0), (2, 0, 1)\}\,.$$

Logo, tomando $\mathbf{v} = (1, 1, 0)$ e $\mathbf{w} = (2, 0, 1)$ obtemos para equações cartesianas da recta:

$$\begin{cases} \langle \mathbf{x} - \mathbf{p}_0, \mathbf{v} \rangle = 0 \\ \langle \mathbf{x} - \mathbf{p}_0, \mathbf{w} \rangle = 0 \end{cases} \iff \begin{cases} x + y = -1 \\ 2x + z = 5. \end{cases}$$

No Exemplo 5.13 determinámos equações paramétricas da recta que passa pelo ponto P_0 e tem a direcção do vector $\mathbf{u}$. Eliminando o parâmetro t nessas equações ($x = 1 - t, y = -2 + t$ e $z = 3 + 2t$) obtêm-se equações cartesianas para a recta, equivalentes às equações acima. ♦

Vimos que uma recta de $\mathbb{R}^3$ é o conjunto de pontos $\mathbf{p}_0 + \mathrm{Span}\{\mathbf{u}\}$ ($\mathbf{u} \neq \mathbf{0}$), enquanto que um plano é o conjunto $\mathbf{p}_0 + \mathrm{Span}\{\mathbf{u}, \mathbf{v}\}$, onde $\mathbf{u}$ e $\mathbf{v}$ são linearmente independentes. Ou seja, uma recta é a soma de $\mathbf{p}_0$ com um subespaço S de dimensão 1, e um plano é a soma de $\mathbf{p}_0$ com um subespaço S de dimensão 2.

No caso da recta $\mathbf{p}_0 + S$, a dimensão de $S^\perp$ é dois, e portanto número de equações cartesianas necessárias para definir a recta é igual a 2 (estas equações descrevem o conjunto de vectores X tais que $\overrightarrow{P_0X}$ é ortogonal a cada vector de uma base de $S^\perp$). No caso de um plano $\mathbf{p}_0 + S$, a dimensão de $S^\perp$ é 1, e portanto o número de equações (cartesianas) que define o plano é 1.

Estas noções podem ser generalizadas a espaços lineares de dimensão superior a 3 da seguinte forma:

Definição 5.11. Chama-se *plano de dimensão k*, ou *plano-k* de $\mathbb{R}^n$, a qualquer subconjunto U de $\mathbb{R}^n$ da forma $\mathbf{p}_0 + S$, onde $\mathbf{p}_0 \in \mathbb{R}^n$ e S é um subespaço de $\mathbb{R}^n$ de dimensão k. Diz-se que U é obtido de S por translação segundo o vector $\mathbf{p}_0$. O subespaço S é designado por *subespaço director*, ou *subespaço das direcções*, de U, e $\mathbf{p}_0$ é designado por *vector de translação*.

Um plano-1 é habitualmente designado por *recta*, um plano-2 por *plano*, e um plano-$(n-1)$ de $\mathbb{R}^n$ por *hiperplano*.

Da discussão efectuada para o caso de rectas e planos de $\mathbb{R}^3$, podemos facilmente concluir que um plano-k de $\mathbb{R}^n$, $U = \mathbf{p}_0 + S$, é definido por um sistema de $n - k$ equações lineares da forma $\langle \mathbf{x} - \mathbf{p}_0, \mathbf{v}_i \rangle = 0$, onde $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-k}\}$ é uma base para o complemento ortogonal $S^\perp$ do subespaço S.

Distância de um ponto a um plano de $\mathbb{R}^3$

Consideremos em $\mathbb{R}^3$ um plano $\mathcal{P}$ que passa por P_0, e um ponto Q não pertencente (em geral) ao plano. Pretende-se calcular a distância de Q ao plano $\mathcal{P}$. Geometricamente, esta distância é a medida do segmento de recta de extremos Q

e R, onde R é o ponto do plano que se obtém projectando ortogonalmente Q sobre $\mathcal{P}$ (observe a Figura 5.13).

Se $\mathcal{P} = \mathbf{p}_0 + S$, onde S é um subespaço de dimensão dois de $\mathbb{R}^3$, a distância de Q ao plano $\mathcal{P}$ é a distância $\mathcal{D}$ de $\overrightarrow{P_0Q}$ ao vector de S que lhe está mais próximo, ou seja, a distância de $\overrightarrow{P_0Q}$ ao vector $\operatorname{proj}_S \overrightarrow{P_0Q}$. Usando a Definição 5.2 e o Teorema da decomposição ortogonal, tem-se

$$\mathcal{D} = \operatorname{dist}\left(\overrightarrow{P_0Q}, \operatorname{proj}_S \overrightarrow{P_0Q}\right) = \|\overrightarrow{P_0Q} - \operatorname{proj}_S \overrightarrow{P_0Q}\| = \|\operatorname{proj}_{S^\perp} \overrightarrow{P_0Q}\|.$$

Como a dimensão de $S^\perp$ é igual a 1, a projecção $\operatorname{proj}_{S^\perp} \overrightarrow{P_0Q}$ reduz-se à projecção de $\overrightarrow{P_0Q}$ sobre um vector $\mathbf{n}$ perpendicular ao plano. Isto é, a distância de Q ao plano $\mathcal{P} = \mathbf{p}_0 + S$ é dada por

$$\mathcal{D} = \|\operatorname{proj}_{\mathbf{n}} \overrightarrow{P_0Q}\|.$$

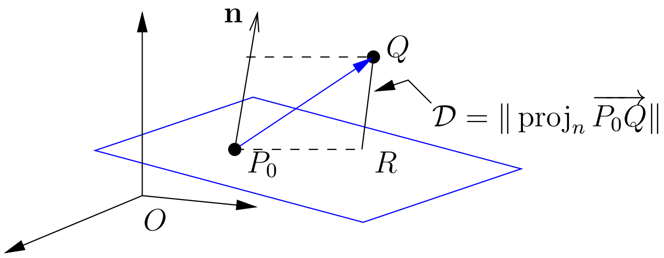


Figura 5.13: Sendo $\mathbf{n}$ um vector normal ao plano, $\mathcal{D}$ designa a distância de Q ao plano.

Suponha-se que é conhecida uma equação cartesiana $a(x-x_0)+b(y-y_0)+c(z-z_0) = 0$ de um plano que passa pelo ponto $P_0 = (x_0, y_0, z_0)$, e que se pretende determinar a distância do ponto $Q = (x_1, y_1, z_1)$ a este plano. Da equação cartesiana do plano sabemos que $\mathbf{n} = (a, b, c)$ é normal ao plano. Por conveniência de cálculo, consideremos a normal unitária $\hat{\mathbf{n}} = \dfrac{1}{\sqrt{a^2+b^2+c^2}}\mathbf{n}$.

A distância do ponto $Q = (x_1, y_1, z_1)$ ao plano é a norma da projecção ortogonal do vector $\overrightarrow{P_0Q}$ sobre $\hat{\mathbf{n}}$. Ou seja, a distância $\mathcal{D}$ pretendida é

$$\begin{aligned}
\mathcal{D} &= \| \operatorname{proj}_{\hat{\mathbf{n}}} \overrightarrow{P_0Q} \| = \| \langle \hat{\mathbf{n}}\,,\, \mathbf{q} - \mathbf{p}_0 \rangle \hat{\mathbf{n}} \| = | \langle \hat{\mathbf{n}}\,,\, \mathbf{q} - \mathbf{p}_0 \rangle | \, \| \hat{\mathbf{n}} \| \\
&= | \langle \hat{\mathbf{n}}\,,\, \mathbf{q} - \mathbf{p}_0 \rangle | = \frac{1}{\sqrt{a^2+b^2+c^2}} \, | \langle (a,b,c)\,,\, (x_1 - x_0, y_1 - y_0, z_1 - z_0) \rangle | \\
&= \frac{|a(x_1 - x_0) + b(y_1 - y_0) + c(z_1 - z_0)|}{\sqrt{a^2+b^2+c^2}}.
\end{aligned}$$

Exemplo 5.17. Considere-se o plano $\mathbf{p}_0 + \operatorname{Span}\{\mathbf{u}, \mathbf{v}\}$, com $\mathbf{p}_0 = (1, -2, 3)$, $\mathbf{u} = (-1, 5, -3)$ e $\mathbf{v} = (2, 2, -2)$. Determinemos a distância do ponto $Q = (1, 0, 1)$ a este plano. Esta distância $\mathcal{D}$ é igual à norma da projecção ortogonal do vector $\overrightarrow{P_0Q} = \mathbf{q} - \mathbf{p}_0 = (0, 2, -2)$ sobre um vector $\mathbf{n}$ ortogonal ao plano. Podemos tomar para $\mathbf{n}$ o vector $\mathbf{u} \times \mathbf{v} = (-4, -8, -12)$. Assim,

$$\begin{aligned}
\mathcal{D} = \| \operatorname{proj}_{(-4,-8,-12)} (0, 2, -2) \| &= \left\| \frac{\langle (-4,-8,-12)\,,\, (0,2,-2) \rangle}{\|(-4,-8,-12)\|^2} (-4,-8,-12) \right\| \\
&= \frac{|-16+24|}{\|(-4,-8,-12)\|} = \frac{8}{\sqrt{224}} = \sqrt{\frac{2}{7}}.
\end{aligned}$$

Como exercício, confirme o resultado obtido calculando a distância $\mathcal{D}$ da seguinte forma: (i) determine uma recta r que passe por Q e seja perpendicular ao plano; (ii) determine o ponto R de intersecção da recta r com o plano; (iii) determine o comprimento do segmento de recta de extremos Q e R. ♦

5.7.2 Mínimos quadrados

É frequentemente necessário determinar uma curva "bem ajustada" a um conjunto de dados obtidos experimentalmente. Por exemplo, suponha que como resultado de uma certa experiência se obteve a seguinte tabela de dados:

$$\begin{array}{c|c|c|c} x_1 & x_2 & \cdots & x_n \\ \hline y_1 & y_2 & \cdots & y_n \end{array} \tag{5.25}$$

Os dados (x_i, y_i) podem representar-se por pontos do plano. Pretende-se encontrar uma curva $y = f(x)$ que "melhor aproxima" os dados (pontos) obtidos. Na Figura 5.14 indicam-se algumas possibilidades.

Comecemos por analisar o caso em que se pretende determinar uma recta de equação $y = ax + b$ que se ajuste bem aos dados da Tabela (5.25). Em Estatística,

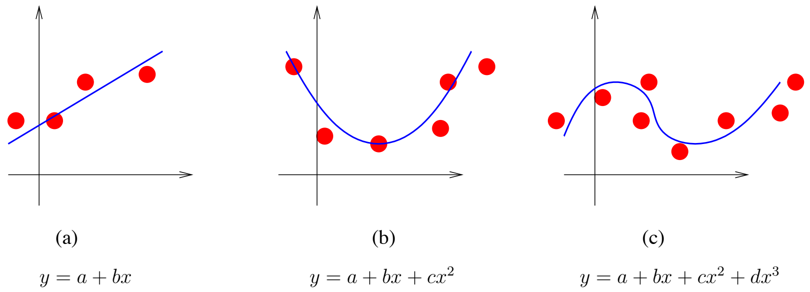


Figura 5.14: Ajuste de curvas a um conjunto de pontos

este problema é conhecido sob a designação de *regressão linear*. Se os pontos (x_i, y_i) forem colineares, os coeficientes a e b de $y = a + bx$ satisfazem o sistema de equações lineares:

$$\begin{cases} y_1 = a + bx_1 \\ y_2 = a + bx_2 \\ \quad\vdots \\ y_n = a + bx_n \end{cases} \iff \mathbf{y} = A\mathbf{x}, \tag{5.26}$$

com

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} a \\ b \end{bmatrix}, \quad \text{e} \quad A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}.$$

Se os pontos (x_i, y_i) não são colineares (como no caso da Figura 5.14-(a)), o sistema (5.26) é impossível. Quando o sistema $A\mathbf{x} = \mathbf{y}$ é impossível, podemos procurar um vector $\hat{\mathbf{x}}$ de modo que a distância de $A\hat{\mathbf{x}}$ a $\mathbf{y}$ seja a menor possível. A distância que aqui consideramos é a distância definida pelo produto interno usual em $\mathbb{R}^n$ (ver Definição 5.2). Pretendemos pois encontrar um vector $\hat{\mathbf{x}}$ tal que $A\hat{\mathbf{x}}$ seja uma (boa) aproximação do vector $\mathbf{y}$.

Definição 5.12. Seja A uma matriz $n \times p$ e $\mathbf{y}$ um vector (fixo) de $\mathbb{R}^n$. Uma *solução de mínimos quadrados* do sistema $A\mathbf{x} = \mathbf{y}$ é um vector $\hat{\mathbf{x}}$ de $\mathbb{R}^p$ tal que

$$\|\mathbf{y} - A\hat{\mathbf{x}}\| \leq \|\mathbf{y} - A\mathbf{x}\|, \tag{5.27}$$

para todo o $\mathbf{x}$ de $\mathbb{R}^p$.

Geometricamente, determinar uma solução de mínimos quadrados $\hat{\mathbf{x}} = (\hat{a}, \hat{b})$ do sistema (5.26) é obter a recta $y = \hat{a} + \hat{b}x$ que melhor aproxima os dados (ver Figura 5.14-(a)).

A designação "mínimos quadrados" (ou *"least-squares"* em inglês) está relacionada com o facto de $\|\mathbf{y} - A\mathbf{x}\|^2$ ser uma soma de quadrados e de se pretender minimizar esta quantidade (ou, equivalentemente, minimizar $\|\mathbf{y} - A\mathbf{x}\|$).

Para encontrar uma solução de mínimos quadrados do sistema $A\mathbf{x} = \mathbf{y}$ iremos aplicar alguns dos resultados de capítulos anteriores. Relembremos alguns resultados básicos essenciais:

(i) Dado um vector $\mathbf{x}$, a expressão $A\mathbf{x}$ é uma combinação linear das colunas de A, ou equivalentemente $A\mathbf{x}$ pertence ao espaço das colunas de A. Isto é, $A\mathbf{x} \in EC(A)$.

(ii) A inequação (5.27) significa que $A\hat{\mathbf{x}}$ está mais próximo de $\mathbf{y}$ que qualquer outro vector do espaço das colunas de A. Pelo Teorema da melhor aproximação (Teorema 5.4, pág. 268), o vector de $EC(A)$ mais próximo de $\mathbf{y}$ é a projecção ortogonal de $\mathbf{y}$ sobre $EC(A)$. Assim,

$$A\hat{\mathbf{x}} = \operatorname{proj}_{EC(A)} \mathbf{y}. \tag{5.28}$$

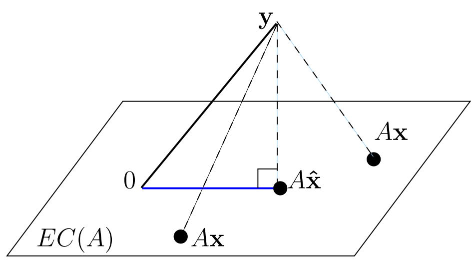


Figura 5.15: O vector $A\hat{\mathbf{x}}$ está mais próximo de $\mathbf{y}$ do que qualquer outro vector $A\mathbf{x} \in EC(A)$.

Nota 34. *A equação* (5.28) *é sempre possível visto que* $A\hat{\mathbf{x}}$ *e* $\operatorname{proj}_{EC(A)} \mathbf{y}$ *pertencem ao espaço das colunas de* A.

O Teorema da decomposição ortogonal (pág. 264) diz-nos que qualquer vector de um espaço linear pode escrever-se (de forma única) como a soma da projecção ortogonal desse vector sobre um subespaço com a projecção ortogonal sobre o seu complemento ortogonal. Assim,

$$\mathbf{y} = \mathrm{proj}_{EC(A)}\,\mathbf{y} + \mathrm{proj}_{(EC(A))^\perp}\,\mathbf{y} \Longrightarrow \mathbf{y} - \mathrm{proj}_{EC(A)}\,\mathbf{y} \overset{(5.28)}{=} \mathbf{y} - A\hat{\mathbf{x}} \in EC(A)^\perp.$$

Logo, o vector $(\mathbf{y} - A\hat{\mathbf{x}})$ é um vector do complemento ortogonal do espaço das colunas de A. Como o complemento ortogonal do espaço das colunas de uma matriz é o núcleo da sua transposta (relembre a igualdade (5.18)), temos:

$$\begin{aligned}(\mathbf{y} - A\hat{\mathbf{x}}) \in (EC(A))^\perp \iff & \quad (\mathbf{y} - A\hat{\mathbf{x}}) \in N(A^T) \iff A^T(\mathbf{y} - A\hat{\mathbf{x}}) = \mathbf{0} \\ \iff & \quad A^T\mathbf{y} - A^T A\hat{\mathbf{x}} = \mathbf{0} \iff A^T A\hat{\mathbf{x}} = A^T\mathbf{y}.\end{aligned} \tag{5.29}$$

Podemos pois concluir que uma solução de mínimos quadrados $\hat{\mathbf{x}}$ satisfaz a seguinte equação.

Equação Normal

$$A^T A\mathbf{x} = A^T\mathbf{y}. \tag{5.30}$$

A equação (5.30) é designada por *equação normal* para o sistema $A\mathbf{x} = \mathbf{y}$.

Acabámos de mostrar que uma solução de mínimos quadrados do sistema $A\mathbf{x} = \mathbf{y}$ satisfaz a equação normal. O resultado recíproco é igualmente válido. De facto, se $\hat{\mathbf{x}}$ é solução da equação normal, tem-se $A^T A\hat{\mathbf{x}} = A^T\mathbf{y}$. Da equivalência (5.29), resulta que $(\mathbf{y} - A\hat{\mathbf{x}}) \in (EC(A))^\perp$. Logo, da unicidade da decomposição ortogonal, temos que $A\hat{\mathbf{x}}$ é necessariamente a projecção ortogonal de $\mathbf{y}$ sobre $EC(A)$. Ou seja, que $\hat{\mathbf{x}}$ é uma solução de mínimos quadrados. No teorema seguinte enunciamos o que acabámos de provar.

Teorema 5.7. O conjunto das soluções de mínimos quadrados do sistema $A\mathbf{x} = \mathbf{y}$ coincide com o conjunto (não vazio) das soluções da equação normal $A^T A\mathbf{x} = A^T\mathbf{y}$.

Exemplo 5.18. Suponhamos que se fizeram as seguintes observações:

x	1	2	3
y	1	2	2

Estas observações correspondem a três pontos no plano, representados na Figura 5.16. Vemos claramente que não existe nenhuma recta $y = a + bx$ que passe

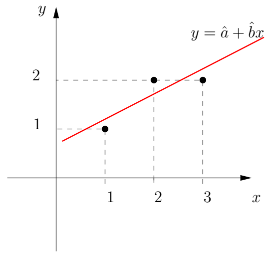


Figura 5.16: Melhor aproximação de mínimos quadrados do Exemplo 5.18.

por estes três pontos. Esta afirmação é equivalente a dizer que é impossível o sistema:

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \iff A\mathbf{x} = \mathbf{y}. \tag{5.31}$$

Pretende-se determinar a recta $y = \hat{a} + \hat{b}x$ que melhor se ajusta aos pontos observados. Isto é, pretende-se determinar um vector $\hat{\mathbf{x}} \in \mathbb{R}^2$ que seja uma solução de mínimos quadrados de $A\mathbf{x} = \mathbf{y}$.

Usando o Teorema 5.7, a equação normal é

$$A^T A\mathbf{x} = A^T\mathbf{y} \iff \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 5 \\ 11 \end{bmatrix}.$$

Como a matriz $A^T A$ é invertível, o sistema de equações normais tem solução única dada por $\hat{\mathbf{x}} = (A^T A)^{-1} A^T\mathbf{y}$. Ou seja,

$$\hat{\mathbf{x}} = \begin{bmatrix} \frac{7}{3} & -1 \\ -1 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 5 \\ 11 \end{bmatrix} = \begin{bmatrix} \frac{2}{3} \\ \frac{1}{2} \end{bmatrix}.$$

Consequentemente, a recta de mínimos quadrados $y = \hat{a} + \hat{b}x$ é

$$y = \frac{2}{3} + \frac{1}{2}x.$$

♦

Sendo $\hat{\mathbf{x}}$ uma solução de mínimos quadrados do sistema $A\mathbf{x} = \mathbf{y}$, o vector $A\hat{\mathbf{x}}$ é uma aproximação de $\mathbf{y}$. Esta aproximação é a melhor aproximação, no sentido em que o *erro de mínimos quadrados*, $\|\mathbf{y} - A\hat{\mathbf{x}}\|$, é o menor possível. O vector $\mathbf{d} = \mathbf{y} - A\hat{\mathbf{x}}$ é denominado por *vector dos desvios*, e portanto o erro de mínimos quadrados é a norma do vector dos desvios.

Para o Exemplo 5.18 calculemos o vector dos desvios, bem como o respectivo erro de mínimos quadrados. Recorde-se que a solução de mínimos quadrados obtida é $\hat{\mathbf{x}} = (2/3, 1/2)$ e que $\mathbf{y} = (1, 2, 2)$. Logo, o vector dos desvios é

$$\mathbf{d} = \mathbf{y} - A\hat{\mathbf{x}} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} - \begin{bmatrix} \frac{7}{6} \\ \frac{5}{3} \\ \frac{13}{6} \end{bmatrix} = \begin{bmatrix} -\frac{1}{6} \\ \frac{1}{3} \\ -\frac{1}{6} \end{bmatrix}.$$

O erro de mínimos quadrados é $\|\mathbf{d}\| = \dfrac{\sqrt{6}}{6}$. Na Figura 5.17 encontram-se representados os desvios, d_1, d_2 e d_3, ou seja as componentes do vector $\mathbf{d} = (d_1, d_2, d_3)$.

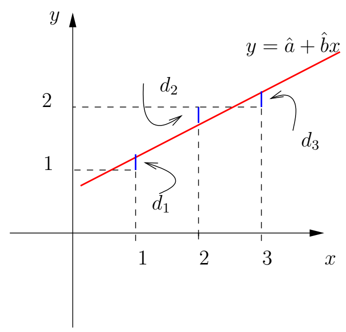


Figura 5.17: Desvios para o Exemplo 5.18.

No Exemplo 5.18 o problema de mínimos quadrados possui solução única, o que nem sempre se verifica. Coloca-se naturalmente a questão de saber em que condições um sistema $A\mathbf{x} = \mathbf{y}$ admite uma única solução de mínimos quadrados. A equação normal é um sistema de equações lineares cuja matriz dos coeficientes $A^T A$ é quadrada, portanto a unicidade de solução de mínimos quadrados é equivalente à existência de inversa da matriz $A^T A$. O resultado que se segue fornece um critério para a unicidade da solução de mínimos quadrados em termos da matriz A.

Teorema 5.8. A matriz $A^T A$ é invertível se e só se as colunas de A são linearmente independentes. Neste caso, o sistema $A\mathbf{x} = \mathbf{y}$ tem solução de mínimos quadrados única, dada por

$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{y}.$$

Demonstração. Quando a matriz $A^T A$ é invertível, da equação normal segue que a solução de mínimos quadrados é dada pela expressão do enunciado. Falta pois mostrar que a matriz $A^T A$ é invertível se e só se as colunas de A são linearmente independentes. A prova deste resultado será realizada em dois passos: (i) mostrar que o núcleo de A é igual ao núcleo de $A^T A$; (ii) usar (i) para provar o resultado.

(i) Mostrar que $N(A) = N(A^T A)$.

Se $\mathbf{x}$ pertence a $N(A)$, isto é $A\mathbf{x} = \mathbf{0}$, o vector $\mathbf{x}$ pertence ao núcleo de $A^T A$, já que

$$A\mathbf{x} = \mathbf{0} \Longrightarrow A^T A\mathbf{x} = \mathbf{0}.$$

Logo, $N(A) \subset N(A^T A)$.

Para a implicação recíproca, suponha-se agora que $\mathbf{x} \in N(A^T A)$, isto é, $A^T A\mathbf{x} = \mathbf{0}$. Multiplicando esta igualdade por $\mathbf{x}^T$, obtém-se

$$A^T A\mathbf{x} = \mathbf{0} \Longrightarrow \mathbf{x}^T A^T A\mathbf{x} = \mathbf{0}.$$

Porém,

$$\mathbf{x}^T A^T A\mathbf{x} = \mathbf{0} \Longleftrightarrow (A\mathbf{x})^T (A\mathbf{x}) = \mathbf{0} \Longleftrightarrow \|A\mathbf{x}\|^2 = \mathbf{0} \Longleftrightarrow A\mathbf{x} = \mathbf{0} \Longleftrightarrow \mathbf{x} \in N(A).$$

Logo, $N(A^T A) \subset N(A)$. Conclui-se portanto que $N(A) = N(A^T A)$.

(ii) Relembre que é condição necessária e suficiente para que uma matriz A tenha colunas linearmente independentes, que a única solução do sistema homogéneo $A\mathbf{x} = \mathbf{0}$ seja a solução nula (ver Proposição 3.14, pág. 152). Assim, a matriz A tem colunas linearmente independentes se e só se $N(A) = \{\mathbf{0}\}$. Como $N(A) = N(A^T A)$, as colunas de $A^T A$ são linearmente independentes se e só se as colunas de A são linearmente independentes.

Além disso, como a matriz $A^T A$ é quadrada, a independência linear das suas colunas é equivalente a dizer que $A^T A$ é invertível. A matriz $A^T A$ é a matriz dos coeficientes do sistema de equações normais, logo a equação normal tem solução única se e só se a matriz $A^T A$ é invertível. $\square$

Nota 35. *No caso de A ser uma matriz quadrada invertível, a solução de mínimos quadrados do sistema $A\mathbf{x} = \mathbf{y}$ reduz-se a $\hat{\mathbf{x}} = A^{-1}(A^T)^{-1}A^T\mathbf{y} = A^{-1}\mathbf{y}$ (usando o Teorema 5.8). Neste caso, o sistema $A\mathbf{x} = \mathbf{y}$ é possível e a solução de mínimos quadrados coincide com a solução do sistema $A\mathbf{x} = \mathbf{y}$.*

A técnica descrita para ajustar uma recta a um conjunto de dados generaliza-se facilmente quando a curva é definida por um polinómio (como nos casos (b) e (c) da Figura 5.14). Por exemplo, suponhamos que se pretende determinar o polinómio $y = a_0 + a_1x + \cdots + a_kx^k$ que melhor se ajusta aos dados da seguinte tabela:

x	x_1	x_2	$\cdots$	x_n
y	y_1	y_2	$\cdots$	y_n

Substituindo em $y = a_0 + a_1x + \cdots + a_kx^k$ os n valores de x e de y tabelados, obtemos n equações lineares:

$$\begin{cases} a_0 + a_1x_1 + \cdots + a_kx_1^k = y_1 \\ a_0 + a_1x_2 + \cdots + a_kx_2^k = y_2 \\ \qquad\cdots\cdots\cdots \\ a_0 + a_1x_n + \cdots + a_kx_n^k = y_n \end{cases} \iff A\mathbf{x} = \mathbf{y}, \quad \text{com } \mathbf{x} = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_k \end{bmatrix}.$$

Resolvendo como anteriormente a equação normal $A^TA\mathbf{x} = A^T\mathbf{y}$, determinamos os coeficientes a_i do polinómio de mínimos quadrados.

Exemplo 5.19. Nos primeiros 5 meses do ano, as vendas (em milhares de euros) de uma certa empresa foram: $2, 2.2, 2.6, 3.2$ e 4. Marcámos estes dados no gráfico que apresentamos na Figura 5.18. A representação gráfica destes dados sugere que as vendas podem ser bem aproximadas por um polinómio de grau 2.

Assim, pretende-se determinar a parábola $y(x) = a_0 + a_1x + a_2x^2$ que melhor se ajusta aos dados, com a finalidade de usarmos esta função para estimar qual o volume de vendas que a empresa irá realizar no último mês do ano (isto é, determinar o valor de $y(12)$).

Vamos portanto determinar a aproximação de mínimos quadrados do sistema

$$\begin{cases} a_0 + a_1 + a_2 = 2 \\ a_0 + 2a_1 + 4a_2 = 2.2 \\ a_0 + 3a_1 + 9a_2 = 2.6 \\ a_0 + 4a_1 + 16a_2 = 3.2 \\ a_0 + 5a_1 + 25a_2 = 4 \end{cases} \iff \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \end{bmatrix} \mathbf{x} = \mathbf{y},$$

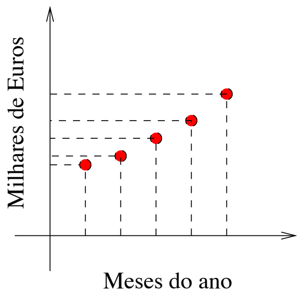


Figura 5.18: Dados do Exemplo 5.19.

com $\mathbf{x} = \begin{bmatrix} a_0 & a_1 & a_2 \end{bmatrix}^T$ e $\mathbf{y} = \begin{bmatrix} 2 & 2.2 & 2.6 & 3.2 & 4 \end{bmatrix}^T$. A equação normal é

$$A^T A\mathbf{x} = A^T\mathbf{y} \iff \begin{bmatrix} 5 & 15 & 55 \\ 15 & 55 & 225 \\ 55 & 225 & 979 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 14 \\ 47 \\ 185.4 \end{bmatrix}.$$

A solução de mínimos quadrados do sistema $A\mathbf{x} = \mathbf{y}$ é

$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{y} = \begin{bmatrix} 2 \\ -0.1 \\ 0.1 \end{bmatrix}.$$

O polinómio procurado é portanto $y(x) = 2 - 0.1x + 0.1x^2$. Assim, prevê-se que as vendas da empresa no último mês do ano sejam de $y(12) = 15.2$ mil euros.

♦

Factorização QR e mínimos quadrados

A decomposição QR de uma matriz é um recurso computacional frequentemente utilizado na resolução de problemas de mínimos quadrados. Esta decomposição apresenta vantagens computacionais, em particular nos casos em que o cálculo de $A^T A$ ou da sua inversa não são aconselháveis, por exemplo devido a propagação de erros de arredondamento. Fazemos por isso uma referência breve à forma como a decomposição QR é utilizada na resolução da equação normal. Relembremos o Teorema 5.6 que diz que uma matriz real A com colunas linearmente independentes admite uma factorização (única) na forma $A = QR$, onde Q é uma matriz que verifica $Q^T Q = I$ e R é uma matriz triangular superior invertível.

Considere-se uma matriz A com colunas linearmente independentes, e o problema de mínimos quadrados modelado pelo sistema linear $A\mathbf{x} = \mathbf{y}$. Como as colunas de A são linearmente independentes, a equação normal $A^T A\mathbf{x} = A^T\mathbf{y}$ tem solução única e a matriz A admite uma factorização QR. Aplicando a factorização $A = QR$ à matriz $A^T A$, temos

$$A^T A = (QR)^T(QR) = R^T \underbrace{Q^T Q}_{I} R = R^T R.$$

Como a matriz R^T é não singular (já que R é não singular), a equação normal pode reescrever-se na forma:

$$A^T A\mathbf{x} = A^T\mathbf{y} \iff R^T R\mathbf{x} = R^T Q^T\mathbf{y} \iff R\mathbf{x} = Q^T\mathbf{y}.$$

A equação normal reduz-se portanto ao sistema $R\mathbf{x} = Q^T\mathbf{y}$, cuja matriz dos coeficientes é uma matriz triangular superior. Logo, a solução deste sistema pode obter-se facilmente por substituição regressiva.

5.7.3 Transformada de Fourier discreta

Os métodos matemáticos adequados ao tratamento de sinais (por exemplo, sinais de audio ou imagem) fazem parte da Análise de Fourier[8]. Nesta secção abordamos o assunto de uma forma sumária, e sob o ponto de vista da álgebra linear.

Supõe-se que o sinal é representado por uma função real de que se conhece apenas um conjunto finito de valores. Pretende-se reconstruir o sinal (contínuo) através da sua amostra discreta. Para o efeito iremos aproximar a função que representa o sinal por um polinómio trigonométrico, isto é uma combinação linear de senos e cosenos, sendo os coeficientes desta combinação linear as incógnitas a determinar. Tal processo conduz-nos à resolução de um sistema linear cuja matriz dos coeficientes é conhecida por *matriz de Fourier*. A resolução deste sistema recebe a designação de *transformada de Fourier discreta* (DFT[9]) da amostra do sinal em causa. Como se verá, as colunas da matriz de Fourier (de ordem n) formam uma base ortogonal de $\mathbb{C}^n$. Esta propriedade de ortogonalidade é essencial em todos os modelos matemáticos dedicados ao processamento de sinal.

Nesta secção fazemos uma breve introdução ao assunto e finalizamos com uma referência ao algoritmo conhecido por *transformada de Fourier rápida* (FFT[10]).

[8] Jean-Baptiste Joseph Fourier (1768 — 1830), matemático e físico francês.

[9] DFT é a abreviatura de "Discrete Fourier Transform".

[10] FFT é a abreviatura de "Fast Fourier Transform".

Este algorimo permite determinar a transformada de Fourier discreta com grande economia computacional. Mostramos ainda que a FFT se deduz facilmente de uma determinada factorização da matriz de Fourier.

Como se disse anteriormente, um sinal é representado por uma função real de variável real f, definida num intervalo $[a, b]$. Por conveniência de cálculo, consideramos $[a, b] = [0, 2\pi]$. Para além disso, consideramos que a amostragem do sinal é realizada em períodos de tempo igualmente espaçados no intervalo referido, isto é, a amostra é realizada nos instantes $x_k = k\dfrac{2\pi}{n}$, onde $k = 0, \ldots, (n-1)$. Ou seja, os dados da amostra são constituídos pelos n valores

$$f(0), f\left(\frac{2\pi}{n}\right), f\left(\frac{4\pi}{n}\right), \ldots, f\left(\frac{2(n-1)\pi}{n}\right).$$

Na prática, o valor de n é bastante grande, pelo que o vector observado,

$$\mathbf{f} = (f(x_0), f(x_1), \ldots, f(x_{n-1})),$$

pode ter dimensões consideráveis.

Pretende-se aproximar f por uma função p tal que p e f coincidam nos pontos da amostra, ou seja, $p(x_k) = f(x_k)$, para $k = 0, 1, \ldots, (n-1)$. Uma função p que verifica esta propriedade diz-se uma *função interpoladora* de f.

As funções interpoladoras que se consideram habitualmente são combinações lineares de senos e cosenos, sendo conhecidas pela designação de polinómios trigonométricos. Para simplificar os cálculos consideramos que a função interpoladora p é uma combinação linear de exponenciais complexas (relembre que $e^{ikx} = \cos(kx) + i\,\mathrm{sen}(kx)$), isto é,

$$p(x) = c_0 + c_1 e^{ix} + c_2 e^{2ix} + \cdots + c_{n-1} e^{(n-1)ix}, \tag{5.32}$$

onde $c_0, \ldots, c_{n-1}$ são n coeficientes a determinar. Estes coeficientes são denominados por *coeficientes de Fourier*. A designação "polinómio trigonométrico" para p, está relacionada com o facto da expressão (5.32) resultar da substituição de z por $e^{ix} = z$ no polinómio $p(z) = c_0 + c_1 z + c_2 z^2 + \cdots + c_{n-1} z^{(n-1)}$. Num polinómio trigonométrico $p(x) = \sum_{k=0}^{n-1} c_k e^{ikx}$, designam-se por *frequências* os inteiros k em e^{ikx}.

A função $p(x)$ é uma função real com valores complexos, e a função f que se pretende aproximar toma valores reais. Por isso, uma vez determinado p, tomamos como aproximação de f a parte real de p, ou seja, consideramos a aproximação $f(x) \approx \mathrm{Re}(p(x))$.

Dada a amostra $\mathbf{f} = (f(x_0), f(x_1), \ldots, f(x_{n-1}))$, com $x_k = \frac{2\pi k}{n}$, determinar a sua *Transformada de Fourier discreta* equivale a determinar os *coeficientes de Fourier* $c_0, c_1, \ldots, c_{n-1}$ que satisfazem as igualdades

$$f(x_0) = p(x_0),\ f(x_1) = p(x_1),\ \ldots,\ f(x_{n-1}) = p(x_{n-1}) \qquad \text{com } \ x_k = \frac{2\pi k}{n}. \tag{5.33}$$

Atendendo à expressão de p em (5.32), tem-se

$$\begin{aligned} p(x_k) = p\left(\frac{2\pi k}{n}\right) &= c_0 + c_1 e^{i\frac{2\pi k}{n}} + c_2 e^{i\frac{4\pi k}{n}} + \cdots + c_{n-1} e^{i\frac{2\pi(n-1)k}{n}} \\ &= c_0 + c_1 \xi_n^k + c_2 \xi_n^{2k} + \cdots + c_{n-1}\xi_n^{(n-1)k} \quad \text{com} \quad \xi_n = e^{i\frac{2\pi}{n}}. \end{aligned} \tag{5.34}$$

O número complexo ξ_n é uma raiz (índice n) da unidade. De facto, $\xi_n^n = e^{2\pi i} = \cos(2\pi) + i\,\mathrm{sen}(2\pi) = 1$. As raízes índice n de um número complexo z estão localizadas numa circunferência de centro na origem e raio $\sqrt[n]{|z|}$, e dividem esta circunferência em n partes iguais. Assim, existem n raízes distintas $\sqrt[n]{z}$, dadas pela expressão (ver (A.4)).

$$\sqrt[n]{|z|}e^{\frac{i(\theta+2k\pi)}{n}}, \quad \text{para } k = 0, \ldots, (n-1),$$

onde θ é o argumento de z. Logo, quando $z = 1$ tem-se $|z| = 1$ e $\theta = 0$, e portanto os n números complexos distintos que representam $\sqrt[n]{1}$ são

$$\xi_n^k = e^{\frac{2k\pi i}{n}}, \quad k = 0, 1, \ldots, (n-1).$$

Usando a expressão (5.34), as condições (5.33) são equivalentes ao seguinte sistema linear, nas incógnitas $c_0, c_1, \ldots, c_{n-1}$,

$$\begin{aligned} f(x_0) &= f(0) &&= c_0 + c_1 + c_2 + \cdots + c_{n-1} \\ f(x_1) &= f\left(\frac{2\pi}{n}\right) &&= c_0 + c_1\xi_n + c_2\xi_n^2 + \cdots + c_{n-1}\xi_n^{n-1} \\ f(x_2) &= f\left(\frac{4\pi}{n}\right) &&= c_0 + c_1\xi_n^2 + c_2\xi_n^4 + \cdots + c_{n-1}\xi_n^{2(n-1)} \\ &\vdots \\ f(x_{n-1}) &= f\left(\frac{2(n-1)\pi}{n}\right) &&= c_0 + c_1\xi_n^{n-1} + c_2\xi_n^{2(n-1)} + \cdots + c_{n-1}\xi_n^{(n-1)^2}. \end{aligned}$$

O sistema anterior pode escrever-se na forma matricial $\mathbf{f} = F_n\mathbf{c}$, onde $\mathbf{c}$ é o vector dos *coeficientes de Fourier* e F_n é a designada *matriz de Fourier* de ordem n. Os vectores coluna da matriz de Fourier, de ordem n, são:

$$F_n = \begin{bmatrix} | & | & & | \\ \mathbf{w}_0 & \mathbf{w}_1 & \cdots & \mathbf{w}_{n-1} \\ | & | & & | \end{bmatrix}, \tag{5.35}$$

com

$$\mathbf{w}_0 = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \quad \mathbf{w}_1 = \begin{bmatrix} 1 \\ \xi_n \\ \vdots \\ \xi_n^{n-1} \end{bmatrix}, \quad \cdots \quad , \quad \mathbf{w}_{n-1} = \begin{bmatrix} 1 \\ \xi_n^{n-1} \\ \vdots \\ \xi_n^{(n-1)^2} \end{bmatrix}.$$

Note-se que os vectores coluna $\mathbf{w}_k$ da matriz de Fourier F_n têm a seguinte forma:

$$\mathbf{w}_k = (1, \xi_n^k, \xi_n^{2k}, \ldots, \xi_n^{(n-1)k}), \qquad \text{com} \quad \xi_n = e^{\frac{2\pi i}{n}}.$$

Por exemplo, a matriz de Fourier de ordem 4 é

$$F_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{bmatrix}, \tag{5.36}$$

uma vez que as colunas de F_4 são $\mathbf{w}_k = (1, e^{\frac{k\pi i}{2}}, e^{k\pi i}, e^{\frac{3k\pi i}{2}})$, para $k = 0, 1, 2, 3$.

O sistema linear $\mathbf{f} = F_n\mathbf{c}$ diz-nos que o vector da amostra $\mathbf{f}$ é uma combinação linear das colunas de F_n sendo os coeficientes desta combinação linear as componentes de $\mathbf{c}$. Ou seja, o vector da amostra $\mathbf{f}$ é uma combinação linear dos vectores do conjunto $S = \{\mathbf{w}_0, \mathbf{w}_1, \ldots, \mathbf{w}_{n-1}\}$. Assim, se provarmos que S é um conjunto ortogonal, o Teorema 5.2 (pág. 260) fornece-nos a expressão dos coeficientes de Fourier. Nomeadamente,

$$c_k = \frac{\langle \mathbf{w}_k, \mathbf{f} \rangle}{\|\mathbf{w}_k\|^2}, \tag{5.37}$$

onde $\langle\, , \,\rangle$ designa o produto interno usual em $\mathbb{C}^n$.

Mostremos agora que o conjunto dos vectores coluna, $\{\mathbf{w}_0, \mathbf{w}_1, \ldots, \mathbf{w}_{n-1}\}$, da matriz de Fourier F_n é um conjunto ortogonal.

Proposição 5.6. As colunas da matriz de Fourier de ordem n formam um conjunto ortogonal para o produto interno usual em $\mathbb{C}^n$. Isto é, o conjunto $S = \{\mathbf{w}_0, \mathbf{w}_1, \ldots, \mathbf{w}_{n-1}\}$, com

$$\mathbf{w}_k = (1, \xi_n^k, \xi_n^{2k}, \ldots, \xi_n^{(n-1)k}) \quad \text{e} \quad \xi_n = e^{\frac{2\pi i}{n}},$$

é ortogonal. Além disso, $\|\mathbf{w}_k\|^2 = n$ para $k = 0, 1, \ldots, (n-1)$.

Demonstração. Qualquer polinómio de grau $n \geq 1$ pode ser factorizado em termos das suas raízes. Como as raízes (distintas) do polinómio $z^n - 1$ são ξ_n^k, com $k = 0, 1, \ldots, n-1$, tem-se

$$z^n - 1 = (z-1)(z-\xi_n)(z-\xi_n^2)\cdots(z-\xi_n^{n-1}).$$

Além disso, é válida a igualdade

$$z^n - 1 = (z-1)(1 + z + z^2 + \cdots + z^{n-1}),$$

como pode facilmente verificar desenvolvendo o membro direito da equação anterior. Comparando as duas últimas igualdades, obtemos

$$1 + z + z^2 + \cdots + z^{n-1} = (z-\xi_n)(z-\xi_n^2)\cdots(z-\xi_n^{n-1}). \tag{5.38}$$

Substituindo z por ξ_n^k em (5.38), tem-se

$$1 + \xi_n^k + \xi_n^{2k} + \cdots + \xi_n^{(n-1)k} = \begin{cases} n, & \text{se} \quad k = 0 \\ 0, & \text{se} \quad 0 < k < n, \end{cases} \tag{5.39}$$

uma vez que $\xi_n^0 = 1$, e para $z = \xi_n^k$, com $0 < k < n$, o membro direito de (5.38) é sempre nulo. Calculemos agora $\langle \mathbf{w}_k \,,\, \mathbf{w}_l \rangle$.

$$\begin{aligned}
\langle \mathbf{w}_k \,,\, \mathbf{w}_l \rangle &= \mathbf{w}_k^H \mathbf{w}_l = \overline{\mathbf{w}}_k^T \mathbf{w}_l \\
&= \begin{bmatrix} 1 & \overline{\xi_n^k} & \overline{\xi_n^{2k}} & \cdots & \overline{\xi_n^{(n-1)k}} \end{bmatrix} \begin{bmatrix} 1 \\ \xi_n^l \\ \xi_n^{2l} \\ \vdots \\ \xi_n^{(n-1)l} \end{bmatrix} \\
&= 1 + \xi_n^{-k}\xi_n^l + \xi_n^{-2k}\xi_n^{2l} + \cdots + \xi_n^{-(n-1)k}\xi_n^{(n-1)l} \qquad (\text{visto que } \overline{\xi_n^s} = \xi_n^{-s}) \\
&= \sum_{j=0}^{n-1} \xi_n^{-jk}\xi_n^{jl} = \sum_{j=0}^{n-1} \xi_n^{j(l-k)} = \begin{cases} n, & \text{se} \quad k = l \\ 0, & \text{se} \quad k \neq l, \end{cases}
\end{aligned}$$

onde na última igualdade aplicámos (5.39). □

A proposição anterior diz-nos que S é um conjunto ortogonal que não é ortonormado. Porém, como todos os vectores de S têm norma igual a $\sqrt{n}$, resulta da equação (5.37) que os coeficientes de Fourier são

$$c_k = \frac{\langle \mathbf{w}_k, \mathbf{f} \rangle}{\|\mathbf{w}_k\|^2} = \frac{\langle \mathbf{w}_k, \mathbf{f} \rangle}{n}, \qquad k = 0, 1, \dots, n-1. \tag{5.40}$$

Exemplo 5.20. Consideremos a amostra da função $f(x) = 2\pi x - x^2$ em 4 pontos, e calculemos o respectivo polinómio interpolador de Fourier. O vector $\mathbf{f}$ correspondente à amostra é

$$\mathbf{f} = (f_0, f_1, f_2, f_3) = (f(0), f(\pi/2), f(\pi), f(3\pi/2)) = (0, 3\pi^2/4, \pi^2, 3\pi^2/4).$$

Os vectores coluna $\mathbf{w}_k$ da matriz de Fourier F_4 são dados por (5.36). Calculando os coeficientes de Fourier obtemos os seguintes valores (arredondados para 4 casas decimais)

$$\begin{aligned} c_0 &= \langle \mathbf{w}_0, \mathbf{f} \rangle = 6.1685 & \qquad c_2 &= \langle \mathbf{w}_2, \mathbf{f} \rangle = -1.2337 \\ c_1 &= \langle \mathbf{w}_1, \mathbf{f} \rangle = -2.4674 & \qquad c_3 &= \langle \mathbf{w}_3, \mathbf{f} \rangle = -2.4674. \end{aligned}$$

Assim, o polinómio interpolador (com valores complexos) é

$$p(x) = 6.1685 - 2.4674e^{ix} - 1.2337e^{2ix} - 2.4674e^{3ix}.$$

Consequentemente, a parte real de p é dada por

$$\mathrm{Re}(p(x)) = 6.1685 - 2.4674\cos x - 1.2337\cos(2x) - 2.4674\cos(3x).$$

Na Figura 5.19 é ilustrado o gráfico de f, os pontos de amostragem, e a parte real do polinómio interpolador. ♦

Nos gráficos da Figura 5.19 observa-se que embora a aproximação $\mathrm{Re}(p(x))$ coincida com a função f nos pontos de amostragem, fora desses pontos $\mathrm{Re}(p(x))$ toma valores bastante diferentes dos da função f, ou seja, $\mathrm{Re}(p(x))$ não é uma boa aproximação de f. É no entanto possível obter uma aproximação melhor de f substituindo no polinómio trigonométrico p as exponenciais complexas de frequências mais elevadas por exponenciais de frequências mais baixas tomando os mesmos valores nos pontos da amostra. As observações que se seguem justificam a razão de ser desta substituição.

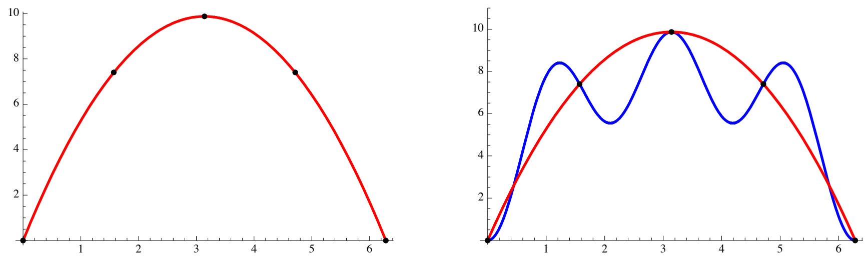


Figura 5.19: O gráfico de $f(x) = 2\pi x - x^2$ está representado a vermelho, e o gráfico da parte real do polinómio interpolador da amostra de f em 4 pontos a cor azul.

Considerem-se as exponenciais $e^{3ix} = \cos(3x) + i\,\mathrm{sen}(3x)$ e $e^{-ix} = \cos x - i\,\mathrm{sen}\,x$, e uma amostra em 4 pontos $x_k = \frac{2\pi k}{n}$. Como se observa na Figura 5.20, os gráficos de $\mathrm{Re}(e^{3ix}) = \cos(3x)$ e $\mathrm{Re}(e^{-ix}) = \cos x$ são bem diferentes no intervalo $[0, 2\pi]$ embora coincidam nos pontos x_k. A parte real (ou a parte imaginária) da exponencial e^{i3x} tem uma frequência de oscilação maior que a exponencial e^{-ix}. É portanto natural que a exponencial e^{-ix} seja mais conveniente em termos de aproximação que e^{3ix}. Este fenómeno é designado por *distorção* ("aliasing" em inglês).

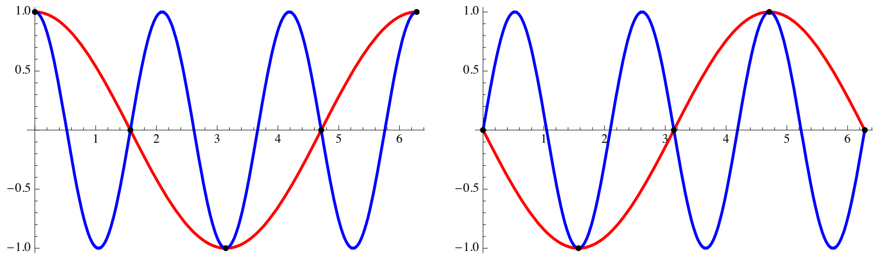


Figura 5.20: A parte real (à esquerda) e a parte imaginária (à direita) de e^{3ix} representada a cor azul, e a cor vermelha as respectivas partes real e imaginária de e^{-ix}. Nos pontos da amostra os valores de e^{3ix} e de e^{-ix} coincidem.

Note-se ainda que as raízes de $z^n - 1$ ocorrerem em pares de conjugados. Por exemplo, para as raízes quartas da unidade $\xi_4^k = e^{\frac{ik\pi}{2}}$, temos

$$\xi_4^0 = 1, \quad \xi_4^1 = e^{\frac{i\pi}{2}} = i, \quad \xi_4^2 = e^{i\pi} = -1 = \overline{\xi_4^2}, \quad \xi_4^3 = e^{-\frac{i\pi}{2}} = -i = \overline{\xi_4^1}.$$

Na Figura 5.21 ilustramos este facto para as raízes quartas e cúbicas da unidade.

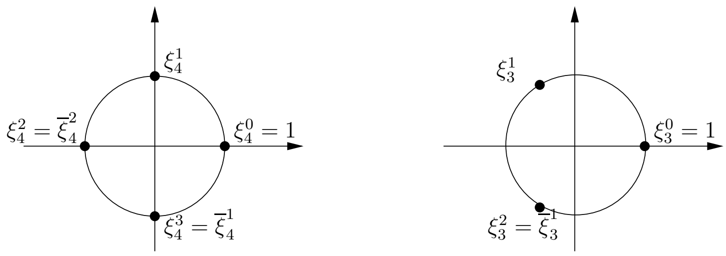


Figura 5.21: As raízes quartas e cúbicas da unidade.

Assim, podemos substituir em $p(x) = c_0 + e^{ix} + \cdots + c_{n-1}e^{i(n-1)x}$ metade das exponenciais por exponenciais de frequências mais baixas, sem alterar os valores de p nos pontos da amostra.

Podemos pois considerar uma versão alternativa ao polinómio p, que envolva frequências mais baixas. Nomeadamente, conforme o número de pontos n de amostragem é par ou ímpar, toma-se para polinómio interpolador:

$$\hat{p}(x) = c_{-m}e^{-imx} + \cdots + c_{-1}e^{-ix} + c_0 + c_1e^{ix} + \cdots + c_me^{imx} \quad \text{se } n = 2m+1$$

ou

$$\hat{p}(x) = c_{-m}e^{-imx} + \cdots + c_{-1}e^{-ix} + c_0 + c_1e^{ix} + \cdots + c_{m-1}e^{i(m-1)x} \quad \text{se } n = 2m.$$

Em ambos os casos os coeficientes de Fourier de índices negativos são iguais aos coeficientes de Fourier dos seus "alternativos" de frequências mais elevadas, isto é,

$$c_{-k} = c_{n-k} = \frac{\langle \mathbf{w}_{n-k}, \mathbf{f} \rangle}{\|\mathbf{w}_{n-k}\|} = \frac{\langle \mathbf{w}_{-k}, \mathbf{f} \rangle}{\|\mathbf{w}_{-k}\|},$$

uma vez que $\xi_n^{-k} = \xi_n^{n-k}$, e portanto $\mathbf{w}_{n-k} = \mathbf{w}_{-k}$.

Exemplo 5.21. No Exemplo 5.20 determinámos o polinómio interpolador $p(x)$ para a função $f(x) = 2\pi x - x^2$ para 4 pontos de amostragem. Determinemos agora o polinómio $\hat{p}(x)$.

$$\hat{p}(x) = c_{-2}e^{-2ix} + c_{-1}e^{-ix} + c_0 + c_1e^{ix}.$$

Os coeficientes de Fourier, calculados no exemplo referido, são os seguintes:

$$c_0 = 6.1685 \qquad c_2 = c_{-2} = -1.2337$$
$$c_1 = -2.4674 \qquad c_{-1} = c_3 = -2.4674.$$

Logo,

$$\hat{p}(x) = -1.2337e^{-2ix} - 2.4674e^{-ix} + 6.1685 - 2.4674e^{ix}. \tag{5.41}$$

A parte real deste polinómio é

$$\mathrm{Re}(\hat{p}(x)) = 6.1685 - 4.9348\cos x - 1.2337\cos(2x).$$

Na Figura 5.22 encontra-se representada a função f e as partes reais dos polinómios interpoladores $p(x)$ e $\hat{p}(x)$, respectivamente a azul e a verde. Podemos verificar que $\mathrm{Re}(\hat{p}(x))$ aproxima melhor a função f do que $\mathrm{Re}(p(x))$. ♦

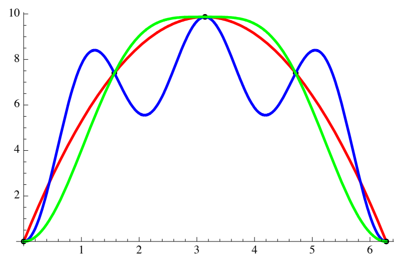


Figura 5.22: O gráfico de $f(x) = 2\pi x - x^2$ está representado a cor vermelha, e nas cores azul e verde, respectivamente, os gráficos das partes reais dos polinómios interpoladores $p(x)$ e $\hat{p}(x)$.

Observamos a seguir algumas propriedades interessantes da representação de Fourier $\hat{p}(x)$ da função real f.

- Como a função f é real e os vectores $\mathbf{w}_{-k}$ verificam $\mathbf{w}_{-k} = \overline{\mathbf{w}}_k$, os coeficientes de Fourier satisfazem a igualdade $c_{-k} = \overline{c}_k$, já que

$$\langle \mathbf{w}_{-k}, \mathbf{f}\rangle = \langle \overline{\mathbf{w}}_k, \mathbf{f}\rangle = \mathbf{w}_k^T \mathbf{f} = \overline{\overline{\mathbf{w}}_k^T \overline{\mathbf{f}}} = \overline{\langle \mathbf{w}_k, \mathbf{f}\rangle},$$

e portanto

$$c_{-k} = \frac{\langle \mathbf{w}_{-k}, \mathbf{f}\rangle}{\|\mathbf{w}_{-k}\|^2} = \frac{\overline{\langle \mathbf{w}_k, \mathbf{f}\rangle}}{\|\mathbf{w}_k\|^2} = \overline{c}_k.$$

- As parcelas de $\hat{p}(x)$ da forma $c_{-k}e^{-ikx}+c_k e^{ikx}$ são reais. De facto, tomando $c_k = a_k + ib_k$, tem-se

$$\begin{aligned} c_{-k}e^{-ikx} + c_k e^{ikx} &= \overline{c}_k e^{-ikx} + c_k e^{ikx} = \overline{c_k e^{ikx}} + c_k e^{ikx} \\ &= 2\operatorname{Re}\left(c_k e^{ikx}\right) = 2\operatorname{Re}\left((a_k + ib_k)(\cos(kx) + i\operatorname{sen}(kx))\right) \\ &= 2\left(a_k\cos(kx) - b_k\operatorname{sen}(kx)\right), \end{aligned}$$

  onde aplicámos a igualdade $z + \overline{z} = 2\operatorname{Re} z$, válida para qualquer número complexo z.

  Assim, se o número de pontos n da amostra for ímpar, o polinómio interpolador $\hat{p}$ é um polinómio real, e no caso em que $m = 2n$ somente o termo $c_{-m}e^{-imx}$ pode produzir um número complexo não real, conforme pode verificar através do polinómio (5.41).

Transformada de Fourier rápida (FFT)

Como vimos, a determinação da transformada de Fourier discreta consiste em resolver (em ordem a $\mathbf{c}$) o sistema linear $\mathbf{f} = F_n\mathbf{c}$. A matriz de Fourier F_n é invertível e a sua inversa é $F_n^{-1} = \dfrac{1}{n}\overline{F}_n$. Assim, a transformada de Fourier discreta pode obter-se multiplicando o vector de amostragem $\mathbf{f}$ por uma matriz de ordem n, isto é, $\mathbf{c} = (1/n)\overline{F}_n\mathbf{f}$. Esta abordagem para o cálculo dos coeficientes de Fourier é apenas satisfatória quando n é relativamente pequeno. Para grandes amostras, como é o caso do processamento de imagens de vídeo ou imagens de vários meios de diagnóstico médico, esta via não é conveniente já que requer n^2 multiplicações ($\overline{F}_n$ tem n^2 entradas).

Em 1965, J. Cooley e J. Tukey[11] introduziram um algoritmo que reduz para $\frac{n}{2}\log_2 n$ o número de multiplicações necessárias para o cálculo da Transformada de Fourier discreta. Este algoritmo é conhecido pela designação de *Transformada de Fourier rápida*[12] (FFT).

Quando n é par, o algoritmo FFT pode descrever-se usando uma factorização da matriz de Fourier de ordem n envolvendo matrizes de Fourier de ordem $n/2$. Não pretendemos aqui descrever completamente este algoritmo mas apenas ilustrar os seus passos essenciais. O leitor interessado em aprofundar este assunto pode consultar, por exemplo, Olver & Shakiban [10].

[11] James Cooley, matemático americano nascido em 1926, e John Wilder Tukey (1915 – 2000) estatístico americano.

[12] A Transformada de Fourier rápida é o nome genérico atribuído a uma classe de algoritmos que inclui o algoritmo de Cooley e Tukey.

Comecemos por observar que quando n é par, o conjunto das raízes n-ésimas da unidade contém o conjunto das raízes $(n/2)$-ésimas da unidade. Por exemplo, $\{1, i, -1, -i\}$ são as raízes quartas da unidade e $\{1, -1\}$ o conjunto das raízes quadradas da unidade. Este facto é essencial para entender a factorização de F_n em termos da matriz $F_{n/2}$ que passamos a enunciar.

Factorização da matriz de Fourier

Se $n = 2^r$, a matriz de Fourier de ordem n pode escrever-se na forma

$$F_n = \begin{bmatrix} I & D_{\frac{n}{2}} \\ I & -D_{\frac{n}{2}} \end{bmatrix} \begin{bmatrix} F_{\frac{n}{2}} & \mathbf{0} \\ \mathbf{0} & F_{\frac{n}{2}} \end{bmatrix} P_n, \tag{5.42}$$

onde I é a matriz identidade, $\mathbf{0}$ a matriz nula, $D_{\frac{n}{2}} = \operatorname{diag}\left(1, \xi_n, \xi_n^2, \ldots, \xi_n^{n/2-1}\right)$, com $\xi_n^k = e^{\frac{2k\pi i}{n}}$, e P_n a seguinte matriz de permutação:

$$P_n \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_{n-2} \\ x_{n-1} \end{bmatrix} = \begin{bmatrix} x_0 \\ x_2 \\ \vdots \\ x_{n-2} \\ x_1 \\ x_3 \\ \vdots \\ x_{n-1} \end{bmatrix}.$$

A matriz $D_{n/2}$ tem na diagonal principal metade das raízes n-ésimas da unidade, e o produto $P_n\mathbf{x}$ é o vector cujas primeiras $\frac{n}{2}$ componentes são as componentes de $\mathbf{x}$ com índices pares e as últimas $\frac{n}{2}$ componentes são as componentes com índices ímpares.

Por exemplo, vejamos qual é a factorização da matriz de Fourier F_4. As matrizes de Fourier F_4 e F_2 são

$$F_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{bmatrix} \quad \text{e} \quad F_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \tag{5.43}$$

A matriz diagonal D_2 tem na diagonal principal as primeiras duas raízes quartas da unidade (começando em 1 e percorrendo a circunferência unitária no sentido anti-horário), isto é, $D_2 = \operatorname{diag}(1, i)$. Além disso, uma vez que para $\mathbf{x}_4 =$

(x_0, x_1, x_2, x_3) se tem $P_4\mathbf{x}_4 = (x_0, x_2, x_1, x_3)$, a matriz de permutação P_4 é a matriz que se obtém da identidade de quarta ordem trocando a segunda linha com a terceira linha. Isto é,

$$P_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Usando estas matrizes, é fácil verificar a validade da factorização (5.42) para F_4. Nomeadamente,

$$F_4 = \begin{bmatrix} I & D_2 \\ I & -D_2 \end{bmatrix} \begin{bmatrix} F_2 & 0 \\ 0 & F_2 \end{bmatrix} P_4 = \begin{bmatrix} F_2 & D_2F_2 \\ F_2 & -D_2F_2 \end{bmatrix} P_4.$$

Ilustremos agora, através de um exemplo, como se utiliza a factorização (5.42) para implementar o algoritmo FFT. Consideremos, por exemplo, a determinação de $F_8\mathbf{x}_8$.

$$\text{Para} \quad \mathbf{x}_8 = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix}, \quad \text{tem-se} \quad P_8\mathbf{x}_8 = \left[\begin{array}{c} x_0 \\ x_2 \\ x_4 \\ x_6 \\ \hline x_1 \\ x_3 \\ x_5 \\ x_7 \end{array}\right] = \left[\begin{array}{c} \mathbf{x}_4^{(0)} \\ \hline \mathbf{x}_4^{(1)} \end{array}\right].$$

Assim,

$$F_8\mathbf{x}_8 = \begin{bmatrix} F_4 & D_4F_4 \\ F_4 & -D_4F_4 \end{bmatrix} \begin{bmatrix} \mathbf{x}_4^{(0)} \\ \mathbf{x}_4^{(1)} \end{bmatrix} = \begin{bmatrix} F_4\mathbf{x}_4^{(0)} + D_4F_4\mathbf{x}_4^{(1)} \\ F_4\mathbf{x}_4^{(0)} - D_4F_4\mathbf{x}_4^{(1)} \end{bmatrix}.$$

A matriz F_4 pode ainda ser factorizada na forma (5.42), obtendo-se:

$$P_4\mathbf{x}_4^{(0)} = \left[\begin{array}{c} x_0 \\ x_4 \\ \hline x_2 \\ x_6 \end{array}\right] = \left[\begin{array}{c} \mathbf{x}_2^{(0)} \\ \hline \mathbf{x}_2^{(1)} \end{array}\right] \quad \text{e} \quad P_4\mathbf{x}_4^{(1)} = \left[\begin{array}{c} x_1 \\ x_5 \\ \hline x_3 \\ x_7 \end{array}\right] = \left[\begin{array}{c} \mathbf{x}_2^{(2)} \\ \hline \mathbf{x}_2^{(3)} \end{array}\right]. \tag{5.44}$$

Logo,

$$F_4\mathbf{x}_4^{(0)} = \begin{bmatrix} F_2 & D_2F_2 \\ F_2 & -D_2F_2 \end{bmatrix} \begin{bmatrix} \mathbf{x}_2^{(0)} \\ \mathbf{x}_2^{(1)} \end{bmatrix} = \begin{bmatrix} F_2\mathbf{x}_2^{(0)} + D_2F_2\mathbf{x}_2^{(1)} \\ F_2\mathbf{x}_2^{(0)} - D_2F_2\mathbf{x}_2^{(1)} \end{bmatrix}$$

e

$$F_4\mathbf{x}_4^{(1)} = \begin{bmatrix} F_2 & D_2F_2 \\ F_2 & -D_2F_2 \end{bmatrix} \begin{bmatrix} \mathbf{x}_2^{(2)} \\ \mathbf{x}_2^{(3)} \end{bmatrix} = \begin{bmatrix} F_2\mathbf{x}_2^{(2)} + D_2F_2\mathbf{x}_2^{(3)} \\ F_2\mathbf{x}_2^{(2)} - D_2F_2\mathbf{x}_2^{(3)} \end{bmatrix}.$$

Como F_2 tem a forma $F_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$, é agora trivial calcular os vectores $F_2\mathbf{x}_2^{(i)}$ para $i = 0, 1, 2, 3$. O algoritmo FFT para calcular $F_8\mathbf{x}_8$ começa por determinar os vectores $F_2\mathbf{x}_2^{(i)}$ para $i = 0, 1, 2, 3$, a partir destes calcula $F_4\mathbf{x}_4^{(0)}$ e $F_4\mathbf{x}_4^{(1)}$, e finalmente $F_8\mathbf{x}_8$, conforme está representado no diagrama seguinte.

$$\begin{tikzcd}
F_2\mathbf{x}_2^{(0)} \arrow[dr] & & F_2\mathbf{x}_2^{(1)} \arrow[dl] & & F_2\mathbf{x}_2^{(2)} \arrow[dr] & & F_2\mathbf{x}_2^{(3)} \arrow[dl] \\
& F_4\mathbf{x}_4^{(0)} \arrow[drr] & & & & F_4\mathbf{x}_4^{(1)} \arrow[dll] & \\
& & & F_8\mathbf{x}_8 & & &
\end{tikzcd}$$

É evidente que para iniciar o algoritmo é necessário conhecer os vectores $\mathbf{x}_2^{(0)}, \mathbf{x}_2^{(1)}, \mathbf{x}_2^{(2)}$ e $\mathbf{x}_2^{(3)}$, ou seja, obter o agrupamento correcto das componentes de $\mathbf{x}_8$ separando em cada grupo as componentes nas posições pares das componentes nas posições ímpares. Por exemplo, no caso apresentado de $n = 8$, efectuaram-se os agrupamentos

$$(0, 1, 2, 3, 4, 5, 6, 7)$$

$$(0, 2, 4, 6) \qquad\qquad (1, 3, 5, 7)$$

$$(0, 4) \quad (2, 6) \qquad (1, 5) \quad (3, 7).$$

Este processo de separação em par/ímpar, necessário para se iniciar o algoritmo, é por vezes designado por *"baralhar perfeito"*. O agrupamento dos índices das componentes de $\mathbf{x}$ tem uma interpretação simples se for usada a representação binária dos índices das componentes de $\mathbf{x}$. Deixa-se como exercício verificar qual é a interpretação correspondente ao agrupamento par/ímpar efectuado para o vector $\mathbf{x}_8$. Para tal, escreva os elementos do conjunto $\{0, 1, 2, 3, 4, 5, 6, 7\}$ na base 2 e proceda aos agrupamentos anteriormente efectuados. Compare a disposição dos dígitos na representação binária final com a disposição inicial.

Para finalizar, ilustramos agora os diversos passos do algoritmo FFT no cálculo dos coeficientes de Fourier para uma amostragem em quatro pontos. Isto é, pretendemos calcular $\mathbf{c} = \frac{1}{4}\overline{F}_4\mathbf{f}$, onde $\mathbf{f} = (f_0, f_1, f_2, f_3)$ e $f_k = f(x_k) = f(\frac{2\pi k}{n})$.

Em primeiro lugar, faz-se a reordenação correcta das componentes do vector da amostra $\mathbf{f}$. Neste caso, como a amostra apenas foi realizada em 4 pontos, obtém-se $\tilde{\mathbf{f}} = (f_0, f_2, f_1, f_3)$. O algoritmo decorre em dois passos, que passamos a descrever no pseudocódigo seguinte (onde o símbolo "$\longleftarrow$" significa uma atribuição).

Passo 1: (Cálculo de $\overline{F}_2\mathbf{f}$)

Dados: $\tilde{\mathbf{f}} = (f_0, f_2, f_1, f_3)$
$\mathbf{f} \longleftarrow P_4\tilde{\mathbf{f}} = (f_0, f_1, f_2, f_3)$
$D \longleftarrow \overline{D}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
$\mathbf{f}^{(0)} \longleftarrow (f_0, f_1)$
$\mathbf{f}^{(1)} \longleftarrow (f_2, f_3)$
Calcular:
$D\,\mathbf{f}^{(1)} = (f_2, f_3)$
$\overline{F}_2\mathbf{f} = \begin{bmatrix} \mathbf{f}^{(0)} + D\,\mathbf{f}^{(1)} \\ \mathbf{f}^{(0)} - D\,\mathbf{f}^{(1)} \end{bmatrix} = (f_0 + f_2, f_1 + f_3, f_0 - f_2, f_1 - f_3)$

Passo 2: (Cálculo de $\overline{F}_4\mathbf{f}$)

$\mathbf{f} \longleftarrow P_4\overline{F}_2\mathbf{f} = (f_0 + f_2, f_0 - f_2, f_1 + f_3, f_1 - f_3)$
$D \longleftarrow \overline{D}_2 = \begin{bmatrix} 1 & 0 \\ 0 & -i \end{bmatrix}$
$\mathbf{f}^{(0)} \longleftarrow (f_0 + f_2, f_0 - f_2)$
$\mathbf{f}^{(1)} \longleftarrow (f_1 + f_3, f_1 - f_3)$
Calcular:
$D\,\mathbf{f}^{(1)} = (f_1 + f_3, -if_1 + if_3)$
$$\overline{F}_4\mathbf{f} = \begin{bmatrix} \mathbf{f}^{(0)} + D\,\mathbf{f}^{(1)} \\ \mathbf{f}^{(0)} - D\,\mathbf{f}^{(1)} \end{bmatrix} = \begin{bmatrix} f_0 + f_2 + f_1 + f_3 \\ f_0 - f_2 - if_1 + if_3 \\ f_0 + f_2 - f_1 - f_3 \\ f_0 - f_2 + if_1 - if_3 \end{bmatrix}.$$

Usando a expressão de F_4 dada em (5.43) confirma-se que o vector obtido é de facto $\overline{F}_4\mathbf{f}$, sendo o vector dos coeficientes de Fourier:

$$\mathbf{c} = \frac{1}{4}\overline{F}_4\mathbf{f} = \frac{1}{4}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -i & -1 & i \\ 1 & -1 & 1 & -1 \\ 1 & i & -1 & -i \end{bmatrix}\begin{bmatrix} f_0 \\ f_1 \\ f_2 \\ f_3 \end{bmatrix} = \frac{1}{4}\begin{bmatrix} f_0 + f_2 + f_1 + f_3 \\ f_0 - f_2 - if_1 + if_3 \\ f_0 + f_2 - f_1 - f_3 \\ f_0 - f_2 + if_1 - if_3 \end{bmatrix}.$$

Para uma amostra em $n = 2^r$ pontos, o número de multiplicações efectuadas por este algoritmo é estimado do seguinte modo. O algoritmo tem r passos e em cada passo realizam-se (no máximo) $n/2$ multiplicações (correspondentes à multiplicação pela matriz diagonal D). Logo, para $n = 2^r$ efectuam-se 2^{r-1} multiplicações. Assim, o número total de multiplicações é $r2^{r-1}$ (uma vez o algoritmo tem r passos). Como $n = 2^r$ (ou seja, $r = \log_2 n$) e $2^{r-1} = n/2$, o número total de multiplicações é (no máximo) $r2^{r-1} = \frac{n}{2}(\log_2 n)$.

Comparando as n^2 multiplicações necessárias para efectuar o produto $\frac{1}{n}\overline{F}_n\mathbf{f}$ com o número de multiplicações usando o algoritmo FFT, nota-se uma economia de operações considerável. Por exemplo, para $n = 2^9 = 512$, temos $n^2 = 262144$ e $\frac{n}{2}\log_2 n = 2304$.

Exercícios

Sempre que não se especifique o produto interno, considere o produto interno usual.

1. Para $\mathbf{u} = (4,1,4,5)$, $\mathbf{v} = (1,2,0,2)$, calcule as expressões:

a) $||\mathbf{u}+\mathbf{v}||$ b $\|\mathbf{u}\|+\|\mathbf{v}\|$ c) $-3\|\mathbf{u}\|+\|3\mathbf{v}\|$
d) $\frac{1}{\|\mathbf{v}\|}\mathbf{v}$ e) $\left\|\frac{1}{\|\mathbf{v}\|}\mathbf{v}\right\|$ f) $\measuredangle(\mathbf{u},\mathbf{v})$.

2. Verifique a desigualdade de Cauchy-Schwarz para os seguintes vectores:

a) $\mathbf{u} = (-3,1,0)$ e $\mathbf{v} = (2,-1,3)$
b) $\mathbf{u} = (-4,2,1,6,-1)$ e $\mathbf{v} = (8,-4,-2,2,1)$.

3. Diga para que valores de k os vectores seguintes são ortogonais:

a) $\mathbf{u} = (2,1,3), \mathbf{v} = (k,1,2)$
b) $\mathbf{u} = (2,1,3,4), \mathbf{v} = (k,-1,2k,1)$.

4. Determine todos os vectores $\mathbf{u}$ de $\mathbb{R}^2$ tais que o ângulo que $\mathbf{u}$ faz com o vector $(1,2)$ é de $\dfrac{\pi}{3}$ radianos e $\|\mathbf{u}\| = 1$.

5. Para $\mathbf{u},\mathbf{v} \in \mathbb{R}^n$ mostre a igualdade

$$\|\mathbf{u}+\mathbf{v}\|^2 + \|\mathbf{u}-\mathbf{v}\|^2 = 2\|\mathbf{u}\|^2 + 2\|\mathbf{v}\|^2.$$

Interprete geometricamente esta igualdade para vectores de $\mathbb{R}^2$.

6. Mostre que se $\mathbf{u}$ e $\mathbf{v}$ são vectores ortogonais tais que $\|\mathbf{u}\| = 1$ e $\|\mathbf{v}\| = 2$, então a distância de $\mathbf{u}$ a $\mathbf{v}$ é $d(\mathbf{u},\mathbf{v}) = \sqrt{\|\mathbf{u}\|^2+\|\mathbf{v}\|^2} = \sqrt{5}$. Interprete geometricamente este resultado para vectores de $\mathbb{R}^2$.

7. Sejam $\mathbf{x} = (\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}})$ e $\mathbf{y} = \left(\frac{2}{\sqrt{k}}, \frac{2}{\sqrt{k}}\right)$. Diga para que valores de $k \neq 0$ o conjunto $\{\mathbf{x},\mathbf{y}\}$ é um conjunto ortonormado.

8. Calcule:

a) $\operatorname{proj}_{(1,2,3)}(3,0,1)$.
b) $\operatorname{proj}_{(1,2,3)}(-\pi,-2\pi,-3\pi)$.
c) $\operatorname{proj}_{(1,2,3)}(-2000,1000,0)$.

9. Considere a matriz:

$$A = \begin{bmatrix} 1 & 2 & -1 & 2 \\ 3 & 5 & 0 & 4 \\ 1 & 1 & 2 & 0 \end{bmatrix}.$$

a) Sem efectuar quaisquer cálculos, justifique se são verdadeiras ou falsas as seguintes afirmações:
   i) $\dim EL(A) + \dim N(A^T) = 4$.
   ii) $\dim EC(A) + \dim N(A^T) = 4$.
b) Determine uma base para o complemento ortogonal do núcleo de A.
c) Determine uma base para o complemento ortogonal do núcleo de A^T.

10. Decida quais das expressões seguintes definem um produto interno. Em caso negativo indique quais dos axiomas da definição do produto interno não são satisfeitos.

a) $\langle (u_1,u_2),(v_1,v_2)\rangle = 4u_1v_1+u_2v_1+u_1v_2+4u_2v_2$, em $\mathbb{R}^2$.
b) $\langle (u_1,u_2,u_3),(v_1,v_2,v_3)\rangle = u_1v_1 - u_2v_2 + u_3v_3$, em $\mathbb{R}^3$.
c) $\langle (u_1,u_2,u_3),(v_1,v_2,v_3)\rangle = u_1^2v_1^2 + u_2^2v_2^2 + u_3^2v_3^2$, em $\mathbb{R}^3$.
d) $\langle (u_1,u_2,u_3),(v_1,v_2,v_3)\rangle = u_1v_1 + u_3v_3$, em $\mathbb{R}^3$.

11. Encontre uma base para o complemento ortogonal de U.
a) $U = \operatorname{Span}\{(1,1,-1),(1,2,3),(2,1,-6)\}$.
b) $U = \operatorname{Span}\{(1,2,3,4),(2,4,6,8)\}$.

12. Seja U o plano de $\mathbb{R}^3$ definido pela equação $x-2y+z=0$. Encontre uma equação para $U^{\perp}$.

13. Determine a distância de $(2,0,1)$ ao plano $x-2y-z=0$, bem como o vector do plano mais próximo de $(2,0,1)$.

14. Determine uma equação para:

a) A recta que passa pelos pontos $(2,2,0)$ e $(1,2,1)$.
b) O plano que passa por $(1,1,-1), (0,1,0)$ e $(0,0,1)$.

15. Seja U a recta de $\mathbb{R}^3$ definida pelas equações paramétricas $x=2t, y=-t, z=4t$ ($t\in\mathbb{R}$). Determine uma equação para $U^{\perp}$.

16. Considere o subespaço S de $\mathbb{R}^3$ gerado pelos vectores $\mathbf{u}_1=(0,1,0)$ e $\mathbf{u}_2=(\frac{4}{5},0,-\frac{3}{5})$.

Exprima $\mathbf{w}=(1,2,3)$ na forma $\mathbf{w}=\mathbf{w}_1+\mathbf{w}_2$, em que $\mathbf{w}_1\in S$ e $\mathbf{w}_2\in S^{\perp}$.

17. Considere em $\mathbb{C}^2$ o produto interno usual e $\mathbf{u}=(i,i)$ e $\mathbf{v}=(1+2i,1-2i)$.

a) Diga se $\mathbf{u}$ e $\mathbf{v}$ são ou não ortogonais.
b) Determine $\mathrm{proj}_{\mathbf{v}}(\mathbf{u}+\mathbf{v})$.

18. Considere o subespaço de $\mathbb{R}^3$:

$$S=\left\{(x,y,z)\in\mathbb{R}^3:\ x-y+z=0\right\}.$$

Para o produto interno usual, uma base ortogonal para S é:

a) $\{(-1/2,1/2,1)\}$
b) $\{(1,1,-2),\ (-1,1,0)\}$
c) $\{(1,1,0)\}$
d) $\{(1,1,0),\ (-1/2,1/2,1)\}$.

19. Considere $\mathbb{R}^2$ munido do produto interno $\langle\mathbf{x}\,,\,\mathbf{y}\rangle=\mathbf{x}^T A\,\mathbf{y}$ onde $A=\begin{bmatrix}1&0\\0&2\end{bmatrix}$ e $\mathbf{x},\mathbf{y}\in\mathbb{R}^2$. Diga qual das afirmações seguintes é verdadeira:

a) Os vectores $(1,1)$ e $(-1,1)$ formam uma base ortogonal para $\mathbb{R}^2$.
b) Os vectores $(1,\frac{-1}{2})$ e $(1,1)$ formam uma base ortogonal para $\mathbb{R}^2$.
c) Os vectores $(1,\frac{-1}{2})$ e $(1,1)$ formam uma base ortonormada para $\mathbb{R}^2$.
d) Os vectores $(1,1)$ e $(-1,1)$ formam uma base ortonormada para $\mathbb{R}^2$.

20. Determine uma base ortonormada para $U=\mathrm{Span}\,\{(1,0,-1,0),(-1,2,0,1),(2,0,2,1),(2,2,1,1)\}$.

21. Utilize o processo de Gram-Schmidt e a normalização para obter uma base ortonormada a partir da base: $\{(1,1,1),(1,1,0),(1,0,0)\}$.

22. Seja S o subespaço de $\mathbb{R}^2$ dado pela equação $y=2x$. Determine $S^{\perp}$ e a distância do vector $(1,-1)$ aos subespaços S e $S^{\perp}$.

23. Seja $M_{2\times2}(\mathbb{R})$ o espaço linear real das matrizes reais 2×2, e U o subespaço das matrizes anti-simétricas (isto é, que satisfazem a igualdade $A=-A^T$). Em $M_{2\times2}(\mathbb{R})$ considere o produto interno

$$\langle A,B\rangle=\mathrm{tr}(AB^T), \tag{5.45}$$

onde tr designa o traço de uma matriz.

a) Mostre que $\langle\cdot,\cdot\rangle$ definido em (5.45) é um produto interno.
b) Determine a dimensão de U e de $U^{\perp}$.
c) Determine bases ortonormadas para U e para $U^{\perp}$.
d) Determine as projecções ortogonais de $A=\begin{bmatrix}1&1\\-1&2\end{bmatrix}$ sobre U e sobre $U^{\perp}$.
e) Qual a matriz anti-simétrica mais próxima de $A=\begin{bmatrix}1&1\\-1&2\end{bmatrix}$?
f) Determine a distância de $A=\begin{bmatrix}1&1\\-1&2\end{bmatrix}$ a U.

24. Considere

$$U = \operatorname{Span}\{(1,0,-1,0),(-1,2,0,1)\}.$$

Use o Exercício 5.11 para determinar a projecção ortogonal de $(2,1,0,1)$ sobre U.

25. Diga se existe algum valor de $k \in \mathbb{R}$ para o qual o conjunto das colunas das matrizes seguintes é ortonormado.

a) $\begin{bmatrix} 1 & k \\ k & 2 \end{bmatrix}$ b) $\begin{bmatrix} 1 & -k \\ k & 1 \end{bmatrix}$

c) $\begin{bmatrix} 2 & -k \\ k & 2 \end{bmatrix}$.

26. Mostre que qualquer matriz real 2×2 de determinante igual a 1, e cujas colunas formam um conjunto ortonormado, se escreve na forma $\begin{bmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{bmatrix}$.

27. Considere $\mathbf{u} = (1,0,-1,2)$. Use a matriz $P_{\mathbf{u}}$, definida por (5.12), para determinar a distância de $\mathbf{v} = (1,0,1,1)$ ao hiperplano $\mathbf{u}^{\perp}$ (isto é o complemento ortogonal ao espaço gerado por $\mathbf{u}$). Indique ainda o vector de $\mathbf{u}^{\perp}$ mais próximo de $\mathbf{v}$.

28. Considere o plano-3 de $\mathbb{R}^4$ definido por $\mathbf{p}_0 + S$ onde $\mathbf{p}_0 = (1,-1,0,1)$ e S o espaço gerado pelos vectores $(0,1,1,0),(1,0,1,0)$ e $(0,0,0,1)$. Determine a distância de $(2,0,1,0)$ a este plano-3.

29. Considere um paralelipípedo com arestas $\mathbf{u} = (1,1,0)$, $\mathbf{v} = (0,1,1)$ e $\mathbf{w} = (1,0,1)$.

a) Determine a área do paralelogramo com arestas $\mathbf{u}$ e $\mathbf{v}$.
b) Determine a altura do paralelipípedo considerando para base o paralelogramo da alínea anterior.
c) Determine o volume do paralelipípedo usando a noção de determinante de uma matriz. Use ainda os resultados das alíneas anteriores para confirmar o resultado obtido.

30. Considere o tetraedro definido pelos vectores $\mathbf{u} = (1,1,0)$, $\mathbf{v} = (0,1,1)$ e $\mathbf{w} = (1,0,1)$. Determine o volume do tetraedro.

31. Seja $A = \begin{bmatrix} 4 & 1 \\ 3 & -1 \\ 4 & 1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 10 \\ 2 \\ 2 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} -2 \\ 3 \end{bmatrix}$ e $\mathbf{v} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$. Calcule $A\mathbf{u}$ e $A\mathbf{v}$ bem como a distância destes vectores a $\mathbf{b}$. Diga se $\mathbf{u}$ pode ou não ser uma solução de mínimos quadrados do sistema $A\mathbf{x} = \mathbf{b}$.

32. Determine todas as soluções de mínimos quadrados do sistema $A\mathbf{x} = \mathbf{b}$, sendo

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 3 \\ 1 \\ 2 \\ 4 \end{bmatrix}.$$

33. Determine todas as soluções de mínimos quadrados do sistema:

$$\begin{cases} 2a + b = 1 \\ 2a + b = 2. \end{cases}$$

34. Indique o valor lógico das afirmações seguintes:

a) Uma solução de mínimos quadrados para $A\mathbf{x} = \mathbf{b}$ é um vector $\mathbf{x}^*$ tal que $\|\mathbf{b} - A\mathbf{x}\| \leq \|\mathbf{b} - A\mathbf{x}^*\|$ para todo $\mathbf{x} \in \mathbb{R}^n$.
b) Se as colunas de A são linearmente independentes, a solução de mínimos quadrados para $A\mathbf{x} = \mathbf{b}$ é única.
c) Uma solução de mínimos quadrados para $A\mathbf{x} = \mathbf{b}$ é um vector do espaço das colunas de A que minimiza a distância a $\mathbf{b}$.

35. Determine a equação de uma recta de mínimos quadrados, $y = \alpha + \beta x$, que melhor se ajusta aos dados: $(2, 3), (2, 5), (6, 4), (8, 1)$.

36. Determine o polinómio interpolador (5.32) para uma amostra da função $f(x) = x - 2\pi$ nos pontos $0, 2\pi/3$ e $4\pi/3$. Esboce o gráfico de f e compare-o com o gráfico da parte real do polinómio interpolador obtido.

37. Indique o valor lógico das afirmações seguintes relativas à matriz de Fourier F_n.

a) F_n é simétrica.
b) O conjunto das colunas de F_n é ortonormado.
c) O conjunto das colunas da matriz $\frac{1}{\sqrt{n}} F_n$ é ortonormado.
d) Verifica-se a igualdade $F_n^H F_n = I$.
e) Verifica-se a igualdade $\dfrac{1}{n} F_n^H F_n = I$.

38. Seja F_n a matriz de Fourier de ordem n e

$$\mathbf{x}_8 = [10, 1, 2, 4, 7, 10, 6, 2]^T.$$

a) Determine os vectores $\mathbf{x}_2^{(i)}$ definidos por (5.44), para $i = 0, 1, 2, 3$.
b) Use os vectores $\mathbf{x}_2^{(i)}$ e a factorização (5.42) para determinar $F_4\mathbf{x}_4^{(0)}$ e $F_4\mathbf{x}_4^{(1)}$. A partir destes vectores calcule $F_8\mathbf{x}_8$.
c) Confirme o resultado obtido na alínea anterior usando para tal a definição da matriz de Fourier F_8.

Capítulo 6

Funções Lineares

Neste capítulo estudamos um tipo particular de funções, as chamadas funções lineares. O principal resultado a reter é que uma função linear $f : V \to W$ entre dois espaços lineares de dimensões respectivamente, $\dim V = n$ e $\dim W = p$, é sempre representada por uma matriz A do tipo $p \times n$. Tal significa que escolhendo bases ordenadas B e B', respectivamente em V e em W se pode escrever $(f(x))_{B'} = A\mathbf{x}_B$, para todo $x \in V$.

Em primeiro lugar começamos por rever alguns conceitos básicos sobre funções. Definem-se depois funções lineares entre espaços vectoriais. Mostra-se que a função composta de funções lineares é ainda uma função linear, e que a inversa de uma função linear invertível ainda é uma função linear. Introduzimos ainda na Secção 6.1 a noção de isomorfismo, e mostramos que qualquer espaço linear real de dimensão n é isomorfo a $\mathbb{R}^n$. Na secção seguinte estudamos a representação matricial de uma função linear em relação a bases fixadas no domínio e no espaço de chegada da função, e relacionamos representações matriciais distintas de uma dada função linear. Na Secção 6.3 caracterizam-se as funções lineares invertíveis através do seu núcleo e relacionam-se os valores e vectores próprios de funções lineares com os valores e vectores próprios de matrizes que representam a função. Finalizamos o capítulo introduzindo a noção de espaço dual de um espaço linear, e deduzindo a representação matricial da função dual induzida por uma dada função linear entre espaços vectoriais.

6.1 Definição de função linear

Comecemos por recordar que uma função f é uma "lei"que a cada elemento x de um conjunto V associa um e um só elemento y de outro conjunto W. Quando f aplica x em w escrevemos $w = f(x)$ e dizemos que w é a *imagem* de x por f. A notação $f : V \to W$ indica que f aplica os elementos de V em elementos de W, e $f : x \mapsto w$ usa-se para indicar que o elemento $x \in V$ é aplicado por f no elemento $w \in W$. Para a função $f : V \to W$, o conjunto V é chamado *conjunto de partida*, ou *domínio* de f, e o conjunto W é denominado *conjunto de chegada*. O subconjunto de W que é formado pelas imagens por f dos elementos de V é designado por *contradomínio*, ou *conjunto imagem* de f, e será denotado por $\mathrm{Im}(f)$, ou $f(V)$. Isto é,

$$Im(f) = \{w \in W : \ w = f(x), \ \text{para algum } x \in V\}\,.$$

Na literatura, os termos *aplicação* e *transformação* são igualmente usados para designar uma função.

Consideremos $V = \mathbb{R}^n$, $W = \mathbb{R}^p$ e $f : \mathbb{R}^n \to \mathbb{R}^p$. A função f aplica cada vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ num vector $\mathbf{w} = f(\mathbf{x}) \in \mathbb{R}^p$, digamos $\mathbf{w} = (w_1, w_2, \ldots, w_p)$. Para funções f entre os espaços lineares $\mathbb{R}^n$ e $\mathbb{R}^p$, dizer que f é uma função linear significa que as equações que relacionam $w_1, w_2, \ldots, w_p$ com $x_1, x_2, \ldots, x_n$ são equações lineares (nas variáveis $x_1, x_2, \ldots, x_n$). Por exemplo, consideremos as funções:

1. $f : \mathbb{R}^2 \to \mathbb{R}^3$ definida por $f(x, y) = (x + y, xy, x^2 + 1)$;
2. $f : \mathbb{R}^2 \to \mathbb{R}$ definida por $f(x, y) = x^2 + 2xy + y^2$;
3. $f : \mathbb{R} \to \mathbb{R}$ definida por $f(x) = \operatorname{sen} x$;
4. $f : \mathbb{R}^3 \to \mathbb{R}^2$ definida por $f(x, y, z) = (x + y - z, 2y - 3z)$.

Destas funções somente a função do item 4 é linear uma vez que as equações $w_1 = x + y - z$ e $w_2 = 2y - 3z$ são lineares.

Como veremos adiante, qualquer função linear f pode ser escrita na forma $f(\mathbf{x}) = A\mathbf{x}$, onde A é uma matriz. Por exemplo, a função do item 4 acima pode escrever-se nesta forma, dado que

$$f(x, y, z) = (x + y - z, 2y - 3z) \Longleftrightarrow f(x, y, z) = \begin{bmatrix} 1 & 1 & -1 \\ 0 & 2 & -3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

Pretendemos agora definir funções lineares $T : V \to W$ no caso geral em que V e W são espaços lineares. É claro que a definição deve generalizar o caso em que $V = \mathbb{R}^n$ e $W = \mathbb{R}^p$. Para tal, note-se que quando $T : \mathbb{R}^n \to \mathbb{R}^p$ é definida por $T(\mathbf{x}) = A\mathbf{x}$, onde A é uma matriz real $p \times n$, são satisfeitas as condições

$$T(\mathbf{x}+\mathbf{y}) = A(\mathbf{x}+\mathbf{y}) = A\mathbf{x} + A\mathbf{y} = T(\mathbf{x}) + T(\mathbf{y});$$
$$T(\alpha\mathbf{x}) = A(\alpha\mathbf{x}) = \alpha A\mathbf{x} = \alpha T(\mathbf{x}),$$

para quaisquer vectores $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ e todo o escalar real α. Estas são precisamente as propriedades que iremos usar para definir função linear entre dois espaços vectoriais.

Definição 6.1. Função linear

Seja $T : V \to W$ uma função entre dois espaços lineares V e W sobre o mesmo corpo $\mathbb{K}$. Diz-se que T é uma *função linear*, ou *transformação linear*, se para quaisquer vectores $x, y \in V$, e qualquer escalar $\alpha \in \mathbb{K}$, são verificadas as condições:

i) $T(x+y) = T(x) + T(y)$.

ii) $T(\alpha x) = \alpha T(x)$.

Note-se que as condições i) e ii) da Definição 6.1 são equivalente à condição:

$$T(\alpha x + \beta y) = \alpha T(x) + \beta T(y), \quad \text{para todos os } \alpha, \beta \in \mathbb{K} \text{ e todos } x, y \in V.$$

Esta condição, que caracteriza igualmente uma função linear, significa que uma função é linear se aplica combinações lineares de vectores em combinações lineares dos respectivos vectores imagem.

Exercício 6.1. Sejam V e W espaços lineares sobre $\mathbb{K}$, e $T_1 : V \to W$ e $T_2 : V \to W$ duas funções lineares. Mostre que as funções $(T_1 + T_2)$ e (αT_1), com $\alpha \in \mathbb{K}$, são lineares. Recorde para tal que $(T_1 + T_2)$ e αT_1 são definidas por:

$$(T_1 + T_2)(x) = T_1(x) + T_2(x) \quad \text{e} \quad (\alpha T_1)(x) = \alpha T_1(x).$$

▲

Apresentemos alguns exemplos de funções lineares e não lineares.

Exemplo 6.1. Exemplos de funções lineares

1. **Produto interno**: Seja $T : V \to \mathbb{C}$ a função definida por $T(x) = \langle v, x \rangle$, onde $\langle \, , \, \rangle$ denota um produto interno no espaço linear complexo V e v um vector de V.

   Para mostrar que T é linear basta verificar que, por definição de produto interno, esta função goza das propriedades P2 e P3 da Definição 5.1 (pág. 243). Ou seja,

   $$T(x+y) = \langle v, x+y \rangle = \langle v, x \rangle + \langle v, y \rangle = T(x) + T(y),$$

   e

   $$T(\alpha x) = \langle v, \alpha x \rangle = \alpha \langle v, x \rangle = \alpha T(x),$$

   onde α é um escalar complexo.

2. **A projecção ortogonal sobre um subespaço:** Seja S um subespaço de um espaço linear V munido de um produto interno, e $T : V \to S$ definida por $T(x) = \operatorname{proj}_S x$.

   Para mostrar que T é uma função linear, basta escolher uma base ortogonal de S e utilizar a expressão (5.15) que define $\operatorname{proj}_S x$ (ver pág. 263). Essa expressão, diz-nos que $\operatorname{proj}_S x$ é igual à soma das projecções ortogonais de x sobre os vectores da base, ou seja, uma soma em que as parcelas são da forma $\operatorname{proj}_{v_i} x = \frac{\langle v_i \, , \, x \rangle}{\|v_i\|^2}$ (sendo v_i um vector da base ortogonal escolhida em S). O produto interno é linear (na variável x), e portanto a função $g(x) = \operatorname{proj}_{v_i} x$ é linear. Como a soma de funções lineares é uma função linear (ver Exercício 6.1), tem-se que T é linear.

3. **A operação de derivação**: Considere-se o subespaço $\mathcal{D}$ do espaço linear $\mathcal{F}$ das funções reais de variável real, formado pelas funções diferenciáveis. Seja $T : \mathcal{D} \to \mathcal{F}$ definida por $T(f) = f'$, onde f' designa a derivada de f. Como,

   $$T(f+g) = (f+g)' = f' + g' = T(f) + T(g)$$

   e

   $$T(\alpha f) = (\alpha f)' = \alpha f' = \alpha T(f),$$

   tem-se que T é linear.

4. **O integral de uma função**: Seja V o espaço linear das funções reais integráveis no intervalo $[a, b]$, e T a função definida em V por $T(f) = \int_a^b f(x)\, dx$.

Como,

$$\begin{aligned} T(f+g) &= \int_a^b (f+g)(x)\,dx = \int_a^b f(x)\,dx + \int_a^b g(x)\,dx = T(f)+T(g), \\ T(\alpha f) &= \int_a^b (\alpha f)(x)\,dx = \alpha \int_a^b f(x)\,dx = \alpha T(f), \end{aligned}$$

tem-se que T é linear.

5. **O traço de uma matriz**: Seja $\mathcal{M}_{n\times n}$ o espaço linear das matrizes reais de ordem $n > 1$, e $T : \mathcal{M}_{n\times n} \to \mathbb{R}$ definida por $T(A) = \mathrm{tr}(A)$. Usando a definição de traço de uma matriz (soma das entradas da diagonal principal), é fácil verificar que

$$\begin{aligned} T(A+B) &= \mathrm{tr}(A+B) = \mathrm{tr}(A) + \mathrm{tr}(B) = T(A) + T(B) \\ T(\alpha A) &= \mathrm{tr}(\alpha A) = \alpha\,\mathrm{tr}(A) = \alpha T(A). \end{aligned}$$

Logo, o traço é uma função linear.

6. **Escolha de coordenadas**: Seja V um espaço linear real de dimensão n e $B = (v_1, \ldots, v_n)$ uma base ordenada de V. Verifiquemos que a função $\mathbb{I} : V \to \mathbb{R}^n$, que a cada vector $x \in V$ faz corresponder o vector das coordenadas $\mathbf{x}_B$, isto é, $\mathbb{I}(x) = \mathbf{x}_B$, é linear.

Se $x, y \in V$ são tais que $x = x_1v_1 + \cdots + x_nv_n$ e $y = y_1v_1 + \cdots + y_nv_n$, tem-se

$$\begin{aligned} \mathbb{I}(x) &= \mathbb{I}(x_1v_1 + \cdots + x_nv_n) = (x_1, \ldots, x_n) \\ \mathbb{I}(y) &= \mathbb{I}(y_1v_1 + \cdots + y_nv_n) = (y_1, \ldots, y_n). \end{aligned}$$

Logo,

$$\begin{aligned} \mathbb{I}(x+y) &= \mathbb{I}((x_1+y_1)v_1 + \cdots + (x_n+y_n)v_n) = (x_1+y_1, \ldots, x_n+y_n) \\ &= (x_1, \ldots, x_n) + (y_1, \ldots, y_n) = \mathbb{I}(x) + \mathbb{I}(y) \\ \mathbb{I}(\alpha x) &= \mathbb{I}((\alpha x_1)v_1 + \cdots + (\alpha x_n)v_n) = (\alpha x_1, \ldots, \alpha x_n) \\ &= \alpha(x_1, \ldots, x_n) = \alpha\mathbb{I}(x). \end{aligned}$$

♦

Exemplo 6.2. Exemplos de funções não lineares

1. **Determinante de uma matriz**: Seja $\mathcal{M}_{n\times n}$ o espaço linear das matrizes reais $n \times n$ (com $n > 1$), e $T : \mathcal{M}_{n\times n} \to \mathbb{R}$ definida por $T(X) = \det(X)$.

   Como para qualquer escalar α se tem $\det(\alpha X) = \alpha^n \det(X)$, a função determinante não é linear.

2. **Translação:** Seja V um espaço linear e $T : V \to V$ definida por $T(x) = x + v$, onde v é um vector não nulo fixado em V.

   Para qualquer escalar α, são válidas as igualdades

   $$T(\alpha x) = \alpha x + v \qquad \text{e} \qquad \alpha T(x) = \alpha(x + v).$$

   Logo, $T(\alpha x) \neq \alpha T(x)$, e portanto T não é linear.

♦

Uma propriedade das funções lineares $T : V \to W$, entre espaços lineares V e W, é que a imagem do vector zero de V é o vector zero de W. De facto, como V é um espaço linear, sabemos que: (a) V contém o vector zero; (b) cada vector x de V possui um simétrico $(-x) = (-1)x$, que ainda pertence a V. Logo, usando a definição de função linear, tem-se

$$T(0) = T(x - x) = T(x) - T(x) = 0.$$

Podemos assim enunciar uma condição necessária para uma função ser linear.

Proposição 6.1. Se V e W são espaços lineares e $T : V \to W$ é uma função linear, então

$$T(0) = 0.$$

Exemplo 6.3. A função T definida por $T(x, y) = (x - y, x + y + 1)$ não é linear uma vez que $T(0, 0) = (0, 1) \neq (0, 0)$. ♦

Recordamos a seguir as definições de função composta de duas funções e de inversa de uma função. Mostramos que no caso das funções em causa serem lineares então a função composta e a função inversa também são funções lineares.

Composição e inversão de funções

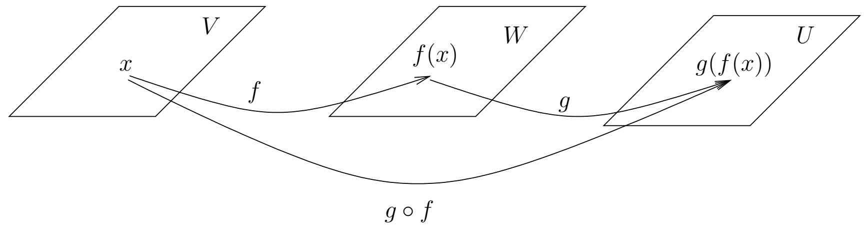


Figura 6.1: A composição $g \circ f$ das funções f e g .

Dadas duas funções $f : V \to W$ e $g : W \to U$ define-se a *composição* $(g \circ f)$ (lê-se "g após f") como sendo a função

$$(g \circ f)(x) = g(f(x)) \quad \text{para todo o } x \in V. \tag{6.1}$$

Por conseguinte, a composição $(g \circ f)$ é uma função de V em U. É claro de (6.1) que $(g \circ f)$ só está definida se o contradomínio de f coincidir com o conjunto de partida de g. Observe o diagrama na Figura 6.1.

Na proposição seguinte mostramos que a composta de duas funções lineares é uma função linear.

Proposição 6.2. Sejam V, W e U espaços lineares. Se $T_1 : V \to W$ e $T_2 : W \to U$ são funções lineares, então $(T_2 \circ T_1) : V \to U$ é uma função linear.

Demonstração. Sejam x e y vectores de V e α um escalar. Da definição (6.1), de função composta, e da definição de função linear, temos

$$\begin{aligned}(T_2 \circ T_1)(x + y) &= T_2(T_1(x + y)) = T_2(T_1(x) + T_1(y)) \\ &= T_2(T_1(x)) + T_2(T_1(y)) = (T_2 \circ T_1)(x) + (T_2 \circ T_1)(y),\end{aligned}$$

e

$$\begin{aligned}(T_2 \circ T_1)(\alpha x) &= T_2(T_1(\alpha x)) = T_2(\alpha T_1(x)) \\ &= \alpha T_2(T_1(x)) = \alpha (T_2 \circ T_1)(x).\end{aligned}$$

Ou seja, $(T_2 \circ T_1)$ satisfaz as duas condições da definição de função linear. □

Dada uma função $f : V \to W$, a inversão de f consiste em determinar uma outra função $g : Im(T) \to V$ tal que a composição $g \circ f$ seja a identidade, isto é, tal que $(g \circ f)(x) = x$ para todo o $x \in V$ (observe a Figura 6.2).

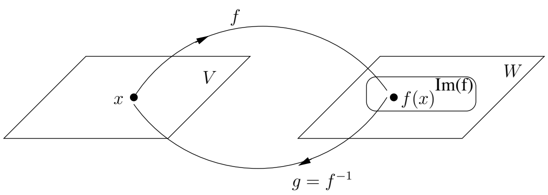


Figura 6.2: A função f e a sua inversa f^{-1}.

Definição 6.2. Inversa de uma função

Uma função $f : V \to W$ diz-se invertível se existe uma função $g : Im(f) \to V$ tal que $(g \circ f)$ é a função identidade em V. Isto é,

$$(g \circ f)(x) = x \text{ para todo } x \in V$$

Uma função g nestas condições designa-se por *função inversa* de f.

Deixamos como exercício mostrar que a inversa de uma função, quando existe, é única.

Exercício 6.2. Suponha que existem duas funções g_1 e g_2 que são funções inversas de uma função invertível f. Mostre que $g_1 = g_2$.

A inversa de uma função invertível é única.

▲

Quando f é invertível, designamos por f^{-1} a função inversa de f.

Note-se que a função inversa de uma função $T : V \to W$ entre espaços lineares, está definida num subespaço de W. De facto, $Im(T) \subseteq W$ é fechado para a adição de vectores e multiplicação por escalares uma vez que, sendo

$$y, w \in Im(T) \Longleftrightarrow y = T(x_1) \text{ e } w = T(x_2) \text{ para alguns } x_1, x_2 \in V, \tag{6.2}$$

se tem

$$y + w = T(x_1) + T(x_2) = T(x_1 + x_2) \Longrightarrow y + w \in Im(T)$$

e

$$\alpha y = \alpha T(x_1) = T(\alpha x_1) \Longrightarrow \alpha y \in Im(T), \text{ com } \alpha \text{ escalar.}$$

Vejamos agora que a inversa de uma função linear ainda é uma função linear.

Proposição 6.3. Seja $T : V \to W$ uma função linear invertível. A função inversa T^{-1} é uma função linear.

Demonstração. Seja $T^{-1} : Im(T) \to V$ a função inversa da função linear T, e $u, v \in Im(T)$, com $u = T(x)$ e $v = T(y)$ para certos $x, y \in V$. Como T é linear, tem-se

$$\begin{aligned} T^{-1}(\alpha u + \beta v) &= T^{-1}(\alpha T(x) + \beta T(y)) = T^{-1}(T(\alpha x + \beta y)) \\ &= \alpha x + \beta y = \alpha T^{-1}(u) + \beta T^{-1}(v), \end{aligned}$$

onde α e β são escalares. Ou seja, T^{-1} é linear. □

A injectividade de uma função f é condição necessária e suficiente para a invertibilidade de f. Recordemos as definições de injectividade[1] e de sobrejectividade de uma função.

Definição 6.3. A função $f : V \to W$ diz-se

(i) *Injectiva* se aplica pontos distintos em pontos distintos. Ou seja, f é injectiva se para todo $x, z \in V$ se tem

$$x \neq z \Longrightarrow f(x) \neq f(z),$$

ou equivalentemente, f é injectiva se: $f(x) = f(z) \Longrightarrow x = z$.

(ii) *Sobrejectiva* se o contradomínio de f é igual a W. Isto é,

$$Im(f) = W.$$

Uma função que é simultaneamente injectiva e sobrejectiva diz-se *bijectiva*.

Proposição 6.4. A função $f : V \to W$ é invertível se e só se f é injectiva.

Demonstração. Suponha-se que f é injectiva e mostremos que f é invertível. Sendo $y \in Im(f)$, existe $x \in V$ tal que $y = f(x)$. Como f é injectiva, a solução x de $y = f(x)$ é única, pelo que a função $g : Im(f) \to V$ tal que $g(y) = x$ está bem definida. Como $g(y) = g(f(x)) = x$ para todo o x, a função g é a inversa de f.

[1] Na terminologia anglo-saxónica uma função injectiva diz-se "one-to-one".

Para a implicação recíproca, suponha-se que f é invertível e mostremos que é injectiva. Sejam $x, z \in V$ tais que $f(x) = f(z)$. Como existe a função inversa, aplicando f^{-1} à igualdade anterior, tem-se $f^{-1}(f(x)) = f^{-1}(f(z))$, ou equivalentemente, $x = z$. Logo, f é injectiva. □

Isomorfismos

Desde o Capítulo 3 que sabemos que, fixada uma base ordenada B num espaço linear V de dimensão n, podemos estabelecer uma correspondência biunívoca entre vectores de um espaço linear real (resp. complexo) V de dimensão n e vectores de $\mathbb{R}^n$ (resp. de $\mathbb{C}^n$). Esta correspondência associa a cada vector $x \in V$ o respectivo vector das coordenadas $\mathbf{x}_B$. No Exemplo 6.1-6 vimos ainda que esta função, $\mathbb{I} : x \mapsto \mathbf{x}_B$, é linear.

A função $\mathbb{I} : x \mapsto \mathbf{x}_B$ é um exemplo de um isomorfismo entre espaços lineares. Como indica a origem etimológica da palavra isomorfo (do grego "iso" quer dizer igual e "morfo" quer dizer forma), dois espaços lineares são isomorfos quando são estruturalmente iguais, no sentido em que qualquer propriedade que é verificada num espaço é igualmente válida no outro espaço. Define-se um isomorfismo entre espaços lineares como sendo uma função bijectiva $f : V \to W$ que preserva a estrutura de espaço linear em V e em W. Ou seja, uma função bijectiva f que preserva as operações de adição e multiplicação por escalares definidas em V e em W, isto é

$$f(x + y) = f(x) + f(y) \quad \text{e} \quad f(\alpha x) = \alpha f(x).$$

As duas condições anteriores significam que f é linear.

Definição 6.4. Isomorfismo entre espaços lineares

Um isomorfismo entre os espaços lineares V e W é uma função $\mathbb{I} : V \to W$ linear e bijectiva.

Quando existe um isomorfismo entre V e W, os espaços lineares V e W dizem-se isomorfos.

Exercício 6.3. Mostre que um isomorfismo entre espaços lineares aplica bases em bases.

▲

Exercício 6.4. Mostre que o espaço linear real $\mathbb{C}$, dos números complexos $z = a + ib$, é isomorfo ao espaço das matrizes reais da forma

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix}.$$

▲

Como a função (linear) que aplica um vector de um espaço linear no respectivo vector das coordenadas, na base ordenada fixada no espaço, é um isomorfismo, conclui-se o que a seguir se enuncia.

Proposição 6.5. Um espaço linear real (resp. complexo) de dimensão n é isomorfo a $\mathbb{R}^n$ (resp. $\mathbb{C}^n$). Um isomorfismo é a função que a cada vector do espaço faz corresponder o vector das coordenadas numa base ordenada fixada no espaço.

Refira-se que o conceito de isomorfismo é mais geral do que o apresentado na Definição 6.4. Podem estabelecer-se isomorfismos entre conjuntos munidos de estruturas distintas da estrutura de espaço linear, como por exemplo, conjuntos munidos da estrutura de grupo, de anel, ou de álgebra de Lie (ver Capítulo 7). Por exemplo, se A e B são conjuntos munidos, respectivamente, das estruturas algébricas $\triangle$ e $\square$, uma função bijectiva do conjunto A no conjunto B que preserve as estruturas, definidas em A e em B diz-se um *isomorfismo*. Isto é, uma função bijectiva $f : A \to B$ que satisfaz

$$f(a \triangle b) = f(a) \square f(b), \quad \text{para todo } a, b \in A\,. \tag{6.3}$$

Se existe um isomorfismo entre os conjuntos $(A, \triangle)$ e $(B, \square)$, munidos das respectivas estruturas, estes conjuntos dizem-se isomorfos.

No caso em que a estrutura em A e em B é a de espaço linear, um isomorfismo é uma função bijectiva que satisfaz a condição (6.3) para as operações de adição de vectores e de multiplicação por escalares. Neste caso, um isomorfismo é uma função linear tal como se considerou na Definição 6.4.

Exercício 6.5. Mostre que a função

$$f : \; x = (x_1, x_2, x_3) \mapsto \begin{bmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{bmatrix} = X$$

é um isomorfismo entre $\mathbb{R}^3$, munido da estrutura de produto vectorial $\times$, e o conjunto $\mathcal{A}_3$ das matrizes reais anti-simétricas de terceira ordem, munido da estrutura $[\,,\,]$ definida por $[A, B] = AB - BA$.

Chama-se *comutador* à operação $[\,,\,]$ que associa ao par de matrizes A, B a matriz $[A, B] = AB - BA$. Como poderá ver no próximo capítulo, $(\mathbb{R}^3, \times)$ e $(\mathcal{A}_3, [\,,\,])$ são álgebras de Lie e a função f é um isomorfismo de álgebras de Lie.

▲

6.2 Matriz que representa uma função linear

O objectivo desta secção é mostrar que qualquer função linear $T : V \to W$, onde V e W são espaços lineares de dimensões n e p respectivamente, pode representar-se por uma (única) matriz A em relação a bases fixadas em V e em W. Isto significa que o vector das coordenadas de $T(x)$ na base fixada em W se pode escrever na forma $A\mathbf{x}_B$, onde $\mathbf{x}_B$ é o vector das coordenadas de x na base fixada em V.

Seja $T : V \to W$ uma função linear, $B = (v_1, v_2, \ldots, v_n)$ uma base ordenada de V e $B' = (w_1, w_2, \ldots, w_p)$ uma base ordenada do espaço de chegada W. Qualquer vector de um espaço linear pode escrever-se (de forma única) como combinação linear dos vectores de uma base desse espaço. Assim, existem escalares c_i únicos tais que

$$x = c_1 v_1 + \cdots + c_n v_n \quad \Longleftrightarrow \quad \mathbf{x}_B = (c_1, \ldots, c_n) \qquad x \in V, \tag{6.4}$$

Como a função T é linear, tem-se

$$w = T(x) = T(c_1 v_1 + \cdots + c_n v_n) = c_1 T(v_1) + c_2 T(v_2) + \cdots + c_n T(v_n).$$

Desta igualdade obtemos

$$\begin{aligned}
\mathbf{w}_{B'} = (\mathbf{T}(x))_{B'} &= c_1(\mathbf{T}(v_1))_{B'} + c_2(\mathbf{T}(v_2))_{B'} + \cdots + c_n(\mathbf{T}(v_n))_{B'} \\
&= \begin{bmatrix} | & | & & | \\ (\mathbf{T}(v_1))_{B'} & (\mathbf{T}(v_2))_{B'} & \cdots & (\mathbf{T}(v_n))_{B'} \\ | & | & & | \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} \quad \text{(Definição 1.11)} \\
&== \begin{bmatrix} | & | & & | \\ (\mathbf{T}(v_1))_{B'} & (\mathbf{T}(v_2))_{B'} & \cdots & (\mathbf{T}(v_n))_{B'} \\ | & | & & | \end{bmatrix} \mathbf{x}_B \quad \text{(por (6.4))} \\
&= A\mathbf{x}_B. \qquad (6.5)
\end{aligned}$$

A matriz A em (6.5) diz-se a matriz que representa a função linear T nas bases B de V e B' de W. Saliente-se que se os espaços lineares V e W são complexos, tem-se

$$\mathbb{C}^n \ni \mathbf{x}_B \longmapsto \mathbf{w}_{B'} = (\mathbf{T}(x))_{B'} = A\mathbf{x}_B \in \mathbb{C}^p,$$

e portanto a matriz A é do tipo $p \times n$.

Para referência futura passamos a enunciar o que acabámos de mostrar.

Definição 6.5. Seja $T : V \to W$ uma função linear, B uma base ordenada do espaço linear V e B' uma base ordenada do espaço linear W.

A matriz A que representa T nas bases, B do espaço de partida e B' do espaço de chegada, é a matriz que para todo $x \in V$ satisfaz

$$(\mathbf{T}(x))_{B'} = A\mathbf{x}_B, \tag{6.6}$$

onde $\mathbf{x}_B$ e $(\mathbf{T}(x))_{B'}$ designam, respectivamente, os vectores das coordenadas de x na base B e de $T(x)$ na base B'. Se $\dim V = n$ e $\dim W = p$, a matriz A é do tipo $p \times n$.

Para designar a matriz A em (6.6) usa-se a notação $M(T, B, B')$.

Nota 36. *Para uma função linear $T : V \to V$ em que se fixa a mesma base B no espaço de partida e de chegada, referimo-nos à matriz que representa T simplesmente como a "matriz que representa T na base B".*

Proposição 6.6. Matriz que representa uma função linear

A matriz que representa a função linear $T : V \to W$ na base ordenada $B = (v_1, \ldots, v_n)$ de V e na base ordenada B' de W, é a matriz A cujas colunas são os vectores das coordenadas $(\mathbf{T}(v_1))_{B'}, (\mathbf{T}(v_2))_{B'}, \ldots, (\mathbf{T}(v_n))_{B'}$ (por esta ordem). Ou seja,

$$A = M(T, B, B') = \begin{bmatrix} | & | & & | \\ (\mathbf{T}(v_1))_{B'} & (\mathbf{T}(v_2))_{B'} & \cdots & (\mathbf{T}(v_n))_{B'} \\ | & | & & | \end{bmatrix}. \tag{6.7}$$

Nota 37. *Existe uma única matriz que representa uma função linear $T : V \to W$ em relação a bases ordenadas fixas em V e em W. A unicidade desta matriz é consequência da unicidade dos vectores das coordenadas que constituem as colunas da matriz.*

Na Figura 6.3 ilustramos a definição de matriz que representa uma função linear em relação a bases fixadas no espaço de partida V e no espaço de chegada W.

Exemplo 6.4. Considere-se a função linear $T(x, y, z) = (2x-y, 3z+y)$. Podemos escrever

$$T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} 2 & -1 & 0 \\ 0 & 1 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = A \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

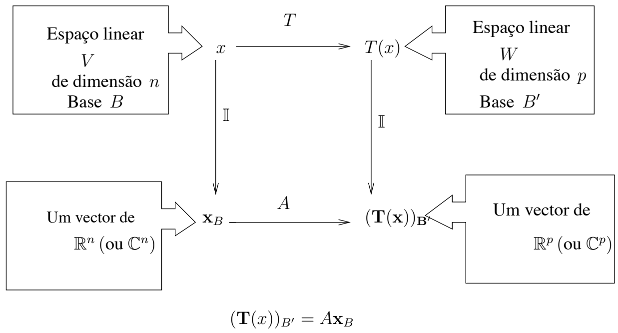


Figura 6.3: A matriz A, do tipo $p \times n$, representa a função linear $T : V \to W$ em relação à base B de V e à base B' de W. Os espaços V e W têm dimensões n e p, respectivamente. O isomorfismo $\mathbb{I}$ é a função (linear) que aplica cada vector de um espaço linear no vector das coordenadas na base fixada nesse espaço.

Ou seja, a imagem por T de um qualquer vector de $\mathbb{R}^3$ pode obter-se multiplicando a matriz A por esse vector.

A matriz A representa T em relação à base canónica de $\mathbb{R}^3$ (espaço de partida) e à base canónica de $\mathbb{R}^2$ (espaço de chegada), uma vez que as colunas da matriz A são as imagens dos vectores da base canónica de $\mathbb{R}^3$,

$$T(1,0,0) = (2,0), \quad T(0,1,0) = (-1,1) \quad \text{e} \quad T(0,0,1) = (0,3).$$

♦

Exemplo 6.5. Consideremos a função $T : P_3 \to P_2$, definida por $T(p(t)) = 3p'(t)$, onde P_n designa o espaço linear dos polinómios de grau menor ou igual a n, e $p'(t)$ designa a derivada de p. Fixemos as bases $B_1 = (1-t, t^2, 2t^3, 1+t)$ em P_3, e $B_2 = (2, t+1, t-t^2)$ em P_2. Pretende-se determinar a matriz que representa T em relação a estas bases. Para tal, calculemos as coordenadas dos vectores

$$T(1-t) = -3, \qquad T(t^2) = 6t, \qquad T(2t^3) = 18t^2, \qquad T(1+t) = 3,$$

na base B_2. As coordenadas de $T(1-t) = -3$ podem obter-se da seguinte forma:

$$\begin{aligned} -3 &= 2\alpha + \beta(t+1) + \gamma(t-t^2) = (2\alpha+\beta) + (\beta+\gamma)t - \gamma t^2 \\ &\iff \begin{cases} 2\alpha + \beta = -3 \\ \beta + \gamma = 0 \\ -\gamma = 0 \end{cases} \iff \alpha = -3/2 \ \text{e} \ \beta = \gamma = 0. \end{aligned}$$

Por conseguinte, $(\mathbf{T}(1-t))_{B_2} = (-3/2, 0, 0)$. Calculando de igual modo as restantes colunas da matriz pretendida, obtemos

$$(\mathbf{T}(t^2))_{B_2} = (-3, 6, 0),\ (\mathbf{T}(2t^3))_{B_2} = (-9, 18, -18) \text{ e } (\mathbf{T}(1+t))_{B_2} = (3/2, 0, 0).$$

Assim, a matriz que representa T nas bases fixadas, é

$$M(T, B_1, B_2) = \begin{bmatrix} -3/2 & -3 & -9 & 3/2 \\ 0 & 6 & 18 & 0 \\ 0 & 0 & -18 & 0 \end{bmatrix}.$$

Vejamos agora como usar esta matriz para calcular a imagem por T de um certo polinómio de P_3. Por exemplo, pretende-se usar $M(T, B_1, B_2)$ para determinar $T(q)$ onde $q(t) = 2 + 2t - 5t^2 + 4t^3$.

Comecemos por notar que o vector das coordenadas de q na base B_1 é $\mathbf{q}_{B_1} = (0, -5, 2, 2)$. Por definição de matriz que representa uma função linear, tem-se

$$(\mathbf{T}(q))_{B_2} = \begin{bmatrix} -3/2 & -3 & -9 & 3/2 \\ 0 & 6 & 18 & 0 \\ 0 & 0 & -18 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ -5 \\ 2 \\ 2 \end{bmatrix}_{B_1} = \begin{bmatrix} 0 \\ 6 \\ -36 \end{bmatrix}_{B_2}.$$

Assim, como $(\mathbf{T}(q))_{B_2} = (0, 6, -36)$, a imagem de q por T é

$$T(q(t)) = 6(t+1) - 36(t - t^2) = 6 - 30t + 36t^2.$$

De facto, usando a expressão de T podemos confirmar este resultado:

$$T(q(t)) = 3q'(t) = 3(2 + 2t - 5t^2 + 4t^3)' = 3(2 - 10t + 12t^2) = 6 - 30t + 36t^2.$$

♦

Exemplo 6.6. Considere-se a função $T(f) = \int_0^t f(x)dx$, definida no espaço P_2 dos polinómios de grau menor ou igual a 2. É claro que a imagem de um polinómio de P_2 é um polinómio de grau menor ou igual a 3, e portanto podemos tomar para espaço de chegada o espaço P_3, ou seja, considerar $T : P_2 \to P_3$. Fixemos em P_2 e em P_3 as respectivas bases canónicas: $BC_{P_2} = (1, t, t^2)$ e $BC_{P_3} = (1, t, t^2, t^3)$. Como

$$T(1) = \int_0^t 1\, dx = t, \quad T(t) = \int_0^t x\, dx = \frac{t^2}{2} \quad \text{e} \quad T(t^2) = \int_0^t x^2\, dx = \frac{t^3}{3},$$

as colunas da matriz que representa T nas bases fixadas são os vectores das coordenadas de $T(1), T(t)$ e $T(t^2)$ na base BC_{P_3}, ou seja, respectivamente os vectores $(0,1,0,0), (0,0,1/2,0)$ e $(0,0,0,1/3)$. Assim, a matriz que representa o integral indefinido (como função de P_2 em P_3) relativamente às bases consideradas, é

$$M = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/3 \end{bmatrix}. \tag{6.8}$$

♦

Alguns exemplos geométricos de funções lineares

Nesta secção pretendemos identificar as matrizes que representam (na base canónica) certas funções lineares definidas em $\mathbb{R}^2$ e $\mathbb{R}^3$. Como vimos, para obter essa matriz basta conhecermos as imagens dos vectores da base canónica. Em todos os exemplos que se seguem consideramos fixada a base canónica, quer no espaço de partida quer no espaço de chegada.

Exemplo 6.7. 1. $T : \mathbb{R}^2 \to \mathbb{R}^2$ **é uma reflexão em relação ao eixo dos** xx (Figura 6.4).

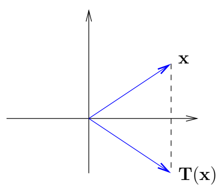


Figura 6.4: Reflexão em relação ao eixo dos xx.

Como

$$T(1,0) = (1,0) \quad \text{e} \quad T(0,1) = (0,-1),$$

a matriz que representa T é

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

Ou seja, $T(x,y) = A \begin{bmatrix} x \\ y \end{bmatrix} = (x,-y)$.

2. $T : \mathbb{R}^2 \to \mathbb{R}^2$ **é uma reflexão em relação à recta** $y = x$ (Figura 6.5).

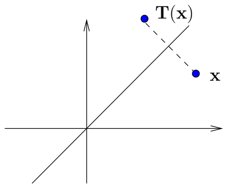


Figura 6.5: Reflexão em relação à recta $y = x$.

Como $T(1,0) = (0,1)$ e $T(0,1) = (1,0)$, a matriz que representa T é

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

Portanto, $T(x, y) = (y, x)$.

3. $T : \mathbb{R}^2 \to \mathbb{R}^2$ **é a projecção ortogonal sobre o eixo dos** yy (Figura 6.6).

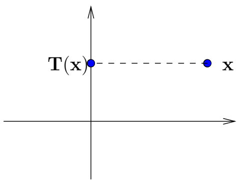


Figura 6.6: Projecção ortogonal sobre o eixo dos yy.

Como $T(1,0) = (0,0)$ e $T(0,1) = (0,1)$, a matriz que representa T é

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

Assim, $T(x, y) = (0, y)$.

4. $T : \mathbb{R}^2 \to \mathbb{R}^2$ **é uma rotação em torno da origem de um ângulo** θ **no sentido directo** (Figura 6.7).

A imagem de $(1,0)$ por T é obviamente $T(1,0) = (\cos\theta, \operatorname{sen}\theta)$ e a de $(0,1)$ é $T(0,1) = (-\operatorname{sen}\theta, \cos\theta)$. Logo, a matriz que representa T é

$$R_\theta = \begin{bmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{bmatrix}.$$

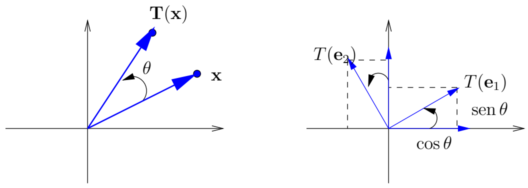


Figura 6.7: Rotação em torno da origem de um ângulo θ no sentido directo.

Por exemplo, a rotação em torno da origem de um ângulo $\pi/4$ no sentido directo (ou anti-horário, ou positivo) é

$$T(\mathbf{x}) = R_{\pi/4}\mathbf{x} = \frac{\sqrt{2}}{2}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}\mathbf{x},$$

ou seja, $T(x, y) = \frac{\sqrt{2}}{2}(x - y, x + y)$.

♦

Exercício 6.6. Considere as funções lineares $T : \mathbb{R}^3 \to \mathbb{R}^3$ definidas a seguir. Mostre que as matrizes M indicadas representam T em relação à base canónica de $\mathbb{R}^3$.

1) T é uma reflexão em relação ao plano xy, $M = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$.

2) T é a projecção ortogonal sobre o plano yz, $M = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

3) T é uma rotação em torno do eixo dos zz de um ângulo θ no sentido positivo, $M = \begin{bmatrix} \cos\theta & -\operatorname{sen}\theta & 0 \\ \operatorname{sen}\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

Verifique ainda que a imagem de um qualquer ponto $\mathbf{x} = (x, y, z)$ é

$$T(x, y, z) = (x\cos\theta - y\operatorname{sen}\theta, x\operatorname{sen}\theta + y\cos\theta, z).$$

O *sentido directo* (ou positivo) de uma rotação em torno de um eixo é estabelecido da seguinte forma: um observador, alinhado com o eixo, com os pés na origem e a cabeça direccionada no sentido positivo do eixo, vê a rotação no plano perpendicular ao eixo realizar-se no sentido anti-horário.

4) T é uma rotação em torno do eixo dos yy de um ângulo θ no sentido directo,

$$M = \begin{bmatrix} \cos\theta & 0 & \operatorname{sen}\theta \\ 0 & 1 & 0 \\ -\operatorname{sen}\theta & 0 & \cos\theta \end{bmatrix}.$$

▲

Para finalizar esta secção refira-se que o conjunto

$$L(V,W) = \{T : V \to W;\ T \text{ é linear}\}$$

é um espaço linear. O conjunto $L(V,W)$ é um subconjunto do espaço linear $\mathcal{F}$ das funções $f : V \to W$, entre espaços lineares sobre o mesmo corpo $\mathbb{K}$, com as operações de adição e multiplicação por escalares definidas por

$$(f+g)(v) = f(v) + g(v) \quad (\alpha f)(v) = \alpha f(v), \quad \text{para todo } f, g \in \mathcal{F} \text{ e } \alpha \in \mathbb{K}.$$

Como a soma de funções lineares é uma função linear e a multiplicação de uma função linear por um escalar é uma função linear (ver Exercício 6.1), tem-se que $L(V,W)$ é fechado para estas operações, e portanto um subespaço de $\mathcal{F}$.

Sejam V e W espaços lineares de dimensões $\dim V = n$ e $\dim W = p$. Fixadas bases ordenadas B e B', respectivamente, em V e em W, é fácil concluir que a aplicação que a cada $T \in L(V,W)$ faz corresponder a matriz $M(T,B,B')$ do tipo $p \times n$, é uma função linear bijectiva. Ou seja, $L(V,W)$ é um espaço linear isomorfo ao espaço $\mathcal{M}_{p\times n}$, das matrizes $p\times n$ (reais se V e W são espaços lineares reais ou complexas se estes espaços forem complexos).

Enunciamos na proposição seguinte este resultado.

Proposição 6.7. Sejam V e W espaços lineares sobre o mesmo corpo $\mathbb{K}$, de dimensões $\dim V = n$ e $\dim W = p$. Fixadas bases ordenadas B em V e B' em W, o espaço linear

$$L(V, W) = \{T : V \to W; \quad T \text{ é linear}\}$$

é isomorfo ao espaço linear $\mathcal{M}_{p\times n}(\mathbb{K})$, das matrizes do tipo $p \times n$ com entradas em $\mathbb{K}$.

O isomorfismo é a bijecção que a cada função $T : V \to W$ de $L(V, W)$ faz corresponder a matriz $M(T, B, B')$, que representa T nas bases ordenadas B de V e B' de W. Ou seja, o isomorfismo $\mathbb{I} : L(V, W) \to \mathcal{M}_{p\times n}$, é definido por

$$T \longmapsto M(T, B, B').$$

Matrizes que representam uma função linear em relação a bases distintas

Pretendemos agora relacionar matrizes que representam uma dada função linear em relação a bases distintas. Para tal, suponha-se que a função linear $T : V \to W$ é representada pela matriz A em relação às bases B_1 de V e $\tilde{B}_1$ de W, e que T é representada pela matriz C em relação às bases B_2 de V e $\tilde{B}_2$ de W. Isto é,

$$(\mathbf{T}(x))_{\tilde{B}_1} = A\mathbf{x}_{B_1} \quad \text{e} \quad (\mathbf{T}(x))_{\tilde{B}_2} = C\mathbf{x}_{B_2} \quad \text{para todo } x \in V. \tag{6.9}$$

Sendo $y = T(x)$, então $(\mathbf{T}(x))_{\tilde{B}_1} = \mathbf{y}_{\tilde{B}_1}$ e $(\mathbf{T}(x))_{\tilde{B}_2} = \mathbf{y}_{\tilde{B}_2}$ representam o mesmo vector y, respectivamente, nas bases $\tilde{B}_1$ e $\tilde{B}_2$. Assim, os vectores $\mathbf{y}_{\tilde{B}_1}$ e $\mathbf{y}_{\tilde{B}_2}$ estão relacionados através da matriz $M = M_{\tilde{B}_2 \leftarrow \tilde{B}_1}$ que realiza a mudança da base $\tilde{B}_1$ para $\tilde{B}_2$ (ver Definição 3.9 na página 157). Nomeadamente,

$$\mathbf{y}_{\tilde{B}_2} = M\mathbf{y}_{\tilde{B}_1}.$$

De igual modo, os vectores $\mathbf{x}_{B_1}$ e $\mathbf{x}_{B_2}$ estão relacionados através da matriz $S = S_{B_2 \leftarrow B_1}$ que realiza a mudança da base B_1 para B_2:

$$\mathbf{x}_{B_2} = S\mathbf{x}_{B_1}.$$

Usando (6.9), temos

$$\mathbf{y}_{\tilde{B}_2} = C\mathbf{x}_{B_2} \Longleftrightarrow M\mathbf{y}_{\tilde{B}_1} = CS\mathbf{x}_{B_1} \Longleftrightarrow MA\mathbf{x}_{B_1} = CS\mathbf{x}_{B_1}.$$

Como as matrizes de mudança de base são invertíveis e a equivalência acima é verificada para todo $x \in V$, as matrizes A e C (que representam a mesma função linear em relação a bases distintas) satisfazem a igualdade

$$MA = CS \Longrightarrow A = M^{-1}CS.$$

Relembremos que a matriz de mudança da base ordenada $B = (v_1, \ldots, v_k)$ para a base ordenada $\tilde{B}$ é a matriz cujas colunas são os vectores das coordenadas, $(\mathbf{v}_1)_{\tilde{B}}, \ldots, (\mathbf{v}_k)_{\tilde{B}}$ (por esta ordem). Tendo em conta a forma da matriz em (6.7), que define uma função linear em relação a bases fixadas no espaço de partida e de chegada, a matriz de mudança de base $M_{\tilde{B}\leftarrow B}$ pode interpretar-se, em termos de funções lineares, como sendo a matriz que representa a função (linear) identidade $I : V \to V$, $x \mapsto x$, em relação à base B no espaço de partida e à base $\tilde{B}$ no espaço de chegada. Ou seja,

Num espaço linear W, a matriz de mudança da base ordenada B de W para a base ordenada $\tilde{B}$ de W, é a matriz que representa a função identidade $I : W \to W$, $x \mapsto x$, em relação à base B no espaço de partida e à base $\tilde{B}$ no espaço de chegada. Isto é,

$$M_{\tilde{B}\leftarrow B} = M(I, B, \tilde{B}).$$

Considere-se o exemplo seguinte.

Exemplo 6.8. Seja $T : \mathbb{R}^3 \to \mathbb{R}^2$ a função linear definida por

$$\begin{aligned} T : \mathbb{R}^3 &\longrightarrow \mathbb{R}^2 \\ (x, y, z) &\longmapsto (x + y, y - z). \end{aligned}$$

Como

$$T(x, y, z) = (x + y, y - z) = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix},$$

a matriz que representa T em relação à base canónica de $\mathbb{R}^3$ e à base canónica de $\mathbb{R}^2$ é a matriz

$$A = M(T, BC_{\mathbb{R}^3}, BC_{\mathbb{R}^2}) = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & -1 \end{bmatrix}.$$

Fixemos agora em $\mathbb{R}^3$ e em $\mathbb{R}^2$ as bases ordenadas

$$B_1 = ((1, 1, 0), (0, 1, 1), (1, 1, 1)), \qquad B_2 = ((2, 1), (1, -1)).$$

Determinemos a matriz que representa T em relação a estas bases. Como

$$T(1,1,0) = (2,1) = \mathbf{u}, \quad T(0,1,1) = (1,0) = \mathbf{v}, \quad T(1,1,1) = (2,0) = \mathbf{w},$$

e

$$\begin{aligned} (2,1) &= 1 \times (2,1) + 0 \times (1,-1) \\ (1,0) &= \tfrac{1}{3}(2,1) + \tfrac{1}{3}(1,-1) \\ (2,0) &= \tfrac{2}{3}(2,1) + \tfrac{2}{3}(1,-1), \end{aligned}$$

temos $\mathbf{u}_{B_2} = (1,0)$, $\mathbf{v}_{B_2} = (1/3, 1/3)$ e $\mathbf{w}_{B_2} = (2/3, 2/3)$. Logo,

$$C = M(T, B_1, B_2) = \begin{bmatrix} 1 & \frac{1}{3} & \frac{2}{3} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{bmatrix}.$$

As matrizes A e C estão relacionadas através de matrizes de mudança de base S e M. O diagrama que apresentamos a seguir ajuda a estabelecer a relação entre estas matrizes.

$$\begin{tikzcd} \mathbb{R}^3_{BC} \arrow[r, "T", "A"'] \arrow[d, "S=M(I,BC,B_1)"', "I_{\mathbb{R}^3}"] & \mathbb{R}^2_{BC} \arrow[d, "I_{\mathbb{R}^2}"', "M=M(I,BC,B_2)"] \\ \mathbb{R}^3_{B_1} \arrow[r, "T", "C"'] & \mathbb{R}^2_{B_2} \end{tikzcd} \tag{6.10}$$

Na horizontal temos a função linear T com duas representações matriciais distintas A e C. Na vertical, à esquerda, temos a matriz S que realiza a mudança da base BC de $\mathbb{R}^3$ para a base B_1, que é igualmente a matriz que representa a função identidade $I_{\mathbb{R}^3} : \mathbb{R}^3 \to \mathbb{R}^3$ em relação à base BC no espaço de partida e à base B_1 no espaço de chegada. Na vertical, à direita, temos a matriz M que realiza a mudança da base BC de $\mathbb{R}^2$ para a base B_2, que também é a matriz que representa a função identidade $I_{\mathbb{R}^2} : \mathbb{R}^2 \to \mathbb{R}^2$ em relação à base BC no espaço de partida e à base B_2 no espaço de chegada.

Dado um qualquer vector $\mathbf{x} = \mathbf{x}_{BC} \in \mathbb{R}^3$, a sua imagem por T, $\mathbf{y}_{B_2} = (\mathbf{T}(\mathbf{x}))_{B_2}$, pode obter-se do diagrama (6.10) de duas formas distintas:

$$\mathbf{x}_{BC} \longmapsto A\mathbf{x}_{BC} \longmapsto MA\mathbf{x}_{BC} = \mathbf{y}_{B_2} \tag{6.11}$$

ou

$$\mathbf{x}_{BC} \longmapsto S\mathbf{x}_{BC} \longmapsto CS\mathbf{x}_{BC} = \mathbf{y}_{B_2}. \tag{6.12}$$

Isto é, $MA\mathbf{x}_{BC} = CS\mathbf{x}_{BC}$ para todo $\mathbf{x} \in \mathbb{R}^3$. Ou seja, $MA = CS$.

As aplicações em (6.11), correspondem a "ler" o diagrama (no sentido das setas) primeiro na horizontal e em seguida na vertical. Enquanto que em (6.12) o diagrama é percorrido primeiro na vertical e depois na horizontal.

Um diagrama (direccionado) deste tipo diz-se *comutativo* se o resultado da composição de aplicações correspondentes a qualquer percurso entre dois vértices é igual.

Determinemos agora as matrizes S e M de mudança de base. Como

$$\begin{aligned}
(1,0,0) &= 0 \times (1,1,0) - (0,1,1) + (1,1,1), & (1,0) &= \frac{1}{3}(2,1) + \frac{1}{3}(1,-1)\\
(0,1,0) &= (1,1,0) + (0,1,1) - (1,1,1), & (0,1) &= \frac{1}{3}(2,1) - \frac{2}{3}(1,-1)\\
(0,0,1) &= -(1,1,0) + 0 \times (0,1,1) + (1,1,1)
\end{aligned}$$

temos

$$S = M(I_{\mathbb{R}^3}, BC, B_1) = \begin{bmatrix} 0 & 1 & -1 \\ -1 & 1 & 0 \\ 1 & -1 & 1 \end{bmatrix}$$

$$M = M(I_{\mathbb{R}^2}, BC, B_2) = \begin{bmatrix} 1/3 & 1/3 \\ 1/3 & -2/3 \end{bmatrix}.$$

Deixamos como exercício a confirmação da igualdade $A = M^{-1}CS$. ♦

No caso geral de uma função linear $T : V \to W$, com V e W espaços lineares reais de dimensões $\dim V = n$ e $\dim W = p$, usamos também um *diagrama comutativo* para ilustrar a relação entre as matrizes que representam T em relação a bases distintas fixadas em V e em W. Designamos por $\mathbb{R}^n_{(U,B)}$ uma cópia de $\mathbb{R}^n$ cujos elementos são vistos como vectores das coordenadas de vectores de U na base B.

Na Figura 6.8 apresentamos à esquerda o *diagrama comutativo* que relaciona as matrizes A e C que representam a função linear $T : V \to W$ em relação a bases fixadas em V e W, e na mesma figura à direita é ilustrado o significado da comutatividade do diagrama.

Obviamente que se T é uma função linear em que o espaço de partida e o de chegada são o mesmo espaço (isto é, $T : V \to V$), e se neste espaço fixarmos a mesma base, o diagrama comutativo anterior reduz-se ao diagrama da Figura 6.9.

$$\begin{tikzcd} \mathbb{R}^n_{(V,B_1)} \arrow[r, "A"'] \arrow[d, "S"'] & \mathbb{R}^p_{(W,\tilde{B}_1)} \arrow[d, "M"] \\ \mathbb{R}^n_{(V,B_2)} \arrow[r, "C"'] & \mathbb{R}^p_{(W,\tilde{B}_2)} \end{tikzcd} \qquad \begin{tikzcd} \mathbf{x}_{B_1} \arrow[r, mapsto] \arrow[d, mapsto] & A\mathbf{x}_{B_1} \arrow[d, mapsto] \\ S\mathbf{x}_{B_1} \arrow[r, mapsto] & CS\mathbf{x}_{B_1} = MA\mathbf{x}_{B_1} \end{tikzcd}$$

Figura 6.8: As matrizes A e C, que representam a função linear T em relação a bases distintas, estão relacionadas através das matrizes de mudança de base M e S. O diagrama do lado esquerdo é comutativo, e portanto $CS = MA$, ou equivalentemente $C = MAS^{-1}$.

$$\begin{tikzcd} \mathbb{R}^n_{(V,B_1)} \arrow[r, "A"'] \arrow[d, "M"'] & \mathbb{R}^n_{(V,B_1)} \arrow[d, "M"] \\ \mathbb{R}^n_{(V,B_2)} \arrow[r, "C"'] & \mathbb{R}^n_{(V,B_2)} \end{tikzcd} \qquad \begin{tikzcd} \mathbf{x}_{B_1} \arrow[r, mapsto] \arrow[d, mapsto] & A\mathbf{x}_{B_1} \arrow[d, mapsto] \\ M\mathbf{x}_{B_1} \arrow[r, mapsto] & CM\mathbf{x}_{B_1} = MA\mathbf{x}_{B_1} \end{tikzcd}$$

Figura 6.9: As matrizes A e C que representam $T : V \to V$ são semelhantes: $C = MAM^{-1}$.

Ou seja, as matrizes $A = M(T, B_1, B_1)$ e $C = M(T, B_2, B_2)$, que representam $T : V \to V$, respectivamente, em relação às bases B_1 e B_2 fixadas em V são matrizes semelhantes, isto é, $C = MAM^{-1}$ (recorde a Definição 4.8 de matrizes semelhantes, página 188). Para referência futura, enunciamos este resultado.

Proposição 6.8. As matrizes $A = M(T, B_1, B_1)$ e $C = M(T, B_2, B_2)$, que representam a função linear $T : V \to V$ em relação às bases B_1 e B_2 fixadas em V, são matrizes semelhantes. Isto é,

$$C = MAM^{-1},$$

onde M é a matriz que realiza a mudança da base B_1 para a base B_2.

Nota 38. *Uma função linear $T : V \to V$ diz-se um endomorfismo. Ou seja, um endomorfismo é uma função linear do espaço linear V em si próprio.*

Matriz que representa a função composta

A partir da definição de matriz que representa uma função linear, é fácil determinar a matriz que representa a composta de duas funções lineares. Sejam $T_1 : V \to W$ e $T_2 : W \to U$ duas funções lineares, $A = M(T_1, B_1, B_2)$ a matriz que representa T_1 em relação à base B_1 de V e à base B_2 de W, e $C = M(T_2, B_2, B_3)$ a matriz que representa T_2 em relação à base B_2 de W e B_3 de U. Se A é a matriz que representa T_1, e C é a matriz que representa T_2, tem-se $(\mathbf{T_1}(x))_{B_2} = A\mathbf{x}_{B_1}$ e $(\mathbf{T_2}(y))_{B_3} = C\mathbf{y}_{B_2}$. Além disso, como $(T_2 \circ T_1)(x) = T_2(T_1(x))$, resulta

$$((\mathbf{T_2} \circ \mathbf{T_1})(x))_{B_3} = C(A\mathbf{x}_{B_1}) = (CA)\mathbf{x}_{B_1}.$$

Conclui-se portanto que a matriz CA representa $(T_2 \circ T_1)$ sempre que A representa T_1 e C representa T_2. Os diagramas seguintes ilustram este facto.

$$\begin{tikzcd} (\mathbf{V}, B_1) \arrow[r, "T_1"] \arrow[dr, "(T_2 \circ T_1)"'] & (\mathbf{W}, B_2) \arrow[d, "T_2"] \\ & (\mathbf{U}, B_3) \end{tikzcd} \qquad \begin{tikzcd} \mathbb{R}^n_{(\mathbf{V}, B_1)} \arrow[r, "A"] \arrow[dr, "CA"'] & \mathbb{R}^p_{(\mathbf{W}, B_2)} \arrow[d, "C"] \\ & \mathbb{R}^k_{(\mathbf{U}, B_3)} \end{tikzcd}$$

Proposição 6.9. Sejam $T_1 : V \to W$ e $T_2 : W \to U$ duas funções lineares, $A = M(T_1, B_1, B_2)$ a matriz que representa T_1 em relação à base B_1 de V e à base B_2 de W, e $C = M(T_2, B_2, B_3)$ a matriz que representa T_2 em relação à base B_2 de W e B_3 de U.

A matriz que representa $(T_2 \circ T_1) : V \to U$ em relação à base B_1 de V e B_3 de U, é a matriz CA, isto é,

$$CA = M(T_2 \circ T_1, B_1, B_3).$$

6.3 Núcleo e contradomínio de uma função linear

Vimos que existe um isomorfismo entre o espaço das funções lineares e o espaço das matrizes (ver Proposição 6.7). Este isomorfismo aplica uma dada função linear $T : V \to W$ na matriz $M(T, B, B')$, que representa T em relação às bases B e B' fixadas em V e em W. É assim de presumir que este isomorfismo (função linear bijectiva) permita traduzir vários resultados obtidos para matrizes em termos de funções lineares. A seguir estabelecemos algumas destas relações para o contradomínio e para o núcleo de uma função linear. Comecemos por definir núcleo de uma função linear entre espaços vectoriais.

Definição 6.6. Seja $T : V \to W$ uma função linear entre espaços lineares. Chamamos *núcleo* de T ao conjunto dos elementos de V que são aplicados em $0 \in W$. Designamos por $N(T)$ o núcleo de T.

Note-se que se $A = M(T, BC_{\mathbb{R}^n}, BC_{\mathbb{R}^p})$ é a matriz que representa $T : \mathbb{R}^n \to \mathbb{R}^p$ em relação às bases canónicas de $\mathbb{R}^n$ e de $\mathbb{R}^p$, isto é, $T(\mathbf{x}) = A\mathbf{x}$, então a definição de núcleo de T está de acordo com a definição de núcleo da matriz A (ver Exemplo 6.9 a seguir).

O núcleo de uma função linear é um subconjunto do espaço de partida, e o contradomínio um subconjunto do espaço de chegada (observe a Figura 6.10).

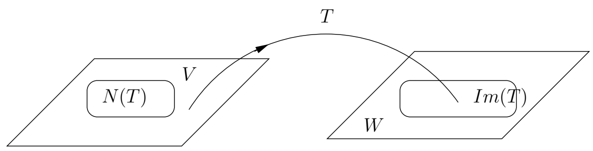


Figura 6.10: O núcleo e o contradomínio de uma função linear T.

Exemplo 6.9. Seja $T : \mathbb{R}^n \to \mathbb{R}^p$ definida por $T(\mathbf{x}) = A\mathbf{x}$. Por definição, o núcleo de T é

$$\begin{aligned} N(T) &= \{\mathbf{x} \in \mathbb{R}^n : T(\mathbf{x}) = \mathbf{0}\} \\ &= \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{0}\} = N(A), \end{aligned}$$

e o contradomínio

$$\begin{aligned} Im(T) &= \{\mathbf{y} \in \mathbb{R}^p : \mathbf{y} = T(\mathbf{x}) \text{ para algum } \mathbf{x} \in \mathbb{R}^n\} \\ &= \{\mathbf{y} \in \mathbb{R}^p : \mathbf{y} = A\mathbf{x} \text{ para algum } \mathbf{x} \in \mathbb{R}^n\} = EC(A). \end{aligned}$$

Conclui-se que o núcleo de T é o núcleo da matriz A, e o contradomínio de T é o conjunto dos vectores $\mathbf{y} \in \mathbb{R}^p$ que são combinação linear das colunas de A, ou seja, o espaço das colunas de A. Na Figura 6.11 esquematizam-se as noções de núcleo e de contradomínio de T.

♦

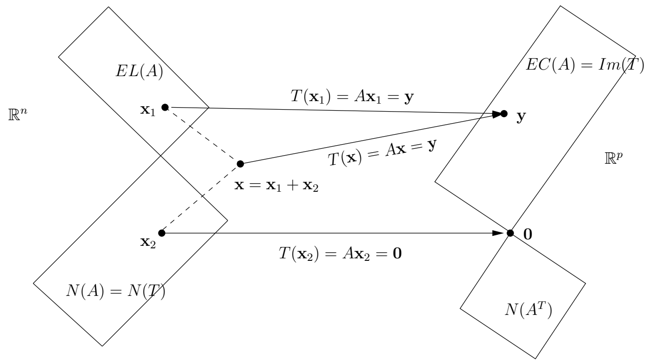


Figura 6.11: O núcleo e o contradomínio da função linear T definida por $T(\mathbf{x}) = A\mathbf{x}$.

Exemplo 6.10. a) Seja $T : \mathbb{R}^2 \to \mathbb{R}^2$ a rotação de um ângulo θ em torno da origem. Como o único vector de $\mathbb{R}^2$ que é aplicado por T em $(0,0)$ é a origem, tem-se $N(T) = \{\mathbf{0}\}$. Além disso, qualquer vector de $\mathbb{R}^2$ é imagem por T de algum outro vector de $\mathbb{R}^2$, logo $Im(T) = \mathbb{R}^2$.

b) Seja $T : \mathbb{R}^3 \to \mathbb{R}^3$ a projecção ortogonal sobre o plano xy.

Como os únicos vectores de $\mathbb{R}^3$ que são aplicados por T na origem são os vectores do eixo dos zz, o núcleo de T é o eixo dos zz. Por outro lado, qualquer vector de $\mathbb{R}^3$ é aplicado num vector do plano xy, logo o conjunto imagem é o plano xy. Ou seja,

$$N(T) = \{(0,0,z) : \ z \in \mathbb{R}\} \quad \text{e} \quad Im(T) = \{(x,y,0) : \ x,y \in \mathbb{R}\}.$$

c) Seja $T : \mathbb{R}^2 \to \mathbb{R}^2$ a reflexão em relação ao eixo dos xx. É fácil ver que $N(T) = \{\mathbf{0}\}$ e $Im(T) = \mathbb{R}^2$.

♦

Proposição 6.10. Sejam V e W espaços lineares. O núcleo e o contradomínio de uma função linear $T : V \to W$ são subespaços lineares de V e de W, respectivamente.

Demonstração. Já mostrámos, na página 330, que $Im(T)$ é um subespaço de W. Para provar que $N(T)$ é um subespaço de V basta verificar que é fechado para a adição de vectores e para a multiplicação por escalares. Sejam $x, v \in N(T)$, isto é

$$x, v \in N(T) \Longleftrightarrow T(x) = 0 \text{ e } T(v) = 0,$$

Da linearidade de T, resulta

$$T(x+v) = T(x) + T(v) = 0 + 0 = 0 \Longrightarrow x + v \in N(T)$$

e

$$T(\alpha x) = \alpha T(x) = 0 \Longrightarrow \alpha x \in N(T),\ \alpha \text{ escalar.}$$

□

Usando a definição de matriz que representa uma função linear T em relação a bases fixadas no espaço de partida e no espaço de chegada, conclui-se facilmente que o núcleo e o contradomínio de uma função linear são (sub)espaços isomorfos, respectivamente, ao núcleo e ao espaço das colunas da matriz que representa T nas bases fixadas (ver também o Exemplo 6.9). Este resultado é enunciado no teorema seguinte.

Teorema 6.1. Seja $T : V \to W$ uma função linear e $A = M(T, B, B')$ a matriz $p \times n$ que representa T em relação à base B de V e à base B' de W. São válidas as afirmações:

a) O isomorfismo de V em $\mathbb{C}^n$ determinado pela base B, aplica o núcleo de T no núcleo de A.

b) O isomorfismo de W em $\mathbb{C}^p$ determinado pela base B', aplica o contradomínio de T no espaço das colunas de A.

c) Verifica-se a igualdade:

$$\dim N(T) + \dim Im(T) = \dim V. \tag{6.13}$$

Passamos a designar a relação (6.13) por "*Teorema da dimensão para funções lineares*".

Demonstração. a) Seja $x \in V$ um qualquer vector do núcleo de T, isto é, x é um vector que verifica $T(x) = 0$. Da definição da matriz que representa T, tem-se $(\mathbf{T}(x))_{B'} = A\mathbf{x}_B$ para todo $x \in V$. Logo,

$$x \in N(T) \Longleftrightarrow T(x) = 0 \Longleftrightarrow (\mathbf{T}(x))_{B'} = \mathbf{0} \Longleftrightarrow A\mathbf{x}_B = \mathbf{0} \Longleftrightarrow \mathbf{x}_B \in N(A).$$

O isomorfismo $N(T) \to N(A)$ é a função (linear) que a cada vector $x \in N(T)$ faz corresponder o vector das coordenadas $\mathbf{x}_B$.

b) Para mostrar que $Im(T)$ é um espaço linear isomorfo a $EC(A)$, considere-se $y \in Im(T)$. Logo, $y = T(x)$ para algum $x \in V$. Da definição de matriz que representa T, temos

$$\begin{aligned} y \in Im(T) &\iff T(x) = y, \text{para algum } x \in V \\ &\iff (\mathbf{T}(x))_{B'} = A\mathbf{x}_B = \mathbf{y}_{B'}, \text{para algum } \mathbf{x}_B \in \mathbb{C}^n \\ &\iff \mathbf{y}_{B'} \in EC(A). \end{aligned}$$

O isomorfismo $Im(T) \to EC(A)$ é a função (linear) que a cada vector $y \in Im(T)$ faz corresponder o vector das coordenadas $\mathbf{y}_{B'}$.

c) Da Proposição 3.9 (pág. 137) sabemos que o isomorfismo que a cada vector de um espaço linear faz corresponder o respectivo vector das coordenadas numa base fixada, aplica bases em bases (ver também o Exercício 6.3). Do Teorema da dimensão para matrizes (Teorema 3.6, pág. 144) temos $\dim N(A) + \dim EC(A) = n$, onde n é o número de colunas de A, ou seja, $n = \dim V$. Logo, como $N(A)$ é isomorfo a $N(T)$ e $Im(T)$ isomorfo a $EC(A)$, tem-se $\dim N(T) = \dim N(A)$ e $\dim Im(T) = \dim EC(A)$. Por conseguinte, $\dim N(T) + \dim Im(T) = \dim V$.

$\square$

O núcleo e o contradomínio de uma função linear $T : V \to W$ são espaços lineares isomorfos, respectivamente, ao núcleo e ao espaço das colunas de uma matriz que represente T, mas não são necessariamente iguais a estes espaços. Os exemplos que apresentamos a seguir ilustram esta situação.

Exemplo 6.11. Consideremos a função linear $T(p(t)) = 3p'(t)$ do Exemplo 6.5. No exemplo referido, foi calculada a matriz que representa $T : P_3 \to P_2$ em relação às bases $B_1 = (1-t, t^2, 2t^3, 1+t)$ de P_3, e $B_2 = (2, t+1, t-t^2)$ de P_2, a saber:

$$A = \begin{bmatrix} -3/2 & -3 & -9 & 3/2 \\ 0 & 6 & 18 & 0 \\ 0 & 0 & -18 & 0 \end{bmatrix}.$$

Esta matriz é tal que $\mathbf{y}_{B_2} = (\mathbf{T}(x))_{B_2} = A\mathbf{x}_{B_1}$. Pretendemos obter o núcleo e o contradomínio de T. Para determinar o núcleo de T e o contradomínio $Im(T)$, vamos calcular $N(A)$ e $EC(A)$. A matriz A já está em escada e portanto uma base para o espaço das colunas é dada pelas três primeiras colunas de A, donde

$$EC(A) = \text{Span}\left\{(-3/2, 0, 0), (-3, 6, 0), (-9, 18, -18)\right\}.$$

A base do espaço das colunas de A é um subconjunto de $\mathbb{R}^3$, e portanto não pode ser um conjunto gerador do contradomínio de T, uma vez que o contradomínio de T é um conjunto de polinómios de grau menor ou igual a 2. No entanto, o contradomínio de T é gerado pelos polinómios cujos vectores das coordenadas na base B_2 geram $EC(A)$. Assim, atendendo a que

$$\begin{array}{lll} (\mathbf{p}_1(t))_{B_2} & = (-3/2, 0, 0) & \Longrightarrow p_1(t) = -3 \\ (\mathbf{p}_2(t))_{B_2} & = (-3, 6, 0) & \Longrightarrow p_2(t) = 6t \\ (\mathbf{p}_3(t))_{B_2} & = (-9, 18, -18) & \Longrightarrow p_3(t) = -18t^2, \end{array}$$

os polinómios p_1, p_2 e p_3 formam uma base de $Im(T)$. Logo,

$$Im(T) = \text{Span}\left\{-3, 6t, -18t^2\right\}.$$

Como $\dim Im(T) = 3$ e $\dim P_2 = 3$, temos $Im(T) = P_2$.

Pelo Teorema da dimensão para funções lineares, sabemos que $\dim N(T) = \dim P_3 - \dim Im(T) = 4 - 3 = 1$. Para determinarmos o núcleo de A, resolvemos o sistema $A\mathbf{x} = \mathbf{0}$:

$$\begin{cases} -3/2x - 3y - 9z + 3/2w = 0 \\ \qquad\qquad 6y + 18z \qquad\quad = 0 \\ \qquad\qquad\quad\; 18z \qquad\quad = 0 \end{cases} \Longleftrightarrow \{x = w, y = 0, z = 0.$$

Consequentemente,

$$N(A) = \text{Span}\left\{(1, 0, 0, 1)\right\}.$$

O núcleo de T é gerado pelos polinómios de P_3 cujos vectores das coordenadas na base B_1 geram o núcleo de A. Assim,

$$N(T) = \text{Span}\left\{2\right\},$$

uma vez que $(\mathbf{p}(t))_{B_1} = (1, 0, 0, 1)$ é o polinómio constante $p(t) = (1 - t) + (1 + t) = 2$.

Podemos confirmar este resultado usando a definição de núcleo de T. O núcleo de T é o conjunto dos polinómios p que verificam $T(p) = 3p'(t) = 0$, ou equivalentemente, os polinómios cuja derivada é zero. Estes polinómios são os constantes, isto é, os múltiplos do polinómio constante 2. ♦

Exemplo 6.12. No Exemplo 6.6 obtivemos a matriz que representa $T : P_2 \to P_3$ definida por $T(f) = \int_0^t f(x)dx$, em relação às bases canónicas fixadas nos espaços dos polinómios P_2 e P_3. Esta matriz (ver (6.8) na página 338), é

$$M = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/3 \end{bmatrix}.$$

A matriz M possui três colunas linearmente independentes, e portanto as colunas de M formam uma base para $EC(M)$ (o qual é um subespaço de $\mathbb{R}^4$ de dimensão 3). Logo, o contradomínio de T (que é um subespaço de P_3) é gerado pelos polinómios cujos vectores das coordenadas na base BC_{P_3} são:

$$(0,1,0,0), (0,0,1/2,0) \text{ e } (0,0,0,1/3).$$

Ou seja, $Im(T) = \text{Span}\{t, t^2/2, t^3/3\}$. O contradomínio de T é assim o espaço dos polinómios cujo valor em zero é igual a zero:

$$Im(T) = \{at + bt^2 + ct^3 : a, b, c \in \mathbb{R}\}.$$

Como $\dim Im(T) = 3$, resulta do Teorema da dimensão para funções lineares que $\dim N(T) = \dim P_2 - \dim Im(T) = 3 - 3 = 0$. Por conseguinte, o núcleo de T é o polinómio constante igual a zero, $N(T) = \{0\}$. Sugere-se como exercício que confirme que o núcleo de T é o vector zero de P_2. ♦

O núcleo de uma função linear e invertibilidade

A injectividade de uma função é uma condição necessária e suficiente para que a função seja invertível (cf. Proposição 6.4). No caso da função T ser uma função linear (entre espaços lineares), ainda se tem que T é invertível se e só se $N(T) = \{0\}$. A Proposição seguinte caracteriza as funções lineares invertíveis.

Proposição 6.11. Sejam V e W espaços lineares e $T : V \to W$ uma função linear. São equivalentes as afirmações:

(i) T é injectiva.

(ii) T é invertível.

(iii) O núcleo de T é $N(T) = \{0\}$.

Demonstração. A equivalência $(i) \iff (ii)$ é precisamente a Proposição 6.4. Vamos mostrar a sequência de implicações: $(ii) \Rightarrow (iii) \Rightarrow (i)$.
$(ii) \Rightarrow (iii)$: Seja x um elemento qualquer do núcleo de T, isto é, tal que $T(x) = 0$. Aplicando T^{-1} à última igualdade, vem $T^{-1}(T(x)) = x = T^{-1}(0)$. Como T^{-1} é uma função linear (cf. Proposição 6.3), a imagem de zero é zero (cf. Proposição 6.1), e portanto $x = T^{-1}(0) = 0$ para qualquer $x \in N(T)$. Ou seja, $N(T) = \{0\}$.
$(iii) \Rightarrow (i)$: Suponhamos que $N(T) = \{0\}$ e sejam $x, y \in V$ tais que $T(x) = T(y)$. Pela linearidade de T, tem-se

$$T(x) = T(y) \iff T(x) - T(y) = 0 \iff T(x - y) = 0 \Longrightarrow x - y = 0,$$

onde a última implicação segue do facto do núcleo de T apenas conter o elemento zero. Conclui-se que, se $T(x) = T(y)$ então $x = y$, ou seja, que T é injectiva. $\square$

Pela Proposição 6.11, a dimensão do núcleo de uma função linear ser zero é uma condição necessária e suficiente para a função linear ser invertível. Consequentemente, pelo Teorema da dimensão para funções lineares, para que uma função linear $T : V \to W$ seja invertível é necessário que a dimensão do contradomínio seja igual à dimensão do domínio (isto é, $\dim Im(T) = \dim V$). No caso de endomorfismos $T : V \to V$, é fácil ver que o Teorema da dimensão para funções lineares, a Proposição 6.11, e o facto do $N(T)$ ser isomorfo ao núcleo da matriz que representa T, permitem estabelecer as equivalências seguintes.

Seja V um espaço linear e $T : V \to V$ uma função linear. São válidas as seguintes equivalências:

- $N(T) = \{0\}$ se e só se $Im(T) = V$.
- T é injectiva se e só se é sobrejectiva.
- $N(T) = \{0\}$ se e só se qualquer matriz que represente T é invertível.

Exemplo 6.13. Vejamos quais das funções lineares do Exemplo 6.7 são invertíveis.

a) Quer a reflexão $T : \mathbb{R}^2 \to \mathbb{R}^2$ em relação ao eixo dos xx, quer em relação à recta $y = x$, são funções lineares invertíveis. A função linear inversa destas funções lineares coincide com a própria função, isto é, $T = T^{-1}$.

b) A projecção ortogonal sobre o eixo dos yy não é invertível. O núcleo desta função linear é o eixo dos xx (e portanto diferente de $\{(0, 0)\}$).

c) É fácil verificar geometricamente que a rotação $T : \mathbb{R}^2 \to \mathbb{R}^2$ em torno da origem de um ângulo θ é uma função invertível, e que a sua inversa é a rotação em torno da origem de um ângulo $(-\theta)$. A função T^{-1} é representada pela matriz de rotação de um ângulo $(-\theta)$, isto é,

$$R_{(-\theta)} = \begin{bmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}.$$

Verifique que $R_\theta^{-1} = R_{(-\theta)} = R_\theta^T$. Note-se ainda que $\det R_\theta = \cos^2\theta + \sin^2\theta = 1 \neq 0$ e portanto a matriz R_θ é sempre invertível qualquer que seja θ.

♦

Exemplo 6.14. a) Pretende-se saber se existem funções lineares $T : \mathbb{R}^3 \to \mathbb{R}^5$ que sejam injectivas ou sobrejectivas.
Uma tal função linear é representada por uma matriz A do tipo 5×3. A matriz A tem característica menor ou igual a 3, ou seja, a dimensão do espaço das colunas de A é no máximo 3. Consequentemente, a dimensão do contradomínio de T é menor ou igual a 3. Assim, como $\dim \mathbb{R}^5 = 5$ o contradomínio de T nunca pode ser $\mathbb{R}^5$, pelo que não existem funções lineares $T : \mathbb{R}^3 \to \mathbb{R}^5$ sobrejectivas.
Por outro lado, do Teorema da dimensão para funções lineares, resulta $\dim N(T) + \dim Im(T) = \dim \mathbb{R}^3 = 3$. Como $\dim Im(T)$ pode ser igual a 3, a igualdade anterior diz-nos que a dimensão do núcleo pode ser zero. Assim, existem funções lineares injectivas (logo, invertíveis) $T : \mathbb{R}^3 \to \mathbb{R}^5$.

b) Pretendemos saber o mesmo que na alínea anterior mas para uma função linear $T : \mathbb{R}^5 \to \mathbb{R}^3$.
Uma matriz que represente uma tal função linear é do tipo 3×5, e portanto tem característica menor ou igual a 3. Se tiver característica 3, então $\dim Im(T) = \dim \mathbb{R}^3 = 3$, e portanto T será sobrejectiva. Além disso, como $\dim N(T) + \dim Im(T) = 5$ e $\dim Im(T)$ é no máximo 3, resulta que $\dim N(T)$ é no mínimo 2, e portanto nunca se poderá ter $N(T) = \{\mathbf{0}\}$. Ou seja, não existem funções lineares injectivas $T : \mathbb{R}^5 \to \mathbb{R}^3$.

♦

Vejamos agora qual é a matriz que representa uma função linear invertível em relação a bases fixadas no espaço de partida e de chegada. Consideremos uma função linear $T : V \to W$ e $T^{-1} : W \to V$ a sua inversa. Sendo T invertível, tem-se que $\dim V = \dim W$. Seja $A = M(T, B_1, B_2)$ a matriz (quadrada) que representa T em relação às bases B_1 de V e B_2 de W, e $C = M(T^{-1}, B_2, B_1)$ a matriz que representa a função (linear) inversa T^{-1}. Da definição de matriz que representa uma função linear, e da definição de inversa, obtém-se

$$T^{-1}(T(x)) = x \Longleftrightarrow C(A\mathbf{x}_{B_1}) = \mathbf{x}_{B_1}, \quad \text{para todo } x \in V.$$

Como $CA\mathbf{x}_{B_1} = \mathbf{x}_{B_1}$ para todo $x \in V$, tem-se $CA = I$. Ou seja, a matriz C é a inversa de A.

Enunciamos na proposição seguinte o que acabámos de mostrar.

Proposição 6.12. Sejam V e W espaços lineares da mesma dimensão, B_1 uma base ordenada de V e B_2 uma base ordenada de W.

Se $A = M(T, B_1, B_2)$ é a matriz que representa a função linear invertível $T : V \to W$, então A^{-1} é a matriz que representa a função (linear) inversa $T^{-1} : W \to V$. Isto é, $A^{-1} = M(T^{-1}, B_2, B_1)$.

Valores próprios de funções lineares

No Capítulo 4 estudámos valores e vectores próprios de matrizes. Como o espaço $L(V, V)$, das funções lineares de um espaço linear V em si próprio, é isomorfo a um espaço de matrizes quadradas (cf. Proposição 6.7), é natural definir valores e vectores próprios de uma função linear. Como veremos, os valores próprios e vectores próprios de uma função linear $T : V \to V$ são valores e vectores próprios de uma matriz que representa T em relação a bases fixas em V. Contudo o recíproco não é sempre válido, em particular se V é um espaço linear real e a matriz A que representa T tem valores próprios complexos, os valores próprios de A não são valores próprios de T (ver o Exemplo 6.15 abaixo).

Definição 6.7. Seja V um espaços linear sobre o corpo $\mathbb{K}$, e $T : V \to V$ uma função linear.

Um escalar $\lambda \in \mathbb{K}$ diz-se um valor próprio de T se existe um vector não nulo $v \in V$ tal que

$$T(v) = \lambda v. \tag{6.14}$$

Equivalentemente, $\lambda \in \mathbb{K}$ é um valor próprio de T se $N(T - \lambda I) \neq \{0\}$, onde I é a função identidade definida em V.

Quando um vector não nulo $v \in V$ satisfaz a igualdade (6.14) para um certo $\lambda \in \mathbb{K}$, diz-se que v é um vector próprio de T associado ao valor próprio λ.

Usando a definição de matriz que representa uma função linear, relacionam-se agora os valores e vectores próprios de uma matriz que represente a função linear com os valores e vectores próprios da função linear dada.

Proposição 6.13. Seja $T : V \to V$ uma função linear, onde V é um espaço vectorial sobre $\mathbb{K}$, e A a matriz que representa T em relação a uma base fixada em V. São válidas as afirmações:

- Os valores próprios de T são valores próprios de A.
- Os valores próprios de A pertencentes a $\mathbb{K}$ são valores próprios de T.
- Se $\dim V = n$, então T tem no máximo n valores próprios (pode não ter nenhum).

Demonstração. Da definição de matriz que representa uma função linear T tem-se

$$T(x) = \lambda x \iff (\mathbf{T}(x))_B = (\lambda x)_B \iff A\mathbf{x}_B = \lambda \mathbf{x}_B.$$

Por conseguinte, os valores próprios de T são valores próprios de A. Os valores próprios da matriz A, só serão valores próprios de T se pertencerem ao corpo $\mathbb{K}$.

Como a matriz A que representa T tem n valores próprios, dos itens anteriores conclui-se que T tem no máximo n valores próprios. □

No exemplo a seguir, apresentamos uma função linear $T : V \to V$ que não tem valores próprios enquanto que a matriz que representa a função tem sempre $n = \dim V$ valores próprios.

Exemplo 6.15. Seja $T : \mathbb{R}^2 \to \mathbb{R}^2$ definida por $T(x, y) = (-y, x)$. A matriz que representa T na base canónica de $\mathbb{R}^2$ é

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

Os valores próprios de A são $\lambda = \pm i$. Claramente, não existe nenhum vector $\mathbf{v} \neq \mathbf{0}$ do domínio da função T (isto é, de $\mathbb{R}^2$) que verifique $T(\mathbf{v}) = i\mathbf{v}$. Logo, T não tem valores próprios. Dito de outra forma, os valores próprios de A não são escalares do corpo $\mathbb{K} = \mathbb{R}$ sobre o qual $V = \mathbb{R}^2$ (o domínio de T) é um espaço linear. ♦

Define-se espaço próprio $E(\lambda)$ de um valor próprio λ da função linear $T : V \to V$ como sendo o espaço gerado pelos vectores próprios de T associados a λ. Ou seja,

$$\lambda \text{ é valor próprio de } T \Longrightarrow E(\lambda) = \mathrm{Span}\{x \in V : T(x) = \lambda x\} = N(T - \lambda I),$$

onde $I : V \to V$ é a função identidade.

Seja A a matriz que representa a função linear $T : V \to V$ em relação a uma base fixada em V. Os espaços próprios correspondentes a valores próprios λ de A (isto é, o núcleo da matriz $(A - \lambda I)$) são isomorfos aos espaços próprios de T sempre que λ é valor próprio de T.

Os espaços próprios de funções lineares são exemplos de espaços invariantes por T.

Definição 6.8. Um subespaço S do espaço linear V diz-se um *subespaço invariante* por $T : V \to V$ se $T(S) \subset S$. Isto é, S é invariante por T se a imagem por T de qualquer vector de S ainda é um vector de S.

É consequência da definição de espaço próprio de uma função linear o resultado seguinte.

Seja V um espaço linear sobre o corpo $\mathbb{K}$ e T uma função linear $T : V \to V$.

Se λ é um valor próprio de T, o espaço próprio $E(\lambda)$ é um subespaço de V invariante por T.

6.4 O espaço dual de um espaço linear

É frequentemente útil considerar funções lineares definidas num espaço vectorial que tomam valores escalares. Estas funções constituem um espaço linear que se designa por dual.

A noção de dualidade é importante em várias áreas da matemática como a geometria projectiva, a lógica e teoria de conjuntos, a geometria diferencial, etc. Abordamos nesta secção a noção de espaço dual de um espaço linear de dimensão finita, definimos a base dual de uma dada base e mostramos que a função dual de uma função linear entre espaços lineares é representada pela matriz transposta da matriz que representa a função dada.

Passamos a definir o espaço dual de um espaço linear de dimensão finita.

Definição 6.9. Espaço dual

Se V é um espaço linear (de dimensão finita) sobre o corpo $\mathbb{K}$, o *espaço dual* de V é o conjunto $V^* = L(V, \mathbb{K})$ das funções lineares de V em $\mathbb{K}$.

Os elementos de V^* são designados por *funcionais lineares*, ou *formas-1*, ou *covectores*.

Se $\xi \in V^*$ usamos indistintamente $\xi(v)$ ou $\langle\langle \xi\,,\, v\rangle\rangle$, para designar a imagem em $\mathbb{K}$ do vector $v \in V$ pela função ξ.

Exemplo 6.16.

1. Um elemento de $(\mathbb{R}^n)^*$ é uma função da forma $(x_1, \ldots, x_n) \longmapsto \alpha_1 x_1 + \cdots + \alpha_n x_n$, com $\alpha_i \in \mathbb{R}$.

2. Seja $V = P_n$ o espaço dos polinómios reais de variável real de grau menor ou igual a n. A função de P_n em $\mathbb{R}$, definida por
$$p \longmapsto \int_a^b p(x)\, dx,$$
é um funcional linear, isto é, um elemento de $V^* = (P_n)^*$.

3. Se $f : \mathbb{R}^n \to \mathbb{R}$ é uma função diferenciável e Df a sua derivada, então a função $\mathbb{R}^n \to \mathbb{R}$, definida por
$$\mathbf{x} \longmapsto Df(\mathbf{x})$$
é um elemento de $(\mathbb{R}^n)^*$.

♦

Da Proposição 6.7 (pág. 342), sabemos que $V^* = L(V, \mathbb{K})$ é um espaço linear e que, se V é um espaço linear de dimensão n, V^* é isomorfo ao espaço das matrizes $\mathcal{M}_{1\times n}(\mathbb{K})$, do tipo $1 \times n$, com entradas em $\mathbb{K}$ (a dimensão de $\mathbb{K}$ é igual a 1). Ou seja, V^* é isomorfo a $\mathcal{M}_{1\times n}(\mathbb{K})$ enquanto que V é isomorfo a $\mathcal{M}_{n\times 1}(\mathbb{K})$. Por conseguinte, a dimensão de V^* é igual à dimensão de V.

Dada uma base B do espaço linear V podemos definir uma base B^* de V^* a partir de B do seguinte modo. Seja $\{v_1, v_2, \ldots, v_n\}$ uma base de V e definam-se n elementos $v^1, v^2, \ldots, v^n$ de V^* através das igualdades
$$v^i(v_j) = \delta_{ij} = \begin{cases} 1 & \text{se} \quad i = j \\ 0 & \text{se} \quad i \neq j. \end{cases} \tag{6.15}$$
Recordemos que na definição anterior, a expressão $v^i(v_j)$ significa a imagem do vector $v_j \in V$ pela função linear v^i. O conjunto $\{v^1, v^2, \ldots, v^n\}$ é uma base de V^* como se mostra na proposição seguinte.

Proposição 6.14. Se $\{v_1, v_2, \ldots, v_n\}$ é uma base de V, então o conjunto $\{v^1, v^2, \ldots, v^n\}$ formado pelos vectores de V^* que satisfazem (6.15), é uma base de V^*.

A base $\{v^1, v^2, \ldots, v^n\}$ é designada por *base dual*.

Demonstração. Como $\dim V^* = \dim V$, o espaço V^* tem dimensão n e portanto basta mostrar que o conjunto $\{v^1, v^2, \ldots, v^n\}$ é linearmente independente. Escrevendo a função linear nula, $0 : V \to \mathbb{K}$, como combinação linear de $v^1, v^2, \ldots, v^n$, pretende-se mostrar que a única solução de $c_1v^1 + \cdots + c_nv^n = 0$ é $c_1 = \cdots = c_n = 0$. Avaliando esta combinação linear de funções lineares em v_j, tem-se

$$c_1v^1 + \cdots + c_nv^n = 0 \Longrightarrow c_1v^1(v_j) + \cdots + c_nv^n(v_j) = 0,$$

Usando a definição (6.15), a expressão anterior é

$$0 = c_1v^1(v_j) + \cdots + c_nv^n(v_j) = c_jv^j(v_j) = c_j,$$

para todo $j = 1, \ldots, n$. Logo, o conjunto $\{v^1, v^2, \ldots, v^n\}$ é linearmente independente e portanto uma base de V^*. $\square$

Se $B = (v_1, v_2, \ldots, v_n)$ é uma base ordenada de V e $B^* = (v^1, v^2, \ldots, v^n)$ é a base dual, qualquer elemento ξ de V^* escreve-se de forma única como combinação linear dos covectores v^i. As componentes de ξ na base dual são os escalares $\xi_1, \ldots, \xi_n$ que satisfazem $\xi(v_i) = \xi_i$, para $i = 1, \ldots, n$. De facto, usando a definição de base dual, o valor de $\xi = \xi_1v^1 + \cdots + \xi_nv^n$ em qualquer vector v_i da base B é

$$\xi(v_i) = \xi_1v^1(v_i) + \cdots + \xi_nv^n(v_i) = \xi_i.$$

Assim, o vector das coordenadas de $\xi \in V^*$ na base $B^* = (v^1, v^2, \ldots, v^n)$ é $(\xi)_{B^*} = (\xi_1, \ldots, \xi_n)$ onde $\xi_i = \xi(v_i)$, para $i = 1, \ldots, n$.

Consideremos $u \in V$ e $\xi \in V^*$ tais que $\mathbf{u}_B = (u_1, \ldots u_n)$ e $\xi_{B^*} = (\xi_1, \ldots \xi_n)$. Isto é, $u = u_1v_1 + \cdots + u_nv_n$ e $\xi = \xi_1v^1 + \cdots + \xi_nv^n$. A imagem de u por ξ é dada por

$$\begin{aligned} \xi(u) &= \xi_1v^1(u_1v_1 + \cdots + u_nv_n) + \cdots + \xi_nv^n(u_1v_1 + \cdots + u_nv_n) \\ &= \xi_1u_1 + \cdots + \xi_nu_n, \end{aligned} \tag{6.16}$$

onde aplicámos o facto dos v^j serem funções lineares bem como a definição de base dual. Usando o isomorfismo que a cada vector de um espaço linear associa o vector das coordenadas, a expressão (6.16) pode escrever-se matricialmente na forma

$$\xi(u) = \xi_{B^*}\ \mathbf{u}_B = \begin{bmatrix}\xi_1 & \cdots & \xi_n\end{bmatrix}\begin{bmatrix}u_1\\ \vdots \\ u_n\end{bmatrix}. \tag{6.17}$$

Nota 39. *Se considerarmos os elementos de $\mathbb{R}^n$ como vectores coluna de n componentes reais, o espaço dual $(\mathbb{R}^n)^*$ identifica-se como os vectores linha de n componentes reais. De acordo com* (6.17)*, o valor de um funcional linear $\xi \in (\mathbb{R}^n)^*$ num vector $\mathbf{u} \in \mathbb{R}^n$ é obtido por multiplicação matricial (do vector linha ξ pelo vector coluna $\mathbf{u}$).*

Exemplo 6.17. Considere-se a base canónica $(\mathbf{e}_1, \ldots, \mathbf{e_n})$ de $\mathbb{R}^n$, e $(\mathbf{e}^1, \ldots, \mathbf{e}^n)$ a respectiva base dual. Se $\mathbf{x} = (x_1, \ldots, x_n)$ é um vector de $\mathbb{R}^n$, então

$$\mathbf{e}^i(\mathbf{x}) = \mathbf{e}^i(x_1\mathbf{e}_1 + \cdots x_n\mathbf{e}_n) = x_i.$$

Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função diferenciável e Df a sua derivada, então o funcional linear $Df(\mathbf{x})$ de $(\mathbb{R}^n)^*$ é

$$Df(\mathbf{x}) = \frac{\partial f}{\partial x_1}\mathbf{e}^1 + \cdots + \frac{\partial f}{\partial x_n}\mathbf{e}^n.$$

♦

Exercício 6.7. Mostre que a aplicação que a um vector $v \in V$ associa o funcional $\eta \in (V^*)^*$ (função linear de V^* em $\mathbb{K}$), definida por $\eta(\xi) = \xi(v)$, para todo $\xi \in V^*$, é um isomorfismo entre V e $(V^*)^*$. ▲

Dada uma função linear T entre dois espaços lineares V e W, esta função induz uma função linear entre os espaços duais W^* e V^*, conforme se explicita a seguir.

Proposição 6.15. Sejam V e W espaços lineares e $T : V \to W$ uma função linear. Existe uma única função linear $T^* : W^* \to V^*$ que satisfaz

$$\langle\langle T^*(\xi)\,,\, v\rangle\rangle = \langle\langle \xi\,,\, T(v)\rangle\rangle, \quad \xi \in W^*,\ v \in V. \tag{6.18}$$

Deixamos como exercício verificar que a função T^* é linear e está bem definida.

Exercício 6.8. Verifique que (6.18) é independente de v e que T^* é linear.

A função T^* na proposição anterior é designada por *função dual* de T.

Na proposição seguinte mostramos que a matriz que representa T^* é a matriz transposta da matriz que representa T.

Proposição 6.16. Seja $B_1 = (e_1, \ldots, e_n)$ uma base ordenada do espaço linear V, $B^1 = (e^1, \ldots, e^n)$ a respectiva base dual, $B_2 = (f_1, \ldots, f_p)$ uma base ordenada de W e $B^2 = (f^1, \ldots, f^p)$ a base dual respectiva.

Se $A = M(T, B_1, B_2)$ é a matriz que representa a função linear $T : V \to W$, então $A^T = M(T^*, B^2, B^1)$ é a matriz que representa a função dual $T^* : W^* \to V^*$.

Demonstração. Dizer que a matriz $A = M(T, B_1, B_2) = [a_{ij}]$ representa T significa que a coluna j de A (isto é, o vector $(a_{1j}, a_{2j}, \ldots, a_{pj})$) é o vector das coordenadas de $T(e_j)$ na base B_2. Ou seja,

$$T(e_j) = a_{1j} f_1 + a_{2j} f_2 + \cdots + a_{pj} f_p, \quad j = 1, \ldots, n. \tag{6.19}$$

Se $C = M(T^*, B^2, B^1) = [c_{ij}]$ é a matriz $n \times p$ que representa T^*, pretende-se mostrar que $c_{ij} = a_{ji}$, para todo $i = 1, \ldots, n$ e $j = 1, \ldots, p$. Dizer que C representa T^* nas bases B^2 de W^* e B^1 de V^*, significa que a coluna j de C é o vector das coordenadas de $T^*(f^j)$ na base B^1. Ou seja,

$$T^*(f^j) = c_{1j} e^1 + c_{2j} e^2 + \cdots + c_{nj} e^n, \quad j = 1, \ldots p. \tag{6.20}$$

Usando a definição de função dual e a expressão (6.19), temos

$$\begin{aligned}
\langle\langle T^*(f^j)\,,\, e_i\rangle\rangle &= \langle\langle f^j\,,\, T(e_i)\rangle\rangle = \langle\langle f^j\,,\, a_{1i} f_1 + a_{2i} f_2 + \cdots + a_{pi} f_p\rangle\rangle \\
&= a_{1i}\langle\langle f^j\,,\, f_1\rangle\rangle + a_{2i}\langle\langle f^j\,,\, f_2\rangle\rangle + \cdots + a_{pi}\langle\langle f^j\,,\, f_p\rangle\rangle \\
&= a_{ji}, \qquad \text{para } i = 1, \ldots, n \text{ e } j = 1, \ldots, p
\end{aligned}$$

onde na terceira igualdade usámos a linearidade de f^j e na última igualdade a definição da base dual.

Por outro lado, usando a igualdade (6.20), a expressão $\langle\langle T^*(f^j)\,,\, e_i\rangle\rangle$ é também igual a

$$\begin{aligned}
\langle\langle T^*(f^j)\,,\, e_i\rangle\rangle &= \langle\langle c_{1j} e^1 + c_{2j} e^2 + \cdots + c_{nj} e^n\,,\, e_i\rangle\rangle \\
&= c_{1j}\langle\langle e^1\,,\, e_i\rangle\rangle + c_{2j}\langle\langle e^2\,,\, e_i\rangle\rangle + \cdots + c_{nj}\langle\langle e^n\,,\, e_i\rangle\rangle \\
&= c_{ij}.
\end{aligned}$$

Por conseguinte, $c_{ij} = a_{ji}$, o que significa que $C = A^T$. □

Exercício 6.9. Considere $\varphi : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ definida por

$$\varphi(\mathbf{u}, \mathbf{v}) = \mathbf{u}^T \mathbb{J} \mathbf{v}, \quad \text{com } \mathbb{J} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

a) Mostre que a aplicação $\varphi^\flat : \mathbb{R}^2 \to (\mathbb{R}^2)^*$, definida por $\langle\langle \varphi^\flat(\mathbf{v}), \mathbf{u} \rangle\rangle = \varphi(\mathbf{u}, \mathbf{v})$, é linear.
b) Determine a representação matricial de $\varphi^\flat$ em relação à base canónica de $\mathbb{R}^2$ e à respectiva base dual.
c) Justifique que $\varphi^\flat$ é invertível.
d) Seja $\varphi^\sharp : (\mathbb{R}^2)^* \to \mathbb{R}^2$ a função inversa de $\varphi^\flat$. Determine a matriz que representa $\varphi^\sharp$ em relação à base dual da base canónica de $\mathbb{R}^2$ na partida, e à base canónica na chegada.

▲

Exercícios

1. Diga, justificando, quais das seguintes funções $T : \mathbb{R}^n \to \mathbb{R}^p$ (com n e p apropriados) são lineares.

a) $T(x_1, x_2) = (x_1 + x_2, 3x_1 - x_2)$
b) $T(x_1, x_2) = (1 + x_2, 3x_1 - 1)$.
c) $T(x_1, x_2, x_3) = (2x_1 - x_2 + x_3, x_2 - 4x_3)$.
d) $T(x_1, x_2, x_3) = (x_1 - x_2x_3, 3x_2^2, x_1 - 4x_3)$.
e) $T(x_1, x_2, x_3) = (x_1 + x_2x_3, 5x_2^2)$.
f) $T(x_1, x_2) = (x_1 - x_2, 3x_2, x_1 + 5x_2)$.

2. Sejam U e V espaços lineares e $T : U \to V$ uma função linear.

a) Mostre que $T(0) = 0$.
b) Use a alínea anterior para concluir que $T : \mathbb{R}^2 \to \mathbb{R}^3$, definida por $T(x, y) = (2x - y, x + y + 1, x)$, não é uma função linear.

3. Diga se as funções seguintes são ou não um isomorfismo de espaços lineares.

a) $T(x, y) = (x, x + y, x - y)$.
b) $T(x, y) = (2x - y, -4x + 2y)$.
c) $T(x, y) = (x + 1, x - y)$.
d) $T(x, y) = (x - y, x + y)$.
e) $T : \mathbb{R}^2 \to M$, onde M é o espaço linear das matrizes 2×2 anti-simétricas, e T é definida por
$T(x, y) = \begin{bmatrix} 0 & x \\ -x & 0 \end{bmatrix}$.

4. Determine a matriz que representa cada uma das funções lineares $T : \mathbb{R}^n \to \mathbb{R}^p$ (com n e p apropriados), em relação à base canónica no espaço de partida e à base canónica no espaço de chegada.

a) $T(x_1, x_2) = (2x_1 - x_2, x_1 + x_2)$.
b) $T(x_1, x_2) = (x_1, x_2)$.
c) $T(x_1, x_2, x_3) = (x_1 + 2x_2 + x_3, x_1 + 5x_2, x_3)$.
d) $T(x_1, x_2, x_3) = (0, 0, 0, 0)$.
e) $T(x_1, x_2, x_3, x_4) = (x_4, x_1, x_3, x_2, x_1 - x_3)$.

5. Seja $T : W \to \mathbb{R}^3$ uma função linear e v_1, v_2 e v_3 vectores do espaço linear W tais que

$T(v_1) = (1, -1, 2)$, $T(v_2) = (0, 3, 2)$, $T(v_3) = (-3, 1, 2)$.
Determine $T(3v_1 - v_2 + 10v_3)$.

6. Para cada uma das funções lineares $T : \mathbb{R}^2 \to \mathbb{R}^2$ seguintes, determine a matriz A tal que $T(\mathbf{x}) = A\mathbf{x}$. Use a matriz A para obter as imagens por T do triângulo I de vértices $(-1, 2)$, $(0, 0)$ e $(1, 1)$.

a) T é a reflexão em relação ao eixo dos xx;
b) T é a reflexão em relação à recta $y = x$;
c) T é a projecção ortogonal sobre a recta $y = -x$;
d) T é a rotação em torno da origem no sentido anti-horário de um ângulo de $\pi/2$.

7. Para cada uma das funções lineares $T : \mathbb{R}^3 \to \mathbb{R}^3$ seguintes, determine a matriz A tal que $T(\mathbf{x}) = A\mathbf{x}$. Use a matriz A para obter as imagens por T do triângulo I de vértices $(-1, 2, 0)$, $(0, 0, 1)$ e $(1, 0, 1)$, onde T é :

a) T é a reflexão em relação ao plano xz;
b) T é a rotação em torno do eixo dos zz, no sentido positivo, de um ângulo de $\pi/2$.

8. Sejam B_1 e B_2 as seguintes bases ordenadas de $\mathbb{R}^2$

$$B_1 = ((1,1),(1,-1))\ ,\ B_2 = ((0,1),(2,3))\ ,$$

e $T : \mathbb{R}^2 \to \mathbb{R}^2$ a função linear definida por $T(x,y) = (x-2y, 2x-y)$. Determine as matrizes seguintes que representam nas bases indicadas.

a) $M(T, B_1, B_1)$.
b) $M(T, B_2, B_2)$.
c) $M(T, B_1, B_2)$.
d) $M(T, B_2, B_1)$.

9. Seja $A = \begin{bmatrix} 1 & 3 \\ -2 & -2 \end{bmatrix}$ a matriz que representa a transformação $T : \mathbb{R}^2 \to \mathbb{R}^2$ em relação à base ordenada $B = (\mathbf{v}_1, \mathbf{v}_2)$ onde $\mathbf{v}_1 = (1,3), \mathbf{v}_2 = (-1,4)$.

a) Obtenha as coordenadas de $T(\mathbf{v}_1)$ e de $T(\mathbf{v}_2)$ na base B.
b) Obtenha as coordenadas de $T(\mathbf{v}_1)$ e de $T(\mathbf{v}_2)$ na base canónica de $\mathbb{R}^2$.
c) Determine a matriz que representa T na base canónica de $\mathbb{R}^2$, e encontre uma fórmula para $T(x_1, x_2)$.
d) Use a fórmula obtida em c) para calcular $T(1,1)$.

10. Seja $B = \left(\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 2 & 0 \end{bmatrix} \right)$ uma base ordenada do espaço $Sim(2)$ das matrizes simétricas reais 2×2, e $T : Sim(2) \to Sim(2)$ definida por $T(A) = A + A^T$.

Diga quais das afirmações seguintes são verdadeiras.

a) A matriz que representa T em relação à base B é uma matriz diagonal 2×2.
b) A matriz que representa T em relação à base B é uma matriz diagonal 3×3
c) T é uma função bijectiva.

11. Sejam $T : \mathbb{R}^2 \to \mathbb{R}^3$ e $S : \mathbb{R}^3 \to \mathbb{R}^2$ definidas por $T(x,y) = (x+y, y-x, 2y)$ $S(x,y,z) = (x-y, y+z)$.

a) Determine a expressão: $(S \circ T)(x,y)$.
b) Determine a matriz A que representa $(S \circ T)$ em relação à base canónica de $\mathbb{R}^2$, e diga se $(S \circ T)$ é um isomorfismo.
c) Seja $M = M_{BC \leftarrow B} = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$ a matriz que realiza a mudança da base ordenada B para a base canónica de $\mathbb{R}^2$. Determine a matriz $C = M(S \circ T, B, B)$.

12. Considere em $\mathbb{R}^3$ a base canónica $BC = (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ e a base ordenada

$$B = ((-1,1,1),(-1,-1,1),(0,0,1))\ .$$

a) Determine a matriz F que realiza a mudança da base BC para B, isto é, a matriz F tal que $\mathbf{x}_B = F\mathbf{x}_{BC}$.
b) Dado um vector $\mathbf{u} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3$, isto é, tal que $\mathbf{u}_{BC} = (x_1, x_2, x_3)$, determine $\mathbf{u}_B = (y_1, y_2, y_3)$.
c) Considere a função linear $T : \mathbb{R}^3 \to \mathbb{R}^3$, cuja representação matricial na base canónica é $\begin{bmatrix} 2 & -4 & -4 \\ 0 & 5 & 1 \\ 0 & 1 & 5 \end{bmatrix}$. Determine a matriz que representa T na base B.

13. Considere as funções $T : \mathbb{R}^3 \to \mathbb{R}^3$ e $S : \mathbb{R}^3 \to \mathbb{R}^2$ definidas por:

$$T(x,y,z) = (x+1, y, 2z+x)$$

$$S(x,y,z) = (2x-2, y-z).$$

Indique o valor lógico das afirmações seguintes:

a) S e T são funções lineares.
b) $S \circ T$ é uma função linear.
c) O espaço de chegada de $S \circ T$ é $\mathbb{R}^2$.
d) T é uma função linear invertível.

14. Seja $T : V \to V$ uma função linear e $B = (u, v, w)$ uma base ordenada para V tal que $T(u - v) = 2u, T(2u + v) = v$ e $T(w) = u - v + w$. Então, a matriz que representa T em relação à base B no espaço de partida e de chegada é:

a) $\begin{bmatrix} \frac{4}{3} & \frac{2}{3} & 1 \\ \frac{-1}{3} & \frac{-1}{3} & -1 \\ 0 & 0 & 1 \end{bmatrix}$

b) $\begin{bmatrix} \frac{2}{3} & \frac{-4}{3} & 1 \\ \frac{1}{3} & \frac{1}{3} & -1 \\ 0 & 0 & 1 \end{bmatrix}$

c) $\begin{bmatrix} 2 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}$ d) $\begin{bmatrix} 0 & 0 & -1 \\ 2 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$.

15. Seja $T : \mathbb{R}^3 \to \mathbb{R}^2$ a função linear definida por:

$$T(1,1,0) = (2,1), \quad T(0,1,1) = (1,1),$$
$$T(1,0,0) = (1,1).$$

Então, T é dada por:

a) $T(x, y, z) = (x + y, x + z)$
b) $T(x, y, z) = (y - x, x + 2z)$
c) $T(x, y, z) = (2z + x + y, x - z)$
d) $T(x, y, z) = (2x, x + z)$

16. Seja V o espaço linear real das matrizes reais 2×2, de entradas a_{ij}, satisfazendo $a_{11} + a_{22} = 0$ e $a_{12} + a_{21} = 0$. Considere as seguintes matrizes de V:

$$H = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad J = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

a) Mostre que H e J são linearmente independentes. Determine a dimensão e indique uma base para V.
b) Dada a função linear $T : V \to V$ definida por

$$T(H) = J \ , \ T(J) = -H \ ,$$

determine a matriz que representa T em relação a uma base que contenha H e J.
c) Determine a dimensão do núcleo de T, e diga (justificando) se T é invertível.

17. Diga, justificando, em que casos se tem $T_1 \circ T_2 = T_2 \circ T_1$.

a) T_1 é a projecção ortogonal de $(x, y) \in \mathbb{R}^2$ no eixo dos xx, e T_2 é a projecção ortogonal de $(x, y) \in \mathbb{R}^2$ no eixo dos yy;
b) T_1 é a reflexão de $(x, y) \in \mathbb{R}^2$ relativamente ao eixo dos xx, e T_2 é a reflexão de $(x, y) \in \mathbb{R}^2$ relativamente ao eixo dos yy;

18. Seja $T : \mathbb{R}^2 \to \mathbb{R}^2$ a função linear definida por:

$$T(x_1, x_2) = (4x_1 - 2x_2, 2x_1 - x_2).$$

a) Determine o núcleo e a imagem da função linear.
b) Indique um vector de $\mathbb{R}^2$ que não esteja na imagem da função.
c) Verifique a fórmula do Teorema da dimensão para este exemplo.

19. Seja $T : M \to M$ uma função linear definida no espaço linear M, das matrizes 2×2, por $T(A) = A - A^T$. Indique o valor lógico das afirmações seguintes:

a) $\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ pertence ao conjunto imagem de T.
b) T é sobrejectiva.
c) $\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ pertence ao núcleo de T.
d) T é invertível.

20. Considere as funções

$$T_1(x,y) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}, \qquad \text{e}$$

$$T_2(x,y) = (x-1, y-x).$$

Indique o valor lógico das afirmações seguintes:

a) T_1 é invertível.
b) T_1 é linear e T_2 não é linear.
c) T_1 é não linear e T_2 é linear.
d) T_1 e T_2 são funções lineares.

21. Seja $M_{2\times 2}$ o espaço linear das matrizes 2×2 e $T : M_{2\times 2} \to M_{2\times 2}$ a aplicação definida por $T(A) = A + A^T$.

a) Mostre que T é uma função linear.
b) Determine a matriz K que representa T em relação à base ordenada B de $M_{2\times 2}$ dada por:

$$B = \left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right).$$

c) Determine bases para o núcleo e para a imagem de T.
d) Diga, justificando, se T é injectiva ou sobrejectiva.
e) Sendo $C = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$ determine $T^{-1}(C)$.

Observação: *Dada uma função $f : A \to B$ e $b \in B$, define-se a imagem inversa de b como sendo $f^{-1}(b) = \{a \in A : f(a) = b\}$. Note que o conjunto $f^{-1}(b)$ está definido mesmo que não exista a função inversa de f.*

22. Seja $T : P_2 \to P_2$ a aplicação definida por $T(p(x)) = 2p(x) + p'(x)$, onde p' a derivada de p, eP_2 designa o espaço dos polinómios de grau menor ou igual a 2.

a) Determine a matriz que representa T em relação a alguma base ordenada de P_2.
b) Decida se T é invertível.
c) Calcule $T^{-1}(2 - x^2)$.

23. Seja $T : \mathbb{R}^3 \to \mathbb{R}^3$ uma função linear com dois valores próprios $\lambda_1 = 2$ e $\lambda_2 = -1$. Considere a base ordenada $\mathcal{B} = (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$ formada pelos vectores que verificam: (a) o vector $v_1 = (1,0,0)$ é um vector próprio de T associado a λ_1; (b) os vectores $\mathbf{v}_2 = (1,1,0)$ e $\mathbf{v}_3 = (0,1,1)$ são vectores próprios de T associados a λ_2.

a) Determine $T(2,1,2)$.
b) Determine a matriz $A = M(T, BC, BC)$ que representa T em relação à base canónica de $\mathbb{R}^3$. Use A para verificar se T é ou não injectiva.
c) Determine a matriz $B = M(T, BC, \mathcal{B})$ que representa T em relação à base canónica de $\mathbb{R}^3$ na partida e à base $\mathcal{B}$ na chegada.
d) Determine a matriz $C = M(T, \mathcal{B}, BC)$ que representa T em relação à base $\mathcal{B}$ na partida e à base canónica na chegada.
e) Determine a matriz $D = M(T, \mathcal{B}, \mathcal{B})$ que representa T em relação à base $\mathcal{B}$ de $\mathbb{R}^3$.
f) Seja M a matriz que realiza a mudança da base BC para a base $\mathcal{B}$. Sem de-

terminar M, apenas usando um diagrama comutativo, escreva A em termos de cada uma das matrizes B, C e D, determinadas nas alíneas anteriores.

24. Seja $T : \mathbb{R}^2 \to \mathbb{R}^3$ a função linear definida por $T(x_1, x_2) = (x_1 + 2x_2, -x_1, 0)$, e considere as bases ordenadas $B_1 = (\mathbf{v}_1, \mathbf{v}_2)$ de $\mathbb{R}^2$ e $B_2 = (\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$ de $\mathbb{R}^3$, onde $\mathbf{v}_1 = (1, 3)$, $\mathbf{v}_2 = (-2, 4)$, $\mathbf{u}_1 = (1, 1, 1)$, $\mathbf{u}_2 = (2, 2, 0)$, e $\mathbf{u}_3 = (3, 0, 0)$.

a) Calcule, a matriz $A = M(T, B_1, B_2)$, que representa T relativamente às bases B_1 e B_2.
b) Determine a matriz $B = M(T, BC, BC)$ que representa T relativamente às bases canónicas de $\mathbb{R}^2$ e de $\mathbb{R}^3$, e relacione-a com a matriz A da alínea anterior através das matrizes mudança de base.
c) Represente o respectivo diagrama comutativo.

25. Considere a base ordenada $B = (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$ de $\mathbb{R}^3$, onde $\mathbf{v}_1 = (1, 1, 1), \mathbf{v}_2 = (1, 0, 0)$ e $\mathbf{v}_3 = (1, 1, 0)$. Seja, $T : \mathbb{R}^3 \to \mathbb{R}^3$ uma transformação linear tal que

$$T(\mathbf{v}_1) = (2, 0, 1), \quad T(\mathbf{v}_2) = (5, 1, 1),$$
$$T(\mathbf{v}_3) = (2, 1, 1).$$

a) Determine a matriz $A = M(T, BC, BC)$ que representa T em relação à base canónica de $\mathbb{R}^3$, na partida e na chegada.
b) Considere a base ordenada

$$B' = (\mathbf{v}_2, \mathbf{v}_1, \mathbf{v}_3).$$

Determine a matriz $C = M(T, B, B')$ que representa T em relação à base B na partida e à base B' na chegada.

c) Sendo A e C as matrizes das alíneas anteriores, determine matrizes invertíveis S e P tais que $SA = CP$. Apresente o diagrama comutativo correspondente.

26. Indique o valor lógico das seguintes proposições:

a) Existem funções lineares injectivas de $\mathbb{R}^4$ para $\mathbb{R}^3$.
b) Existem funções lineares sobrejectivas de $\mathbb{R}^4$ para $\mathbb{R}^3$.
c) Existem funções lineares injectivas de $\mathbb{R}^3$ para $\mathbb{R}^4$.
d) Existem funções lineares sobrejectivas de $\mathbb{R}^3$ para $\mathbb{R}^4$.
e) Existem funções lineares injectivas do espaço dos polinómios de grau menor ou igual a 3 para o espaço das matrizes 2×2.
f) Existem funções lineares sobrejectivas do espaço dos polinómios de grau menor ou igual a 3 para o espaço das matrizes 2×2.
g) Qualquer função linear injectiva de $\mathbb{R}^4$ para $\mathbb{R}^4$ é sobrejectiva.
h) Qualquer função linear injectiva de $\mathbb{R}^4$ para $\mathbb{R}^5$ é sobrejectiva.
i) Existe uma função linear bijectiva do espaço dos polinómios de grau menor ou igual a 3 para o das matrizes 2×2.

27. Sejam U, V e W espaços lineares e: B_1 e $\tilde{B}_1$ bases de U; B_2 e $\tilde{B}_2$ bases de V; B_3 e $\tilde{B}_3$ bases de W. Considere ainda as seguintes matrizes de mudança de base

$$K = M_{\tilde{B}_1 \leftarrow B_1}, \; P = M_{\tilde{B}_2 \leftarrow B_2},$$
$$Q = M_{B_3 \leftarrow \tilde{B}_3}.$$

Considere as funções lineares $T : U \to V$ e $S : V \to W$, e as seguintes matrizes A

e D, que representam T e S,

$$A = M(T, B_1, B_2), \;\; D = M(S, \tilde{B}_2, \tilde{B}_3).$$

Determine, em termos das matrizes A, D, K, P e Q, as matrizes seguintes que representam a função composta $S \circ T$.

a) $M(S \circ T, B_1, \tilde{B}_3)$.
b) $M(S \circ T, \tilde{B}_1, \tilde{B}_3)$.
c) $M(S \circ T, B_1, B_3)$.

28. Sejam $f : V \to W$ e $g : W \to U$ duas funcões lineares entre espaços vectoriais, e $f^* : W^* \to V^*$ e $g^* : U^* \to W^*$ as respectivas funções duais. Mostre que $(g \circ f)^* = f^* \circ g^*$.

29. Seja $B = (v_1, v_2, v_3, v_4)$ uma base ordenada do espaço linear V. Considere a aplicação linear $\phi : V \to V$ definida por
$\phi(v_1) = 2v_1 - 3v_3$,
$\phi(v_2) = v_1 - v_3 + 2v_4$, $\phi(v_3 + v_4) = v_1 - v_3 + 2v_4$,
$\phi(v_3 - v_4) = 3v_1 + 2v_2 + v_3$.

a) Determine a característica da matriz M que representa ϕ em relação à base B.
b) Use a matriz da alínea anterior para determinar uma base para o núcleo da função dual ϕ^*.

Capítulo 7

Matrizes ortogonais, unitárias, simétricas e hermitianas

As matrizes ortogonais, unitárias, simétricas e hermitianas revestem-se de particular importância nas mais diversas áreas da matemática aplicada e por isso justificam o destaque que lhes é dado neste capítulo. Eis os principais tópicos de que nos ocuparemos:

(1) Triangularização unitária de Schur.
(2) Diagonalização unitária de matrizes.
(3) Classificação de matrizes simétricas reais.
(4) Decomposição em valores singulares (SVD).
(5) Grupos de matrizes e as suas álgebras de Lie.

Na Secção 7.1 estabelecem-se as propriedades básicas das matrizes unitárias, ortogonais, hermitianas e simétricas reais incluindo as propriedades espectrais destas matrizes. Nas duas secções seguintes, estudam-se tópicos relevantes na identificação de matrizes que são diagonalizáveis por uma matriz unitária.

Na Secção 7.4 lidamos com a diagonalização ortogonal de matrizes simétricas reais, culminando com a obtenção da forma canónica de uma forma bilinear simétrica real e, em particular, de uma forma quadrática. Os resultados obtidos são aplicados depois na identificação de cónicas e superfícies em $\mathbb{R}^3$ definidas por equações quadráticas.

Na Secção 7.5, estudamos a decomposição em valores singulares (SVD) de uma matriz $p \times n$. Interpretamos geometricamente esta decomposição, e obtemos bases ortonormadas para os subespaços fundamentais associados à matriz dada. Na secção seguinte são abordadas algumas aplicações da SVD, a saber: número de condição e sistemas mal condicionados; aproximação de uma matriz por outra de característica inferior; compressão de imagem; pseudoinversa de uma matriz e mínimos quadrados.

Finalizamos o capítulo estudando grupos de matrizes e suas álgebras de Lie, com especial ênfase para os grupos ditos clássicos. A exponencial de matrizes, estudada no Capítulo 4, é aqui utilizada para relacionar álgebras de Lie de matrizes e grupos de matrizes.

7.1 Noções gerais

7.1.1 Matrizes ortogonais e unitárias

No Exemplo 5.10 (pág. 260) vimos que sendo M uma matriz real do tipo $p \times n$, cujos vectores colunas são $\mathbf{u}_1, \ldots, \mathbf{u}_n$, a matriz $M^T M$ tem a forma

$$M^T M = \begin{bmatrix} \|\mathbf{u}_1\|^2 & \langle \mathbf{u}_1, \mathbf{u}_2 \rangle & \cdots & \langle \mathbf{u}_1, \mathbf{u}_n \rangle \\ \langle \mathbf{u}_2, \mathbf{u}_1 \rangle & \|\mathbf{u}_2\|^2 & \cdots & \langle \mathbf{u}_2, \mathbf{u}_n \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{u}_n, \mathbf{u}_1 \rangle & \langle \mathbf{u}_n, \mathbf{u}_2 \rangle & \cdots & \|\mathbf{u}_n\|^2 \end{bmatrix}, \tag{7.1}$$

onde o produto interno é o produto interno usual em $\mathbb{R}^p$. Assim, o conjunto dos vectores coluna de M é ortonormado se e só se $M^T M$ é a matriz identidade. No caso em que a matriz real M é quadrada e as colunas formam um conjunto ortonormado (isto é, $M^T M = I$) dizemos que M é uma matriz ortogonal.

Definição 7.1. Uma matriz quadrada real M diz-se uma *matriz ortogonal* se satisfaz a igualdade

$$M^T M = I.$$

Da igualdade (7.1) resulta que as colunas de uma matriz ortogonal de ordem n formam uma base ortonormada de $\mathbb{R}^n$.

Exercício 7.1. Mostre que o produto de matrizes ortogonais ainda é uma matriz ortogonal. ▲

É consequência imediata da definição de matriz ortogonal que qualquer matriz ortogonal é invertível e que a respectiva inversa é a sua transposta. Além disso, tendo em conta as propriedades do determinante, se M é uma matriz ortogonal, obtém-se

$$\det(MM^T) = \det(M)\det(M^T) = (\det(M))^2 = \det I = 1.$$

Ou seja, o determinante de matrizes ortogonais é igual a ± 1. Resumindo:

Se M é uma matriz ortogonal de ordem n, tem-se:

- $M^{-1} = M^T$.
- $\det(M) = \pm 1$.
- As colunas de M formam uma base ortonormada de $\mathbb{R}^n$.

Exemplo 7.1. Deixamos como exercício verificar que uma matriz ortogonal 2×2 tem uma das seguintes formas:

$$R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \qquad \text{ou} \qquad S_\theta = \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}.$$

A matriz de rotação R_θ é a matriz que realiza a mudança da base ortonormada $B = (\mathbf{u}, \mathbf{v})$, com $\mathbf{u} = (\cos\theta, \sin\theta)$ e $\mathbf{v} = (-\sin\theta, \cos\theta)$, para a base canónica de $\mathbb{R}^2$. Geometricamente, a base B é obtida da base canónica de $\mathbb{R}^2$ por rotação de um ângulo θ em torno da origem. ♦

Os vectores coluna de uma matriz complexa M do tipo $p \times n$ pertencem a $\mathbb{C}^p$. Designe-se por $A^H = \overline{A}^T$ a transposta da matriz cujas entradas são os conjugados das entradas de A. Se calcularmos $M^H M$, onde M é uma matriz complexa com colunas $\mathbf{u}_1, \ldots, \mathbf{u}_n$, obtemos exactamente a matriz (7.1) onde o produto interno é agora o produto interno usual em $\mathbb{C}^p$ ($\langle \mathbf{z}, \mathbf{w} \rangle = \overline{\mathbf{z}}^T \mathbf{w} = \mathbf{z}^H \mathbf{w}$). Assim, se as colunas de uma matriz complexa M de ordem n formarem uma base ortonormada de $\mathbb{C}^n$, é válida a igualdade $M^H M = I$. Esta condição permite definir para matrizes complexas o conceito de matriz unitária (análogo ao de matriz ortogonal).

Definição 7.2. Uma matriz quadrada complexa M diz-se *unitária* se $\overline{M}^T M = I$, ou seja, se $M^{\mathrm{H}} M = I$.

Exemplo 7.2. Verifiquemos que a matriz

$$M = \frac{1}{\sqrt{3}} \begin{bmatrix} 1+i & -i \\ -i & 1-i \end{bmatrix}$$

é unitária.

$$M^{\mathrm{H}} M = \frac{1}{3} \begin{bmatrix} 1-i & i \\ i & 1+i \end{bmatrix} \begin{bmatrix} 1+i & -i \\ -i & 1-i \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1-i^2-i^2 & 0 \\ 0 & -i^2+1-i^2 \end{bmatrix} = I.$$

♦

A condição $M^H M = I$ garante que as colunas de uma matriz unitária formam uma base ortonormada de $\mathbb{C}^n$ (relativamente ao produto interno usual). Note-se ainda que o conjunto das matrizes unitárias de entradas reais coincide com o conjunto das matrizes ortogonais, já que se M é uma matriz unitária real tem-se $M^{\mathrm{H}} = M^T$ (o conjugado de um número real é ele próprio). Por conseguinte, todos os resultados que obtivermos para matrizes unitárias são válidos substituindo a palavra "unitária" por "ortogonal" (ou simbolicamente, substituindo o símbolo "H" de Hermite pelo "T" de transposto).

Proposição 7.1. Sejam U e V matrizes unitárias de ordem n. São válidas as afirmações:

a) Qualquer matriz unitária U é invertível e $U^{-1} = U^{\mathrm{H}}$.
b) $(UV)^{\mathrm{H}} = V^{\mathrm{H}} U^{\mathrm{H}}$.
c) O produto de matrizes unitárias é uma matriz unitária.
d) Qualquer matriz unitária U preserva normas, isto é,

$$\|U\mathbf{x}\| = \|\mathbf{x}\| \quad \text{para todo } \mathbf{x} \in \mathbb{C}^n.$$

e) Qualquer matriz unitária U preserva distâncias, ou seja,

$$\|U\mathbf{x} - U\mathbf{y}\| = \|\mathbf{x} - \mathbf{y}\| \quad \text{para todo } \mathbf{x}, \mathbf{y} \in \mathbb{C}^n.$$

Demonstração. a) É imediato das definições de matriz unitária e de inversa de uma matriz, a equivalência

$$U^H U = I \iff U^{-1} = U^H.$$

b) $(UV)^H = \overline{(UV)}^T = (\overline{U}\,\overline{V})^T = \overline{V}^T \overline{U}^T = V^H U^H$.

c) Sendo U e V matrizes unitárias temos $U^H U = I$ e $V^H V = I$. Logo, pela alínea anterior resulta

$$(UV)^H (UV) = V^H U^H U V = V^H V = I.$$

d) Da definição de norma, temos

$$\|U\mathbf{x}\|^2 = (U\mathbf{x})^H (U\mathbf{x}) = \mathbf{x}^H U^H U \mathbf{x} = \mathbf{x}^H \mathbf{x} = \|\mathbf{x}\|^2,$$

onde na penúltima igualdade aplicámos o facto de U ser unitária.

e) $\|U\mathbf{x} - U\mathbf{y}\| = \|U(\mathbf{x} - \mathbf{y})\| \overset{\text{d)}}{=} \|\mathbf{x} - \mathbf{y}\|$.

□

As funções $f : \mathbb{R}^n \to \mathbb{R}^n$ que preservam distâncias recebem a designação de *isometrias*. Esta noção generaliza-se a funções definidas num espaço euclidiano, do seguinte modo:

Definição 7.3. Seja V um espaço linear munido de um produto interno.

Uma função $f : V \to V$ diz-se uma *isometria* se, para todo $x, y \in V$, verifica

$$\operatorname{dist}(f(x), f(y)) = \operatorname{dist}(x, y) \text{ ou, equivalentemente, } \|f(x) - f(y)\| = \|x - y\|.$$

Em particular, resulta da alínea e) da Proposição 7.1 que, sendo M uma matriz ortogonal, a função $f(\mathbf{x}) = M\mathbf{x}$ é uma isometria em $\mathbb{R}^n$. No entanto existem isometrias em $\mathbb{R}^n$ que não são funções lineares, por exemplo, $f(\mathbf{x}) = M\mathbf{x} + \mathbf{a}$, onde M é uma matriz ortogonal e $\mathbf{a}$ é um vector não nulo de $\mathbb{R}^n$.

Exercício 7.2. Mostre que se f é uma isometria num espaço linear real V (ou complexo), então f preserva o produto interno. Isto é,

$$\langle f(\mathbf{x}), f(\mathbf{y})\rangle = \langle \mathbf{x}, \mathbf{y}\rangle, \quad \text{para todo } \mathbf{x}, \mathbf{y} \in V.$$

Sugestão: Use o Exercício 5.5, página 250. ▲

Exemplo 7.3. Matriz de Fourier

A matriz de Fourier de ordem n foi definida em (5.35) (pág. 307) como sendo a matriz F_n cujas colunas são:

$$\mathbf{w}_0 = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \quad \mathbf{w}_1 = \begin{bmatrix} 1 \\ e^{\frac{2\pi i}{n}} \\ \vdots \\ e^{\frac{2(n-1)\pi i}{n}} \end{bmatrix}, \quad \cdots, \quad \mathbf{w}_{n-1} = \begin{bmatrix} 1 \\ e^{\frac{2(n-1)\pi i}{n}} \\ \vdots \\ e^{\frac{2(n-1)^2\pi i}{n}} \end{bmatrix}.$$

Mostrámos ainda na Secção 5.7.3 que as colunas de F_n formam um conjunto ortogonal (para o produto interno usual em $\mathbb{C}^n$) e que todos os vectores coluna de F_n têm norma igual a $\sqrt{n}$ (ver Proposição 5.6, pág. 308). Logo, a matriz de Fourier de ordem n não é uma matriz unitária, uma vez que, apesar das suas colunas serem ortogonais duas a duas, elas não têm norma igual a 1.

Porém, como todas as colunas de F_n têm norma $\sqrt{n}$, a matriz $\frac{1}{\sqrt{n}}F_n$ é uma matriz unitária e, sendo unitária, a sua inversa é $\left(\frac{1}{\sqrt{n}}F_n\right)^H$ (cf. Proposição 7.1). Logo,

$$\left(\frac{1}{\sqrt{n}}F_n\right)^{-1} = \left(\frac{1}{\sqrt{n}}F_n\right)^H = \frac{1}{\sqrt{n}}\overline{F}_n^T = \frac{1}{\sqrt{n}}\overline{F}_n,$$

onde na última igualdade aplicámos o facto da matriz F_n ser simétrica ($F_n^T = F_n$). A igualdade anterior é ainda equivalente a

$$\left(\frac{1}{\sqrt{n}}F_n\right)^{-1} = \frac{1}{\sqrt{n}}\overline{F}_n \iff \sqrt{n}F_n^{-1} = \frac{1}{\sqrt{n}}\overline{F}_n \iff F_n^{-1} = \frac{1}{n}\overline{F}_n.$$

♦

Valores próprios de matrizes unitárias e ortogonais

A preservação de normas pelas matrizes unitárias (resp. ortogonais) impõe restrições nos valores próprios destas matrizes. De facto, se $(\lambda, \mathbf{x})$ é um par próprio de uma matriz unitária U (isto é, $U\mathbf{x} = \lambda\mathbf{x}$), então

$$U\mathbf{x} = \lambda\mathbf{x} \text{ e } \|U\mathbf{x}\| = \|\mathbf{x}\| \Longrightarrow \|U\mathbf{x}\| = \|\lambda\mathbf{x}\| = |\lambda|\|\mathbf{x}\| = \|\mathbf{x}\|.$$

Como $\mathbf{x}$ é um vector próprio, logo não nulo, a igualdade anterior é equivalente a

$$|\lambda| = 1.$$

Note-se que para $\lambda \in \mathbb{C}$, o módulo $|\lambda|$ é a distância de λ à origem. Assim,

Os valores próprios λ de uma matriz unitária (ou ortogonal) pertencem à circunferência em $\mathbb{C}$, de centro na origem e raio 1. Isto é, os valores próprios satisfazem

$$|\lambda| = 1.$$

Exemplo 7.4. Verifiquemos que uma matriz ortogonal do tipo 3×3 com determinante igual a 1, representa uma rotação em torno de um eixo.

Seja M uma matriz ortogonal de terceira ordem tal que $\det(M) = 1$. Como o determinante de uma matriz é igual ao produto dos seus valores próprios, a contribuição dos valores próprios complexos λ para o determinante de M é 1 (já que $\lambda\overline{\lambda} = |\lambda|^2 = 1$). Uma vez que a matriz M é do tipo 3×3, existe pelo menos um valor próprio (real) igual a 1. Seja $\mathbf{u} \in \mathbb{R}^3$ um vector próprio de norma unitária associado a $\lambda = 1$, isto é, tal que $M\mathbf{u} = \mathbf{u}$. Considere-se $(\mathbf{u}, \mathbf{v}, \mathbf{w})$ uma base ordenada ortonormada de $\mathbb{R}^3$, e P a matriz (ortogonal) cujas colunas são os vectores $\mathbf{u}, \mathbf{v}$ e $\mathbf{w}$ (por esta ordem). Então,

$$\begin{aligned} P^T M P &= \begin{bmatrix} — & \mathbf{u}^T & — \\ — & \mathbf{v}^T & — \\ — & \mathbf{w}^T & — \end{bmatrix} \begin{bmatrix} | & | & | \\ M\mathbf{u} & M\mathbf{v} & M\mathbf{w} \\ | & | & | \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{u}^T M\mathbf{u} & \mathbf{u}^T M\mathbf{v} & \mathbf{u}^T M\mathbf{w} \\ \mathbf{v}^T M\mathbf{u} & \mathbf{v}^T M\mathbf{v} & \mathbf{v}^T M\mathbf{w} \\ \mathbf{w}^T M\mathbf{u} & \mathbf{w}^T M\mathbf{v} & \mathbf{w}^T M\mathbf{w} \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \tilde{Q} \end{bmatrix} = Q, \end{aligned} \tag{7.2}$$

onde a penúltima igualdade resulta dos seguintes factos:

- Como $M\mathbf{u} = \mathbf{u}$, tem-se: $\mathbf{u}^T M\mathbf{u} = \mathbf{u}^T\mathbf{u} = \|\mathbf{u}\|^2 = 1$ e $\mathbf{x}^T M\mathbf{u} = \mathbf{x}^T\mathbf{u} = \langle \mathbf{x}, \mathbf{u}\rangle = 0$, para qualquer vector $\mathbf{x} \in \mathbb{R}^3$ ortogonal a $\mathbf{u}$ (em particular, para $\mathbf{v}$ e $\mathbf{w}$).

- As condições $M\mathbf{u} = \mathbf{u}$ e $M^T M = I$ implicam que $\mathbf{u} = M^T\mathbf{u}$. Portanto, qualquer vector $\mathbf{x}$ ortogonal a $\mathbf{u}$ satisfaz $\mathbf{u}^T M\mathbf{x} = (M^T\mathbf{u})^T\mathbf{x} = \mathbf{u}^T\mathbf{x} = 0$.

Como M, P e P^T são matrizes ortogonais e o produto de matrizes ortogonais é ainda uma matriz ortogonal, a matriz $Q = P^T M P$ é ortogonal. Sendo Q uma matriz ortogonal, a matriz $\tilde{Q}$ em (7.2) também o é. Por outro lado, como Q é semelhante a M, tem-se $\det Q = \det(M) = 1$, e consequentemente $\det(\tilde{Q}) = 1$.

Do Exemplo 7.1 sabemos que uma matriz ortogonal $\tilde{Q}$ do tipo 2×2 que tenha determinante igual a 1 é uma matriz de rotação. Ou seja, podemos escrever (7.2) na forma

$$P^T M P = Q = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix}, \quad \text{com } P \text{ ortogonal.}$$

A matriz Q representa portanto uma rotação em torno do eixo dos xx (ver Exercício 6.6, pág. 340).

A matriz P é a matriz que realiza a mudança da base $B = (\mathbf{u}, \mathbf{v}, \mathbf{w})$ para a base canónica BC de $\mathbb{R}^3$, e as matrizes Q e M são matrizes que representam a mesma função linear, de $\mathbb{R}^3$ em si próprio, em relação a bases distintas (ver Proposição 6.8, pág. 346). Ou seja, é válido o seguinte diagrama comutativo:

$$\begin{tikzcd} \mathbb{R}^3_B \arrow[r, "Q"] \arrow[d, "P"'] & \mathbb{R}^3_B \arrow[d, "P"] \\ \mathbb{R}^3_{BC} \arrow[r, "M"] & \mathbb{R}^3_{BC} \end{tikzcd}$$

Em conclusão, a matriz M representa uma rotação em torno do eixo que tem a direcção de um vector próprio $\mathbf{u}$ associado ao valor próprio $\lambda = 1$. ♦

7.1.2 Matrizes simétricas e hermitianas

Nesta secção estudam-se várias propriedades de matrizes hermitianas e, em particular, de matrizes simétricas reais. Como veremos na Secção 7.3.1, as matrizes hermitianas possuem a importante propriedade de serem diagonalizáveis. Além disso, é sempre possível escolher uma matriz unitária para matriz diagonalizante da matriz hermitiana.

Definição 7.4. Uma matriz M diz-se *simétrica* se $M^T = M$ e diz-se *anti-simétrica* se $M^T = -M$.

Uma matriz complexa M diz *hermitiana*, ou *hermítica*, se $M^H = M$, e diz-se *anti-hermitiana* se $M^H = -M$.

Resulta imediato da definição que uma matriz hermitiana real é simétrica (já que neste caso $\overline{M}^T = M^T = M$). Contudo, existem matrizes complexas simétricas que não são hermitianas como, por exemplo, a matriz

$$A = \begin{bmatrix} 2+i & i \\ i & 2-i \end{bmatrix}. \tag{7.3}$$

Dada uma matriz A do tipo $p \times n$, a matriz $M = A^T A$ é sempre simétrica, visto que

$$M^T = (A^T A)^T = A^T A = M.$$

De igual modo, a matriz $M = A^H A$ é hermitiana, como se verifica facilmente.

Proposição 7.2. Qualquer matriz quadrada M pode escrever-se como a soma de uma matriz hermitiana (resp. simétrica) com uma matriz anti-hermitiana (resp. anti-simétrica). Mais precisamente,

$$M = \frac{1}{2}\underbrace{(M + M^H)}_{\text{hermitiana}} + \frac{1}{2}\underbrace{(M - M^H)}_{\text{anti-hermitiana}}, \qquad M = \frac{1}{2}\underbrace{(M + M^T)}_{\text{simétrica}} + \frac{1}{2}\underbrace{(M - M^T)}_{\text{anti-simétrica}}.$$

Para demonstrar a proposição anterior verifique que $M + M^H$ é hermitiana e $M - M^H$ é anti-hermitiana, analogamente para a decomposição em parte simétrica e anti-simétrica.

Valores próprios de matrizes hermitianas e simétricas reais

Mostramos em seguida que o espectro de matrizes hermitianas (em particular, de matrizes simétricas reais) é sempre real.

Proposição 7.3. Os valores próprios de uma matriz hermitiana (em particular, de uma matriz simétrica real) são reais. Além disso, a valores próprios distintos de uma matriz hermitiana correspondem vectores próprios ortogonais.

Demonstração. Seja M uma matriz hermitiana e $(\lambda, \mathbf{x})$ um par próprio de M. Isto é, supõe-se que $\mathbf{x} \neq \mathbf{0}$ e que

$$M\mathbf{x} = \lambda\mathbf{x}. \tag{7.4}$$

Multiplicando a equação (7.4) por $\mathbf{x}^H$, temos

$$\mathbf{x}^H M\mathbf{x} = \lambda\mathbf{x}^H\mathbf{x}. \tag{7.5}$$

Conjugando e transpondo (7.5), obtemos

$$\mathbf{x}^H M^H\mathbf{x} = \overline{\lambda}\mathbf{x}^H\mathbf{x} \iff \mathbf{x}^H M\mathbf{x} = \overline{\lambda}\mathbf{x}^H\mathbf{x}, \tag{7.6}$$

onde na equivalência aplicámos o facto de M ser hermitiana. Comparando (7.5) e (7.6), temos

$$\lambda\mathbf{x}^H\mathbf{x} = \overline{\lambda}\mathbf{x}^H\mathbf{x}.$$

Esta igualdade é equivalente a $\lambda = \overline{\lambda}$, uma vez que, sendo $\mathbf{x} \neq \mathbf{0}$ o valor de $\mathbf{x}^H\mathbf{x}$ é não nulo, pois $\mathbf{x}^H\mathbf{x} = \|\mathbf{x}\|^2 \neq \mathbf{0}$.

Sejam λ_1 e λ_2 dois valores próprios distintos de M, e $\mathbf{x}$ e $\mathbf{y}$ vectores próprios associados respectivamente a λ_1 e λ_2. Ou seja, $M\mathbf{x} = \lambda_1\mathbf{x}$ e $M\mathbf{y} = \lambda_2\mathbf{y}$. Então,

$$\begin{aligned} &\mathbf{y}^H M\mathbf{x} = \mathbf{y}^H\lambda_1\mathbf{x} = \lambda_1\mathbf{y}^H\mathbf{x} \\ &\mathbf{y}^H M\mathbf{x} = \mathbf{y}^H M^H\mathbf{x} = (M\mathbf{y})^H\mathbf{x} = \lambda_2\mathbf{y}^H\mathbf{x} \quad \text{(já que } \lambda_2 \text{ é real).} \end{aligned}$$

Comparando as últimas igualdades, tem-se

$$\lambda_1\mathbf{y}^H\mathbf{x} = \lambda_2\mathbf{y}^H\mathbf{x} \iff (\lambda_1 - \lambda_2)\mathbf{y}^H\mathbf{x} = 0.$$

Como $\lambda_1 \neq \lambda_2$, resulta $\mathbf{y}^H\mathbf{x} = \langle\mathbf{y}, \mathbf{x}\rangle = 0$. Ou seja, $\mathbf{x} \perp \mathbf{y}$. □

Nota 40. *Só podemos garantir que os valores próprios de uma matriz simétrica são reais se essa matriz tiver entradas reais. Por exemplo, a matriz A em* (7.3) *é simétrica e tem valores próprios $2 + i\sqrt{2}$ e $2 - i\sqrt{2}$.*

Proposição 7.4. Matrizes anti-hermitianas (em particular, matrizes anti-simétricas reais) têm valores próprios imaginários puros, isto é, valores próprios de parte real nula ($\lambda = -\overline{\lambda}$).

A prova desta proposição é deixada como exercício já que é inteiramente análoga à da proposição anterior (no passo em que se usou $M = M^H$ deve agora usar-se $M = -M^H$).

Como vimos no Capítulo 6, o espaço vectorial $L(V, V)$, das funções lineares do espaço vectorial V (sobre o corpo $\mathbb{K}$) em si próprio, é isomorfo ao espaço $\mathcal{M}_{n\times n}(\mathbb{K})$, das matrizes de ordem n (onde $n = \dim V$) de entradas em $\mathbb{K}$. É portanto natural que as matrizes (anti)hermitianas, e unitárias representem funções lineares (definidas em espaços euclidianos) com propriedades particulares. De facto, estas matrizes representam funções lineares T que satisfazem a propriedade: $T(S) \subset S \Longrightarrow T(S^\perp) \subset S^\perp$, para qualquer subespaço (invariante) S do domínio de T (recorde a Definição 6.8 de subespaço invariante). No exercício seguinte estabelecem-se algumas destas relações.

Exercício 7.3. Seja V um espaço linear complexo de dimensão n munido de um produto interno $\langle\ ,\ \rangle$, e B uma base ortonormada de V. Considere a função linear $T : V \to V$ e $A = M(T, B, B)$ a matriz que representa T na base B.

a) Justifique que o produto interno em V pode ser escrito na forma:

$$\langle x, y\rangle = \mathbf{x}_B^H \mathbf{y}_B, \text{ para todo } x, y \in V.$$

b) Mostre as equivalências seguintes.

(i) $\langle T(x), y\rangle = \langle x, T(y)\rangle \Longleftrightarrow A$ é hermitiana.

(ii) $\langle T(x), y\rangle = -\langle x, T(y)\rangle \Longleftrightarrow A$ é anti-hermitiana.

(iii) $\langle T(x), T(y)\rangle = \langle x, y\rangle \Longleftrightarrow A$ é unitária.

c) Mostre que se T verifica uma das condições (i)-(iii) da alínea anterior e S é um subespaço de V tal que $T(S) \subseteq S$, então $T(S^\perp) \subset S^\perp$.

▲

7.2 Triangularização unitária

Como se viu na Secção 4.2 (Exemplo 4.7, pág. 190) nem toda a matriz é diagonalizável (ou seja, semelhante a uma matriz diagonal). Nesta secção mostramos que qualquer matriz quadrada A é unitariamente semelhante a uma matriz triangular superior. Isto é, que existe uma matriz unitária U e uma matriz triangular superior T tais que $A = UTU^{-1} = UTU^H$. A matriz T é conhecida pela designação de *forma triangular de Schur*[1] da matriz A.

No próximo capítulo mostramos que qualquer matriz quadrada é semelhante (não necessariamente por matrizes unitárias) a uma matriz triangular superior possuindo todas as entradas nulas excepto (possivelmente) na diagonal principal e na diagonal acima desta. Uma tal matriz triangular é conhecida por *forma canónica de Jordan* da matriz dada.

Teorema 7.1. Triangularização de Schur

Dada uma matriz quadrada A, existe uma matriz unitária U e uma matriz triangular superior T, tais que

$$A = UTU^H.$$

Demonstração. A prova é feita por indução sobre a ordem da matriz. Seja A do tipo $n \times n$. O resultado é trivial quando a ordem de A é $n = 1$. Suponha-se que a igualdade é válida para $(n-1)$ e mostre-se que é verdadeira para n. Seja λ um valor próprio de A e $\mathbf{u}$ um vector próprio associado de norma igual a 1. Podemos acrescentar a $\{\mathbf{u}\}$ um conjunto V com $(n-1)$ vectores por forma a que $\{\mathbf{u}\} \cup V$ seja uma base ortonormada. Seja K a matriz que tem nas colunas os vectores de V, e P a matriz (em blocos) $P = [\mathbf{u}|K]$ cuja primeira coluna é o vector $\mathbf{u}$. A matriz P é uma matriz unitária, já que as suas colunas formam o conjunto ortonormado $\{\mathbf{u}\} \cup V$. Como o produto AP é a matriz cujas colunas são os produtos de A pelas colunas de P, e $A\mathbf{u} = \lambda\mathbf{u}$, tem-se

$$P^H AP = \begin{bmatrix} \mathbf{u}^H \\ \hline K^H \end{bmatrix} \left[\ \lambda\mathbf{u} \ \middle|\ AK\ \right] = \begin{bmatrix} \lambda\mathbf{u}^H\mathbf{u} & \mathbf{u}^H AK \\ \lambda K^H\mathbf{u} & K^H AK \end{bmatrix} = \begin{bmatrix} \lambda & \mathbf{u}^H AK \\ \mathbf{0} & K^H AK \end{bmatrix}, \tag{7.7}$$

onde a última igualdade resulta do facto de $\|\mathbf{u}\|^2 = \mathbf{u}^H\mathbf{u} = 1$ e dos vectores coluna de K serem todos ortogonais a $\mathbf{u}$ (pelo que $K^H\mathbf{u}$ é igual à matriz nula).

[1] Issai Schur (1875 – 1941), matemático com trabalhos em teoria de representação de grupos, combinatória, teoria de números, e em física teórica.

A matriz $B = K^H AK$ é uma matriz de ordem $(n-1)$, logo por hipótese de indução, a matriz B admite uma triangularização de Schur. Ou seja, existe uma matriz unitária Q_1, tal que $Q_1^H BQ_1 = T_1$ é triangular superior. Considere-se a seguinte matriz (unitária)

$$Q = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Q_1 \end{bmatrix}.$$

Efectuando o produto $Q^H P^H APQ$, obtém-se a seguinte matriz triangular superior

$$Q^H P^H APQ = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Q_1^H \end{bmatrix} \begin{bmatrix} \lambda & \mathbf{u}^H AK \\ 0 & K^H AK \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & Q_1 \end{bmatrix} = \begin{bmatrix} \lambda & \mathbf{u}^H AKQ_1 \\ \mathbf{0} & T_1 \end{bmatrix}.$$

Como o produto de matrizes unitárias é ainda uma matriz unitária, a matriz unitária $U = PQ$ é a matriz pretendida. □

Como U é uma matriz unitária ($U^H = U^{-1}$), a matriz $A = UTU^H$ é semelhante a T. Logo, as matrizes A e T possuem os mesmos valores próprios. Por conseguinte, a matriz triangular superior T, da forma triangular de Schur, possui na diagonal principal os valores próprios de A (recorde que os valores próprios de uma matriz triangular são as entradas da sua diagonal principal).

No teorema anterior não são feitas quaisquer hipóteses sobre a matriz A. Essa matriz pode ser real ou complexa. De facto, mesmo que A seja real, esta matriz pode ter valores próprios complexos e nesse caso a matriz U será complexa.

Exercício 7.4. Seja A uma matriz real do tipo $p \times n$, de característica $\operatorname{car}(A) = k$. Considere as seguintes bases ortonormadas dos subespaços fundamentais associados a A:

$$B_{EC(A)} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\} \qquad B_{N(A^T)} = \{\mathbf{u}_{k+1}, \mathbf{u}_{k+2}, \ldots, \mathbf{u}_p\}.$$

e

$$B_{EL(A)} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\} \qquad B_{N(A)} = \{\mathbf{v}_{k+1}, \mathbf{v}_{k+2}, \ldots, \mathbf{v}_n\}.$$

a) Justifique que $B_{EC(A)} \cup B_{N(A^T)}$ é uma base ortonormada de $\mathbb{R}^p$, e que $B_{EL(A)} \cup B_{N(A)}$ é uma base ortonormada de $\mathbb{R}^n$.

b) Seja U a matriz cujas colunas são os vectores $\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots \mathbf{u}_p$ (por esta ordem) e V a matriz cujas colunas são os vectores $\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{v}_{k+1}, \ldots \mathbf{v}_n$ (por esta ordem). Justifique que U e V são matrizes ortogonais.

c) Calcule o produto $U^T AV$, e mostre que esta matriz é da forma

$$U^T AV = \begin{bmatrix} S_{k\times k} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

onde $S_{k\times k}$ é uma matriz invertível de ordem k, e $\mathbf{0}$ designa matrizes nulas.

d) Compare a decomposição obtida na alínea anterior com forma normal de uma matriz de característica k enunciada na Proposição 3.11.

▲

7.3 Diagonalização unitária de matrizes

Como sabemos, nem todas as matrizes são diagonalizáveis. Mesmo que uma dada matriz seja diagonalizável nem sempre é possível escolher uma matriz unitária que diagonalize a matriz dada. Por exemplo, a matriz

$$A = \begin{bmatrix} -1 & 2 \\ 0 & 1 \end{bmatrix} \tag{7.8}$$

é diagonalizável (tem valores próprios distintos) mas não existe uma matriz unitária (ou ortogonal) que diagonalize esta matriz. De facto, como facilmente se verifica, não existe uma base ortonormada de $\mathbb{R}^2$ formada por vectores próprios da matriz A.

O objectivo desta secção é identificar as matrizes que são diagonalizáveis por uma matriz unitária. Como veremos, as matrizes hermitianas são diagonalizáveis por uma matriz unitária e, em particular, as matrizes simétricas reais são diagonalizáveis por uma matriz ortogonal (ver Teorema 7.2). Em geral, as matrizes unitariamente diagonalizáveis são as matrizes ditas *normais*.

Decomposição espectral

Comecemos por referir alguns factos gerais sobre matrizes diagonalizáveis, e em particular sobre matrizes unitariamente diagonalizáveis.

Seja A uma matriz diagonalizável de ordem n e $\sigma(A) = \{\lambda_1, \ldots, \lambda_k\}$ o seu espectro. Dizer que A é diagonalizável significa que existe uma base de $\mathbb{C}^n$ constituída por vectores próprios de A ou, equivalentemente, que $\mathbb{C}^n$ se escreve como a soma directa dos subespaços próprios de A, isto é,

$$\mathbb{C}^n = N(A - \lambda_1 I) \oplus \cdots \oplus N(A - \lambda_k I) = E(\lambda_1) \oplus \cdots \oplus E(\lambda_k). \tag{7.9}$$

Se U e V são subespaços complementares num espaço linear W (isto é, $W = U \oplus V$), podemos definir uma projecção de W sobre U da seguinte forma. Todo o vector $x \in W$ pode escrever-se de forma única como $x = u + v$, com $u \in U$ e

$v \in V$. A função $P : W \to W$ definida por $Px = u$, designa-se por *projector*, ou *projecção de W sobre U paralelamente a V*. Deixamos como exercício verificar que P é linear e idempotente (ver Capítulo 5, pág. 265).

Exercício 7.5. Sejam U e V subespaços complementares no espaço linear W e $x = u + v$ com $u \in U$ e $v \in V$ a decomposição de um vector $x \in W$. Mostre que a função $P : W \to W$ definida por $Px = u$, satisfaz as propriedades:

a) P é uma função linear.
b) P é *idempotente*, isto é, $P^2 = P \circ P = P$.

▲

Exercício 7.6. Mostre que uma função linear $P : W \to W$ é um projector se e só se $P^2 = P$ (isto é, P é idempotente).
Sugestão: Verifique que $Im(P)$ e $N(P)$ são subespaços complementares em W.
▲

Se $B = B_U \cup B_V$ é uma base de W, com B_U e B_V bases dos subespaços complementares U e V respectivamente, então a matriz que representa a projecção $P_U : W \to W$ sobre o subespaço U na base B é

$$M(P_U, B, B) = \begin{bmatrix} I & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

onde a ordem da matriz identidade I é igual à dimensão de U.

É portanto natural interpretar a diagonalização de uma matriz em termos de projectores sobre subespaços próprios. Seja A uma matriz diagonalizável de ordem n e S uma matriz que diagonaliza A, isto é, tal que $A = SDS^{-1}$ com D diagonal. Consideremos ainda a matriz S particionada em blocos S_i, do tipo $n \times p_i$, de tal forma que as colunas de cada bloco S_i formam uma base para o espaço próprio $E(\lambda_i)$. Particione-se a matriz S^{-1} em k blocos R_i^T, do tipo $p_i \times n$. Ou seja,

$$A = SDS^{-1} = \left[S_1 \,\middle|\, S_2 \,\middle|\, \cdots \,\middle|\, S_k \right] \begin{bmatrix} \lambda_1 I_{p_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \lambda_2 I_{p_2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \lambda_k I_{p_k} \end{bmatrix} \begin{bmatrix} R_1^T \\ \hline R_2^T \\ \hline \vdots \\ \hline R_k^T \end{bmatrix}, \quad (7.10)$$

onde I_{p_i} designa a matriz identidade de ordem p_i (note que p_i é a multiplicidade geométrica de λ_i) e $\mathbf{0}$ denota matrizes nulas. Refira-se ainda, que sendo A diagonalizável, então $\sum_{i=1}^{k} p_i = n$.

De (7.10) conclui-se que a matriz A se pode escrever na forma

$$A = \lambda_1 S_1 R_1^T + \lambda_2 S_2 R_2^T + \cdots + \lambda_k S_k R_k^T. \qquad (7.11)$$

A decomposição (7.11) é designada por *decomposição espectral* de A.

No caso de A ser diagonalizável por uma matriz unitária $U = [S_1|S_2|\cdots|S_k]$, a decomposição espectral de A tem a forma

$$A = UDU^H = \lambda_1 S_1 S_1^H + \lambda_2 S_2 S_2^H + \cdots + \lambda_k S_k S_k^H, \qquad (7.12)$$

onde as colunas das matrizes S_i formam uma base ortonormada do espaço próprio $E(\lambda_i)$. Do Exercício 5.12 (pág. 272) conclui-se que as matrizes $S_i S_i^H$ são matrizes de projecção ortogonal sobre os espaços próprios $E(\lambda_i) = N(A - \lambda_i I)$.

No caso geral de uma matriz diagonalizável A, as matrizes $P_i = S_i R_i^T$ na decomposição espectral de A em (7.11), são também matrizes de projecção (não necessariamente ortogonal) sobre os subespaços próprios $E(\lambda_i)$.

No exercício guiado seguinte, indicam-se os passos que permitem mostar que as matrizes $P_i = S_i R_i^T$ em (7.11), são matrizes de projecção sobre os subespaços próprios da matriz A.

Exercício 7.7. Mostre que as matrizes $P_i = S_i R_i^T$ na decomposição espectral (7.11), verificam:

a) $P_1 + \cdots + P_k = I$.
b) $P_i P_j = \mathbf{0}$ para $i \neq j$ e $P_i^2 = P_i$.
c) $EC(P_i) = N(A - \lambda_i I)$.
Sugestão: Use o facto de $R_i^T S_i = I_{p_i}$ e a seguinte propriedade do espaço das colunas do produto de matrizes: $EC(MN) \subseteq EC(M)$.
d) Use a decomposição espectral de A, para mostrar que $P_i(A - \lambda_i I) = \mathbf{0}$. Justifique ainda que esta igualdade implica a inclusão: $EC(A - \lambda_i I) \subseteq N(P_i)$.
e) Use o teorema da dimensão para matrizes (Teorema 3.6, pág. 144) e as alíneas c) e d) para concluir a igualdade: $EC(A - \lambda_i I) = N(P_i)$.
f) Conclua das alíneas anteriores que as matrizes P_i são matrizes de projecção sobre o subespaço próprio $E(\lambda_i)$, paralelamente a $EC(A - \lambda_i I)$.

▲

Exercício 7.8. Seja A uma matriz diagonalizável do tipo 3×3, com valores próprios $\lambda_1 = 2$ e $\lambda_2 = -5$ com multiplicidades algébricas 2 e 1, respectivamente. Considere a decomposição espectral: $A = \lambda_1 P_1 + \lambda_2 P_2$. Mostre que

$$P_1 = \frac{(A + 5I)}{7}, \qquad P_2 = \frac{(A - 2I)}{-7}.$$

Sugestão: Calcule $(A - \lambda_i I)$ usando a decomposição espectral de A e a propriedade enunciada na alínea a) do exercício anterior. ▲

Nota 41. *Usando um procedimento análogo ao do exercício anterior, é possível mostrar que as matrizes P_i, na decomposição espectral $A = \lambda_1 P_1 + \cdots + \lambda_k P_k$, verificam,*

$$P_i = \frac{\prod_{\substack{j=1 \\ j \neq i}}^{k} (A - \lambda_j I)}{\prod_{\substack{j=1 \\ j \neq i}}^{k} (\lambda_i - \lambda_j)}.$$

7.3.1 Diagonalização unitária de matrizes hermitianas

Nesta secção mostramos que as matrizes hermitianas são unitariamente diagonalizáveis, e na secção seguinte prova-se que a classe das matrizes unitariamente diagonalizáveis é constituída pelas matrizes ditas *normais*.

Algumas das demonstrações nesta secção são realizadas para matrizes hermitianas reais (ou seja, matrizes simétricas reais) sendo a prova no caso de matrizes complexas completamente análoga e por isso deixada como exercício.

Seja M uma matriz ortogonalmente diagonalizável, isto é, tal que existe uma matriz ortogonal P satisfazendo

$$M = PDP^{-1}, \text{ com } D \text{ diagonal e } P^{-1} = P^T.$$

Então,

$$M = PDP^T \Longrightarrow M^T = (PDP^T)^T = PDP^T = M.$$

Ou seja, M é simétrica.

De forma análoga, se conclui que uma matriz unitariamente diagonalizável e com valores próprios reais é hermitiana. Podemos assim resumir:

Uma matriz ortogonalmente diagonalizável é simétrica.

Uma matriz unitariamente diagonalizável e com valores próprios reais, é hermitiana.

Vejamos agora que o recíproco é igualmente válido.

Teorema 7.2. Toda a matriz hermitiana M é unitáriamente diagonalizável. Isto é, existe uma matriz unitária U tal que $U^H MU$ é uma matriz diagonal (real).

Em particular, toda a matriz simétrica real é ortogonalmente diagonalizável.

Demonstração. Do teorema da triangularização de Schur (Teorema 7.1), a matriz M é unitariamente semelhante a uma matriz triangular superior T. Isto é, existe uma matriz unitária U tal que $M = UTU^H$. Como M é hermitiana, tem-se

$$M^H = M \iff UT^H U^H = UTU^H \implies T^H = T,$$

onde a última implicação segue do facto da matriz unitária U ser invertível e de $U^H = U^{-1}$ também ser invertível.

Como T é triangular superior, a igualdade $T^H = T$ implica que T é uma matriz diagonal, e portanto M é unitariamente diagonalizável. Além disso, sendo M hermitiana os seus valores próprios são reais (cf. Proposição 7.3), e por isso a matriz diagonal T é real (matrizes semelhantes têm os mesmos valores próprios).

Se M é (simétrica) real, então podemos escolher as colunas de U reais. Por conseguinte, uma matriz simétrica real é diagonalizável por uma matriz ortogonal U. □

De forma inteiramente análoga à demonstração do teorema anterior mostra-se que as matrizes anti-hermitianas são unitariamente diagonalizáveis. Deixamos como exercício esta demonstração.

Exercício 7.9. Mostre que uma matriz anti-hermitiana é unitariamente diagonalizável. Justifique que as matrizes anti-simétricas reais (não nulas) são diagonalizáveis por matrizes unitárias mas não são ortogonalmente diagonalizáveis. ▲

Resumimos a seguir num único enunciado algumas propriedades fundamentais de matrizes simétricas reais.

Proposição 7.5. Uma matriz simétrica real M, de ordem n, goza das seguintes propriedades:

1. M é ortogonalmente diagonalizável;
2. M tem n valores próprios reais (contando as multiplicidades);
3. As multiplicidades algébrica e geométrica de cada valor próprio de M são iguais. Ou seja, todos os valores próprios de M são semi-simples;
4. Os espaços próprios de M são ortogonais dois a dois, no sentido em que vectores próprios correspondentes a valores próprios distintos são ortogonais.

Demonstração. O item 1) é o Teorema 7.2. Os itens 2) e 4) decorrem da Proposição 7.3 desta secção, e o item 3) do Corolário 4.4 (pág. 197) . ☐

Vejamos agora como calcular uma matriz ortogonal que diagonaliza uma certa matriz simétrica real.

Seja M uma matriz de ordem n e $M = PDP^T$, com P ortogonal e D diagonal. Sendo a matriz P ortogonal, as suas colunas formam uma base ortonormada de $\mathbb{R}^n$ constituída por vectores próprios de A. Da Proposição 7.3 sabemos que se todos os valores próprios de M são distintos, o cálculo dos espaços próprios fornece-nos automaticamente n vectores próprios ortogonais entre si (e portanto linearmente independentes). Logo, dividindo cada vector pela sua norma, obtém-se uma base ortonormada de $\mathbb{R}^n$ constituída por vectores próprios (note que se $\mathbf{u}$ é um vector próprio de M então $\frac{\mathbf{u}}{\|\mathbf{u}\|}$ também é) com a qual construímos a matriz P.

Resta portanto o caso de matrizes simétricas reais com valores próprios de multiplicidade algébrica superior a 1. Neste caso, dada uma base de um espaço próprio $E(\lambda)$ podemos aplicar o processo de Gram-Schmidt a esta base e obtemos uma base ortogonal para esse subespaço próprio. Normalizando esta base, tem-se uma base ortonormada para $E(\lambda)$. Procedendo de igual modo para cada um dos espaços próprios de M é assim possível encontrar uma base ortonormada de $\mathbb{R}^n$, constituída por vectores próprios, com a qual construimos a matriz P que diagonaliza ortogonalmente a matriz simétrica dada.

Exercício 7.10. Sejam $\mathbf{u}$ e $\mathbf{v}$ dois vectores próprios de uma matriz real A associados ao valor próprio λ. Mostre que se $\mathbf{u}$ e $\mathbf{v}$ não são colineares, o vector $\mathbf{w} = \mathbf{v} - \mathrm{proj}_{\mathbf{u}}\, \mathbf{v}$ é um vector próprio de A associado ao valor próprio λ. Diga ainda por que razão precisa da hipótese de não colinearidade de $\mathbf{u}$ e $\mathbf{v}$. ▲

No exemplo seguinte ilustramos como obter uma matriz que diagonaliza ortogonalmente de uma matriz simétrica.

Exemplo 7.5. Determinemos uma diagonalização ortogonal da matriz

$$A = \begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{bmatrix}.$$

A matriz A é simétrica e, além disso, a soma de cada linha é igual a 5. Logo, 5 é um valor próprio da matriz e $\mathbf{u} = (1, 1, 1)$ é um vector próprio associado (ver Proposição 4.10, pág. 212). Como $\det(A) = 20 = 5\lambda_1\lambda_2$, o produto dos restantes valores próprios é $\lambda_1\lambda_2 = 4$. Atendendo a que $\operatorname{tr}(A) = 9$, tem-se $\lambda_1 + \lambda_2 = 9 - 5 = 4$. Substituindo $\lambda_1 = \frac{4}{\lambda_2}$ em $\lambda_1 + \lambda_2 = 4$ e resolvendo a equação resultante, obtém-se: $\lambda_1 = \lambda_2 = 2$ (ou seja, 2 tem multiplicidade algébrica 2). Calculemos o núcleo de $(A - 2I)$:

$$(A - 2I)\mathbf{x} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \iff a = -b - c.$$

Assim, podemos considerar $\{(-1, 1, 0), (-1, 0, 1)\}$ como base para o espaço próprio $E(2)$. Este conjunto não é ortogonal, pelo que vamos aplicar o processo de Gram-Schmidt para obter uma base ortogonal para $E(2)$.

$$\mathbf{v}_1 = (-1, 1, 0), \qquad \tilde{\mathbf{v}}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} = \frac{1}{\sqrt{2}}(-1, 1, 0),$$
$$\mathbf{v}_2 = (-1, 0, 1) - \frac{1}{2}\langle(-1, 1, 0), (-1, 0, 1)\rangle(-1, 1, 0) = \frac{1}{2}(-1, -1, 2).$$

Atendendo ao Exercício 7.10, o vector $\mathbf{v}_2$ é um vector próprio associado ao valor próprio $\lambda = 2$. Por conseguinte, o conjunto $\{\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2\}$, com $\tilde{\mathbf{v}}_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|} = \frac{1}{\sqrt{6}}(-1, -1, 2)$, é uma base ortonormada para $N(A - 2I)$. Logo, o conjunto $\{\tilde{\mathbf{u}}, \tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2\}$, com $\tilde{\mathbf{u}} = \frac{\mathbf{u}}{\|\mathbf{u}\|} = \frac{1}{\sqrt{3}}(1, 1, 1)$, é uma base ortonormada formada por vectores próprios de A. Podemos portanto escolher para P e D as matrizes:

$$D = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \quad \text{e} \quad P = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & -\sqrt{3}/\sqrt{2} & -1/\sqrt{2} \\ 1 & \sqrt{3}/\sqrt{2} & -1/\sqrt{2} \\ 1 & 0 & \sqrt{2} \end{bmatrix}.$$

Confirme que as colunas de P formam uma base ortonormada de $\mathbb{R}^3$ e que $A = PDP^T$. ♦

7.3.2 Diagonalização unitária e matrizes normais

Nesta secção mostramos que a classe das matrizes unitariamente diagonalizáveis é constituída pelas matrizes normais.

Definição 7.5. Uma matriz A diz-se uma *matriz normal* se satisfaz a igualdade

$$A^H A = AA^H.$$

Usando a definição de matriz normal, mostra-se facilmente que as matrizes hermitianas e as matrizes unitárias são exemplos de matrizes normais.

Exercício 7.11. Seja A uma matriz quadrada. Mostre as implicações seguintes.

a) A (anti)simétrica real $\Longrightarrow$ A (anti)hermitiana $\Longrightarrow$ A normal.
b) A ortogonal $\Longrightarrow$ A unitária $\Longrightarrow$ A normal.

Dê exemplos de matrizes 2×2 para as quais as implicações recíprocas das implicações anteriores sejam falsas. ▲

Comecemos por mostrar que se A é uma matriz unitariamente diagonalizável, então A é uma matriz normal.

Seja $A = UDU^H$ com U unitária e $D = \text{diag}(\lambda_1, \ldots, \lambda_n)$. Então

$$\begin{aligned} A^H A &= (UD^H U^H)(UDU^H) = UD^H DU^H \\ AA^H &= (UDU^H)(UD^H U^H) = UDD^H U^H. \end{aligned} \tag{7.13}$$

Como D é diagonal, tem-se

$$D^H D = \overline{D} D = D \overline{D} = DD^H = \text{diag}(|\lambda_1|^2, \ldots, |\lambda_n|^2),$$

e portanto, de (7.13) resulta $A^H A = AA^H$.

Resumindo,

Proposição 7.6. Toda a matriz unitariamente diagonalizável é uma matriz normal.

Vejamos agora que uma matriz normal e triangular, é diagonal.

Lema 7.1. Uma matriz normal que seja triangular, é uma matriz diagonal.

Demonstração. Seja $A = [a_{ij}]_{i,j=1,\dots,n}$ uma matriz triangular superior, isto é, com $a_{ij} = 0$ para $i > j$. Sendo $A^H A = [c_{ij}]$, $AA^H = [d_{ij}]$ e $A^H = [b_{ij}]$ com $b_{ij} = \overline{a}_{ji}$, usando a expressão (1.10) para as entradas do produto de duas matrizes, tem-se

$$c_{ij} = \sum_{k=1}^{n} b_{ik} a_{kj} = \sum_{k=1}^{n} a_{kj} \overline{a}_{ki}, \qquad d_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} = \sum_{k=1}^{n} a_{ik} \overline{a}_{jk}.$$

Em particular, as entradas da diagonal principal de $A^H A$ e de AA^H são

$$c_{ii} = \sum_{k=1}^{n} a_{ki} \overline{a}_{ki} = \sum_{k=1}^{n} |a_{ki}|^2 = \sum_{k=1}^{i} |a_{ki}|^2 \tag{7.14}$$

$$d_{ii} = \sum_{k=1}^{n} a_{ik} \overline{a}_{ik} = \sum_{k=1}^{n} |a_{ik}|^2 = \sum_{k=i}^{n} |a_{ik}|^2, \tag{7.15}$$

onde na última igualdade das somas anteriores se aplicou o facto de $a_{ij} = 0$ para $i > j$.

Sendo A uma matriz normal, tem-se $c_{ij} = d_{ij}$. Por conseguinte, as somas (7.14), (7.15) e a igualdade $c_{ii} = d_{ii}$, implicam $a_{ij} = 0$ para $i \neq j$. Ou seja, A é diagonal. □

Mostramos em seguida que ser uma matriz normal é condição necessária e suficiente para a diagonalização unitária da matriz.

Teorema 7.3. Uma matriz é diagonalizável por uma matriz unitária se e só se é normal.

Demonstração. Pela Proposição 7.6, uma matriz unitariamente diagonalizável é uma matriz normal. Falta provar o recíproco. Seja A uma matriz normal. Pelo Teorema 7.1 da triangularização de Schur, a matriz A é unitariamente semelhante a uma matriz triangular (superior). Isto é, existem matrizes U e T, respectivamente, unitária e triangular superior, tais que $A = UTU^H$. Como a matriz A é normal tem-se

$$A^H A = AA^H \Longleftrightarrow UT^H TU^H = UTT^H U^H \Longrightarrow T^H T = TT^H,$$

onde a implicação resulta do facto de uma matriz unitária U ser invertível.

Ou seja, a matriz triangular T é uma matriz normal. Se T é uma matriz normal e triangular, pelo Lema 7.1, a matriz T é diagonal. Por conseguinte, A é unitariamente semelhante à matriz diagonal T. □

7.4 Classificação de matrizes simétricas reais

Nesta secção classificamos matrizes simétricas reais em termos dos seus valores próprios. As matrizes simétricas reais com valores próprios todos positivos (ou todos negativos) surgem em vários tipos de aplicações, nomeadamente em problemas de optimização e no estudo de pontos de máximo e mínimo de funções $f : \mathbb{R}^n \to \mathbb{R}$. Nesta secção estabelecem-se testes que permitem obter informação sobre os valores próprios de uma matriz simétrica real sem recurso ao cálculo explícito desses valores próprios.

Comecemos por obter um critério para que uma matriz 2×2, real e simétrica, tenha valores próprios positivos.

Proposição 7.7. Os valores próprios da matriz real $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ são positivos se e só se $ac - b^2 > 0$ e $a > 0$.

Demonstração. ($\Longrightarrow$) Suponha-se que λ_1 e λ_2 são valores próprios positivos de A. Como o determinante de A é igual ao produto dos valores próprios e o traço de A é igual à soma dos valores próprios, tem-se que

$$\det(A) = ac - b^2 > 0 \qquad \text{e} \qquad \operatorname{tr} A = a + c > 0.$$

Da primeira expressão conclui-se que c não pode ser nulo. Mostremos que $a > 0$:

(i) Se $a \leq 0$ e $c > 0$ então $\det(A) = ac - b^2 < 0$ (contradição).

(ii) Se $a \leq 0$ e $c < 0$ então $\operatorname{tr} A = a + c < 0$ (contradição).

Logo $a > 0$.

($\Longleftarrow$) Suponha-se que $ac - b^2 > 0$, $a > 0$ e λ_1, λ_2 são valores próprios de A. Mostremos que λ_1 e λ_2 são positivos.

Como $\det(A) = ac - b^2 > 0$, tem-se $\lambda_1\lambda_2 > 0$. Ou seja, λ_1 e λ_2 possuem o mesmo sinal.

Suponha-se que $\lambda_1 < 0$ e $\lambda_2 < 0$, então $\operatorname{tr} A = \lambda_1 + \lambda_2 = a + c < 0$. Por conseguinte, sendo $a > 0$ temos $c < 0$. Mas $a > 0$ e $c < 0$ implica também que $\det(A) = ac - b^2 < 0$, o que é uma contradição. Conclui-se portanto que λ_1 e λ_2 são positivos.

□

Exemplo 7.6. As matrizes

$$\begin{bmatrix} 3 & 2 \\ 2 & 1 \end{bmatrix} \qquad \text{e} \qquad \begin{bmatrix} -1 & 1 \\ 1 & -2 \end{bmatrix}$$

não têm valores próprios positivos, já que no primeiro caso o determinante é negativo, e no segundo caso a entrada da diagonal principal $a = -1$ é negativa. ♦

Nota 42. *O critério da Proposição 7.7 para que uma matriz simétrica 2×2 tenha valores próprios positivos traduz-se nas duas condições: (i) a matriz tem determinante positivo; (ii) é também positivo o determinante da submatriz que se obtém da matriz dada suprimindo a segunda linha e a segunda coluna.*

Como veremos adiante, este critério em termos de determinantes de submatrizes, generaliza-se a matrizes de ordem superior a 2. Por isso, e para referência futura, definimos a seguir o conceito de menor principal central[2].

Definição 7.6. Chamam-se *menores principais centrais* de uma matriz A, do tipo $n \times n$, aos determinantes das submatrizes A_k que são obtidas de A suprimindo as últimas $(n-k)$ linhas e colunas de A. Isto é, para $A = [a_{ij}]_{i,j=1,\ldots,n}$, os menores principais centrais são:

$$\det(A_1) = |a_{11}|\,, \det(A_2) = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \ldots, \det(A_k) = \begin{vmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{k1} & \cdots & a_{kk} \end{vmatrix}, \ldots .$$

São designados simplesmente por *menores principais* os determinantes de qualquer submatriz da matriz dada.

Nota 43. *Em termos desta nomenclatura, a Proposição 3.12 (pág. 150) enuncia-se da seguinte forma: a característica de uma matriz é igual à ordem do maior menor principal (não necessariamente um menor principal central) não nulo da matriz.*

Outra forma de reconhecer matrizes reais simétricas com valores próprios positivos é baseada na observação seguinte. Seja A uma matriz quadrada e $(\lambda, \mathbf{x})$ um par próprio de A. Da igualdade $A\mathbf{x} = \lambda\mathbf{x}$, obtemos

$$\mathbf{x}^T A\mathbf{x} = \lambda\mathbf{x}^T\mathbf{x} = \lambda\|\mathbf{x}\|^2.$$

[2] Na literatura anglo-saxónica usa-se a designação "leading principal minor" para menor principal central.

Assim, se todos os valores próprios de A são positivos (resp. negativos) tem-se $\mathbf{x}^T A\mathbf{x} > 0$ (resp. $\mathbf{x}^T A\mathbf{x} < 0$) para todos os vectores próprios $\mathbf{x}$ de A.

O sinal dos valores assumidos pela função $\mathbf{x} \mapsto \mathbf{x}^T A\mathbf{x}$, com $\mathbf{x} \in \mathbb{R}^n$, será usado para classificar matrizes simétricas reais, como se enuncia a seguir.

Definição 7.7. Classificação de matrizes simétricas
Uma matriz simétrica e real A, de ordem n, diz-se:

- *Definida positiva* se $\mathbf{x}^T A\mathbf{x} > 0$ para todo $\mathbf{x} \neq \mathbf{0}$.
- *Definida negativa* se $\mathbf{x}^T A\mathbf{x} < 0$ para todo $\mathbf{x} \neq \mathbf{0}$.
- *Semidefinida positiva* se $\mathbf{x}^T A\mathbf{x} \geq 0$ para todo $\mathbf{x} \in \mathbb{R}^n$.
- *Semidefinida negativa* se $\mathbf{x}^T A\mathbf{x} \leq 0$ para todo $\mathbf{x} \in \mathbb{R}^n$.
- *Indefinida* se $\mathbf{x}^T A\mathbf{x}$ toma valores positivos e negativos.

Nota 44. *A expressão $\mathbf{x}^T A\mathbf{x}$ não é linear. De facto, por exemplo, para uma matriz 2×2, temos*

$$\mathbf{x}^T A\mathbf{x} = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = ax^2 + 2bxy + cy^2. \tag{7.16}$$

Exercício 7.12. Mostre que se a matriz A, de ordem n, é uma matriz real, simétrica e definida positiva, a expressão

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T A\mathbf{y}$$

define um produto interno em $\mathbb{R}^n$. ▲

A função $q : \mathbb{R}^n \to \mathbb{R}$ definida por $q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$, com A uma matriz real de ordem n, chama-se *forma quadrática*.

Vejamos agora que dada uma forma quadrática $q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$, podemos sempre supor que a matriz A é uma matriz simétrica. Sendo A uma matriz real, a função q é uma função de $\mathbb{R}^n$ em $\mathbb{R}$. Assim, ao transpor $q(\mathbf{x})$ estamos transpondo uma matriz 1×1, e portanto $q(\mathbf{x})^T = q(\mathbf{x})$. Ou seja,

$$\begin{aligned} q(\mathbf{x}) = q(\mathbf{x})^T &\iff \mathbf{x}^T A\mathbf{x} = (\mathbf{x}^T A\mathbf{x})^T \iff \mathbf{x}^T A\mathbf{x} = \mathbf{x}^T A^T\mathbf{x} \\ &\iff \mathbf{x}^T (A^T - A)\mathbf{x} = 0. \end{aligned} \tag{7.17}$$

Atendendo à Proposição 7.2, qualquer matriz A pode escrever-se como a soma de uma matriz simétrica com uma anti-simétrica, ou seja,

$$A = \frac{1}{2}\left[\underbrace{A + A^T}_{\text{simétrica}} + \underbrace{A - A^T}_{\text{anti-simétrica}}\right].$$

Logo, usando a equação (7.17) tem-se

$$\mathbf{x}^T A\mathbf{x} = \frac{1}{2}\left[\mathbf{x}^T(A + A^T)\mathbf{x} + \mathbf{x}^T(A - A^T)\mathbf{x}\right] = \mathbf{x}^T\frac{(A + A^T)}{2}\mathbf{x}.$$

Por conseguinte, se A não é uma matriz simétrica a forma quadrática $\mathbf{x}^T A\mathbf{x}$ é igual à forma quadrática associada à parte simétrica de A. Não há portanto perda de generalidade em supor que a matriz que define uma forma quadrática é simétrica.

Para referência futura, definimos forma quadrática.

Definição 7.8. Chama-se *forma quadrática* à função $q : \mathbb{R}^n \to \mathbb{R}$ definida por $q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ com A uma matriz real de ordem n.

Quando A é uma matriz simétrica, denotamos a forma quadrática q por $q_A(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$.

A nomenclatura na Definição 7.7 para uma matriz A ser definida positiva, negativa, etc. é igualmente adoptada na classificação de formas quadráticas $q_A(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ (com A simétrica).

Teorema 7.4. Seja A uma matriz simétrica real.

1) A matriz A é semidefinida positiva (resp. semidefinida negativa) se e só se todos os valores próprios de A são não negativos (resp. não positivos).

2) A matriz A é definida positiva (resp. definida negativa) se e só se todos os valores próprios de A são positivos (resp. negativos).

3) A matriz A é indefinida se e só se tem valores próprios positivos e negativos.

Demonstração. 1)
($\Longrightarrow$): Mostremos que se A é semidefinida positiva, então os seus valores próprios são não negativos. Sendo A semidefinida positiva, tem-se $\mathbf{x}^T A\mathbf{x} \geq 0$ para todo

$\mathbf{x}$. Em particular, se $\mathbf{x}$ é um vector próprio de A (logo, $\mathbf{x} \neq \mathbf{0}$) associado ao valor próprio λ, resulta

$$\mathbf{x}^T A\mathbf{x} = \mathbf{x}^T \lambda \mathbf{x} = \lambda \mathbf{x}^T \mathbf{x} = \lambda \|\mathbf{x}\|^2 \geq 0 \Longleftrightarrow \lambda \geq 0.$$

($\Longleftarrow$): Mostremos agora que se todos os valores próprios de A são não negativos, a matriz A é semidefinida positiva. Sendo A uma matriz simétrica, esta matriz é ortogonalmente diagonalizável, isto é, $A = PDP^T$ com P ortogonal e D diagonal. Considere-se $D = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ em que os valores próprios λ_i estão ordenados por ordem decrescente: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ (tal ordenação é possível uma vez que uma matriz simétrica real tem valores próprios reais). Logo,

$$\begin{aligned} \mathbf{x}^T A\mathbf{x} &= \mathbf{x}^T PDP^T \mathbf{x} = (P^T\mathbf{x})^T D (P^T \mathbf{x}) = \mathbf{y}^T D\mathbf{y} \\ &= \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2 \geq \lambda_n (y_1^2 + \cdots + y_n^2) = \lambda_n \|\mathbf{y}\|^2 \geq 0. \end{aligned} \tag{7.18}$$

Ou seja, A é semidefinida positiva.

A demonstração para o caso de matrizes semidefinidas negativas inteiramente análoga e fica como exercício.

2) A prova para o caso de matrizes definidas positivas (resp. definidas negativas) é igual à do item 1) bastando substituir nessa prova os sinais de $\geq$ (resp. $\geq$) por $>$ (resp. $<$), conforme apropriado.

3) O item 3) segue como consequência do item 2).

$\square$

Da demonstração do Teorema 7.4 é fácil concluir que o máximo e o mínimo da forma quadrática $q_A(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ na esfera de $\mathbb{R}^n$, definida por $\|\mathbf{x}\| = 1$, é respectivamente igual ao maior e ao menor valor próprio da matriz A. De facto, sendo A uma matriz real e simétrica, e P e D matrizes como na demonstração anterior, tem-se

$$\mathbf{x}^T A\mathbf{x} = \mathbf{x}^T PDP^T \mathbf{x} = \mathbf{y}^T D\mathbf{y} = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2, \tag{7.19}$$

com $\mathbf{y} = P^T\mathbf{x} = (y_1, \ldots, y_n)$. Como a matriz P é ortogonal, tomando para $\mathbf{x}$ um vector de norma unitária tem-se $\|\mathbf{y}\| = \|P^T\mathbf{x}\| = \|\mathbf{x}\| = 1$. Logo, de (7.19) obtemos

$$\lambda_n = \lambda_n \|\mathbf{y}\|^2 \leq \mathbf{x}^T A\mathbf{x} = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2 \leq \lambda_1 \|\mathbf{y}\|^2 = \lambda_1,$$

onde λ_1 é o maior valor próprio de A e λ_n o menor valor próprio de A. É óbvio que há vectores do conjunto $\{\mathbf{x} \in \mathbb{R}^n : \ \|\mathbf{x}\| = 1\}$ para os quais a forma quadrática toma os valores λ_1 e λ_n, uma vez que existem vectores próprios de norma unitária associados respectivamente a λ_1 e λ_n.

Proposição 7.8. O máximo e o mínimo da forma quadrática $q_A(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$, definida pela matriz real e simétrica A, no conjunto $\{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| = 1\}$ são, respectivamente, o maior e o menor valor próprio de A. Isto é,

$$\max_{\|\mathbf{x}\|=1} q_A(\mathbf{x}) = \lambda, \qquad \min_{\|\mathbf{x}\|=1} q_A(\mathbf{x}) = \mu,$$

com λ e μ, respectivamente, o maior e o menor valor próprio de A.

Em seguida obtemos uma série de resultados que permitem classificar matrizes simétricas sem recurso ao cálculo dos seus valores próprios.

Lema 7.2. Se A é uma matriz real simétrica e definida positiva, então:

(a) Todas as entradas da diagonal principal de A são positivas;

(b) Todos os menores principais centrais de A são positivos.

Demonstração. **a)**: Sendo A definida positiva, verifica-se $\mathbf{x}^T A\mathbf{x} > 0$ para todo $\mathbf{x} \neq \mathbf{0}$. Tomando, em particular, para $\mathbf{x}$ os vectores $\mathbf{e}_k$ da base canónica de $\mathbb{R}^n$, a expressão $\mathbf{x}^T A\mathbf{x}$ é igual à entrada a_{kk} da diagonal principal de A. Logo, todas as entradas da diagonal principal de A são positivas.
b): Como $\mathbf{x}^T A\mathbf{x} > 0$ para todo $\mathbf{x} \neq \mathbf{0}$, em particular $\mathbf{x}_k^T A\mathbf{x}_k > \mathbf{0}$ para vectores não nulos $\mathbf{x}_k = (x_1, \ldots, x_k, 0, \ldots, 0)$. Porém,

$$\mathbf{x}_k^T A\mathbf{x}_k = \mathbf{x}_k^T A_k\mathbf{x}_k,$$

onde A_k é a submatriz (simétrica) que se obtém de A suprimindo as últimas $(n-k)$ linhas e colunas. Logo, se A é definida positiva, as submatrizes A_k são definidas positivas, para $k = 0, \ldots, (n-1)$. Como o determinante de uma matriz é igual ao produto dos valores próprios, o Teorema 7.4 (pág. 395) garante a positividade do determinante de todas as matrizes A_k. Por conseguinte, todos os menores principais centrais de A são positivos. □

Pretendemos agora mostrar que o critério dos menores principais centrais (enunciado na alínea b) do Lema 7.2) é necessário e suficiente para que uma matriz simétrica seja definida positiva. Tal vai ser mostrado recorrendo à factorização LU de uma matriz.

No Capítulo 1 (pág. 67) vimos que a factorização $A = LU$ é uma factorização em que L é triangular inferior com 1's na diagonal principal e U é triangular superior com todas as entradas na diagonal principal não nulas. Passaremos a designar por pivôs as entradas da diagonal principal de U, uma vez que estas entradas são precisamente os pivôs do método de eliminação de Gauss.

É óbvio que se uma matriz é factorizável na forma LU, então essa matriz é invertível. Contudo a condição de invertibilidade da matriz não é suficiente para a existência da factorização LU, como se pode verificar pelo contra-exemplo apresentado no exercício seguinte.

Exercício 7.13. Mostre que a matriz invertível (e simétrica)

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 5 & 3 \end{bmatrix}$$

não admite uma factorização $A = LU$.
Sugestão: Escreva $A = LU$ com L e U matrizes gerais da forma pretendida e mostre que o sistema obtido é impossível. ▲

Mostramos de seguida que a existência da factorização LU para matrizes simétricas é equivalente à propriedade da matriz possuir todos os menores principais centrais não nulos.

Proposição 7.9. Uma matriz real invertível A admite uma factorização LU se e só se todos os menores principais centrais de A são não nulos.

Demonstração. Mostremos que se A admite uma factorização $A = LU$, qualquer menor principal central é não nulo. Para tal, considere-se a seguinte partição em blocos:

$$A = LU = \begin{bmatrix} L_{11} & \mathbf{0} \\ L_{21} & L_{22} \end{bmatrix} \begin{bmatrix} U_{11} & U_{12} \\ \mathbf{0} & U_{22} \end{bmatrix},$$

onde L_{11} e U_{11} são matrizes $k \times k$. A submatriz (central) $A_k = L_{11}U_{11}$ é não singular, uma vez que L_{11} é triangular inferior com 1's na diagonal principal e U_{11} é triangular superior com todas as entradas da diagonal não nulas. Logo, qualquer menor principal central é não nulo.

Para a implicação recíproca, vamos usar indução sobre a ordem da matriz A e mostrar que esta matriz admite uma factorização LU. Para $n = 1$ a verificação é trivial já que sendo $A = [a_{11}]$ e $\det A \neq 0$, tem-se que $A = [1][a_{11}]$ é a factorização pretendida. Assuma-se agora que qualquer matriz A_{n-1}, de ordem

$(n-1)$, admite uma factorização LU e que todos os menores principais centrais da matriz A, de ordem n, são não nulos. A matriz A pode escrever-se na forma

$$A = \begin{bmatrix} A_{n-1} & \mathbf{b} \\ \mathbf{c}^T & \alpha_n \end{bmatrix},$$

onde $\mathbf{b}$ e $\mathbf{c}^T$ são vectores cujas componentes são as primeiras $(n-1)$ entradas, respectivamente, da coluna e da linha n de A. Como por hipótese de indução A_{n-1} admite uma factorização LU, digamos $A_{n-1} = L_{n-1}U_{n-1}$, podemos escrever

$$A = \begin{bmatrix} A_{n-1} & \mathbf{b} \\ \mathbf{c}^T & \alpha_n \end{bmatrix} = \underbrace{\begin{bmatrix} L_{n-1} & \mathbf{0} \\ \mathbf{c}^T U_{n-1}^{-1} & 1 \end{bmatrix}}_{L_n} \underbrace{\begin{bmatrix} U_{n-1} & L_{n-1}^{-1}\mathbf{b} \\ \mathbf{0} & \alpha_n - \mathbf{c}^T A_{n-1}^{-1}\mathbf{b} \end{bmatrix}}_{U_n},$$

uma vez que as submatrizes U_{n-1}, L_{n-1} e A_{n-1} são invertíveis. Para mostrar que L_nU_n é uma factorização LU de A, resta verificar que

$$\alpha_n - \mathbf{c}^T A_{n-1}^{-1}\mathbf{b} \neq 0.$$

Como $\det(A) \neq 0$ (por hipótese todos os menores principais centrais de A são não nulos), tem-se

$$0 \neq \det A = \det(L_n)\det(U_n) = \det(U_n) = \big(\alpha_n - \mathbf{c}^T A_{n-1}^{-1}\mathbf{b}\big)\det(U_{n-1}),$$

e portanto $\alpha_n - \mathbf{c}^T A_{n-1}^{-1}\mathbf{b} \neq 0$. □

Note-se que a demonstração anterior permite concluir que todas as submatrizes centrais A_k (de ordem k) admitem uma factorização LU. Além disso, se $A = LU$, os pivôs podem exprimir-se em termos dos menores principais centrais da seguinte forma:

$$u_{kk} = \begin{cases} \det A_1 & \text{para } k = 1 \\[2ex] \dfrac{\det A_{k+1}}{\det A_k} & \text{para } k = 2, \ldots, n. \end{cases} \tag{7.20}$$

Tomando em consideração os resultados anteriormente obtidos, o Teorema 7.5 abaixo fornece condições necessárias e suficientes para que uma matriz simétrica real seja definida positiva.

Teorema 7.5. Cada uma das condições seguintes é necessária e suficiente para que uma matriz simétrica real A seja definida positiva.

(i) Todos os valores próprios de A são positivos;

(ii) Todos os menores principais centrais de A são positivos;

(iii) A matriz A admite uma factorização LU com todos os pivôs positivos.

Demonstração. **(i)**: Segue da demonstração do Teorema 7.4.

(ii): Tendo em conta o Lema 7.2, falta mostrar que se a matriz A tem todos os menores principais centrais positivos, então é definida positiva. De facto, se todos os menores principais centrais de A são positivos, pela Proposição 7.9, a matriz A admite uma factorização LU. Além disso, a igualdade (7.20) diz-nos que todos os pivôs são positivos (isto é, todas as entradas da diagonal principal de U são positivas).

Qualquer matriz triangular superior pode ser escrita na forma $U = DW$ onde D é uma matriz diagonal cuja diagonal principal é igual à diagonal principal de U, e W é uma matriz triangular superior com 1's na diagonal principal. Assim, a factorização LU de A pode reescrever-se na forma $A = LDW$. Uma vez que A é simétrica e as matrizes L e W são invertíveis, resulta

$$A^T = A \Longleftrightarrow W^T D L^T = LDW \Longrightarrow L^{-1}W^T = DW(L^T)^{-1}D^{-1}.$$

Como $L^{-1}W^T$ é uma matriz triangular inferior com 1's na diagonal e $DW(L^T)^{-1}D^{-1}$ é uma matriz triangular superior, da igualdade anterior obtém-se $L^{-1}W^T = I$, ou equivalentemente $L = W^T$. Consequentemente $A = LDL^T$. Assim,

$$\mathbf{x}^T A\mathbf{x} = \mathbf{x}^T LDL^T\mathbf{x} = (L^T\mathbf{x})^T D(L^T\mathbf{x}) = \mathbf{y}^T D\mathbf{y}.$$

Como as entradas da diagonal principal de D são positivas, a forma quadrática $\mathbf{y}^T D\mathbf{y}$ é definida positiva, isto é,

$$\mathbf{y}^T D\mathbf{y} > 0 \quad \text{para todo } \mathbf{y} \neq \mathbf{0}.$$

Finalmente, como L^T é uma matriz invertível (visto que $\det(L) = 1$), o vector $\mathbf{y} = L^T\mathbf{x}$ é não nulo se e só se o vector $\mathbf{x}$ é não nulo. Assim, $\mathbf{x}^T A\mathbf{x} > 0$ para todo $\mathbf{x} \neq \mathbf{0}$, ou seja, A é definida positiva.

(iii): Pelo item (ii) uma matriz é definida positiva se e só se todos os menores principais centrais são positivos. Logo, pela Proposição 7.9, uma matriz é definida positiva se e só se admite uma decomposição LU e tem todos os menores principais positivos. Como as entradas da diagonal principal de U são dadas pela expressão (7.20), podemos concluir que A é definida positiva se e só se as entradas de U (os pivôs) são positivas. □

Exercício 7.14. Mostre que uma matriz real e simétrica A é definida positiva se e só se pode ser factorizada na forma $A = R^T R$, onde R é uma matriz triangular superior com todas as entradas da diagonal principal positivas.
Sugestão: Use a factorização $A = LDL^T$ obtida na demonstração anterior, e mostre que pode tomar $R = D^{1/2}L^T$, onde $D^{1/2}$ é a matriz diagonal cujas entradas da diagonal principal são as raízes quadradas das respectivas entradas de D.

A factorização $A = R^T R$ é designada por *factorização de Cholesky*[3] e a matriz R por *factor de Cholesky*. ▲

Atendendo à definição, uma matriz A é definida negativa se e só se a matriz $(-A)$ é definida positiva. Assim, como corolário do Teorema 7.5, obtemos a seguinte caracterização de matrizes definidas negativas.

Corolário 7.1. Cada uma das condições seguintes é necessária e suficiente para que uma matriz simétrica real A seja definida negativa.

(i) Todos os valores próprios de A são negativos;

(ii) Todos os menores principais centrais de A de ordem par são positivos e os de ordem ímpar negativos;

(iii) A matriz A admite uma factorização LU com todos os pivôs negativos.

Demonstração. Basta aplicar o Teorema 7.5 à matriz $(-A)$ e ter em conta que: $\det(-A_k) = (-1)^k \det(A_k)$; da igualdade (7.20) obtém-se $u_{kk} < 0$ para todo o k. □

Poderíamos ser tentados a pensar que a caracterização de matrizes semidefinidas, por exemplo positivas, podia ser realizada substituindo na caracterização deste tipo de matrizes a palavra "positiva" pelo termo "não negativo". Porém, se levarmos em linha de conta as demonstrações realizadas (em particular as baseadas na factorização LU) vemos que tal caracterização não pode ser feita com essa generalidade, uma vez que as matrizes semidefinidas são singulares e portanto não admitem uma factorização LU. O leitor interessado em caracterizações de matrizes

[3] Esta factorização foi descoberta pelo matemático francês André-Louis Cholesky (1875 – 1918) para matrizes reais.

semidefinidas distintas da caracterização enunciada no Teorema 7.4, (em termos, por exemplo, de menores principais) pode consultar, entre outros, Meyer [9].

7.4.1 Forma canónica de uma forma bilinear simétrica real

Para finalizar esta secção, deduzimos a forma canónica (ou forma normal) de uma forma bilinear simétrica e, em particular, de uma forma quadrática. Na base dessa forma canónica encontra-se um resultado conhecido por *lei de inércia de Sylvester*[4].

Comecemos por definir o que se entende por uma forma bilinear simétrica num espaço linear real W.

Definição 7.9. Seja W um espaço linear real. Uma forma φ em W é uma função

$$\begin{array}{rccl} \varphi: & W \times W & \to & \mathbb{R} \\ & (u,v) & \mapsto & \varphi(u,v). \end{array}$$

A forma φ diz-se:

- *Bilinear*: φ é linear nas duas variáveis, isto é, para todos $u_1, v_1, u_2, v_2 \in W$ e todos os escalares c_1, c_2, c_3 e c_4 de $\mathbb{R}$, verifica-se:

$$\begin{aligned} \varphi(c_1u_1 + c_2u_2, c_3v_1 + c_4v_2) &= c_1c_3\varphi(u_1, v_1) + c_1c_4\varphi(u_1, v_2) \\ &\quad + c_2c_3\varphi(u_2, v_1) + c_2c_4\varphi(u_2, v_2). \end{aligned}$$

- *Simétrica*: $\varphi(u, v) = \varphi(v, u)$, para todo $u, v \in W$;
- *Anti-simétrica*: $\varphi(u, v) = -\varphi(v, u)$, para todo $u, v \in W$.

Note-se que uma forma num espaço vectorial real que seja simétrica e linear apenas numa variável é automaticamente bilinear.

Um exemplo de uma forma bilinear simétrica é um produto interno num espaço linear real.

Nota 45. *Num espaço linear complexo a noção de forma bilinear simétrica é substituída pela noção de forma hermitiana, ou forma sesquilinear. Uma forma hermitiana no espaço linear complexo* W *é uma função* $\varphi : W \times W \to \mathbb{C}$ *que é linear numa das variáveis e tal que* $\varphi(u, v) = \overline{\varphi(v, u)}$, *para todo* $u, v \in W$.

[4] James Joseph Sylvester (1814 – 1897), matemático inglês.

É fácil deduzir que fixada uma base ordenada $B = (u_1, \ldots, u_n)$ no espaço linear real W, uma forma bilinear simétrica (resp. anti-simétrica) é representada por uma matriz simétrica real (resp. anti-simétrica) A. Nomeadamente,

$$
\begin{aligned}
&\varphi(x_1u_1 + \cdots + x_nu_n, y_1u_1 + \cdots + y_nu_n) \\
&\qquad = \mathbf{x}_B^T \begin{bmatrix} \varphi(u_1, u_1) & \varphi(u_1, u_2) & \cdots & \varphi(u_1, u_n) \\ \varphi(u_2, u_1) & \varphi(u_2, u_2) & \cdots & \varphi(u_2, u_n) \\ \vdots & \vdots & & \vdots \\ \varphi(u_n, u_1) & \varphi(u_n, u_2) & \cdots & \varphi(u_n, u_n) \end{bmatrix} \mathbf{y}_B \\
&\qquad = \mathbf{x}_B^T A \mathbf{y}_B .
\end{aligned}
$$

A Definição 7.8 de forma quadrática em $\mathbb{R}^n$ generaliza-se a qualquer espaço vectorial real W da seguinte forma. Uma forma quadrática q em W é uma função $q : W \times W \to \mathbb{R}$ que em qualquer base ordenada B de W se escreve na forma $q(x) = \mathbf{x}_B^T A \mathbf{x}_B$, onde A é uma matriz real. Como vimos anteriormente, podemos sempre supor que a matriz A é simétrica.

Se φ é uma forma bilinear simétrica no espaço vectorial real W, podemos associar-lhe uma forma quadrática q definindo a função $q : W \to \mathbb{R}$ por $q(x) = \varphi(x, x)$. Reciprocamente, sendo q uma forma quadrática num espaço linear real, podemos definir uma forma bilinear simétrica do seguinte modo:

$$\varphi(x, y) = \frac{1}{2} \left(q(x + y) - q(x) - q(y) \right).$$

A forma φ é designada por *forma polar* de q.

Exercício 7.15. Seja q uma forma quadrática definida num espaço linear real W. Mostre que

$$\frac{1}{2} \left(q(x + y) - q(x) - q(y) \right),$$

é uma forma bilinear simétrica em W. ▲

Exercício 7.16. Mostre que o conjunto $\Sigma_2(W)$, das formas bilineares simétricas definidas num espaço vectorial real W, é um espaço linear para as operações:

$$(\varphi_1 + \varphi_2)(x, y) = \varphi_1(x, y) + \varphi_2(x, y), \qquad (\alpha\varphi)(x, y) = \alpha\varphi(x, y) \text{ com } \alpha \in \mathbb{R}.$$

▲

Suponha-se que q é uma forma quadrática num espaço linear real W, representada pela matriz A em relação a uma certa base ordenada B de W, isto é, $q_A(\mathbf{x}_B) = \mathbf{x}_B^T A \mathbf{x}_B$. Se considerarmos outra base ordenada B_1 de W, a forma quadrática será representada por outra matriz. De facto, se M é a matriz de mudança da base B_1 para a base B, ou seja, $\mathbf{x}_B = M\mathbf{x}_{B_1}$, tem-se

$$q_A(\mathbf{x}_B) = \mathbf{x}_{B_1}^T M^T A M \mathbf{x}_{B_1} = q_C(\mathbf{x}_{B_1}), \quad \text{com } C = M^T A M.$$

Ou seja, as matrizes A e C que representam a mesma forma quadrática, em relação a bases distintas, satisfazem a relação $C = M^T A M$, onde M é uma matriz invertível.

Quando existe uma matriz invertível K tal que $A = KCK^T$, as matrizes A e C dizem-se *congruentes*. Escrevemos $A \simeq C$ quando A e C são congruentes.

Exercício 7.17. Mostre que se A e C são matrizes que representam uma forma bilinear simétrica $\varphi : W \times W \to \mathbb{R}$, em relação a bases distintas fixadas em W, então $A \simeq C$.

▲

Exercício 7.18. Mostre que:

a) A relação de congruência[5] de matrizes é:
   - *Reflexiva*: $A \simeq A$,
   - *Simétrica*: $A \simeq C \Longrightarrow C \simeq A$;
   - *Transitiva*: ($A \simeq C$ e $C \simeq D$) $\Longrightarrow A \simeq D$.

b) Matrizes congruentes têm a mesma característica.

▲

Em 1852, o matemático J. J. Sylvester descobriu que a inércia de uma matriz é invariante por congruência.

Definição 7.10. A inércia de uma matriz simétrica real é a sequência (ρ, ν, ζ) onde ρ, ν e ζ são, respectivamente, o número de valores próprios positivos, negativos, e nulos, contando as suas multiplicidades algébricas.

[5] Uma relação que seja reflexiva, simétrica e transitiva, diz-se uma relação de equivalência.

Teorema 7.6. Lei de inércia de Sylvester
Sejam A e B matrizes simétricas reais.

$$A \simeq B \text{ se e só se } A \text{ e } B \text{ têm a mesma inércia.}$$

Demonstração. Como A é simétrica e real, esta matriz é ortogonalmente diagonalizável, isto é, $A = PDP^T$ com P ortogonal e D uma matriz diagonal cujas entradas da diagonal principal são os valores próprios de A. Se a inércia de A é (p, j, s), podemos tomar para D a matriz

$$D = \operatorname{diag}(\lambda_1, \ldots, \lambda_p, -\lambda_{p+1}, \ldots, -\lambda_{p+j}, 0, \ldots, 0),$$

onde $\lambda_i > 0$, para $i = 1, \ldots, p + j$.

Considere-se agora a matriz (invertível) $\tilde{D} = \operatorname{diag}(\lambda_1^{-1/2}, \ldots, \lambda_{p+j}^{-1/2}, 1 \ldots, 1)$. São válidas as igualdades

$$\tilde{D}^T P^T A P \tilde{D} = \tilde{D} D \tilde{D} = \begin{bmatrix} I_{p\times p} & & \\ & -I_{j\times j} & \\ & & \mathbf{0}_{s\times s} \end{bmatrix} = E, \tag{7.21}$$

onde $I_{k\times k}$ designa a matriz identidade de ordem k, $\mathbf{0}_{s\times s}$ a matriz nula de ordem s e os blocos omissos são matrizes nulas. A igualdade (7.21) diz-nos que a matriz A é congruente com a matriz E.

Se a matriz simétrica B tem a mesma inércia de A, a matriz B é congruente com a matriz E acima. Ou seja, $B \simeq E$. Como a relação de congruência é transitiva e reflexiva (ver Exercício 7.18), tem-se que $A \simeq E$ e $B \simeq E$ implica $A \simeq B$.

Para a implicação recíproca, suponha-se que $A \simeq B$ e mostre-se que A e B têm a mesma inércia. Admita-se que a inércia de A é (p, j, s) e a de B é (q, k, t). Então $A \simeq E$ onde E é a matriz em (7.21) e $B \simeq F$, com

$$F = \begin{bmatrix} I_{q\times q} & & \\ & -I_{k\times k} & \\ & & \mathbf{0}_{t\times t} \end{bmatrix}.$$

Como $A \simeq B$, $A \simeq E$ e $B \simeq F$, tem-se $E \simeq F$ (transitividade da relação de congruência). Ou seja, existe uma matriz invertível K tal que $F = K^T E K$. Sendo $E \simeq F$, as matrizes E e F têm a mesma característica (cf. Lema 3.1, pág. 148), e portanto $t = s$. Usando a igualdade $F = K^T E K$ é fácil ver que $p = q$. $\square$

Como corolário do teorema anterior (e da sua demonstração) obtemos o resultado conhecido por *forma canónica* (isto é, uma forma particularmente simples) de uma forma quadrática real.

Teorema 7.7. Forma canónica de uma forma quadrática real

Seja A uma matriz simétrica real e $q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ uma forma quadrática em $\mathbb{R}^n$. Existe uma base em $\mathbb{R}^n$ em relação à qual a forma q se exprime na forma

$$q(\mathbf{x}) = \mathbf{y}^T E\mathbf{y} = (y_1)^2 + \cdots + (y_p)^2 - (y_{p+1})^2 - \cdots - (y_{p+j})^2,$$

onde a matriz E é a matriz em (7.21).

A uma representação deste tipo chama-se *forma canónica* da forma quadrática q. O número de valores próprios positivos de A é p, e o número de valores próprios negativos é j.

Chama-se *característica* da forma quadrática a $p + j$ e *assinatura* a $p - j$.

Note-se que a característica de uma forma quadrática coincide com a característica da matriz simétrica que define a forma quadrática (uma vez que matrizes semelhantes têm a mesma característica) bem como com o número de valores próprios não nulos dessa matriz.

As formas bilineares simétricas são representadas por matrizes congruentes, em relação a bases distintas fixadas no seu domínio (cf. Exercício 7.17). Segue como corolário do Teorema 7.6 o seguinte resultado que dá uma forma canónica para uma forma bilinear simétrica real.

Teorema 7.8. Forma canónica de uma forma bilinear simétrica real

Seja $\varphi : W \times W \to \mathbb{R}$ uma forma bilinear simétrica. Existe uma base B de W em relação à qual φ tem a forma

$$\varphi(x, y) = \mathbf{x}_B^T E\mathbf{y}_B = x_1y_1 + \cdots + x_py_p - x_{p+1}y_{p+1} - \cdots - x_{p+j}y_{p+j},$$

onde $\mathbf{x}_B = (x_1, \ldots, x_n)$, $\mathbf{y}_B = (y_1, \ldots, y_n)$ e E é a matriz em (7.21).

7.4.2 Cónicas e quádricas

Os resultados obtidos relativos a matrizes simétricas reais (e formas quadráticas) são aplicados nesta secção ao estudo de equações envolvendo formas quadráticas em $\mathbb{R}^2$ e $\mathbb{R}^3$. Dada uma matriz simétrica real A pretende-se identificar o lugar geométrico definido por uma equação quadrática, ou seja, uma equação do tipo

$$q_A(\mathbf{x}) + \mathbf{k}^T\mathbf{x} + j = 0, \quad j \in \mathbb{R}, \tag{7.22}$$

onde $q_A(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$, para $\mathbf{x} \in \mathbb{R}^2$ (ou $\mathbf{x} \in \mathbb{R}^3$) e $\mathbf{k}$ um vector (fixo) de $\mathbb{R}^2$ ou de $\mathbb{R}^3$.

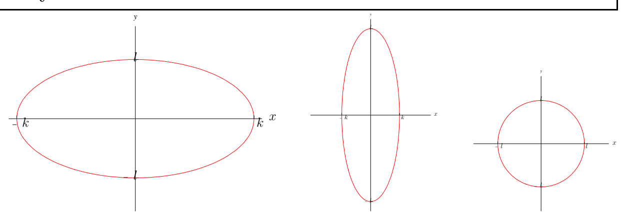
Chama-se forma quadrática associada à equação quadrática (7.22) ao termo $q_A(\mathbf{x})$. Por exemplo,

Equação quadrática	$q_A(\mathbf{x})$	Matriz A
$8x^2 - 4xy + 5y^2 - 36y = 0$	$8x^2 - 4xy + 5y^2$	$\begin{bmatrix} 8 & -2 \\ -2 & 5 \end{bmatrix}$
$xy - y + 2 = 0$	xy	$\begin{bmatrix} 0 & 1/2 \\ 1/2 & 0 \end{bmatrix}$
$x^2 + 2yz + 2y^2 - 3z^2 - 12 = 0$	$x^2 + 2yz + 2y^2 - 3z^2$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & -3 \end{bmatrix}$

Os lugares geométricos determinados por equações quadráticas de duas variáveis são designados por *cónicas* ou *secções cónicas*. As cónicas mais importantes são as elipses, as circunferências, as hipérboles e as parábolas. Estas cónicas dizem-se *cónicas não degeneradas*.

Uma cónica não degenerada diz-se na *posição padrão* (relativa aos eixos coordenados) se a sua equação pode ser expressa numa das formas indicadas nas Figuras 7.1 a 7.3.

$$\boxed{\frac{x^2}{k^2} + \frac{y^2}{l^2} = 1; \quad k, l > 0 \quad \text{Elipse ou circunferência}}$$

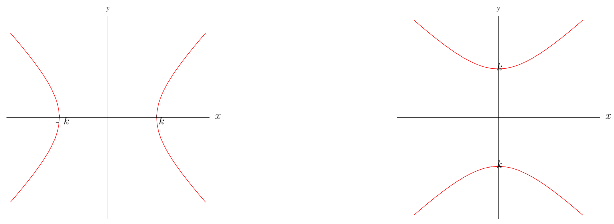


$\boxed{k > l}$ $\boxed{k < l}$ $\boxed{k = l}$

Figura 7.1: Elipse ou circunferência.

Note-se que as formas quadráticas associadas às equações que definem as cónicas não degeneradas das figuras 7.1 a 7.3 são representadas por uma matriz diagonal. Quando tal acontece dizemos que a equação que define a cónica está

Hipérbole



$$\boxed{\frac{x^2}{k^2}-\frac{y^2}{l^2}=1; \quad k,l>0} \qquad \boxed{\frac{y^2}{l^2}-\frac{x^2}{k^2}=1; \quad k,l>0}$$

Figura 7.2: Hipérbole.

Parábola $\qquad$ Parábola

$y^2 = kx$ $\qquad$ $x^2 = ky$

$\boxed{k>0}$ $\quad$ $\boxed{k<0}$ $\quad$ $\boxed{k>0}$ $\quad$ $\boxed{k<0}$

Figura 7.3: Parábola.

na forma reduzida. As equações reduzidas das cónicas não degeneradas são, para $k, l > 0$ e $p \neq 0$:

Circunferência:	$\frac{x^2}{k^2}+\frac{y^2}{k^2}=1$	$\frac{1}{k^2}I$.
Elipse:	$\frac{x^2}{k^2}+\frac{y^2}{l^2}=1$	$\operatorname{diag}\left(\frac{1}{k^2},\frac{1}{l^2}\right)$.
Hipérbole:	$\frac{x^2}{k^2}-\frac{y^2}{l^2}=1$	$\operatorname{diag}\left(\frac{1}{k^2},-\frac{1}{l^2}\right)$.
Parábola:	$y^2 = px$	$\operatorname{diag}(0,1)$.

As *cónicas degeneradas* são o conjunto vazio, um ponto, uma recta, duas rectas concorrentes, e duas rectas paralelas. As equações reduzidas das cónicas degeneradas são:

Um ponto: $\frac{x^2}{k^2} + \frac{y^2}{k^2} = 0$

Uma recta: $x^2 = 0.$

Duas rectas paralelas: $\frac{x^2}{k^2} = 1.$

Duas rectas concorrentes: $\frac{x^2}{k^2} - \frac{y^2}{l^2} = 0.$

Para identificar a cónica definida por (7.22), usa-se a seguinte estratégia: (a) reduz-se a forma quadrática (associada à equação) à sua forma canónica (ver Teorema 7.7); (b) "eliminam-se" da equação o máximo de termos lineares possível, usando para tal um processo conhecido por "*completar quadrados*".

Comecemos por relembrar que se A é uma matriz simétrica real, existe uma matriz ortogonal P que diagonaliza A. Isto é, tal que $A = PDP^T$ com $P^{-1} = P^T$ e $D = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$. Assim,

$$q_A(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} = (P^T\mathbf{x})^T DP^T\mathbf{x} = \mathbf{y}^T D\mathbf{y} = q_D(\mathbf{y}), \quad \text{com } \mathbf{y} = P^T\mathbf{x}.$$

Uma matriz ortogonal P tem $\det P = \pm 1$, no entanto podemos sempre escolher P por forma a que $\det(P) = 1$ (basta trocar duas colunas de P e a ordem dos valores próprios nas entradas da diagonal de D). No caso de formas quadráticas em $\mathbb{R}^2$, a matriz P tem uma das formas indicadas no Exemplo 7.1 (pág. 373), e portanto pode sempre ser escolhida como sendo uma matriz de rotação R_θ.

A matriz P tem nas colunas vectores próprios de A de norma igual a 1, e é a matriz que realiza a mudança da base formada pelos vectores próprios (definida pelas colunas de P) para a base canónica de $\mathbb{R}^n$. Assim a mudança de variáveis $P^T\mathbf{x} = \mathbf{y}$ aplica direcções definidas pelos vectores da base canónica de $\mathbb{R}^n$ em direcções definidas por vectores próprios de A. Estas direcções são designadas por *direcções principais*.

Redução a direcções principais

Efectuando a mudança de variáveis $P^T\mathbf{x} = \mathbf{y}$, onde P é uma matriz ortogonal que diagonaliza a matriz simétrica real A, a forma quadrática $q_A(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ transforma-se em $q_D(\mathbf{y}) = \mathbf{y}^T D\mathbf{y}$, onde $D = P^T AP$ é uma matriz diagonal. Ou seja,

$$q_A(\mathbf{x}) = q_D(\mathbf{y}) = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2,$$

onde $\lambda_1, \dots, \lambda_n$ são os valores próprios de A. Além disso, a mudança de variáveis $P^T\mathbf{x} = \mathbf{y}$ pode sempre efectuar-se usando para P uma matriz com determinante igual a 1.

Exemplo 7.7. Considere-se a equação

$$8x_1^2 - 4x_1x_2 + 5x_2^2 - 36 = 0. \tag{7.23}$$

A forma quadrática associada a esta equação é $q_A(x_1, x_2) = 8x_1^2 - 4x_1x_2 + 5x_2^2$, ou seja,

$$q_A(x_1, x_2) = \begin{bmatrix} x_1 & x_2 \end{bmatrix} A \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 8 & -2 \\ -2 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

A matriz A tem valores próprios $\lambda_1 = 4$ e $\lambda_2 = 9$. Como $\lambda_1 \neq \lambda_2$, quaisquer vectores próprios associados a λ_1 e a λ_2 são ortogonais. Tome-se, por exemplo, para vector próprio associado a λ_1 o vector $\mathbf{v}_1 = (1, 2)$, e para vector próprio associado a λ_2 o vector $\mathbf{v}_2 = (-2, 1)$ (verifique que estes vectores são vectores próprios de A). Uma matriz que tenha nas colunas $\mathbf{v}_1$ e $\mathbf{v}_2$ diagonaliza A, porém não é uma matriz ortogonal já que estes vectores não têm norma unitária. Se se pretende uma matriz que diagonalize ortogonalmente A, devemos normalizar estes vectores. Por exemplo, uma matriz P que diagonaliza ortogonalmente A é

$$P = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}. \tag{7.24}$$

A matriz P tem determinante igual a 1, e representa por isso uma rotação dos eixos coordenados (ver Exemplo 7.1, pág. 373 e Exemplo 6.7-4, pág. 338).

A mudança de coordenadas $P^T\mathbf{x} = \mathbf{y}$ transforma a equação (7.23) na equação

$$q_D(\mathbf{y}) = 4\,y_1^2 + 9\,y_2^2 = 36 \iff \frac{y_1^2}{9} + \frac{y_2^2}{4} = 1.$$

Em conclusão, a cónica é uma elipse centrada na origem do sistema de eixos coordenados y_1y_2 (eixos com direcções dos vectores próprios de A), cujos semieixos medem 3 e 2. O sistema de eixos coordenados y_1y_2 é obtido por rotação do sistema de eixos x_1x_2, rotação esta definida pela matriz P. Na Figura 7.4 é apresentado o gráfico da cónica. ♦

A equação que define a cónica do exemplo anterior não possui termos lineares em $\mathbf{x}$ uma vez que é uma equação do tipo

$$q_A(\mathbf{x}) + \mathbf{k}^T\mathbf{x} + j = 0,$$

com $\mathbf{k} = \mathbf{0}$. Se essa equação tivesse um termo linear $\mathbf{k}^T\mathbf{x}$ com $\mathbf{k} \neq \mathbf{0}$, depois de reduzir a forma quadrática associada à equação à sua forma canónica, deveríamos "completar os quadrados", eliminando o máximo de termos lineares possível. No exemplo seguinte ilustramos este procedimento.

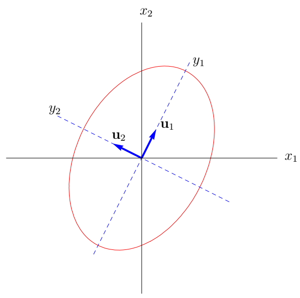


Figura 7.4: Elipse cujos eixos têm as direcções dos vectores próprios (unitários) $\mathbf{u}_1$ e $\mathbf{u}_2$ da matriz A do Exemplo 7.7.

Exemplo 7.8. Considere a modificação seguinte da equação do Exemplo 7.7:

$$8x_1^2 - 4x_1x_2 + 5x_2^2 - \frac{64}{\sqrt{5}}x_1 + \frac{52}{\sqrt{5}}x_2 + 4 = 0. \tag{7.25}$$

Do Exemplo 7.7 sabemos que a forma quadrática associada a esta equação pode ser diagonalizada usando a mudança de variáveis $\mathbf{x} = P\mathbf{y}$, onde P é a matriz em (7.24). Esta mudança de variáveis, $x_1 = \dfrac{1}{\sqrt{5}}(y_1 - 2y_2)$ e $x_2 = \dfrac{1}{\sqrt{5}}(2y_1 + y_2)$, reduz a equação (7.25) à forma

$$0 = 4\,y_1^2 + 9\,y_2^2 + 8y_1 + 36y_2 + 4.$$

Completem-se agora os quadrados:

$$\begin{aligned} 4y_1^2 + 8y_1 &= 4\left(y_1^2 + 2y_1\right) = 4\left(y_1^2 + 2y_1 + 1 - 1\right) = 4\left(y_1 + 1\right)^2 - 4 \\ 9y_2^2 + 36y_2 &= 9\left(y_2^2 + 4y_2\right) = 9\left(y_2^2 + 4y_2 + 4 - 4\right) = 9\left(y_2 + 2\right)^2 - 36. \end{aligned}$$

Assim,

$$\begin{aligned} 4\,y_1^2 + 9\,y_2^2 + 8y_1 + 36y_2 + 4 = 0 &\iff 4\left(y_1 + 1\right)^2 - 4 + 9\left(y_2 + 2\right)^2 - 36 + 4 = 0 \\ &\iff 4\left(y_1 + 1\right)^2 + 9\left(y_2 + 2\right)^2 = 36 \\ &\iff \frac{\left(y_1 + 1\right)^2}{9} + \frac{\left(y_2 + 2\right)^2}{4} = 1 \end{aligned}$$

Ou seja, a cónica definida por (7.25) é obtida da cónica dada por (7.23) mediante uma translação (nas coordenadas y_1, y_2) definida pelo vector $\mathbf{a} = (-1, -2)$. Na

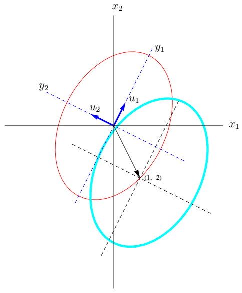


Figura 7.5: A elipse do Exemplo 7.8 é uma translação da elipse do Exemplo 7.7.

Figura 7.5 apresenta-se a cónica definida pela equação (7.23) bem como a cónica transladada definida por (7.25).

♦

Exercício 7.19. Seja C o conjunto de pontos do plano definido pela equação

$$ax^2 + bxy + cy^2 + dx + ey + f = 0,$$

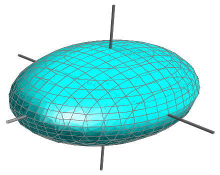
e λ, μ os valores próprios da matriz simétrica que define a forma quadrática associada a esta equação. Mostre que:

a) Se $\lambda\mu > 0$, então C é uma elipse, um ponto, ou o conjunto vazio.
b) Se $\lambda\mu < 0$, então C é uma hipérbole, ou um par de rectas concorrentes.
c) Se $\lambda\mu = 0$, então existem duas possibilidades:
   (i) Se $\lambda \neq 0$ ou $\mu \neq 0$, então C é uma parábola, um par de rectas paralelas, uma recta, ou o conjunto vazio.
   (ii) Se $\lambda = \mu = 0$, então C é uma recta, ou o conjunto vazio.

▲

Quádricas

As equações quadráticas em 3 variáveis definem superfícies no espaço tridimensional designadas por *quádricas*. As quádricas na posição padrão relativamente aos eixos coordenados são definidas pelas equações indicadas nas Figuras 7.6-7.14. A curva de intersecção de uma superfície com um plano é designada por *traço*.

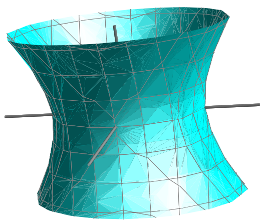


$\boxed{\text{Elipsóide.}}$

$$\frac{x^2}{k^2} + \frac{y^2}{l^2} + \frac{z^2}{n^2} = 1$$

Os traços são elipses, ou circunferências, ou um ponto, ou o conjunto vazio.

Figura 7.6: Elipsóide

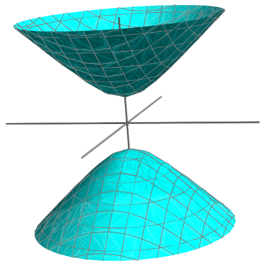


$\boxed{\text{Hiperbolóide de uma folha.}}$

$$\frac{x^2}{k^2} + \frac{y^2}{l^2} - \frac{z^2}{n^2} = 1$$

Os traços são elipses, hipérboles, ou um par de rectas concorrentes.

Figura 7.7: Hiperbolóide de uma folha.

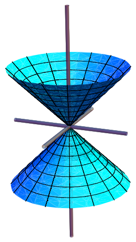


$\boxed{\text{Hiperbolóide de 2 folhas}}$

$$\frac{x^2}{k^2} + \frac{y^2}{l^2} - \frac{z^2}{n^2} = -1$$

Os traços são elipses, ou hipérboles, ou um ponto, ou o conjunto vazio.

Figura 7.8: Hiperbolóide de 2 folhas.

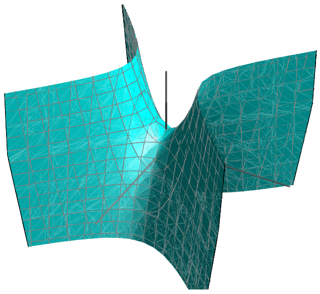


$\boxed{\text{Cone elíptico}}$

$$\frac{x^2}{k^2} + \frac{y^2}{l^2} = z^2$$

Os traços são elipses, ou hipérboles, ou rectas concorrentes, ou um ponto.

Figura 7.9: Cone elíptico.

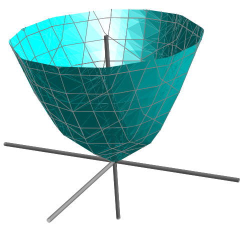


$\boxed{\text{Parabolóide hiperbólico}}$

$$\frac{y^2}{l^2} - \frac{x^2}{k^2} = z$$

Os traços são hipérboles, ou parábolas, ou duas rectas concorrentes, ou o conjunto vazio.

Figura 7.10: Hiperbolóide parabólico.

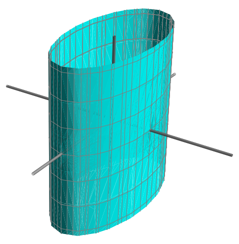


$\boxed{\text{Parabolóide elíptico}}$

$$\frac{x^2}{k^2} + \frac{y^2}{l^2} = z$$

Os traços são elipses, ou parábolas, ou um ponto, ou o conjunto vazio.

Figura 7.11: Parabolóide elíptico.

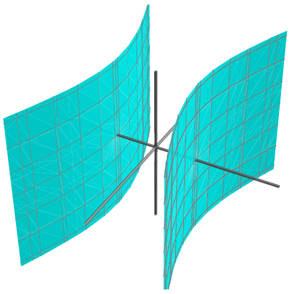


$\boxed{\text{Cilindro elíptico}}$

$$\frac{x^2}{k^2} + \frac{y^2}{l^2} = 1$$

Os traços são elipses, ou circunferências, ou rectas paralelas, ou o conjunto vazio.

Figura 7.12: Cilindro elíptico.

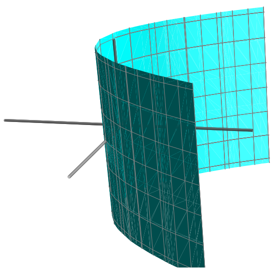


$\boxed{\text{Cilindro hiperbólico}}$

$$\frac{x^2}{k^2} - \frac{y^2}{l^2} = 1$$

Os traços são hipérboles, ou rectas, ou o conjunto vazio.

Figura 7.13: Cilindro hiperbólico.



$\boxed{\text{Cilindro parabólico}}$

$$\frac{x^2}{k^2} = py$$

Os traços são parábolas, ou rectas, ou o conjunto vazio.

Figura 7.14: Cilindro parabólico.

Nota 46. *As superfícies das Figuras 7.6-7.9 dizem-se superfícies de característica 3, as das Figuras 7.10-7.13 superfícies de característica 2 e a da Figura 7.14 é uma superfície de característica 1. Esta classificação está de acordo com o facto de a forma quadrática associada à equação que define a superfície ter respectivamente característica 3, 2 e 1.*

As equações que definem as quádricas das Figuras 7.6-7.14 não têm termos cruzados, isto é, termos envolvendo xy, yz ou xz, pelo que a matriz que define a forma quadrática associada à equação é diagonal. No caso da forma quadrática associada à equação da superfície possuir termos cruzados isso indica que a superfície não se encontra na posição padrão. No entanto, a superfície pode ser obtida por rotação de uma superfície na posição padrão. A rotação é definida por uma matriz ortogonal P do tipo 3×3 com determinante igual a 1 (ver Exemplo 7.4). Esta matriz P diagonaliza a matriz que define a forma quadrática associada à equação da superfície.

Exemplo 7.9. Pretendemos identificar a superfície definida pela equação

$$-15 - 20x_1 + 8x_1^2 + 14y_1 - 4x_1y_1 + 5y_1^2 - 8z_1 + 4z_1^2 = 0. \tag{7.26}$$

A forma quadrática associada a esta equação é definida pela matriz simétrica

$$A = \begin{bmatrix} 8 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 4 \end{bmatrix}.$$

Os valores próprios desta matriz são $\lambda_1 = 9$ e $\lambda_2 = 4$ (com multiplicidade algébrica 2). Os espaços próprios são

$$E(4) = \operatorname{Span}\{(1,2,0),(0,0,1)\}, \qquad E(9) = \operatorname{Span}\{(-2,1,0)\}.$$

Realizando a mudança de variáveis $(x_1, y_1, z_1) = P(x, y, z)$, onde P é a matriz ortogonal

$$P = \begin{bmatrix} 0 & 1/\sqrt{5} & -2/\sqrt{5} \\ 0 & 2/\sqrt{5} & 1/\sqrt{5} \\ 1 & 0 & 0 \end{bmatrix},$$

isto é, $x_1 = 1/\sqrt{5}y - 2/\sqrt{5}z$, $y_1 = 2/\sqrt{5}y + 1/\sqrt{5}z$ e $z_1 = x$, a equação (7.26) é transformada na equação

$$4x^2 + 4y^2 + 9z^2 - 8x + \frac{8}{\sqrt{5}}y + \frac{54}{\sqrt{5}}z - 15 = 0. \tag{7.27}$$

Esta equação tem termos lineares, pelo que vamos completar os quadrados.

$$\begin{aligned}
4x^2 - 8x &= 4(x^2 - 2x + 1 - 1) = 4(x-1)^2 - 4,\\
4y^2 + \frac{8}{\sqrt{5}}y &= 4\left(y^2 + \frac{2}{\sqrt{5}}y + \frac{1}{5} - \frac{1}{5}\right) = 4\left(y + \frac{1}{\sqrt{5}}\right)^2 - \frac{4}{5},\\
9z^2 + \frac{54}{\sqrt{5}}z &= 9\left(z^2 + \frac{6}{\sqrt{5}} + \frac{9}{5} - \frac{9}{5}\right) = 9\left(z + \frac{3}{\sqrt{5}}\right)^2 - \frac{81}{5}.
\end{aligned}$$

Por conseguinte, a equação (7.27) escreve-se

$$4(x-1)^2 + 4\left(y + \frac{1}{\sqrt{5}}\right)^2 + 9\left(z + \frac{3}{\sqrt{5}}\right)^2 = 36,$$

ou equivalentemente,

$$\frac{(x-1)^2}{9} + \frac{(y + \frac{1}{\sqrt{5}})^2}{9} + \frac{(z + \frac{3}{\sqrt{5}})^2}{4} = 1.$$

Concluindo, a equação (7.26) define um elipsóide de semieixos medindo respectivamente $3, 3$ e 2 unidades. Este elipsóide é obtido por translação segundo o vector $(-1, \frac{1}{\sqrt{5}}, \frac{3}{\sqrt{5}})$ de um elipsóide na posição padrão no sistema de eixos x, y, z. ♦

7.5 Decomposição em valores singulares (SVD)

A decomposição de uma matriz em valores singulares (SVD[6]) é actualmente usada em várias áreas da matemática aplicada como a estatística, a álgebra computacional e numérica, a bioquímica, o reconhecimento de voz e imagem, etc.. Uma decomposição SVD de uma matriz real A, do tipo $p \times n$, é uma factorização de A na forma $A = U\Sigma V^T$, com U e V matrizes ortogonais e Σ uma matriz em blocos com todos os blocos nulos à excepção de um único bloco diagonal. Como veremos, toda a matriz admite uma decomposição SVD. Na base da obtenção desta decomposição encontra-se a diagonalização ortogonal (resp. unitária) de matrizes simétricas reais (resp. hermitianas).

Nesta secção mostramos que qualquer matriz admite uma decomposição SVD e fazemos a interpretação geométrica desta decomposição. A partir desta factorização, obtemos bases ortonormadas para os quatro subespaços fundamentais associados a uma matriz.

[6]SVD é a abreviatura de "singular value decomposition".

7.5.1 Factorização SVD

Seja A uma matriz real do tipo $p \times n$. A matriz A^TA é uma matriz simétrica (de ordem n), e portanto ortogonalmente diagonalizável. Assim, existe uma base ortonormada de $\mathbb{R}^n$ formada por vectores próprios de A^TA. Seja $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ uma tal base, e designe-se por λ_i o valor próprio de A^TA associado a $\mathbf{v}_i$. Logo,

$$\begin{aligned}\|A\mathbf{v}_i\|^2 &= (A\mathbf{v}_i)^T A\mathbf{v}_i = \mathbf{v}_i^T A^T A\mathbf{v}_i \\ &= \mathbf{v}_i^T \lambda_i \mathbf{v}_i \qquad\qquad (\mathbf{v}_i \text{ é um vector próprio de } A^TA) \\ &= \lambda_i \mathbf{v}_i^T \mathbf{v}_i = \lambda_i \|\mathbf{v}_i\|^2 = \lambda_i \qquad (\mathbf{v}_i \text{ é um vector próprio de norma } 1).\end{aligned}$$

Conclui-se que os valores próprios de A^TA são todos não negativos visto que, para cada λ_i se tem $\lambda_i = \|A\mathbf{v}_i\|^2 \geq 0$.

Nota 47. *Quando A é uma matriz complexa, a matriz A^HA é hermitiana. Logo, os valores próprios de A^HA são reais e não negativos já que, sendo $(\lambda_i, \mathbf{v}_i)$ um par próprio de A^HA com $\|\mathbf{v}_i\| = 1$, se tem $\|A\mathbf{v}_i\|^2 = \mathbf{v}_i^H A^H A\mathbf{v}_i = \lambda_i$.*

Definição 7.11. Valores singulares

Seja A uma matriz real (resp. complexa) do tipo $p \times n$.

Os *valores singulares* σ_i de A são as raízes quadradas dos valores próprios de A^TA (resp. de A^HA). Isto é, $\sigma_i = \sqrt{\lambda_i}$ para $1 \leq i \leq n$, onde λ_i é um valor próprio de A^TA (resp. de A^HA).

Se λ_i é um valor próprio de A^TA e $\mathbf{v}_i$ é um vector próprio associado com norma unitária, então o valor singular $\sigma_i = \sqrt{\lambda_i}$ é igual à norma de $A\mathbf{v}_i$, visto que

$$\sigma_i = \sqrt{\lambda_i} = \sqrt{\|A\mathbf{v}_i\|^2} = \|A\mathbf{v}_i\|, \quad i = 1, \ldots, n. \tag{7.28}$$

É habitual ordenar os valores singulares de uma matriz por ordem decrescente:

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0.$$

Exemplo 7.10. Seja $A = \begin{bmatrix} 3 & -8 & 10 \\ 4 & 5 & -2 \end{bmatrix}$.

A matriz

$$A^TA = \begin{bmatrix} 25 & -4 & 22 \\ -4 & 89 & -90 \\ 22 & -90 & 104 \end{bmatrix}$$

tem valores próprios $\lambda_1 = 189$ e $\lambda_2 = 29$ e $\lambda_3 = 0$. Portanto, os valores singulares de A são

$$\sigma_1 = \sqrt{189} = 3\sqrt{21}, \quad \sigma_2 = \sqrt{29}, \quad \sigma_3 = 0.$$

♦

Enunciamos agora o teorema da decomposição SVD.

Teorema 7.9. Decomposição SVD
Seja A uma matriz real do tipo $p \times n$ de característica igual a k. Existem matrizes ortogonais U, V e uma matriz Σ, tais que

$$A = U\Sigma V^T.$$

(i) A matriz Σ, do tipo $p \times n$, tem a forma

$$\Sigma = \begin{bmatrix} D & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

onde D é a matriz diagonal $D = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_k)$, sendo os σ_i's os valores singulares positivos de A assim ordenados: $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k > 0$.

(ii) Os vectores coluna de V, $\mathbf{v}_1, \ldots, \mathbf{v}_n$ (por esta ordem), formam uma base ortonormada de $\mathbb{R}^n$. Estes vectores são vectores próprios de A^TA que satisfazem as igualdades $A^TA\mathbf{v}_i = \lambda_i\mathbf{v}_i = \sigma_i^2\mathbf{v}_i$, onde $\sigma_{k+1} = \cdots = \sigma_n = 0$.

(iii) Os vectores coluna de U formam uma base ortonormada de $\mathbb{R}^p$. As suas primeiras k colunas verificam: $\mathbf{u}_i = \dfrac{1}{\sigma_i}A\mathbf{v}_i$, para $i = 1, \ldots k$, onde $\mathbf{v}_i$ é a coluna número i de V.

Uma decomposição $A = U\Sigma V^T$ diz-se uma *decomposição de A em valores singulares* (abreviadamente SVD). Os vectores coluna de U são designados por *vectores singulares esquerdos*, enquanto que os vectores coluna de V se designam por *vectores singulares direitos*.

Nota 48. *A matriz Σ da decomposição SVD é única. No entanto, as matrizes U e V não o são, o que significa que a decomposição SVD não é única.*

Antes de provarmos o Teorema 7.9 mostramos o lema seguinte.

Lema 7.3. Seja A uma matriz real $p \times n$, $\lambda_1, \ldots, \lambda_n$ os valores próprios de $A^T A$, e $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ uma base ortonormada de $\mathbb{R}^n$, tal que cada vector $\mathbf{v}_i$ é um vector próprio associado ao valor próprio λ_i, e $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$.

Se k é o número de valores singulares positivos de A (contando com as multiplicidades), então $\{A\mathbf{v}_1, \ldots, A\mathbf{v}_k\}$ é uma base ortogonal do espaço das colunas de A, e a característica de A é k.

Demonstração. Se mostrarmos que o conjunto $\{A\mathbf{v}_1, \ldots, A\mathbf{v}_k\}$ é uma base ortogonal de $EC(A)$, resulta imediatamente que $\dim EC(A) = k = \operatorname{car}(A)$.

Comecemos por verificar que o conjunto $\{A\mathbf{v}_1, \ldots, A\mathbf{v}_n\}$ é ortogonal. Para $i \neq j$, tem-se

$$\begin{aligned}
(A\mathbf{v}_i)^T(A\mathbf{v}_j) &= \mathbf{v}_i^T A^T A\mathbf{v}_j \\
&= \mathbf{v}_i^T \lambda_j \mathbf{v}_j \qquad (\lambda_j \text{ é valor próprio de } A^T A \text{ associado a } \mathbf{v}_j) \\
&= \lambda_j \mathbf{v}_i^T \mathbf{v}_j = \lambda_j \langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0,
\end{aligned}$$

e portanto $\{A\mathbf{v}_1, \ldots, A\mathbf{v}_n\} \subset \mathbb{R}^p$ é um conjunto ortogonal do espaço das colunas de A. Como $\sigma_i = \|A\mathbf{v}_i\| \neq 0$ para $i \leq k$ e $\|A\mathbf{v}_i\| = 0$ para $i > k$, o subconjunto $\{A\mathbf{v}_1, \ldots, A\mathbf{v}_k\}$ é um conjunto ortogonal que não contém o vector nulo, e portanto linearmente independente (cf. Proposição 5.1, pág. 259). Falta mostrar que este conjunto gera o espaço das colunas de A. Seja $\mathbf{y} \in EC(A)$, isto é, $\mathbf{y} = A\mathbf{x}$ para algum vector $\mathbf{x} \in \mathbb{R}^n$. Escrevendo $\mathbf{x} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$, tem-se

$$\begin{aligned}
\mathbf{y} = A\mathbf{x} &= c_1 A\mathbf{v}_1 + c_2 A\mathbf{v}_2 + \cdots + c_k A\mathbf{v}_k + c_{k+1} A\mathbf{v}_{k+1} + \cdots + c_n A\mathbf{v}_n \\
&= c_1 A\mathbf{v}_1 + c_2 A\mathbf{v}_2 + \cdots + c_k A\mathbf{v}_k, \quad \text{já que } A\mathbf{v}_i = 0 \text{ para } k < i \leq n.
\end{aligned}$$

Assim, o conjunto linearmente independente $\{A\mathbf{v}_1, \ldots, A\mathbf{v}_k\}$ gera $EC(A)$, e portanto é uma base de $EC(A)$. □

Passamos agora à demonstração do Teorema 7.9.

Demonstração. (**do Teorema 7.9**)

Comecemos por notar que sendo $A = U\Sigma V^T$ uma decomposição SVD de A, o produto $A^T A$ tem a forma

$$A^T A = (V\Sigma^T U^T)(U\Sigma V^T) = V\Sigma^T\Sigma V^T,$$

onde $\Sigma^T\Sigma$ é uma matriz diagonal e V é uma matriz ortogonal. Como $A^T A$ é uma matriz simétrica real, a igualdade anterior é uma diagonalização ortogonal de A^TA, uma vez que a matriz diagonal $\Sigma^T\Sigma$ possui na diagonal principal os valores próprios de A^TA. Logo, as colunas de V formam uma base ortonormada de $\mathbb{R}^n$ constituída por vectores próprios de A^TA. Seja $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ uma tal base, ordenada como no Lema 7.3. Deste lema segue que $\{A\mathbf{v}_1, \ldots, A\mathbf{v}_k\} \subset \mathbb{R}^p$ é uma base ortogonal para o espaço da colunas de A. Normalizando cada vector desta base, formamos a base ortonormada $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ de $EC(A)$. Aplicando as igualdades (7.28), temos

$$\mathbf{u}_i = \frac{1}{\|A\mathbf{v}_i\|}A\mathbf{v}_i = \frac{1}{\sigma_i}A\mathbf{v}_i \Longleftrightarrow A\mathbf{v}_i = \sigma_i\mathbf{u}_i, \quad i = 1, \ldots, k. \tag{7.29}$$

Podemos, se necessário, completar o conjunto ortonormado $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ por forma a obter uma base ortonormada $\{\mathbf{u}_1, \ldots, \mathbf{u}_k, \ldots, \mathbf{u}_p\}$ de $\mathbb{R}^p$.

Finalmente, considerando as bases ortonormadas $(\mathbf{u}_1, \ldots, \mathbf{u}_p)$ e $(\mathbf{v}_1, \ldots, \mathbf{v}_n)$, respectivamente de $\mathbb{R}^p$ e $\mathbb{R}^n$, construam-se as seguintes matrizes ortogonais U e V,

$$U = \begin{bmatrix} | & | & & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_p \\ | & | & & | \end{bmatrix} \quad \text{e} \quad V = \begin{bmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \\ | & | & & | \end{bmatrix}.$$

Verifiquemos agora a igualdade $A = U\Sigma V^T$ ou, equivalentemente, $AV = U\Sigma$ (uma vez que $V^{-1} = V^T$):

$$\begin{aligned} AV = A\begin{bmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \\ | & | & & | \end{bmatrix} &= \begin{bmatrix} | & | & & | \\ A\mathbf{v}_1 & A\mathbf{v}_2 & \cdots & A\mathbf{v}_n \\ | & | & & | \end{bmatrix} \\ &= \begin{bmatrix} | & & | & | & & | \\ \sigma_1\mathbf{u}_1 & \cdots & \sigma_k\mathbf{u}_k & \mathbf{0} & \cdots & \mathbf{0} \\ | & & | & | & & | \end{bmatrix}, \end{aligned}$$

e

$$\begin{aligned} U\Sigma &= \begin{bmatrix} | & | & & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_p \\ | & | & & | \end{bmatrix} \left[\begin{array}{cccc|c} \sigma_1 & & & & \\ & \sigma_2 & & & \\ & & \ddots & & \mathbf{0} \\ & & & \sigma_k & \\ \hline & & \mathbf{0} & & \mathbf{0} \end{array}\right] \\ &= \begin{bmatrix} | & & | & | & & | \\ \sigma_1\mathbf{u}_1 & \cdots & \sigma_k\mathbf{u}_k & \mathbf{0} & \cdots & \mathbf{0} \\ | & & | & | & & | \end{bmatrix}. \end{aligned}$$

Logo, $AV = U\Sigma$. $\square$

Exercício 7.20. Use a decomposição SVD para mostrar que qualquer matriz real A, de característica k, se pode escrever na forma:

$$A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_k \mathbf{u}_k \mathbf{v}_k^T. \tag{7.30}$$

Mostre ainda que as matrizes $\mathbf{u}_i \mathbf{v}_i^T$, do tipo $p \times n$, têm característica 1.

Compare a decomposição anterior com a decomposição espectral (7.12) de matrizes unitariamente diagonalizáveis. ▲

Nota 49. *Qualquer matriz complexa admite uma decomposição SVD. Ou seja, se A é uma matriz complexa, existem matrizes unitárias U e V, e uma matriz Σ com um único bloco diagonal não nulo com entradas na diagonal principal positivas, tais que $A = U\Sigma V^H$. A demonstração deste resultado é análoga à realizada para matrizes reais, bastando substituir nessa demonstração o símbolo "T" por "H".*

Exemplo 7.11. Considere-se a matriz $A = \begin{bmatrix} 3 & -8 & 10 \\ 4 & 5 & -2 \end{bmatrix}$.

Viu-se, no Exemplo 7.10 (pág. 418) que os valores singulares de A são $\sigma_1 = \sqrt{189}$, $\sigma_2 = \sqrt{29}$ e $\sigma_3 = 0$. É fácil verificar que $(5, -29, 32)$, $(15, 7, 4)$ e $(-34, 46, 47)$ são vectores próprios de $A^T A$, associados respectivamente a $\lambda_1 = 189$, $\lambda_2 = 29$ e $\lambda_3 = 0$. Estes vectores são ortogonais pois estão associados a valores próprios distintos (de uma matriz simétrica). Normalizando estes vectores, obtemos

$$\mathbf{v}_1 = \frac{1}{3\sqrt{210}}(5, -29, 32), \mathbf{v}_2 = \frac{1}{\sqrt{290}}(15, 7, 4) \quad \text{e} \quad \mathbf{v}_3 = \frac{1}{3\sqrt{609}}(-34, 46, 47).$$

Uma matriz V da decomposição SVD de A é

$$V = \begin{bmatrix} \frac{5}{3\sqrt{210}} & \frac{15}{\sqrt{290}} & \frac{-34}{3\sqrt{609}} \\ \frac{-29}{3\sqrt{210}} & \frac{7}{\sqrt{290}} & \frac{46}{3\sqrt{609}} \\ \frac{32}{3\sqrt{210}} & \frac{4}{\sqrt{290}} & \frac{47}{3\sqrt{609}} \end{bmatrix},$$

uma vez que $\mathbf{v}_1, \mathbf{v}_2$ e $\mathbf{v}_3$ são vectores próprios associados respectivamente a valores próprios λ_i que respeitam a ordenação $\lambda_1 > \lambda_2 > \lambda_3$.

A matriz D é diagonal, tendo na diagonal principal os valores singulares não nulos de A (associados às colunas de V) ordenados de forma decrescente. Conclui-se assim que $D = \operatorname{diag}(\sqrt{189}, \sqrt{29})$, e a matriz Σ correspondente é

$$\Sigma = \begin{bmatrix} \sqrt{189} & 0 & 0 \\ 0 & \sqrt{29} & 0 \end{bmatrix}.$$

Falta determinar uma matriz U que, neste caso, é do tipo 2×2 (uma vez que A é do tipo 2×3). Como a característica de A é 2 (pelo Lema 7.3 a característica é igual ao número de valores singulares não nulos), as 2 primeiras colunas de U podem ser os vectores $A\mathbf{v}_1$ e $A\mathbf{v}_2$ normalizados. Usando a equação (7.29), obtém-se

$$\mathbf{u}_1 = \frac{1}{\|A\mathbf{v}_1\|}A\mathbf{v}_1 = \frac{1}{\sigma_1}A\mathbf{v}_1 = \frac{1}{\sqrt{189}\sqrt{10}}\begin{bmatrix} 9\sqrt{21} \\ -3\sqrt{21}\end{bmatrix} = \frac{1}{\sqrt{10}}\begin{bmatrix} 3 \\ -1\end{bmatrix},$$

$$\mathbf{u}_2 = \frac{1}{\|A\mathbf{v}_2\|}A\mathbf{v}_2 = \frac{1}{\sigma_2}A\mathbf{v}_2 = \frac{1}{\sqrt{10}}\begin{bmatrix} 1 \\ 3\end{bmatrix}.$$

Neste caso não necessitamos de completar $\{\mathbf{u}_1, \mathbf{u}_2\}$ para construir U, uma vez que esta matriz é 2×2. Assim,

$$U = \frac{1}{\sqrt{10}}\begin{bmatrix} 3 & 1 \\ -1 & 3\end{bmatrix}.$$

Sugere-se como exercício que confirme a igualdade $A = U\Sigma V^T$. ♦

Exercício 7.21. Mostre que se A é real e simétrica, uma decomposição SVD para A coincide com uma diagonalização ortogonal de A. ▲

Exercício 7.22. Mostre que toda a matriz quadrada A admite uma *decomposição polar* da forma $A = PQ$, onde P é uma matriz semidefinida positiva com a mesma característica de A, e Q é uma matriz ortogonal.
Sugestão: Sendo $A = U\Sigma V^T$ uma decomposição SVD de A, escreva A na forma $A = (U\Sigma U^T)(UV^T)$.

▲

7.5.2 Bases ortonormadas para os quatro subespaços fundamentais

Uma decomposição SVD de uma matriz real A do tipo $p \times n$ fornece-nos bases ortonormadas para os quatro subespaços fundamentais associados a A, e em particular bases ortonormadas para o núcleo e o contradomínio da função linear $T(\mathbf{x}) = A\mathbf{x}$. De facto, se $A = U\Sigma V^T$ é uma decomposição SVD de A, a demonstração do Teorema 7.9 diz-nos que

$$T(\mathbf{v}_i) = A\mathbf{v}_i = \begin{cases} \sigma_i\mathbf{u}_i & \text{para } i = 1, \dots, k \\ \mathbf{0} & \text{para } k < i \leq n, \end{cases}$$

onde os $\mathbf{v}_i$'s e os $\mathbf{u}_i$'s são, respectivamente, vectores singulares direitos e esquerdos. Desta igualdade concluímos que os últimos $(n-k)$ vectores singulares direitos, $\mathbf{v}_{k+1}, \ldots, \mathbf{v}_n$, formam uma base ortonormada para o núcleo de T, e os primeiros k vectores singulares esquerdos, $\mathbf{u}_1, \ldots, \mathbf{u}_k$, uma base ortonormada para o contradomínio de T (ver ainda Lema 7.3). Atendendo a que, quer os vectores singulares direitos, quer os esquerdos, formam bases ortonormadas respectivamente do espaço de partida ($\mathbb{R}^n$) e de chegada de T (neste caso $\mathbb{R}^p$), tem-se que os k primeiros vectores singulares direitos, $\mathbf{v}_1, \ldots, \mathbf{v}_k$, formam uma base ortonormada para o complemento ortogonal $N(T)^\perp = N(A)^\perp = EL(A)$, e os últimos $(p-k)$ vectores singulares esquerdos, $\mathbf{u}_{k+1}, \ldots, \mathbf{u}_p$, uma base ortonormada para o complemento ortogonal do contradomínio de T, isto é, para $(Im(T))^\perp = EC(A)^\perp = N(A^T)$ (relembre as relações (5.18) na página 270).

Em conclusão, uma decomposição SVD de A fornece bases ortonormadas para os quatro subespaços fundamentais associados a A (núcleo de A, o espaço das colunas de A e os respectivos complementos ortogonais destes subespaços). Na Figura 7.15 completa-se o diagrama apresentado na Figura 5.7 da página 271, agora em termos da função linear T e das bases dos vectores singulares de A (esquerdos e direitos).

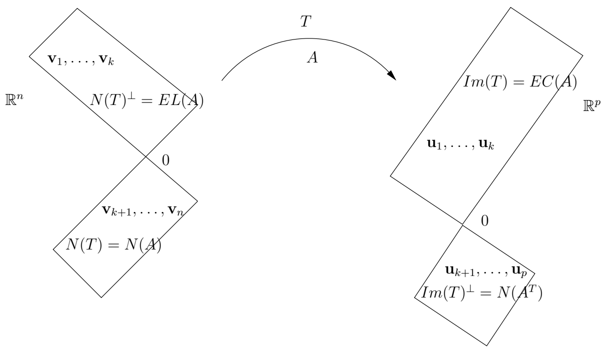


Figura 7.15: Bases ortonormadas formadas por vectores singulares de A para os quatro subespaços fundamentais. A função linear $T : \mathbb{R}^n \to \mathbb{R}^p$ é definida por $T(\mathbf{x}) = A\mathbf{x}$, e $\mathrm{car}(A) = k$.

Interpretação geométrica da decomposição SVD

Interpretamos agora geometricamente a acção de uma função linear $T : \mathbb{R}^n \to \mathbb{R}^p$ sobre vectores de $\mathbb{R}^n$ usando a decomposição SVD de uma matriz que represente T.

Seja $T(\mathbf{x}) = A\mathbf{x}$, com A matriz real do tipo $p \times n$. Consideremos o efeito de T em vectores da esfera de $\mathbb{R}^n$ de raio 1, isto é, em vectores $\mathbf{x} = (x_1, \ldots, x_n)$ satisfazendo a igualdade

$$x_1^2 + \cdots + x_n^2 = 1.$$

A matriz A admite uma decomposição SVD, isto é, $A = U\Sigma V^T$, com U, Σ e V nas condições do enunciado do Teorema 7.9. As colunas de V, $\mathbf{v}_1, \ldots, \mathbf{v}_n$, formam uma base ortonormada de $\mathbb{R}^n$. Logo, qualquer vector $\mathbf{x} \in \mathbb{R}^n$ pode escrever-se como combinação linear dos vectores desta base. Em particular, qualquer vector $\mathbf{x}$ da esfera de raio 1 escreve-se na forma

$$\mathbf{x} = x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n, \quad \text{com} \sum_{i=1}^{n} x_i^2 = 1.$$

Aplicando o Teorema 7.9, a imagem de $\mathbf{x}$ por T é

$$\begin{aligned} T(\mathbf{x}) = A\mathbf{x} &= A\left(x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n\right) \\ &= x_1A\mathbf{v}_1 + \cdots + x_nA\mathbf{v}_n \\ &= x_1A\mathbf{v}_1 + \cdots + x_kA\mathbf{v}_k \qquad (\text{já que } A\mathbf{v}_i = \mathbf{0} \text{ para } i = k, \ldots, n) \\ &= x_1\sigma_1\mathbf{u}_1 + \cdots + x_k\sigma_k\mathbf{u}_k \qquad (\text{por } (7.29)) \\ &= y_1\mathbf{u}_1 + \cdots + y_k\mathbf{u}_k \qquad (\text{com } y_i = x_i\sigma_i). \end{aligned}$$

Para vectores $\mathbf{x}$ da esfera considerada, como $x_i = \dfrac{y_i}{\sigma_i}$ para $i = 1, \ldots, k \leq n$, é satisfeita a desigualdade

$$x_1^2 + \cdots + x_n^2 = 1 \Longrightarrow \frac{y_1^2}{\sigma_1^2} + \cdots + \frac{y_k^2}{\sigma_k^2} \leq 1.$$

A inequação $\dfrac{y_1^2}{\sigma_1^2} + \cdots + \dfrac{y_k^2}{\sigma_k^2} \leq 1$ pode ser vista como um elipsóide em $\mathbb{R}^k$ cujos eixos têm as direcções dos vectores $\mathbf{u}_i$, e tal que os comprimentos dos semieixos são iguais aos valores singulares σ_i.

Podemos assim resumir a acção de $A = U\Sigma V^T$ em $\mathbb{R}^n$: a matriz A colapsa $(n - k)$ direcções do domínio de T; seguidamente a bola unitária de $\mathbb{R}^k$ é expandida, contraída ou mantida nas direcções dos primeiros k vectores singulares direitos, tendo por efeito transformar a referida bola num elipsóide; finalmente este elipsóide é aplicado em $\mathbb{R}^p$. Na Figura 7.16 é ilustrado este comportamento de $T(\mathbf{x}) = A\mathbf{x}$ para uma matriz A do tipo 3×3, de característica igual a 2.

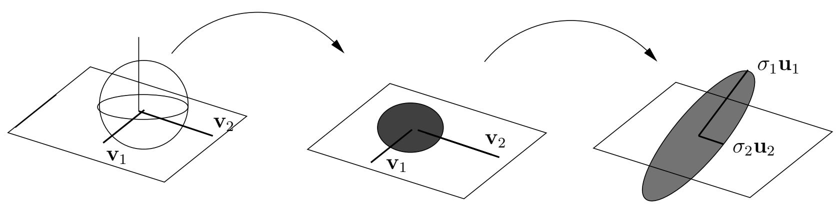


Figura 7.16: Interpretação geométrica da decomposição SVD para uma matriz A do tipo 3×3 e de característica 2.

7.6 Algumas aplicações de valores singulares

Os valores singulares de uma matriz e a decomposição SVD desempenham um papel relevante nos algoritmos computacionais modernos sendo habitualmente utilizados em problemas envolvendo matrizes de grandes dimensões. A grandeza dos valores singulares de uma matriz é relevante em vários contextos como teremos oportunidade de constatar nas aplicações que detalhamos adiante. Por exemplo, se a matriz é quadrada, a existência de valores singulares nulos revela que a matriz é singular. Além disso, mesmo que a matriz (quadrada) não tenha valores singulares nulos se tiver valores singulares próximos de zero, a decomposição SVD indica que essa matriz está próxima de ser uma matriz singular, facto que pode conduzir a sistemas mal condicionados que tenham essa matriz como matriz dos coeficientes (ver Secção 7.6.1).

As aplicações que abordamos neste texto são:

i) Número de condição de uma matriz e sistemas lineares mal condicionados.

ii) Compressão de imagem.

iii) Pseudoinversa de uma matriz e soluções de mínimos quadrados.

Como referências bibliográficas para esta secção aconselhamos a leitura do artigo [6] e referências nele incluídas.

7.6.1 Número de condição e sistemas mal condicionados

Dado um sistema linear $A\mathbf{x} = \mathbf{b}$, em que A é uma matriz invertível, pretende-se saber se pequenas perturbações (ou erros) no segundo membro do sistema se traduzem ou não em pequenas perturbações na solução. Ou seja, se $\mathbf{x}$ e $\mathbf{x}^*$ são soluções, respectivamente do sistema $A\mathbf{x} = \mathbf{b}$ e do sistema perturbado $A\mathbf{x} = \mathbf{b}^*$, pretendemos comparar o erro relativo em $\mathbf{x}^*$, $\frac{\|\mathbf{x}-\mathbf{x}^*\|}{\|\mathbf{x}\|}$, com o erro relativo no segundo membro $\mathbf{b}^*$, isto é, com $\frac{\|\mathbf{b}-\mathbf{b}^*\|}{\|\mathbf{b}\|}$. Considere-se o exemplo seguinte.

Exemplo 7.12. Seja $A\mathbf{x} = \mathbf{b}$, com

$$A = \begin{bmatrix} 1.0001 & 1 \\ 1 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 0.1 \\ 0 \end{bmatrix}.$$

A matriz A é invertível, e a única solução do sistema é $\mathbf{x} = (1000, -1000)$. Considere-se agora o vector perturbado $\mathbf{b}^* = (0.11, 0)$. A única solução do sistema $A\mathbf{x} = \mathbf{b}^*$ é $\mathbf{x}^* = (1100, -1100)$. Comparando as duas soluções, vemos que o sistema é muito sensível a perturbações no segundo membro. Quando tal acontece o sistema diz-se mal condicionado. ♦

O chamado *número de condição* é usado como medida de quão mal condicionado é um sistema. Este número pode ser definido como a razão entre o maior e o menor valor singular (não nulo) de A.

Neste contexto, interessa começar por definir norma num espaço linear, generalizando o conceito de norma proveniente de um produto interno conforme se introduziu no Capítulo 5.

Definição 7.12. Norma num espaço linear

Seja W um espaço linear sobre o corpo $\mathbb{K}$.

Uma *norma* em W é uma função $\|\cdot\|$ definida em W com valores em $\mathbb{R}$, que satisfaz as propriedades seguintes, para todos os $x, y \in W$,

- $\|x\| \geq 0 \quad$ e $\quad \|x\| = 0 \Longleftrightarrow x = 0$.
- $\|\alpha x\| = |\alpha|\|x\|, \quad$ para todos os escalares $\alpha \in \mathbb{K}$.
- $\|x + y\| \leq \|x\| + \|y\| \quad$ (Desigualdade triangular).

Existem normas no espaço linear das matrizes tipo $p \times n$ induzidas por normas definidas nos espaços vectoriais $\mathbb{R}^k$ (ou $\mathbb{C}^k$), com $k = n$ e $k = p$. Estas normas matriciais recebem a designação de *normas induzidas por normas vectoriais*.

Norma matricial induzida por normas vectoriais

Uma norma vectorial em $\mathbb{C}^k$ com $k = p, n$, induz uma norma no espaço das matrizes complexas A, do tipo $p \times n$, dada por

$$\|A\| = \max_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|, \quad \text{para } \mathbf{x} \in \mathbb{C}^n. \tag{7.31}$$

Note-se que a existência do máximo em (7.31) está garantida pelo importante Teorema de Weierstrass da Análise.[7]

Deixamos como exercício a verificação de que uma norma matricial induzida por normas vectoriais, satisfaz as condições da Definição 7.12.

Verifiquemos agora que a norma (7.31) é equivalente a

$$\|A\| = \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|}. \tag{7.32}$$

Sendo $\mathbf{x} \neq \mathbf{0}$ um vector qualquer de $\mathbb{C}^n$, o vector $\frac{\mathbf{x}}{\|\mathbf{x}\|}$ tem norma unitária. Portanto, é válida a desigualdade

$$\|A\| \geq \left\| A\frac{\mathbf{x}}{\|\mathbf{x}\|} \right\| = \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} \Longrightarrow \|A\| \geq \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|}.$$

Para a desigualdade inversa, como $\max_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} \geq \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|}$, para todos os $\mathbf{x} \neq \mathbf{0}$, esta desigualdade é verificada em particular para os vectores $\mathbf{y}$ tais que $\|\mathbf{y}\| = 1$. Ou seja, $\max_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} \geq \|A\mathbf{y}\|$ para todos os vectores $\mathbf{y}$ de norma unitária. Logo,

$$\max_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} \geq \max_{\|\mathbf{y}\|=1} \|A\mathbf{y}\| = \|A\|.$$

Concluimos assim que (7.32) é equivalente a (7.31).

Deixamos como exercício a verificação de que uma norma matricial induzida por uma norma vectorial goza das propriedades seguintes.

[7] Teorema de Weierstrass: "Toda a função contínua de um conjunto limitado e fechado $K \subset W$ com valores em $\mathbb{R}$, atinge um máximo e um mínimo em pontos de K". A esfera unitária em espaços vectoriais de dimensão finita é um conjunto limitado e fechado, e a norma num espaço vectorial é uma função contínua.

Exercício 7.23. Mostre que

$$\text{a) } \|A\mathbf{x}\| \leq \|A\|\|\mathbf{x}\|. \tag{7.33}$$

$$\text{b) } \|AB\| \leq \|A\|\|B\|.$$

▲

Considere-se agora a norma vectorial usual em $\mathbb{R}^n$ (ou em $\mathbb{C}^n$). É frequente designar-se esta norma vectorial por *norma-2* e usar-se a notação $\|\mathbf{x}\|_2 = \sqrt{\mathbf{x}^T\mathbf{x}}$. Usamos igualmente a notação $\|\cdot\|_2$ para a norma matricial induzida pela norma-2.

Proposição 7.10. Seja A uma matriz real e $\|A\|_2 = \max\limits_{\|x\|_2=1} \|A\mathbf{x}\|_2$. Então,

- $\|A\|_2 = \sigma_1$, onde σ_1 é o maior valor singular de A.
- Se A é invertível, então

$$\|A^{-1}\|_2 = \frac{1}{\sigma_k}, \quad \text{onde } \sigma_k \text{ é o menor valor singular (não nulo) de } A.$$

Demonstração. Como $\|A\mathbf{x}\|_2 = \sqrt{\mathbf{x}^T A^T A\mathbf{x}} = \sqrt{q_{A^TA}(\mathbf{x})}$, resulta da Proposição 7.8 que o máximo da forma quadrática $q_{A^TA}(\mathbf{x})$ na esfera unitária, é igual ao maior valor próprio de A^TA. Os valores próprios de A^TA são não negativos e as suas raízes quadradas são os valores singulares de A (cf. Definição 7.11). Por conseguinte, $\|A\|_2 = \sigma_1$, onde σ_1 é o maior valor singular de A.

Sendo A invertível, a matriz AA^T também o é. Além disso, os espectros de A^TA e de AA^T são iguais, uma vez que uma matriz e a sua transposta têm os mesmos valores próprios. Como,

$$\|A^{-1}\mathbf{x}\|_2 = \sqrt{\mathbf{x}^T(A^{-1})^TA^{-1}\mathbf{x}} = \sqrt{\mathbf{x}^T(AA^T)^{-1}\mathbf{x}},$$

e os valores próprios da inversa de uma matriz são os inversos dos valores próprios da matriz (Exercício 4.1), segue de novo pela Proposição 7.8 que $\|A^{-1}\|_2 = \frac{1}{\sigma_k}$, onde σ_k é o menor valor singular de A. □

Exercício 7.24. Sejam U e V matrizes ortogonais, respectivamente, de ordens p e n, e A uma matriz $p \times n$. Mostre as igualdades

$$\text{a) } \|UA\|_2 = \|A\|_2, \qquad \text{b) } \|AV\|_2 = \|A\|_2.$$

▲

Considere-se de novo os sistemas $A\mathbf{x} = \mathbf{b}$ e $A\mathbf{x} = \mathbf{b}^*$, onde A é uma matriz invertível. Sejam $\mathbf{x}$ e $\mathbf{x}^*$, respectivamente, a solução de $A\mathbf{x} = \mathbf{b}$ e de $A\mathbf{x} = \mathbf{b}^*$ (com $\mathbf{b} \neq \mathbf{0}$). Pretendemos comparar o erro relativo $\frac{\|\mathbf{x}-\mathbf{x}^*\|_2}{\|\mathbf{x}\|_2}$ com o erro relativo $\frac{\|\mathbf{b}-\mathbf{b}^*\|_2}{\|\mathbf{b}\|_2}$. Atendendo à desigualdade (7.33), obtém-se:

$$\begin{aligned}
\|\mathbf{b} - \mathbf{b}^*\|_2 &= \|A\mathbf{x} - A\mathbf{x}^*\|_2 = \|A(\mathbf{x} - \mathbf{x}^*)\|_2 \leq \|A\|_2\|\mathbf{x} - \mathbf{x}^*\|_2, \\
\|\mathbf{x} - \mathbf{x}^*\|_2 &= \|A^{-1}\mathbf{b} - A^{-1}\mathbf{b}^*\|_2 = \|A^{-1}(\mathbf{b} - \mathbf{b}^*)\|_2 \leq \|A^{-1}\|_2\|\mathbf{b} - \mathbf{b}^*\|_2, \\
\|\mathbf{b}\|_2 &= \|A\mathbf{x}\|_2 \leq \|A\|_2\|\mathbf{x}\|_2, \\
\|\mathbf{x}\|_2 &= \|A^{-1}\mathbf{b}\|_2 \leq \|A^{-1}\|_2\|\mathbf{b}\|_2.
\end{aligned}$$

Destas desigualdades resulta

$$\frac{\|\mathbf{x} - \mathbf{x}^*\|_2}{\|\mathbf{x}\|_2} \leq \|A\|_2\|A^{-1}\|_2\frac{\|\mathbf{b} - \mathbf{b}^*\|_2}{\|\mathbf{b}\|_2} = k(A)\frac{\|\mathbf{b} - \mathbf{b}^*\|_2}{\|\mathbf{b}\|_2} \tag{7.34}$$

$$\frac{\|\mathbf{x} - \mathbf{x}^*\|_2}{\|\mathbf{x}\|_2} \geq \frac{1}{\|A\|_2\|A^{-1}\|_2}\frac{\|\mathbf{b} - \mathbf{b}^*\|_2}{\|\mathbf{b}\|_2} = \frac{1}{k(A)}\frac{\|\mathbf{b} - \mathbf{b}^*\|_2}{\|\mathbf{b}\|_2}. \tag{7.35}$$

O número $k(A) = \|A\|_2\|A^{-1}\|_2$ é designado por *número de condição* da matriz A. Da Proposição 7.10 conclui-se:

O número de condição $k(A)$, de uma matriz invertível A, é dado por

$$k(A) = \|A\|_2\|A^{-1}\|_2 = \frac{\sigma_1}{\sigma_k},$$

onde σ_1 e σ_k são, respectivamente, o maior e o menor valor singular de A.

As expressões (7.34) e (7.35) podem reescrever-se na forma

$$\frac{1}{k(A)}\frac{\|\mathbf{b} - \mathbf{b}^*\|_2}{\|\mathbf{b}\|_2} \leq \frac{\|\mathbf{x} - \mathbf{x}^*\|_2}{\|\mathbf{x}\|_2} \leq k(A)\frac{\|\mathbf{b} - \mathbf{b}^*\|_2}{\|\mathbf{b}\|_2}. \tag{7.36}$$

Da dupla desigualdade (7.36), vê-se que pode acontecer que um certo vector $\mathbf{b}^*$ possua um erro relativo muito pequeno, e no entanto a solução $\mathbf{x}^*$ possua um erro relativo muito grande, uma vez que o erro relativo em $\mathbf{b}^*$ é multiplicado por $k(A)$. Assim, se $k(A) >> 1$ (k muito superior a 1) o erro relativo de $\mathbf{x}^*$ pode ser muito superior ao erro relativo de $\mathbf{b}^*$, caso em que dizemos que o sistema é mal condicionado.

No Exemplo 7.12 os valores singulares da matriz A são, aproximadamente, $\sigma_1 \simeq 2.00005$ e $\sigma_2 \simeq 0.00005$. O número de condição é portanto $k(A) = \frac{\sigma_1}{\sigma_2} \simeq 400002$.

Distância de uma matriz a outra de característica inferior

Os valores singulares de uma matriz A são igualmente úteis para medir a distância de uma matriz a matrizes de característica inferior. Comecemos por esclarecer o que se entende por *distância entre matrizes*.

Definição 7.13. Sejam A e B matrizes do tipo $p \times n$. Chamamos distância entre A e B a

$$dist(A, B) = \|A - B\|.$$

Mostramos agora que dada uma matriz A, a distância de A a uma matriz de característica inferior r é igual ao valor singular σ_{r+1}.

Proposição 7.11. Seja A uma matriz do tipo $p \times n$, de característica k. Considere-se $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k$ os k valores singulares não nulos de A.

A distância de A ao conjunto das matrizes $p \times n$ de característica $r < k$, é igual ao valor singular σ_{r+1}. Isto é,

$$\sigma_{r+1} = \min_{\mathrm{car}(B)=r} \|A - B\|_2.$$

Demonstração. Seja $A = U\Sigma V^T$ uma decomposição SVD de A, e B uma matriz de característica $r < k \leq n$. Do Exercício 7.24, temos

$$\begin{aligned}\|A - B\|_2 &= \|U\Sigma V^T - B\|_2 = \|U(\Sigma - U^T BV)V^T\|_2 \\ &= \|\Sigma - U^T BV\|_2 = \|\Sigma - X\|_2, \quad \text{com } X = U^T BV.\end{aligned}$$

Do Lema 3.1 (pág. 148), resulta

$$\mathrm{car}(X) = \mathrm{car}(B) = r < k \leq n.$$

Portanto, $\dim N(X) = n - r \geq 1$, já que $n - r > 0$. Considere-se o subespaço S de $\mathbb{R}^n$ formado pelos vectores $\mathbf{x}$ da forma $(x_1, x_2, \ldots, x_{r+1}, 0, \ldots, 0)$. Verifiquemos que a dimensão de $S \cap N(X)$ é maior ou igual a 1. Como $\dim(S + N(X)) \leq n$, da Proposição 3.10, pág. 141, tem-se

$$\begin{aligned}\dim(S \cap N(X)) &= \dim S + \dim N(X) - \dim(S + N(X)) \\ &\geq \dim S + \dim N(X) - n = r + 1 + n - r - n = 1.\end{aligned}$$

Por conseguinte, existe um vector $\mathbf{y} = (y_1, y_2, \ldots, y_{r+1}, 0, \ldots, 0)$ de $S \cap N(X)$ tal que $\|\mathbf{y}\|_2 = 1$. A distância $\|A - B\|_2$ é

$$\begin{aligned}\|A - B\|_2 &= \|\Sigma - X\|_2 \geq \|\Sigma\mathbf{y} - X\mathbf{y}\|_2 = \|\Sigma\mathbf{y}\|_2 \\ &= \sqrt{\sigma_1^2 y_1^2 + \cdots + \sigma_{r+1}^2 y_{r+1}^2} \geq \sqrt{\sigma_{r+1}^2 (y_1^2 + \cdots + y_{r+1}^2)} = \sigma_{r+1}.\end{aligned}$$

Ou seja, $\|A - B\|_2 \geq \sigma_{r+1}$ para qualquer matriz B de característica r.

Para finalizar a demonstração, basta mostrar que existe uma matriz $X = U^T BV$, de característica r, tal que $\|A - B\|_2 = \sigma_{r+1}$.

Tomando $B = \begin{bmatrix} D_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$, com $D_r = \operatorname{diag}(\sigma_1, \ldots, \sigma_r)$, esta matriz satisfaz a igualdade pretendida. □

7.6.2 Compressão de imagem

Ao digitalizarmos uma imagem, usando por exemplo um *scanner*, codifica-se a imagem através de uma matriz contendo $(p \times n)$ píxeis. A cada píxel associa-se um número que corresponde à sua cor. Por exemplo, se a imagem é a preto e branco, associando o número zero à cor branca e 1 à cor preta (números entre zero e um indicam diferentes tonalidades de cinzento), obtemos uma matriz (de números) do tipo $p \times n$.

Admitindo que píxeis correspondentes à mesma cor (ou a cores próximas não distinguiveis pelo olho humano) produzem valores singulares nulos ou próximos de zero, é razoável pensar-se que uma matriz de característica inferior à característica da matriz original contenha a informação necessária para reproduzir a imagem com boa qualidade. O objectivo da compressão de imagem é substituir a matriz original por uma sua aproximação de característica inferior, de tal forma que a imagem correspondente à matriz de característica inferior tenha uma qualidade "próxima" da imagem original. Fazemos aqui uma referência breve à utilização da SVD neste contexto.

Usando a igualdade (7.30), a decomposição em valores singulares de uma matriz A pode ser escrita na forma

$$A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_k \mathbf{u}_k \mathbf{v}_k^T, \tag{7.37}$$

onde $\sigma_1, \ldots, \sigma_k$ designam os valores singulares positivos de A, e os vectores $\mathbf{u}_1, \ldots, \mathbf{u}_k$ e $\mathbf{v}_1, \ldots, \mathbf{v}_k$ os k primeiros vectores singulares de A, respectivamente, esquerdos e direitos. Na expressão (7.37), as matrizes $A_i = \sigma_i \mathbf{u}_i \mathbf{v}_i^T$ são do tipo $p \times n$ (do mesmo tipo de A), uma vez que $\mathbf{u}_i \in \mathbb{R}^p$ e $\mathbf{v}_i \in \mathbb{R}^n$, e têm característica igual a 1 (ver Exercício 7.20, pág. 422). Podemos assim usar diversas aproximações da matriz A, a saber:

$$
\begin{aligned}
&S_1 = A_1 = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T, && \text{uma aproximação de característica 1;}\\
&S_2 = A_1 + A_2 = A_1 + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T, && \text{uma aproximação de característica 2;}\\
&\quad\vdots\\
&S_r = A_1 + \cdots + A_r = S_{r-1} + \sigma_r \mathbf{u}_r \mathbf{v}_r^T, && \text{uma aproximação de característica r.}
\end{aligned}
$$

Dada uma matriz X, do tipo $p \times n$, a norma $\|X\|_F$, designada por norma de Frobenius[8], é definida como a raiz quadrada da soma dos quadrados das entradas de X, isto é,

$$\|X\|_F^2 = \sum_{\substack{1 \le i \le p \\ 1 \le j \le n}} (x_{ij})^2. \tag{7.38}$$

A norma de Frobenius de uma matriz X é igual à norma usual de um vector de $\mathbb{R}^{pn}$ cujas componentes são as entradas da matriz X, e é também igual a

$$\|X\|_F^2 = \sum (x_{ij})^2 = \operatorname{tr}(X^T X).$$

Deixamos como exercício a verificação de que a norma de Frobenius satisfaz todas as propriedades da Definição 7.12.

Considerando uma decomposição em valores singulares da matriz A, tem-se

$$
\begin{aligned}
\|A\|_F^2 &= \operatorname{tr}(V\Sigma^T U^T U \Sigma V^T) = \operatorname{tr}(V \Sigma^T \Sigma V^T) = \operatorname{tr}(\Sigma^T \Sigma V^T V)\\
&= \operatorname{tr}(\Sigma^T \Sigma) = \sum_{i=1}^{k} \sigma_i^2,
\end{aligned}
$$

onde a antepenúltima igualdade resulta da propriedade: $\operatorname{tr}(AB) = \operatorname{tr}(BA)$ (ver Exercício 5.1, pág. 245).

É fácil verificar que a norma da aproximação S_r (de característica r) é

$$\|S_r\|_F^2 = \sum_{i=1}^{r} \sigma_i^2.$$

[8]Ferdinand Georg Frobenius (1849 – 1917), matemático alemão.

O erro relativo (em norma) da aproximação S_r da matriz A é dado por

$$
\begin{aligned}
e(r) = \frac{\|A - S_r\|_F}{\|A\|_F} &= \sqrt{\frac{\sum_{i=r+1}^{k} \sigma_i^2}{\sum_{i=1}^{k} \sigma_i^2}} = \sqrt{\frac{\sigma_1^2 + \cdots + \sigma_k^2 - \sigma_1^2 - \cdots - \sigma_r^2}{\sigma_1^2 + \cdots + \sigma_k^2}} \\
&= \sqrt{1 - \frac{\sigma_1^2 + \cdots + \sigma_r^2}{\sigma_1^2 + \cdots + \sigma_k^2}} = \sqrt{1 - \frac{\sum_{i=1}^{r} \sigma_i^2}{\sum_{i=1}^{k} \sigma_i^2}}. \qquad (7.39)
\end{aligned}
$$

Da expressão do erro $e(r)$ deduz-se que, se os valores singulares $\sigma_{r+1}, \ldots, \sigma_k$ são muito próximos de zero, o erro cometido ao usar S_r para aproximar A é pequeno. Ou seja, S_r é uma boa aproximação de A.

Apresentamos a seguir um exemplo muito simples, ilustrativo de como se usa a decomposição SVD em compressão de imagem. Os cálculos foram efectuados usando o sistema $Mathematica^{®}$.

A imagem da Figura 7.17 é representada por uma matriz A, do tipo 24×24, cujas entradas são zero (correspondente à cor branca) e 1 (para a cor preta). Os valores singulares não nulos de A, com 4 casas decimais, são:

$$\{18, 8212, 5.2511, 4.2539, 3.0832, 1.9935, 1.6293, 1.0429, 0.9328\}.$$

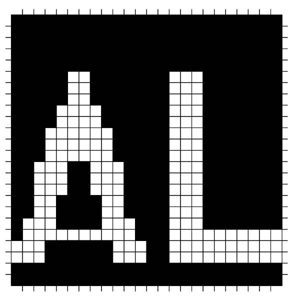


Figura 7.17: Uma imagem com 24^2 pixéis.

Suponha-se que se pretende aproximar A por S_r (de característica r) com um erro relativo não superior a 7%. O cálculo de $e(2)$ dá-nos $e(2) = 0.0442668$ (enquanto que $e(1) > 0.07$). Podemos assim aproximar A por S_2. A imagem correspondente a esta aproximação é indicada na Figura 7.18, do lado esquerdo.

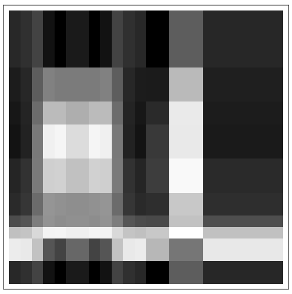


Figura 7.18: Do lado esquerdo uma aproximação de característica 2 da imagem da Figura 7.17, e do lado direito uma aproximação de característica 5.

Se pretendermos um erro relativo inferior a 1%, devemos considerar uma aproximação de característica 5, já que neste caso $e(5) = 0.00553241$. Esta aproximação da imagem original é indicada à direita na Figura 7.18.

Para armazenar uma imagem descrita por uma matriz $p \times n$, necessitamos de armazenar pn números, enquanto que para armazenar uma aproximação de característica r teremos de armazenar os r valores singulares, as rp entradas dos vectores singulares esquerdos e as rn entradas dos vectores singulares direitos. Ou seja, teremos de armazenar $r(p+n+1)$ números. Por exemplo, para armazenar a imagem indicada à direita na Figura 7.18, correspondente a uma aproximação de característica 5 da matriz 24×24 teríamos de armazenar $5(24 + 24 + 1) = 245$ números, enquanto que para a imagem original precisamos de armazenar $24^2 = 576$.

É claro que se considerarmos uma aproximação de característica 8, obtemos exactamente a imagem inicial, já que a matriz original tem característica 8 (há apenas 8 valores singulares não nulos).

Neste exemplo, conforme se observa na Figura 7.18, o resultado da compressão efectuada não é muito satisfatório uma vez que é necessário usar uma aproximação de característica "elevada" para obter uma boa reprodução da imagem original. Isto fica a dever-se ao facto da matriz original não ter valores singulares próximos de zero.

7.6.3 Pseudoinversa e mínimos quadrados

Nas ciências experimentais é frequente que a modelação de um problema conduza a um sistema linear $A\mathbf{x} = \mathbf{b}$ impossível. No entanto é útil encontrar uma "solução" para o problema em questão. É usual escolher-se uma solução de mínimos quadrados do sistema em causa, uma vez que essa solução minimiza a norma $\|A\mathbf{x} - \mathbf{b}\|$ (ver Secção 5.7.2). A solução de mínimos quadrados de um sistema não é necessariamente única, no entanto é possível escolher de forma única uma dessas soluções usando o conceito de *inversa generalizada de Moore-Penrose*[9], ou *pseudoinversa*.

A pseudoinversa de uma matriz generaliza o conceito de inversa no seguinte sentido: se A é uma matriz invertível, a pseudoinversa de A coincide com A^{-1}, e o sistema $A\mathbf{x} = \mathbf{b}$ tem solução $\mathbf{x} = A^{-1}\mathbf{b}$; se A não admite inversa, o sistema $A\mathbf{x} = \mathbf{b}$ tem solução (de mínimos quadrados) $\mathbf{x} = A^{+}\mathbf{b}$, onde a matriz A^{+} é a pseudoinversa de A.

Definição 7.14. A pseudoinversa A^{+} de uma matriz real (resp. complexa) A do tipo $p \times n$, é uma matriz $n \times p$ que verifica as propriedades:

i) $AA^{+}A = A$.

ii) $A^{+}AA^{+} = A^{+}$.

iii) $(AA^{+})^T = AA^{+}$ (resp. $(AA^{+})^H = AA^{+}$).

iv) $(A^{+}A)^T = A^{+}A$ (resp. $(A^{+}A)^H = A^{+}A$).

Como veremos adiante, a decomposição em valores singulares permite calcular uma pseudoinversa, assegurando a sua existência. Mostramos agora a unicidade da pseudoinversa de matrizes reais usando as quatro identidades que a definem. A prova para matrizes complexas é deixada como exercício.

Proposição 7.12. A pseudoinversa de uma matriz é única.

Demonstração. Considerem-se B e C matrizes reais do mesmo tipo que satisfazem as propriedades da pseudoinversa da matriz A, isto é,

1. $ABA = A$ e $ACA = A$.
2. $BAB = B$ e $CAC = C$.
3. $AB = (AB)^T$ e $AC = (AC)^T$.

[9] A pseudoinversa foi independentemente descrita em 1920 por Eliakim Hastings Moore (1862-1932) e em 1950 por Sir Roger Penrose (nascido em 1931). Contudo já em 1903 Erik Ivar Fredholm (1866 – 1927) tinha usado a pseudoinversa no contexto de operadores integrais.

4. $BA = (BA)^T$ e $CA = (CA)^T$.

Comecemos por mostrar que $AB = AC$ e $BA = CA$. De facto,

$$AB \overset{(3)}{=} B^T A^T \overset{(1)}{=} B^T A^T C^T A^T \overset{(3)}{=} ABC^T A^T \overset{(3)}{=} ABAC \overset{(1)}{=} AC,$$

e de forma análoga se mostra que $BA = CA$. Assim, tendo em conta que $AB = AC$ e $BA = CA$, resulta

$$B \overset{(2)}{=} BAB = BAC = CAC \overset{(2)}{=} C.$$

□

Exercício 7.25. Mostre que se A^+ é a pseudoinversa de A, então a pseudoinversa de A^T é $(A^T)^+ = (A^+)^T$. ▲

Consideremos agora uma decomposição em valores singulares de A. Seja $A = U\Sigma V^T$, onde as matrizes U, V são ortogonais e Σ uma matriz rectangular (do mesmo tamanho de A), da forma

$$\Sigma = \begin{bmatrix} D & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad \text{com} \quad D = \operatorname{diag}(\sigma_1, \ldots, \sigma_k).$$

Assim,

$$\begin{aligned} A\mathbf{x} - \mathbf{b} &= U\Sigma V^T \mathbf{x} - \mathbf{b} = U\left(\Sigma V^T \mathbf{x} - U^T \mathbf{b}\right) \\ &= U\left(\Sigma \mathbf{y} - \mathbf{c}\right) \qquad \text{com } \mathbf{y} = V^T \mathbf{x} \text{ e } \mathbf{c} = U^T \mathbf{b}. \end{aligned}$$

Como U é ortogonal, e portanto preserva distâncias (ver Proposição 7.1, pág. 374), tem-se

$$\|A\mathbf{x} - \mathbf{b}\| = \|U\left(\Sigma \mathbf{y} - \mathbf{c}\right)\| = \|\Sigma \mathbf{y} - \mathbf{c}\|.$$

Tal significa que o problema de determinação de uma solução $\hat{\mathbf{x}}$ de mínimos quadrados para $A\mathbf{x} = \mathbf{b}$ é equivalente a encontrar uma solução de mínimos quadrados $\hat{\mathbf{y}}$ para $\Sigma\mathbf{y} = \mathbf{c}$. Porém, dada a forma da matriz Σ, uma solução de mínimos quadrados do sistema $\Sigma\mathbf{y} - \mathbf{c} = \mathbf{0}$ é facilmente obtida. Para esse efeito, note-se que sendo A uma matriz do tipo $p \times n$, a matriz Σ é $p \times n$, U é $p \times p$, $\mathbf{y} \in \mathbb{R}^n$ e $\mathbf{c} \in \mathbb{R}^p$. Sejam $\mathbf{y} = (y_1, \ldots, y_n) \in \mathbb{R}^n$ e $\mathbf{c} = (c_1, \ldots, c_p) \in \mathbb{R}^p$, logo

$$\Sigma\mathbf{y} = \begin{bmatrix} \sigma_1 y_1 \\ \vdots \\ \sigma_k y_k \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \text{e} \quad \mathbf{c} = \begin{bmatrix} c_1 \\ \vdots \\ c_k \\ c_{k+1} \\ \vdots \\ c_p \end{bmatrix}.$$

Para $\hat{\mathbf{y}} = \left(\frac{c_1}{\sigma_1}, \ldots, \frac{c_k}{\sigma_k}, 0, \ldots, 0\right)$, a norma $\|\Sigma\mathbf{y} - \mathbf{c}\|$ é mínima. Portanto $\hat{\mathbf{y}}$ é a solução de mínimos quadrados para $\Sigma\mathbf{y} = \mathbf{c}$. Além disso, esta solução pode escrever-se na forma

$$\hat{\mathbf{y}} = \Sigma^+\mathbf{c} = \begin{bmatrix} D^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}_{n\times p} \mathbf{c}, \qquad \text{com } D^{-1} = \operatorname{diag}\left(\frac{1}{\sigma_1}, \ldots, \frac{1}{\sigma_k}\right). \tag{7.40}$$

Resumindo, se $A = U\Sigma V^T$ é uma decomposição SVD, tomando $\mathbf{c} = U^T\mathbf{b}$ e a solução de mínimos quadrados $\hat{\mathbf{y}}$ do sistema $\Sigma\mathbf{x} = \mathbf{c}$, a solução de mínimos quadrados $\hat{\mathbf{x}}$ de $A\mathbf{x} = \mathbf{b}$ é dada por

$$\hat{\mathbf{x}} = V\hat{\mathbf{y}} = V\Sigma^+\mathbf{c} = V\Sigma^+U^T\mathbf{b} = A^+\mathbf{b}. \tag{7.41}$$

Na proposição seguinte mostramos que a matriz A^+ em (7.41) é a pseudoinversa de A.

Proposição 7.13. Se $A = U\Sigma V^T$ é uma decomposição em valores singulares da matriz real A do tipo $p \times n$, então a pseudoinversa de A é

$$A^+ = V\Sigma^+U^T \quad \text{com} \quad \Sigma^+ = \begin{bmatrix} D^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}_{n\times p},$$

onde D é a matriz diagonal dos valores singulares não nulos de A.

Demonstração. Como vimos na Proposição 7.12, a pseudoinversa de uma matriz é única. Basta pois provar que $A^+ = V\Sigma^+U^T$ verifica as quatro propriedades da Definição 7.14. Em primeiro lugar, é fácil verificar que as identidades seguintes resultam imediatas da definição das matrizes Σ e Σ^+

$$\Sigma\Sigma^+ = \begin{bmatrix} I_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}_{p\times p} \qquad \Sigma^+\Sigma = \begin{bmatrix} I_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}_{n\times n},$$

onde I_k designa a matriz identidade de ordem k. Assim,

$$\begin{aligned}
AA^+A &= U\Sigma V^T V\Sigma^+U^T U\Sigma V^T = U\Sigma\Sigma^+\Sigma V^T = U\Sigma V^T = A \\
A^+AA^+ &= V\Sigma^+U^T U\Sigma V^T V\Sigma^+U^T = V\Sigma^+\Sigma\Sigma^+U^T = V\Sigma^+U^T = A^+ \\
(AA^+)^T &= (U\Sigma\Sigma^+U^T)^T = U\Sigma\Sigma^+U^T = AA^+ \\
(A^+A)^T &= (V\Sigma^+\Sigma V^T)^T = V\Sigma^+\Sigma V^T = A^+A.
\end{aligned}$$

□

Exemplo 7.13. Determinemos a pseudoinversa da matriz $A = \begin{bmatrix} 2 & -1 & 1 \\ 1 & 1 & -1 \end{bmatrix}$.

Os valores singulares de A são $\sqrt{6}$, $\sqrt{3}$ e 0. Portanto a matriz Σ da decomposição SVD de A é

$$\Sigma = \begin{bmatrix} \sqrt{6} & 0 & 0 \\ 0 & \sqrt{3} & 0 \end{bmatrix}.$$

Uma decomposição SVD de A é $A = U\Sigma V^T$ com

$$U = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad V = \begin{bmatrix} \frac{\sqrt{2}}{\sqrt{3}} & \frac{1}{\sqrt{3}} & 0 \\ \frac{-1}{\sqrt{6}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & \frac{-1}{\sqrt{3}} & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

Por conseguinte, a pseudoinversa de A é

$$A^+ = V\Sigma^+ U^T = V\Sigma^+ = \begin{bmatrix} \frac{\sqrt{2}}{\sqrt{3}} & \frac{1}{\sqrt{3}} & 0 \\ \frac{-1}{\sqrt{6}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & \frac{-1}{\sqrt{3}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & 0 \\ 0 & \frac{1}{\sqrt{3}} \\ 0 & 0 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 & 1 \\ \frac{-1}{2} & 1 \\ \frac{1}{2} & -1 \end{bmatrix}.$$

Sugerimos, como exercício, que confirme a unicidade da pseudoinversa calculando outra decomposição SVD para a matriz A. ◆

Exercício 7.26. Mostre que se A é uma matriz quadrada invertível, é válida a igualdade $A^+ = A^{-1}$. ▲

Vejamos agora qual é o efeito da pseudoinversa sobre vectores de $\mathbb{R}^p$.

Proposição 7.14. Seja A^+ a pseudoinversa de A. As matrizes AA^+ e A^+A são matrizes de projecção ortogonal, respectivamente, sobre o espaço das colunas e sobre o espaço das linhas de A. Isto é,

$$AA^+ = \operatorname{proj}_{EC(A)} \qquad \text{e} \qquad A^+A = \operatorname{proj}_{EL(A)}.$$

Demonstração. Para mostrar que um certo operador P_S é um operador de projecção ortogonal sobre o subespaço S basta provar que $P_S y = y$ para todo $y \in S$ e que $P_S y = 0$ para $y \in S^\perp$.

Seja $\mathbf{y} \in EC(A)$, isto é, $\mathbf{y} = A\mathbf{x}$ para algum $\mathbf{x}$. Logo, da definição de pseudoinversa temos

$$AA^+\mathbf{y} = AA^+A\mathbf{x} \overset{\text{i)}}{=} A\mathbf{x} = \mathbf{y}, \quad \text{para } \mathbf{y} \in EC(A).$$

Considerando agora $\mathbf{y} \in (EC(A))^{\perp} = N(A^T)$, resulta

$$AA^+\mathbf{y} \overset{\text{iii)}}{=} (AA^+)^T\mathbf{y} = (A^+)^T A^T\mathbf{y} = \mathbf{0}, \quad \text{para } \mathbf{y} \in N(A^T).$$

Por conseguinte, $AA^+ = \mathrm{proj}_{EC(A)}$.

Para mostrar que $A^+A = \mathrm{proj}_{EL(A)}$, comecemos por observar que $\mathrm{proj}_{EL(A)} = \mathrm{proj}_{EC(A^T)} = A^T(A^T)^+$. Do Exercício 7.25 sabe-se que $(A^T)^+ = (A^+)^T$, e portanto

$$\mathrm{proj}_{EL(A)} = A^T(A^T)^+ = A^T(A^+)^T = (A^+A)^T \overset{\text{iv)}}{=} A^+A.$$

□

A proposição anterior permite-nos agora estabelecer de que forma terá de actuar a pseudoinversa de uma matriz nos quatro subespaços fundamentais associados à matriz. Em primeiro lugar, e uma vez que

$$A^+ \underbrace{A\mathbf{x}}_{\in EC(A)} = \mathrm{proj}_{EL(A)}\,\mathbf{x} \in EL(A),$$

concluimos que a pseudoinversa aplica o espaço das colunas de A no espaço das linhas de A. Além disso, como $AA^+ = \mathrm{proj}_{EC(A)}$, tem-se $AA^+\mathbf{x} = \mathbf{0}$ para todo $\mathbf{x} \in (EC(A))^{\perp} = N(A^T)$. Multiplicando a igualdade $AA^+\mathbf{x} = \mathbf{0}$ por A^+, obtemos

$$A^+AA^+\mathbf{x} = \mathbf{0}, \text{ para todo } \mathbf{x} \in N(A^T) \overset{\text{ii)}}{\Longleftrightarrow} A^+\mathbf{x} = \mathbf{0}, \text{ para todo } \mathbf{x} \in N(A^T).$$

Em resumo, a pseudoinversa de A aplica vectores do espaço das colunas de A em vectores do espaço das linhas de A, e vectores do núcleo de A^T são aplicados em $\mathbf{0}$.

Relembre que a matriz A aplica vectores do espaço das linhas de A em vectores do espaço das colunas de A, e vectores do núcleo de A são aplicados em $\mathbf{0}$. Estamos assim em condições de estabelecer um diagrama comparativo da acção da matriz A e da sua pseudoinversa em vectores do respectivo domínio. Este diagrama é apresentado na Figura 7.19.

7.7 Grupos de matrizes e suas álgebras de Lie

Os grupos de matrizes ocorrem frequentemente em áreas que estudam simetrias. São variadas as áreas da matemática aplicada e das ciências experimentais que

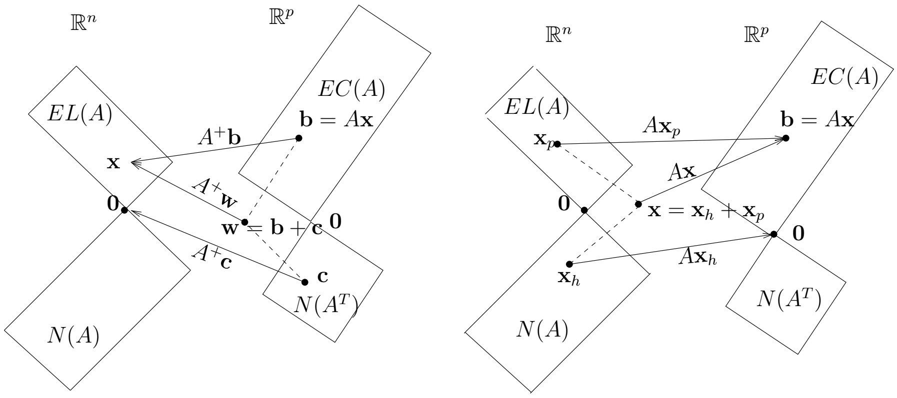


Figura 7.19: A acção da pseudoinversa A^+ e a acção de A.

recorrem a grupos de matrizes, como a geometria algébrica, a análise complexa, a teoria de grupos e anéis, a física quântica, a teoria de números, a relatividade especial de Einstein, a combinatória e muitas outras áreas –ver por exemplo o artigo de Howe [5]. Neste artigo, Howe escreve que grupos de matrizes "touch a tremendous spectrum of mathematical areas... the aplications are astonishing in their pervasiveness and sometimes in their unexpectedness"[10]. Em particular,

- Os programadores gráficos usam o grupo euclidiano para rodarem e transladarem objectos tridimensionais.
- A computação quântica usa o grupo das matrizes unitárias.

O objectivo desta secção é aproveitar vários tópicos anteriormente desenvolvidos para apresentar certos grupos de matrizes e suas álgebras de Lie, com especial ênfase nos denominados grupos clássicos.

7.7.1 Alguns grupos clássicos

Um *grupo* é um conjunto G juntamente com uma operação $\star$ sobre pares de elementos de G que satisfaz as seguintes propriedades:

[10]Traduzindo livremente: Os grupos de matrizes tocam um enorme espectro de áreas da matemática... as aplicações espantam pela sua abrangência e são por vezes completamente inesperadas...

- *Fecho da operação*: $g \star h \in G$, para todo $g, h \in G$.
- *Associatividade*: $(g \star h) \star f = g \star (h \star f)$, para todo $g, h, f \in G$.
- *Existência de identidade*: existe $e \in G$ tal que $e \star g = g \star e = g$, para todo $g \in G$.
- *Existência de inversos*: para cada elemento $g \in G$, existe um elemento $g^{-1} \in G$ tal que $g \star g^{-1} = g^{-1} \star g = e$.

Passamos a designar o grupo G por $(G, \star)$ sempre que seja necessário especificar a operação $\star$ do grupo.

Um *subgrupo* é um subconjunto de um grupo tal que a restrição da operação ao subconjunto satisfaz ainda os axiomas de grupo.

É imediato da definição de grupo que um conjunto de matrizes quadradas fechado para a operação de multiplicação de matrizes constitui um grupo se e só se as matrizes desse conjunto são invertíveis e as suas inversas ainda pertencem ao conjunto.

Definição 7.15. Um grupo cujos elementos são matrizes quadradas com a operação de multiplicação de matrizes, diz-se um *grupo de matrizes*.

Tendo em conta os resultados da Proposição 7.1 (pág. 374), podemos já identificar alguns grupos de matrizes estudados naquela secção. Em particular, relembre-se que:

- O produto de matrizes unitárias (resp. ortogonais) é ainda uma matriz unitária (resp. ortogonal);
- Uma matriz unitária (resp. ortogonal) é invertível e a sua inversa é ainda uma matriz unitária (resp. ortogonal).

Conclui-se portanto, que o conjunto das matrizes unitárias (resp. ortogonais) de ordem n, forma um grupo de matrizes. Este grupo é designado por *grupo unitário* $U(n)$ (resp. *grupo ortogonal* $O(n)$).

Recordando a noção de isomorfismo (ver na página 333), dois grupos $(G, \star)$ e $(H, \Box)$ dizem-se *isomorfos* se existe uma aplicação bijectiva $i : G \to H$ que preserva as operações de G e de H. Isto é, tal que $i(g_1 \star g_2) = i(g_1) \Box i(g_2)$, para quaisquer $g_1, g_2 \in G$.

Qualquer grupo de matrizes pode ser visto como um subgrupo do dito *grupo linear geral* $GL(V)$ das funções lineares invertíveis do espaço linear V em si próprio, onde a operação de grupo é agora a composição de funções lineares (ver Proposição 6.7, pág. 342). De facto, se V é um espaço linear real (resp. complexo) de dimensão n, é fácil de verificar que o grupo $GL(V)$ é isomorfo ao grupo das matrizes reais (resp. complexas) invertíveis ordem n.

O grupo das matrizes reais (resp. complexas) invertíveis, de ordem n, é denotado por $GL(n, \mathbb{R})$ (resp. $GL(n, \mathbb{C})$).

O subconjunto de $GL(n, \mathbb{K})$, com $\mathbb{K} = \mathbb{R}$ ou $\mathbb{K} = \mathbb{C}$, formado pelas matrizes cujo determinante é igual a 1, é um subgrupo de $GL(n, \mathbb{K})$, designado por *grupo linear especial* e denotado por $SL(n, \mathbb{K})$. Isto é,

$$SL(n, \mathbb{K}) = \{A \in GL(n, \mathbb{K}) : \ \det(A) = 1\} .$$

O grupo unitário $U(n)$ e o grupo ortogonal $O(n)$ são assim subgrupos de $GL(n, \mathbb{C})$ e de $GL(n, \mathbb{R})$, respectivamente.

Como subgrupos de $O(n)$ e de $U(n)$ destacamos os subconjuntos formados pelas matrizes cujo determinante vale 1. Estes subgrupos recebem a designação de *grupo unitário especial* $SU(n)$, e de *grupo ortogonal especial* $SO(n)$:

$$\begin{aligned} SU(n) &= \left\{U \in GL(n, \mathbb{C}) : \ U^H U = I \ \text{ e } \ \det(U) = 1\right\} \\ SO(n) &= \left\{O \in GL(n, \mathbb{R}) : \ O^T O = I \ \text{ e } \ \det(O) = 1\right\} . \end{aligned}$$

Exercício 7.27. Mostre que o grupo $SO(2)$ é o grupo das matrizes de rotação em torno da origem de um ângulo θ, isto é, o conjunto das matrizes da forma

$$R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} .$$

Verifique ainda que $SO(2)$ é isomorfo ao *grupo circular*

$$S^1 = \{z \in \mathbb{C} : |z| = 1\},$$

sendo a operação de grupo em S^1 a multiplicação de números complexos.
Sugestão: Verifique que $R_\theta \mapsto e^{i\theta}$ é um isomorfismo. ▲

Geometricamente, o grupo ortogonal $O(n)$ pode ser descrito como o grupo das isometrias de $\mathbb{R}^n$ que fixam a origem, conforme se mostra na próxima proposição. Recordemos que uma isometria em $\mathbb{R}^n$ é uma função $f : \mathbb{R}^n \to \mathbb{R}^n$ que preserva distâncias (cf. Definição 7.3). Salvo menção em contrário, consideramos que a distância em $\mathbb{R}^n$ é a induzida pelo produto interno usual de $\mathbb{R}^n$.

Proposição 7.15. 1) Se $M \in O(n)$, a função $f : \mathbb{R}^n \to \mathbb{R}^n$ definida por $f(\mathbf{x}) = M\mathbf{x}$, é uma isometria.

2) Se $f : \mathbb{R}^n \to \mathbb{R}^n$ é uma isometria tal que $f(\mathbf{0}) = \mathbf{0}$, então existe uma matriz $M \in O(n)$ tal que $f(\mathbf{x}) = M\mathbf{x}$. Em particular, a função f é linear.

Demonstração. 1) Como M é uma matriz ortogonal são preservadas distâncias, e portanto a função f é uma isometria.

Para mostrar 2) note-se que, por definição de isometria, $\operatorname{dist}(f(\mathbf{x}), f(\mathbf{y})) = \operatorname{dist}(\mathbf{x}, \mathbf{y})$, isto é,

$$\|f(\mathbf{x}) - f(\mathbf{y})\| = \|\mathbf{x} - \mathbf{y}\|, \quad \text{para todo } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n. \tag{7.42}$$

Em particular, como $f(\mathbf{0}) = \mathbf{0}$ resulta

$$\|f(\mathbf{x}) - f(\mathbf{0})\| = \|f(\mathbf{x})\| = \|\mathbf{x}\|. \tag{7.43}$$

Do Exercício 7.2, sabemos que uma isometria preserva produtos internos, e portanto

$$\langle \mathbf{x}, \mathbf{y} \rangle = \langle f(\mathbf{x}), f(\mathbf{y}) \rangle.$$

Considere-se agora a matriz M com colunas $f(\mathbf{e}_1), \dots, f(\mathbf{e}_n)$ (por esta ordem), onde os vectores $\mathbf{e}_i$ são os vectores da base canónica de $\mathbb{R}^n$. A matriz M é ortogonal, visto que

$$M^T M = \begin{bmatrix} — & f(\mathbf{e}_1)^T & — \\ — & f(\mathbf{e}_2)^T & — \\ & \vdots & \\ — & f(\mathbf{e}_n)^T & — \end{bmatrix} \begin{bmatrix} | & & | \\ f(\mathbf{e}_1) & \cdots & f(\mathbf{e}_n) \\ | & & | \end{bmatrix}$$

$$= [\langle f(\mathbf{e}_i), f(\mathbf{e}_j) \rangle]_{i,j=1,\dots,n} = [\langle \mathbf{e}_i, \mathbf{e}_j \rangle]_{i,j=1,\dots,n} = I,$$

onde a penúltima igualdade resulta da preservação do produto interno, e a última do facto da base canónica ser ortonormada. Note-se ainda que a igualdade $\langle f(\mathbf{e}_i), f(\mathbf{e}_j) \rangle = \langle \mathbf{e}_i, \mathbf{e}_j \rangle$ garante que $B = \{f(\mathbf{e}_1), \dots, f(\mathbf{e}_n)\}$ é uma base ortonormada de $\mathbb{R}^n$.

Falta mostrar que a matriz M representa f em relação à base canónica de $\mathbb{R}^n$. Ou seja, que $f(\mathbf{x}) = M\mathbf{x}$ para todo $\mathbf{x} \in \mathbb{R}^n$. De facto, como B é uma base ortonormada de $\mathbb{R}^n$, o Teorema 5.2 (pág. 260) garante que $f(\mathbf{x})$ se escreve

$$\begin{aligned} f(\mathbf{x}) &= \langle f(\mathbf{x}), f(\mathbf{e}_1) \rangle f(\mathbf{e}_1) + \cdots + \langle f(\mathbf{x}), f(\mathbf{e}_n) \rangle f(\mathbf{e}_n) \\ &= \langle \mathbf{x}, \mathbf{e}_1 \rangle f(\mathbf{e}_1) + \cdots + \langle \mathbf{x}, \mathbf{e}_1 \rangle f(\mathbf{e}_n) \\ &= x_1 f(\mathbf{e}_1) + \cdots + x_n f(\mathbf{e}_n) = M\mathbf{x}, \end{aligned}$$

onde a última igualdade resulta do facto de $M\mathbf{x}$ ser uma combinação linear das colunas de M com coeficientes iguais às componentes de $\mathbf{x} = (x_1, \ldots, x_n)$. $\square$

Para entender a diferença entre $O(3)$ e $SO(3)$ é necessário entrar em linha de conta com a noção de orientação em $\mathbb{R}^3$.

Proposição 7.16. Seja $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ a base canónica de $\mathbb{R}^3$.

Uma matriz M pertence a $SO(3)$ se e só se $(M\mathbf{e}_1, M\mathbf{e}_2, M\mathbf{e}_3)$ é um triedro directo ortonormado de $\mathbb{R}^3$.

Demonstração. De acordo com a definição de triedro directo (Definição 5.10, pág. 284), $(M\mathbf{e}_1, M\mathbf{e}_2, M\mathbf{e}_3)$ é um triedro directo se o determinante da matriz M é positivo (as colunas de M são $M\mathbf{e}_1, M\mathbf{e}_2, M\mathbf{e}_3$).

Uma matriz M é ortogonal se e só se as suas colunas formam uma base ortonormada de $\mathbb{R}^3$. Ou seja, M é ortogonal se e só se a base ordenada $(M\mathbf{e}_1, M\mathbf{e}_2, M\mathbf{e}_3)$ é uma base ortonormada de $\mathbb{R}^3$. Por conseguinte, a base $(M\mathbf{e}_1, M\mathbf{e}_2, M\mathbf{e}_3)$ é um triedro directo ortonormado se e só se a matriz (ortogonal) M tem determinante igual a 1. $\square$

Pela proposição anterior, $SO(3)$ é o grupo das isometrias de $\mathbb{R}^3$ que fixam a origem e aplicam triedros directos em triedros directos (isto é, preservam a orientação), enquanto que existem matrizes de $O(3)$ que não preservam triedros directos. Por exemplo, para a matriz

$$M = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \in O(3),$$

tem-se $(M\mathbf{e}_1, M\mathbf{e}_2, M\mathbf{e}_3) = (\mathbf{e}_1, \mathbf{e}_3, \mathbf{e}_2)$, que não é um triedro directo.

Verifica-se facilmente que o conjunto das isometrias de $\mathbb{R}^n$ forma um grupo para a operação de composição de funções. Designemos por $\mathrm{Isom}(\mathbb{R}^n)$ este grupo:

$$\mathrm{Isom}(\mathbb{R}^n) = \{f : \mathbb{R}^n \to \mathbb{R}^n;\ f \text{ é uma isometria}\}.$$

Vimos que $O(n)$ é isomorfo ao grupo das isometrias de $\mathbb{R}^n$ que fixam a origem. Uma isometria que não fixe a origem não é uma função linear (pela Proposição 6.1, se f é linear, então $f(\mathbf{0}) = \mathbf{0}$).

Seja f uma isometria de $\mathbb{R}^n$, tal que $f(\mathbf{0}) = \mathbf{a} \neq \mathbf{0}$. A função $g(\mathbf{x}) = f(\mathbf{x}) - \mathbf{a}$ é uma isometria que fixa a origem, uma vez que, sendo f uma isometria, se tem

$$\|g(\mathbf{x}) - g(\mathbf{y})\| = \|f(\mathbf{x}) - f(\mathbf{y})\| = \|\mathbf{x} - \mathbf{y}\|.$$

Resulta da Proposição 7.15 que existe $M \in O(n)$ tal que $g(\mathbf{x}) = M\mathbf{x}$. Portanto f escreve-se na forma

$$f(\mathbf{x}) = M\mathbf{x} + \mathbf{a},$$

com $M \in O(n)$ e $\mathbf{a} \in \mathbb{R}^n$.

Uma isometria em $\mathbb{R}^n$ pode ainda ser representada por uma matriz do tipo $(n{+}1)\times(n{+}1)$. Em computação gráfica usa-se frequentemente esta representação matricial do grupo das isometrias. No que se segue usamos a notação $GL(n+1)$ para $GL(n+1, \mathbb{R})$.

Proposição 7.17. O grupo $\mathrm{Isom}(\mathbb{R}^n)$ é isomorfo ao grupo *euclidiano* $E(n)$ definido por

$$E(n) = \left\{ \begin{bmatrix} M & \mathbf{a} \\ \mathbf{0} & 1 \end{bmatrix} \in GL(n+1) : \; M \in O(n) \text{ e } \mathbf{a} \in \mathbb{R}^n \right\}.$$

Demonstração. Em primeiro lugar, $E(n)$ é um subgrupo de $GL(n+1)$, uma vez que o produto de duas matrizes de $E(n)$ é uma matriz de $E(n)$ (relembre-se que o produto de matrizes ortogonais ainda é uma matriz ortogonal), e a inversa de uma matriz de $E(n)$ ainda pertence a $E(n)$:

$$\begin{bmatrix} M & \mathbf{a} \\ \mathbf{0} & 1 \end{bmatrix}^{-1} = \begin{bmatrix} M^T & -M^T\mathbf{a} \\ \mathbf{0} & 1 \end{bmatrix}.$$

A função $i : \mathrm{Isom}(\mathbb{R}^n) \to E(n)$ definida por

$$i : (\mathbf{x} \mapsto M\mathbf{x} + \mathbf{a}) \longmapsto \begin{bmatrix} M & \mathbf{a} \\ \mathbf{0} & 1 \end{bmatrix}$$

é um isomorfismo de grupos. De facto, a função i é bijectiva e, para quaisquer isometrias $f(\mathbf{x}) = M\mathbf{x} + \mathbf{a}$ e $g(\mathbf{x}) = M_1\mathbf{x} + \mathbf{b}$, tem-se

$$\begin{aligned} i(f \circ g) &= i(MM_1\mathbf{x} + M\mathbf{b} + \mathbf{a}) = \begin{bmatrix} MM_1 & M\mathbf{b} + \mathbf{a} \\ \mathbf{0} & 1 \end{bmatrix} = \begin{bmatrix} M & \mathbf{a} \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} M_1 & \mathbf{b} \\ \mathbf{0} & 1 \end{bmatrix} \\ &= i(f)i(g). \end{aligned}$$

□

Grupo dos quaterniões unitários

Seguidamente fazemos uma referência breve a certos subgrupos do grupo unitário. Em particular, mostramos que o grupo $SU(2)$ é isomorfo ao grupo dos quaterniões unitários. O conjunto dos quaterniões é habitualmente denotado pela letra $\mathbb{H}$ em honra de Hamilton[11] a quem se deve a sua descoberta.

A um par ordenado $(a, b) \in \mathbb{R}^2$ associamos (bijectivamente) um número complexo $z = a + ib$ (com $i^2 = -1$). Esta bijecção preserva as operações de adição e de multiplicação por números reais. De igual modo, existe uma bijecção entre $\mathbb{R}^4$ e o conjunto dos quaterniões que preserva as operações de adição e de multiplicação de números reais, definidas nos dois conjuntos.

Um quaternião é um número $\mathfrak{q} = a + bi + cj + dk$, com $(a, b, c, d) \in \mathbb{R}^4$ e i, j, k satisfazendo as igualdades

$$i^2 = j^2 = k^2 = ijk = -1. \tag{7.44}$$

As operações de adição de vectores de $\mathbb{R}^4$ e de multiplicação por escalares, traduzem-se em termos de quaterniões na forma

$$\begin{aligned}(a + bi + cj + dk) + (a_1 + b_1 i + c_1 j + d_1 k) &= (a + a_1) + (b + b_1)i \\ &\quad + (c + c_1)j + (d + d_1)k, &(7.45)\\ \alpha(a + bi + cj + dk) &= \alpha a + \alpha bi + \alpha cj + \alpha dk, \quad \alpha \in \mathbb{R} &(7.46)\end{aligned}$$

É fácil verificar que a função que a cada vector $(a, b, c, d) \in \mathbb{R}^4$ faz corresponder o quaternião $\mathfrak{q} = a + bi + cj + dk$ é um isomorfismo de espaços lineares. Assim, $\mathbb{H}$ é um espaço linear real de dimensão 4 e uma base de $\mathbb{H}$ é $\{1, i, j, k\}$.

A operação de multiplicação de quaterniões, ao contrário da operação de adição e multiplicação por escalares reais, não é comutativa. No entanto, a multiplicação de quaterniões goza da propriedade associativa. Usando a associatividade da multiplicação de quaterniões, mostra-se facilmente que os quaterniões i, j, k, satisfazem as relações:

$$\begin{array}{l} 1\mathfrak{q} = \mathfrak{q}1 = \mathfrak{q} \quad \text{para todo } \mathfrak{q} \in \mathbb{H} \\ ij = -ji = k \\ jk = -kj = i \\ ki = -ik = j. \end{array} \tag{7.47}$$

[11]Sir William Rowan Hamilton (1805 – 1865) que em 1843 introduziu os quaterniões e os aplicou em estudos da mecânica no espaço tridimensional.

Destas relações segue que a multiplicação de quaterniões não é comutativa, por exemplo, $ij = -ji$.

Exercício 7.28. Use a associatividade da multiplicação de quaterniões e as igualdades (7.44) para mostrar as relações (7.47). ▲

As relações (7.45), (7.46) e (7.47), e a distributividade da multiplicação em relação à adição, sugerem que se adopte a seguinte definição de multiplicação de quaterniões:

$$\begin{aligned}(q_0 + ai + bj + ck)(q_1 + a_1 i + b_1 j + c_1 k) = {} & (q_0 q_1 - aa_1 - bb_1 - cc_1) + \\ & + (q_0 a_1 + q_1 a + bc_1 - cb_1) i \\ & + (q_0 b_1 + q_1 b + a_1 c - ac_1) j \\ & + (q_0 c_1 + q_1 c + ab_1 - a_1 b) k.\end{aligned}$$

Exercício 7.29. Verifique que a multiplicação de quaterniões verifica as propriedades:

$$\begin{aligned}\mathfrak{q}_1(\mathfrak{q}_2\mathfrak{q}_3) &= (\mathfrak{q}_1\mathfrak{q}_2)\mathfrak{q}_3, && \text{(Associatividade)}\\ \mathfrak{q}\mathfrak{q}^{-1} &= 1, \ \text{para } \mathfrak{q} \neq \mathbf{0} && \text{(Existência de inverso)}\\ \mathfrak{q}1 &= \mathfrak{q}, && \text{(Existência de identidade)}\\ \mathfrak{q}_1(\mathfrak{q}_2 + \mathfrak{q}_3) &= \mathfrak{q}_1\mathfrak{q}_2 + \mathfrak{q}_1\mathfrak{q}_3, && \text{(Distributividade à esquerda)}\\ (\mathfrak{q}_1 + \mathfrak{q}_2)\mathfrak{q}_3 &= \mathfrak{q}_1\mathfrak{q}_3 + \mathfrak{q}_2\mathfrak{q}_3, && \text{(Distributividade à direita)}\end{aligned}$$

▲

Exercício 7.30. Identifique um quaternião $\mathfrak{q} = q_0 + ai + bj + ck$ com $\mathfrak{q} = (q_0, \mathbf{q}) \in \mathbb{R} \times \mathbb{R}^3$ onde $\mathbf{q} = (a, b, c) \in \mathbb{R}^3$. Mostre que a multiplicação de quaterniões é dada por

$$\mathfrak{q}\mathfrak{p} = (q_0, \mathbf{q})(r_0, \mathbf{r}) = (q_0 r_0 - \langle \mathbf{q}, \mathbf{r} \rangle \,,\, q_0 \mathbf{r} + r_0 \mathbf{q} + \mathbf{q} \times \mathbf{r}),$$

onde $\langle \, , \, \rangle$ é o produto interno usual em $\mathbb{R}^3$, e $\mathbf{q} \times \mathbf{r}$ é o produto vectorial em $\mathbb{R}^3$. ▲

À semelhança dos números complexos, define-se a conjugação de quaterniões do seguinte modo:

$$\overline{q_0 + ai + bj + ck} := q_0 - (ai + bj + ck).$$

Segue então da definição de produto que

$$\mathfrak{q}\overline{\mathfrak{q}} = q_0^2 + a^2 + b^2 + c^2, \quad \text{para} \quad \mathfrak{q} = q_0 + ai + bj + ck.$$

Define-se *norma* de um quaternião $\mathfrak{q}$ por $|\mathfrak{q}| = \sqrt{\mathfrak{q}\overline{\mathfrak{q}}}$. O inverso de um quaternião não nulo é $\mathfrak{q}^{-1} = \dfrac{1}{\mathfrak{q}} = \dfrac{\overline{\mathfrak{q}}}{|\mathfrak{q}|^2}$.

Um quaternião $\mathfrak{q} = q_0 + ai + bj + ck$ diz-se um *quaternião unitário* quando satisfaz a igualdade

$$q_0^2 + a^2 + b^2 + c^2 = 1,$$

ou seja, quando $|\mathfrak{q}|^2 = 1$.

É fácil concluir que o conjunto dos quaterniões unitários é um grupo para a multiplicação de quaterniões (note-se que o inverso de um quaternião unitário ainda é um quaternião unitário). O *grupo dos quaterniões unitários* designa-se por S^3 (a mesma designação da esfera unitária de $\mathbb{R}^4$), ou seja,

$$S^3 = \left\{\mathfrak{q} \in \mathbb{H} : \ |\mathfrak{q}|^2 = 1\right\}.$$

Proposição 7.18. O grupo $SU(2)$ é isomorfo ao grupo dos quaterniões unitários S^3.

Demonstração. Em primeiro lugar, determine-se a forma de uma matriz Q de $SU(2)$. Seja $Q = \begin{bmatrix} z_1 & w_1 \\ w_2 & z_2 \end{bmatrix}$ uma matriz complexa. Esta matriz pertence a $SU(2)$ se e só se $Q^{-1} = Q^H$. A matriz inversa de $Q \in SU(2)$ é

$$Q^{-1} = \frac{1}{\det Q}\begin{bmatrix} z_2 & -w_1 \\ -w_2 & z_1 \end{bmatrix} = \begin{bmatrix} z_2 & -w_1 \\ -w_2 & z_1 \end{bmatrix},$$

uma vez que $\det(Q) = 1$, por definição de $SU(2)$. A igualdade $Q^{-1} = Q^H$ traduz-se em

$$\begin{bmatrix} z_2 & -w_1 \\ -w_2 & z_1 \end{bmatrix} = Q^{-1} = Q^H = \begin{bmatrix} \overline{z}_1 & \overline{w}_2 \\ \overline{w}_1 & \overline{z}_2 \end{bmatrix}.$$

Logo, Q pertence a $SU(2)$ se e só se $z_2 = \overline{z}_1$ e $w_1 = -\overline{w}_2$. Ou seja, as matrizes $Q \in SU(2)$ são da forma

$$Q = \begin{bmatrix} z_1 & w_1 \\ -\overline{w}_1 & \overline{z}_1 \end{bmatrix},$$

com $\det Q = z_1\overline{z}_1 + w_1\overline{w}_1 = |z_1|^2 + |w_1|^2 = 1$.

Considere-se agora a função $f : S^3 \to SU(2)$ definida por

$$f : S^3 \ni \mathfrak{q} = q_0 + ai + bj + ck \longmapsto \begin{bmatrix} q_0 + ai & -b - ci \\ b - ci & q_0 - ai \end{bmatrix} = Q \in SU(2).$$

Note-se que na definição anterior, $\mathfrak{q} = q_0 + ai + bj + ck \in S^3$ se e só se $\det(Q) = |\mathfrak{q}|^2 = 1$. A função f é obviamente bijectiva. Além disso, f é um isomorfismo entre o grupo S^3 e o grupo $SU(2)$, visto que $f(\mathfrak{q}_1\mathfrak{q}_2) = f(\mathfrak{q}_1)f(\mathfrak{q}_2) = Q_1Q_2$, como se pode verificar facilmente efectuando os cálculos correspondentes. □

7.7.2 Álgebras de Lie

Nesta secção abordamos de forma sumária as álgebras de Lie[12] de alguns grupos de matrizes já estudados.

A álgebra de Lie de um grupo de matrizes é uma ferramenta indispensável ao estudo desse grupo. Esta álgebra é um espaço linear que contém informação relevante acerca do grupo. Neste contexto, a exponencial de matrizes desempenha um papel crucial, pois a exponencial de uma matriz da álgebra de Lie de um grupo de matrizes G é um elemento deste grupo.

São pré-requisitos para esta secção os tópicos sobre equações diferenciais ordinárias por nós abordados na Secção 4.4.2 e, em particular, a exponencial de matrizes definida na página 229.

Definição 7.16. Uma álgebra de Lie $\mathfrak{g}$ é um espaço linear, sobre um corpo de escalares $\mathbb{K}$, munido de uma operação $[\,,\,]$ sobre pares de elementos de $\mathfrak{g}$ que satisfaz as seguintes propriedades:

- *Fecho*: $X, Y \in \mathfrak{g} \Longrightarrow [X, Y] \in \mathfrak{g}$.
- *Linearidade*: $[X, \alpha Y + \beta Z] = \alpha[X, Y] + \beta[X, Z]$, para todo $X, Y, Z \in \mathfrak{g}$ e $\alpha, \beta \in \mathbb{K}$.
- *Anticomutatividade* : $[X, Y] = -[Y, X]$, para todo $X, Y \in \mathfrak{g}$.
- *Identidade de Jacobi*: $[X, [Y\,,\,Z]] + [Y, [Z\,,\,X]] + [Z, [X\,,\,Y]] = 0$, para todo $X, Y, Z \in \mathfrak{g}$.

Quando a operação $[\,,\,]$ goza das propriedades enunciadas diz-se que $[\,,\,]$ é um *parêntesis de Lie*.

[12] O termo 'álgebra de Lie' foi introduzido em 1930 por Hermann Klaus Hugo Weyl (1885 – 1955) para designar o que o matemático Marius Sophus Lie (lê-se "Li" e não "Lai") (1842 - 1899) chamava 'elementos infinitesimais de um grupo contínuo'. Sophus Lie criou e desenvolveu a teoria das simetrias contínuas e aplicou-a em geometria e no estudo de equações diferenciais.

Quando o corpo de escalares $\mathbb{K}$ é $\mathbb{R}$ (resp. $\mathbb{C}$) diz-se que a álgebra de Lie é real (resp. complexa).

Exercício 7.31. Mostre que o produto externo em $\mathbb{R}^3$ é um parêntesis de Lie. Recorde para tal a definição (5.20) na página 280, e as propriedades enunciadas nos exercícios 5.14 e 5.16 (pág. 282).

▲

Uma *álgebra de Lie de matrizes* é um espaço linear de matrizes em que o parêntesis de Lie é o comutador de matrizes, isto é, $[A, B] = AB-BA$. Deixamos como exercício verificar que o comutador de matrizes define um parêntesis de Lie no espaço das matrizes de ordem n sobre o corpo $\mathbb{K}$.

Para definirmos a álgebra de Lie de um grupo de matrizes G é necessário precisar a noção de curva no grupo.

Uma função contínua γ definida num intervalo $I \subset \mathbb{R}$ com valores no grupo G diz-se um caminho em G e a sua imagem $\gamma(I)$ diz-se uma *curva* em G. Um caminho $\gamma : I \to G$ diz-se *regular* se é um caminho C^1 (isto é, γ é contínua com derivada contínua) e tal que $\gamma'(t) \neq 0$ para qualquer $t \in I$.

Um caminho diferenciável em $GL(2, \mathbb{C})$ tem a forma

$$\gamma(t) = \begin{bmatrix} a(t) + ib(t) & c(t) + id(t) \\ e(t) + if(t) & g(t) + ih(t) \end{bmatrix},$$

onde $a(t), b(t), c(t), d(t), e(t), f(t), g(t), h(t)$ são funções reais de variável real diferenciáveis em $t \in]-\epsilon, \epsilon[$. A respectiva derivada é

$$\gamma'(t) = \begin{bmatrix} a'(t) + ib'(t) & c'(t) + id'(t) \\ e'(t) + if'(t) & g'(t) + ih'(t) \end{bmatrix}.$$

Esta noção de caminho diferenciável em $GL(2, \mathbb{C})$ estende-se de forma óbvia a $GL(n, \mathbb{C})$.

Exercício 7.32. Seja $M_n(\mathbb{K})$ o conjunto das matrizes $n \times n$ de entradas em $\mathbb{K}$, com $\mathbb{K} = \mathbb{R}$ ou $\mathbb{K} = \mathbb{C}$. Se $\gamma, \delta :]-\epsilon, \epsilon[\to M_n(\mathbb{K})$ são dois caminhos diferenciáveis, então o produto $(\gamma \cdot \delta)(t) := \gamma(t)\delta(t)$ também é diferenciável e

$$(\gamma \cdot \delta)'(t) = \gamma'(t)\delta(t) + \gamma(t)\delta'(t). \tag{7.48}$$

▲

A álgebra de Lie $\mathfrak{g}$, de um grupo G, é o conjunto dos vectores tangentes às curvas regulares na identidade de G. Mais precisamente, se $\gamma(t)$ é um caminho regular em G que passa pela identidade em $t = 0$ (ou seja, $\gamma(0) = I$), tem-se que $X = \gamma'(0)$ pertence à algebra de Lie $\mathfrak{g}$.

Definição 7.17. Seja $\mathbb{K} = \mathbb{R}$ ou $\mathbb{K} = \mathbb{C}$. A *álgebra de Lie* de um grupo $G \subset GL(n, \mathbb{K})$ é o espaço tangente a G na identidade I. Ou seja,

$$T_I G = \{\gamma'(0) : \ \gamma :]-\epsilon, \epsilon[\to G \text{ é um caminho regular verificando } \gamma(0) = I\}.$$

A álgebra de Lie do grupo G é designada por $\mathfrak{g} := T_I G$.

Mostramos de seguida que a álgebra de Lie de um grupo $G \subset GL(n, \mathbb{K})$, com $\mathbb{K} = \mathbb{R}$ ou $\mathbb{K} = \mathbb{C}$, é uma álgebra de Lie de matrizes. Isto é, trata-se de um subespaço linear do espaço das matrizes com entradas em $\mathbb{K}$ fechado para o comutador de matrizes. Comecemos por provar que a álgebra de Lie de um grupo de matrizes é um subespaço linear.

Proposição 7.19. A álgebra de Lie $\mathfrak{g}$ de um grupo $G \subset GL(n, \mathbb{K})$, com $\mathbb{K} = \mathbb{R}$ ou $\mathbb{K} = \mathbb{C}$, é um subespaço linear <u>real</u> do espaço das matrizes $n \times n$ de entradas em $\mathbb{K}$.

Demonstração. Sejam γ, δ caminhos regulares em G tais que $\gamma(0) = \delta(0) = I$ e $A = \gamma'(0)$ e $B = \delta'(0)$.

Para $\alpha \in \mathbb{R}$, o caminho $\delta(t) = \gamma(\alpha t)$ satisfaz as igualdades $\delta(0) = \gamma(0) = I$ e $\delta'(0) = \alpha\gamma'(0) = \alpha A$. Portanto, $\alpha A \in \mathfrak{g}$.

O caminho produto $\sigma(t) = (\gamma \cdot \delta)(t) = \gamma(t)\delta(t)$ verifica $\sigma(0) = I$ e, pelo Exercício 7.32, $\sigma'(0) = \gamma'(0)\delta(0) + \gamma(0)\delta'(0) = A + B$. Por conseguinte, $A + B \in \mathfrak{g}$.

Assim, a álgebra de Lie $\mathfrak{g}$ é fechada para a adição e multiplicação por escalares reais e portanto é um subespaço linear. $\square$

Para provar que a álgebra de Lie de um grupo de matrizes é uma álgebra de Lie de matrizes, falta mostrar que o comutador de matrizes ainda pertence ao espaço tangente ao grupo na identidade.

Proposição 7.20. Seja G um grupo de matrizes e $\mathfrak{g} = T_I G$ a sua álgebra de Lie. Se $A, B \in \mathfrak{g}$, então $[A, B] = AB - BA$ pertence a $\mathfrak{g}$.

Demonstração. Sejam γ e ρ caminhos regulares em G tais que $\gamma(0) = \rho(0) = I$, $A = \gamma'(0)$ e $B = \rho'(0)$. Considere-se ainda o caminho

$$\delta_s(t) = \gamma(s)\rho(t)\gamma(s)^{-1}, \quad \text{para } s \text{ fixo.}$$

O caminho δ verifica $\delta_s(0) = I$, e portanto $\delta_s'(0) \in \mathfrak{g} = T_I G$. A função (de s)

$$\delta_s'(0) = \gamma(s)\rho'(0)\gamma(s)^{-1} = \gamma(s)B\gamma(s)^{-1}$$

pertence a $T_I G$ e é regular uma vez que γ é regular. A derivada de $\delta_s'(0)$ ainda pertence ao espaço tangente $T_I G$, visto que a derivada de uma função é o limite de uma razão incremental ($f'(x) = \lim_{h\to 0} \frac{1}{h}(f(x+h) - f(x))$, e essa razão incremental ainda pertence a $T_I G$, uma vez que $T_I G$ é um espaço vectorial. Portanto, $\frac{d}{ds}\Big|_{s=0} \gamma(s)\rho'(0)\gamma(s)^{-1}$ pertence à álgebra de Lie $\mathfrak{g} = T_I G$.

Para calcular esta derivada, comecemos por observar que derivando a expressão $\gamma(s)\gamma(s)^{-1} = I$ se obtém

$$\gamma'(0)\gamma(0)^{-1} + \gamma(0)\frac{d}{ds}\Big|_{s=0}\gamma(s)^{-1} = 0 \iff \frac{d}{ds}\Big|_{s=0}\gamma(s)^{-1} = -\gamma'(0),$$

uma vez que $\gamma(0) = I$. Por conseguinte,

$$\frac{d}{ds}\Big|_{s=0}\gamma(s)B\gamma(s)^{-1} = \gamma'(0)B - B\gamma'(0) = AB - BA$$

pertence a $\mathfrak{g}$. $\square$

O facto das álgebras de Lie de grupos de matrizes serem espaços vectoriais permite-nos falar da sua dimensão. A dimensão de um grupo é definida como sendo a dimensão da sua álgebra de Lie.

Definição 7.18. Define-se *dimensão* de um grupo de matrizes como sendo a dimensão da álgebra de Lie do grupo.

Nota 50. *Apesar do espaço das matrizes $n \times n$ de entradas complexas (isomorfo a $\mathbb{C}^{n^2}$) ser um espaço linear complexo, a álgebra de Lie de um grupo de matrizes complexas não é necessariamente um subespaço sobre $\mathbb{C}$ do espaço das matrizes complexas (veja por exemplo o caso da álgebra de Lie $\mathfrak{su}(2)$ apresentado no final desta secção). A dimensão de um grupo de matrizes significa sempre a dimensão da sua álgebra de Lie como espaço linear real.*

Exemplos de álgebras de Lie

Passamos a designar por $M_n(\mathbb{K})$ o conjunto das matrizes de ordem n com entradas em $\mathbb{K}$. Salvo menção em contrário, o corpo $\mathbb{K}$ é $\mathbb{R}$ ou $\mathbb{C}$.

1. **Álgebra de Lie de** $GL(n, \mathbb{K})$

   A álgebra de Lie de $GL(n, \mathbb{K})$ é o conjunto das matrizes $n \times n$ com entradas em $\mathbb{K}$, isto é, $\mathfrak{gl}(n, \mathbb{K}) = M_n(\mathbb{K})$.

   De facto, dada a matriz $A \in M_n(\mathbb{K})$ considere-se a curva $\gamma(t) = I + tA$. A curva γ toma valores em $GL(n, \mathbb{K})$ para t suficientemente pequeno, uma vez que, sendo a função determinante contínua, $\det(\gamma(t))$ está próximo de $\det(\gamma(0)) = \det(I) = 1$ para t suficientemente pequeno. Ou seja, $\det \gamma(t) \neq 0$ para t próximo de zero. Por conseguinte, como $\gamma'(0) = A$, resulta que $A \in \mathfrak{gl}(n, \mathbb{K})$.

   Consequentemente,

   $$\dim GL(n, \mathbb{R}) = n^2 \quad \text{e} \quad \dim GL(n, \mathbb{C}) = 2n^2.$$

2. **Álgebra de Lie do grupo unitário** $U(1)$

   O grupo $U(1)$ é o conjunto das matrizes unitárias do tipo 1×1, ou seja, o conjunto das matrizes que verificam $[a + ib]^H [a + ib] = [1]$. Equivalentemente, $U(1)$ é o conjunto das matrizes $[a+ib]$ tais que $|a+ib|^2 = a^2+b^2 = 1$.

   Seja $\gamma(t) = [a(t) + ib(t)]$ um caminho diferenciável com $a^2(t) + b^2(t) = 1$ e tal que $\gamma(0) = a(0) + ib(0) = 1$ (ou seja, $a(0) = 1$ e $b(0) = 0$).

   Como $a^2(t) + b^2(t) = 1$, tem-se $2a(0)a'(0) + 2b(0)b'(0) = 0$. Ou seja, $a'(0) = 0$, uma vez que $a(0) = 1$ e $b(0) = 0$. Logo, $\gamma'(0) = a'(0)+ib'(0) = ib'(0) \in \operatorname{Span}\{i\}$.

   Conclui-se que $\mathfrak{u}(1) = \operatorname{Span}\{i\}$, e portanto $\dim U(1) = 1$.

3. **Álgebra de Lie do grupo dos quaterniões unitários** S^3

   A álgebra de Lie do grupo dos quaterniões unitários é $\operatorname{Span}\{i, j, k\}$, e portanto $\dim S^3 = 3$.

   A demonstração é deixada como exercício pois é análoga à demonstração apresentada no item anterior.

4. **Álgebra de Lie do grupo linear especial** $SL(n, \mathbb{K})$

   A álgebra de Lie do grupo

$$SL(n, \mathbb{K}) = \{A \in GL(n, \mathbb{K}) : \ \det(A) = 1\},$$

é

$$\mathfrak{sl}(n, \mathbb{K}) = \{A \in GL(n, \mathbb{K}) : \ \mathrm{tr}(A) = 0\}.$$

Para mostrar a afirmação anterior iremos usar o resultado seguinte:

$$\frac{d}{dt}\Big|_{t=0} \det(\gamma(t)) = \mathrm{tr}(\gamma'(0)), \tag{7.49}$$

para um qualquer caminho diferenciável $\gamma(t) \in M_n(\mathbb{K})$ com $\gamma(0) = I$.

A demonstração desta igualdade pode ser realizada conforme se sugere no exercício seguinte.

Exercício 7.33. Considere o caminho diferenciável $\gamma(t) = [a_{ij}(t)]_{i,j=1,\ldots,n}$ em $GL(n, \mathbb{K})$ tal que $\gamma(0) = I$.

a) Use o desenvolvimento de Laplace ao longo da última linha de $\gamma(t)$ para mostrar

$$\frac{d}{dt}\Big|_{t=0} \det(\gamma(t)) = a'_{nn}(0) + \frac{d}{dt}\Big|_{t=0} \det(M_{nn}(t)), \tag{7.50}$$

onde $M_{nn}(t)$ designa o menor da entrada $a_{nn}(t)$.

b) Use a alínea anterior e o desenvolvimento de Laplace (segundo a última linha) no cálculo de $\det(M_{nn}(t))$, para mostrar a igualdade (7.49) por indução no tamanho da matriz.

▲

Se $\gamma(t)$ é um caminho regular em $SL(n, \mathbb{R})$ tal que $\gamma(0) = I$, então o traço de $\gamma'(0)$ é igual a zero, uma vez que por (7.49) se tem

$$0 = \frac{d}{dt}\Big|_{t=0} \det(\gamma(t)) = \mathrm{tr}(\gamma'(0)).$$

Ou seja, todo o elemento da álgebra de Lie $\mathfrak{sl}(n)$ tem traço igual a zero.

Falta mostrar que toda a matriz de traço igual a zero é um elemento de $\mathfrak{sl}(n)$. Para tal, considere-se uma matriz $A \in M_n(\mathbb{R})$ com traço igual a zero. Como a álgebra de Lie de $GL(n)$ é $M_n(\mathbb{R})$, existe um caminho regular $\gamma \in GL(n)$ tal que $\gamma(0) = I$ e $\gamma'(0) = A$.

O caminho $\alpha(t) = \dfrac{1}{\sqrt[n]{\det(\gamma(t))}}\gamma(t)$, onde $\gamma(0) = I$ e $\gamma'(0) = A$, é um caminho regular em $SL(n, \mathbb{R})$ uma vez que $\det(\alpha(t)) = 1$. Além disso, esta curva passa por I quando $t = 0$, isto é, $\alpha(0) = \gamma(0) = I$. Calculando $\alpha'(0)$, obtém-se

$$\begin{aligned}\alpha'(0) &= \frac{-\operatorname{tr}(\gamma'(0))}{n\sqrt[n]{(\det(\gamma(0)))^{1+n}}}\gamma(0) + \frac{1}{\sqrt[n]{\det(\gamma(0))}}\gamma'(0)\\ &= \frac{-\operatorname{tr}(A)\gamma(0)}{n} + \gamma'(0) = A\end{aligned}$$

Logo, qualquer matriz de traço igual a zero pertence a $\mathfrak{sl}(n)$.

A demonstração para $SL(n, \mathbb{C})$ é inteiramente análoga, uma vez que sendo o determinante uma função contínua podemos escolher o ramo da função complexa $h(t) = \sqrt[n]{\det(\gamma(t))}$ por forma a que $h(0) = I$.

Por conseguinte,

$$\dim SL(n, \mathbb{R}) = n^2 - 1 \quad \text{e} \quad \dim SL(n, \mathbb{C}) = 2(n^2 - 1).$$

Aproveitando os tópicos desenvolvidos na Secção 4.4.2, mostramos que a exponencial de matrizes representa um papel fundamental na escolha de caminhos num grupo de matrizes G, porquanto a exponencial de uma matriz da álgebra de Lie $\mathfrak{g}$ é um elemento de G. Pretendemos mostrar que se $A \in \mathfrak{gl}(n, \mathbb{K})$ a curva $\gamma(t) = e^{At}$ (verificando $\gamma(0) = I$ e $\gamma'(0) = A$) está necessariamente contida num grupo $G \subset GL(n, \mathbb{K})$. Em seguida abordamos esta questão e calculamos as álgebras de Lie dos (sub)grupos ortogonais e unitários.

Comecemos por definir o que se entende por homomorfismo entre dois grupos.

Definição 7.19. Um homomorfismo entre o grupo $(G, \star)$ e o grupo $(H, \cdot)$ é uma função $f : G \to H$ que satisfaz

$$f(g_1 \star g_2) = f(g_1) \cdot f(g_2), \quad \text{para todo } g_1, g_2 \in G.$$

Exercício 7.34. Mostre que se f é um homomorfismo entre os grupos $(G, \star)$ e $(H, \cdot)$, então:

$$f(e_G) = f(e_H) \quad \text{e} \quad f(g^{-1}) = (f(g))^{-1},$$

onde e_G e e_H designam, respectivamente, a identidade em G e em H. ▲

Definição 7.20. Um subgrupo a um parâmetro de um grupo de matrizes G é homomorfismo diferenciável $\gamma : \mathbb{R} \to G$.

Note-se que na definição anterior, o homomorfismo $\gamma : \mathbb{R} \to G$ é um homomorfismo entre o grupo (aditivo) $(\mathbb{R}, +)$ e o grupo (multiplicativo) $(G, \cdot)$. Portanto, $\gamma(t + \theta) = \gamma(t) \cdot \gamma(\theta)$.

Considere-se o grupo $SO(2)$, das matrizes ortogonais, do tipo 2×2, cujo determinante é igual a 1. Qualquer matriz deste grupo tem a forma

$$R(t) = \begin{bmatrix} \cos t & -\operatorname{sen} t \\ \operatorname{sen} t & \cos t \end{bmatrix}. \tag{7.51}$$

Como facilmente se verifica $R(t)R(\theta) = R(t + \theta)$, e portanto $f : t \mapsto R(t)$ é um homomorfismo do grupo $(\mathbb{R}, +)$ no grupo $(SO(2), \cdot)$. Ou seja, $SO(2)$ é um exemplo de um subgrupo a um parâmetro.

A função $R(t)$ define um caminho regular em $SO(2)$ cujo valor na identidade é $R(0) = I$. Logo, a matriz

$$R'(0) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = A \tag{7.52}$$

pertence à álgebra de Lie de $SO(2)$, isto é, $A \in \mathfrak{so}(2)$. As matrizes $R(t)$ satisfazem a equação diferencial

$$R'(t) = \begin{bmatrix} -\operatorname{sen} t & -\cos t \\ \cos t & -\operatorname{sen} t \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \cos t & -\operatorname{sen} t \\ \operatorname{sen} t & \cos t \end{bmatrix} = AR(t).$$

A exponencial e^{At} é a única matriz solução fundamental da equação $R'(t) = AR(t)$ que em $t = 0$ é a identidade (ver Definição 4.10, pág. 229). Assim, as matrizes $R(t)$ pertencentes ao grupo $SO(2)$ são da forma $R(t) = e^{At}$, com $A \in \mathfrak{so}(2)$. No Exemplo 4.17 (pág. 230) calculou-se a exponencial da matriz At, e obtivemos a matriz $R(t)$, confirmando-se assim este resultado.

Os subgrupos a um parâmetro de um grupo de matrizes são sempre dados pela exponencial de matrizes como passamos a explicar.

Teorema 7.10. Os subgrupos a um parâmetro de um grupo de matrizes G são da forma $t \mapsto e^{At}$ com A pertencente à álgebra de Lie $\mathfrak{g}$ de G.

Demonstração. Seja $\mathfrak{g} \subset \mathfrak{gl}(n, \mathbb{K})$ a álgebra de Lie de $G \subseteq GL(n, \mathbb{K})$ (com $\mathbb{K} = \mathbb{R}$ e $\mathbb{K} = \mathbb{C}$) e $A \in \mathfrak{g}$. Mostremos que a função $\gamma : \mathbb{R} \to G$ definida por $\gamma(t) = e^{At}$, é um subgrupo a um parâmetro.

Em primeiro lugar refira-se que do Exercício 4.8 (pág.230) podemos concluir que a matriz e^{At} é sempre invertível e portanto pertencente a $GL(n, \mathbb{K})$. Usando os resultados enunciados no referido exercício, sabe-se que se as matrizes A e B comutam, é válida a igualdade $e^{(A+B)t} = e^{At}e^{Bt}$. Em particular, como At comuta com $A\theta$, tem-se $e^{A(t+\theta)} = e^{At}e^{A\theta}$. Por conseguinte, $\gamma(t+\theta) = e^{A(t+\theta)} = e^{At}e^{A\theta} = \gamma(t)\gamma(\theta)$. Ou seja, a função diferenciável γ é um subgrupo a um parâmetro de G.

Suponha-se agora que $\gamma : \mathbb{R} \to G$ é um subgrupo a um parâmetro de G e que $A = \gamma'(0)$. Por definição de derivada, tem-se

$$\begin{aligned}\gamma'(t) &= \lim_{h\to 0} \frac{\gamma(h+t) - \gamma(t)}{h} = \lim_{h\to 0} \frac{\gamma(h) - I}{h}\gamma(t)\\ &= \lim_{h\to 0} \frac{\gamma(h) - \gamma(0)}{h}\gamma(t) = \gamma'(0)\gamma(t) = A\gamma(t),\end{aligned}$$

onde na segunda igualdade se aplicou a definição de subgrupo a um parâmetro, e na terceira igualdade o facto de $\gamma(0)$ ser igual a I (um homomorfismo entre dois grupos preserva a identidade desses grupos, conforme o Exercício 7.34).

Como $\gamma'(t) = A\gamma(t)$ e $\gamma(0) = I$, temos que $\gamma(t) = e^{tA}$. ☐

Do Teorema anterior segue que a exponencial $\mathfrak{g} \to G$ está definida por $A \mapsto e^A$. A imagem desta exponencial é em geral um subconjunto estrito de G. Isto é, podem existir elementos de G que não sejam a exponencial de nenhum elemento de $\mathfrak{g}$. Por exemplo, considere-se a álgebra de Lie $\mathfrak{sl}(2)$, do grupo $SL(2)$. Qualquer matriz não nula $A \in \mathfrak{sl}(2)$ tem traço nulo, pelo que A tem valores próprios simétricos (o traço de uma matriz é igual à soma dos valores próprios). Como A é 2×2, esta matriz é diagonalizável. Logo, se $A \in \mathfrak{sl}(2)$ tem-se $A = PDP^{-1}$ com D uma matriz diagonal. A exponencial e^{At} é também uma matriz diagonalizável, já que $e^{At} = Pe^{Dt}P^{-1}$ com $e^{Dt} = \operatorname{diag}(e^{\lambda t}, e^{-\lambda t})$ onde λ é valor próprio de A (ver Exercício 4.9, na página 231). No entanto, existem matrizes em $SL(2)$ que não são diagonalizáveis, por exemplo, a matriz $B = \begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix}$. Logo, a matriz $B \in SL(2)$ não é a exponencial de nenhuma matriz de $\mathfrak{sl}(2)$.

Determinemos agora as álgebras de Lie de $O(n)$ e $U(n)$.

1. **Álgebra de Lie de** $O(n)$

   A álgebra de Lie do grupo ortogonal é o conjunto das matrizes reais anti-simétricas, isto é,

   $$\mathfrak{o}(n) = \left\{A \in GL(n) : A = -A^T\right\}.$$

   De facto, qualquer caminho $\gamma(t)$ em $O(n)$, com $\gamma(0) = I$, satisfaz a igualdade $\gamma(t)\gamma(t)^T = I$. Logo, diferenciando esta igualdade, obtemos $\gamma'(0) + \gamma'(0)^T = 0$, ou seja, $A = \gamma'(0)$ verifica $A = -A^T$. Assim, a álgebra de Lie de $O(n)$ está contida no conjunto das matrizes anti-simétricas.

   Falta mostrar que o conjunto das matrizes anti-simétricas está contido na álgebra de Lie de $O(n)$. Sendo A uma matriz anti-simétrica verifica-se

   $$(e^{At})^T(e^{At}) = (e^{A^Tt})(e^{At}) = (e^{-At})(e^{At}) = I,$$

   onde a última igualdade resulta do facto da matriz inversa de e^{At} ser e^{-At} (ver Exercício 4.8). Assim, se A é anti-simétrica, $\gamma(t) = e^{At}$ é um caminho em $O(n)$ que satisfaz $\gamma(0) = I$ e $\gamma'(0) = A$, ou seja $A \in \mathfrak{o}(n)$. Por conseguinte, o conjunto das matrizes anti-simétricas está contido na álgebra de Lie de $O(n)$.

   O subgrupo $SO(n)$ de $O(n)$ é formado pelas matrizes ortogonais cujo determinante é igual a 1. Como o determinante toma os valores ± 1, logo é constante igual a 1 ao longo de um caminho em $O(n)$ que passe pela identidade. Assim, os caminhos em $O(n)$ que passam pela identidade estão em $SO(n)$. Conclui-se portanto que $\mathfrak{o}(n) = \mathfrak{so}(n)$.

   Calculemos a dimensão de $\mathfrak{o}(n) = \mathfrak{so}(n)$. Para esse efeito, relembramos que uma matriz anti-simétrica, de ordem n, possui entradas abaixo da diagonal principal iguais aos simétricos das entradas acima da diagonal principal, e que as entradas na diagonal principal são todas nulas (pois $A = -A^T$). Existem $\frac{n^2-n}{2}$ entradas acima da diagonal principal, e uma base para $\mathfrak{o}(n) = \mathfrak{so}(n)$ é constituída pelas seguintes $\frac{n^2-n}{2}$ matrizes anti-simétricas:

   $$\{E_{ij} - E_{ji} : 1 \leq i < j \leq n\},$$

onde E_{ij} designa a matriz cuja entrada ij é igual a -1 e todas as outras entradas são iguais a zero. Por exemplo,

$$\mathfrak{so}(3) = \operatorname{Span}\left\{ \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \right\}. \qquad (7.53)$$

Logo,

$$\dim \mathfrak{o}(n) = \dim \mathfrak{so}(n) = \frac{n^2 - n}{2}.$$

Exercício 7.35. Determine a exponencial de cada uma das matrizes da base de $\mathfrak{so}(3) = \mathfrak{o}(3)$ em (7.53). Verifique que as exponenciais obtidas representam rotações em torno dos eixos coordenados. ▲

2. **Álgebra de Lie de $U(n)$**

A álgebra de Lie de $U(n)$ é o conjunto das matrizes anti-hermitianas, isto é,

$$\mathfrak{u}(n) = \left\{ A \in GL(n, \mathbb{C}) : A = -A^H \right\}.$$

A demonstração é análoga à realizada para a álgebra de Lie de $O(n)$. Note-se que derivando $\gamma(t) \in U(n)$ se obtém $\gamma'(0)^H = -\gamma'(0)$.

Para calcular a dimensão da álgebra de Lie $\mathfrak{u}(n)$, deverá levar-se em consideração que uma matriz anti-hermitiana possui imaginários puros na diagonal principal, e que as entradas abaixo da diagonal principal são os simétricos dos conjugados das entradas acima da diagonal principal. Logo, uma base de $\mathfrak{u}(n)$ é formada por:

i) As n matrizes diagonais P_k que têm todas as entradas nulas excepto a entrada kk que é igual a i;

ii) As $(n^2 - n)/2$ matrizes anti-simétricas $(E_{ij} - E_{ji})$ (com $i < j$), onde E_{ij} são matrizes de ordem n com todas as entradas nulas excepto a entrada ij que é igual a -1;

iii) As $(n^2 - n)/2$ matrizes hermitianas $i(E_{ij} - E_{ji})$ (com $i < j$).

Assim, a dimensão de $\mathfrak{u}(n)$ é

$$\dim \mathfrak{u}(n) = n + 2\frac{n^2 - n}{2} = n^2.$$

Por exemplo, a álgebra de Lie de $U(2)$ é

$$\begin{aligned}\mathfrak{u}(2) &= \left\{ \begin{bmatrix} ai & -b+ci \\ b+ci & di \end{bmatrix} : \ a,b,c,d \in \mathbb{R} \right\} \\ &= \operatorname{Span}_{\mathbb{R}} \left\{ \begin{bmatrix} i & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}, \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & i \end{bmatrix} \right\}.\end{aligned}$$

Esta álgebra de Lie é um subespaço linear real das matrizes 2×2 de entradas complexas, porém não é um subespaço linear complexo das matrizes 2×2 de entradas complexas como se verifica facilmente.

3. **As álgebras de Lie $\mathfrak{su}(2)$ e $\mathfrak{so}(3)$**

   Embora as álgebras de Lie de $O(n)$ e de $SO(n)$ sejam iguais, o mesmo não acontece no caso dos grupos unitários. De facto, como $SU(n)$ é o grupo das matrizes unitárias de determinante igual a 1, a sua álgebra de Lie é o conjunto das matrizes anti-hermitianas de traço igual a zero (a derivada do determinante é igual ao traço). Assim, a dimensão de $\mathfrak{su}(n)$ é igual à dimensão de $\mathfrak{u}(n)$ menos uma unidade (correspondente à condição do traço ser nulo), isto é, $\dim \mathfrak{su}(n) = \dim \mathfrak{u}(n) - 1 = n^2 - 1$. Por exemplo, o espaço das matrizes complexas, de segunda ordem anti-hermitianas de traço nulo, é

$$\mathfrak{su}(2) = \left\{ \begin{bmatrix} ai & -b+ci \\ b+ci & -ai \end{bmatrix} : \ a,b,c \in \mathbb{R} \right\} = \operatorname{Span}_{\mathbb{R}} \{E_1, E_2, E_3\},$$

   com

$$E_1 = \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}, \qquad E_2 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \qquad E_3 = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}. \tag{7.54}$$

   Como $\{E_1, E_2, E_3\}$ é uma base de $\mathfrak{su}(2)$, a dimensão de $\mathfrak{su}(2)$ é igual a 3.

   Vejamos agora que a álgebra de Lie $(\mathfrak{su}(2), [\ ,\ ])$ é isomorfa à álgebra de Lie $(\mathbb{R}^3, \times)$, onde $\times$ é o produto vectorial em $\mathbb{R}^3$. Considere-se a função, $\varphi : \mathfrak{su}(2) \to \mathbb{R}^3$, definida por

$$\varphi : \begin{bmatrix} ai & -b+ci \\ b+ci & -ai \end{bmatrix} \longmapsto (2c, 2b, 2a).$$

   A função φ é obviamente linear e bijectiva. Para mostrar que φ é um isomorfismo de álgebras de Lie, falta mostrar a igualdade $\varphi([A,B]) =$

$\varphi(A) \times \varphi(B)$, para quaisquer $A, B \in \mathfrak{su}(2)$. Para tal, basta verificar esta igualdade para os vectores de uma base de $\mathfrak{su}(2)$.

Considere-se a base de $\mathfrak{su}(2)$ constituída pelas matrizes E_1, E_2, E_3 definidas em (7.54). As imagens por φ destas matrizes são $\varphi(E_1) = 2\mathbf{e}_1, \varphi(E_2) = 2\mathbf{e}_2$ e $\varphi(E_3) = 2\mathbf{e}_3$. Calculando os comutadores respectivos, obtém-se

$$\begin{aligned}
[E_1, E_2] = 2E_3 \Longrightarrow \varphi([E_1, E_2]) &= \varphi(2E_3) = 4\mathbf{e}_3 = 2\mathbf{e}_1 \times 2\mathbf{e}_2 \\
&= \varphi(E_1) \times \varphi(E_2), \\
[E_2, E_3] = 2E_1 \Longrightarrow \varphi([E_2, E_3]) &= \varphi(2E_1) = 4\mathbf{e}_1 = 2\mathbf{e}_2 \times 2\mathbf{e}_3 \\
&= \varphi(E_2) \times \varphi(E_3), \\
[E_3, E_1] = 2E_2 \Longrightarrow \varphi([E_3, E_1]) &= \varphi(2E_2) = 4\mathbf{e}_2 = 2\mathbf{e}_3 \times 2\mathbf{e}_1 \\
&= \varphi(E_3) \times \varphi(E_1).
\end{aligned}$$

Do Exercício 6.5, na página 333, sabemos que as ágebras de Lie $(\mathbb{R}^3, \times)$ e $(\mathfrak{so}(3), [\,,\,])$ são isomorfas, sendo o isomorfismo definido por

$$\mathbb{R}^3 \ni x = (x_1, x_2, x_3) \mapsto \begin{bmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{bmatrix} \in \mathfrak{so}(3).$$

Conclui-se que as álgebras de Lie $(\mathfrak{su}(2), [\,,\,])$, $(\mathfrak{so}(3), [\,,\,])$ e $(\mathbb{R}^3, \times)$ são isomorfas:

$$\mathfrak{su}(2) \simeq \mathfrak{so}(3) \simeq \mathbb{R}^3.$$

Refira-se ainda que a base $\{E_1, E_2, E_3\}$ de $\mathfrak{su}(2)$ pode ser descrita em termos de matrizes hermitianas $\sigma_1, \sigma_2, \sigma_3$ conhecidas como *matrizes spin de Pauli*[13], que desempenham um papel importante em mecânica quântica. Estas matrizes são

$$\sigma_1 = -iE_1, \quad \sigma_2 = iE_2, \quad \sigma_3 = -iE_3.$$

[13] Wolfgang Ernst Pauli (1900 – 1958), físico teórico austríaco, que foi prémio Nobel da Física em 1945.

Exercícios

1. Indique se as matrizes seguintes são ortogonais, unitárias, simétricas ou hermitianas.

a) $\begin{bmatrix} \cos\theta & i\,\mathrm{sen}\,\theta \\ i\,\mathrm{sen}\,\theta & \cos\theta \end{bmatrix}$, para qualquer $\theta \in \mathbb{R}$.

b) $\begin{bmatrix} a & ib \\ -ib & c \end{bmatrix}$, para quaisquer $a, b, c \in \mathbb{R}$.

c) $\begin{bmatrix} i & ib \\ -ib & -i \end{bmatrix}$, para qualquer $b \in \mathbb{R}$.

d) $\begin{bmatrix} \sqrt{3} & \sqrt{2} \\ 1/\sqrt{3} & -\sqrt{2} \end{bmatrix}$.

e) $\begin{bmatrix} \sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{3} & -\sqrt{3} \end{bmatrix}$.

2. Seja $\mathbf{x}$ um vector de $\mathbb{R}^n$, de norma igual a 1, e $B = \mathbf{x}\mathbf{x}^T$.

a) Mostre que $B^2 = B$.

b) Mostre que $\mathbf{x}$ é um vector próprio de B e indique o valor próprio associado.

3. Indique o valor lógico das afirmações seguintes:

a) Uma matriz complexa simétrica é hermitiana.

b) Toda a matriz simétrica tem valores próprios reais.

c) Toda a matriz real e simétrica tem valores próprios reais.

d) $\lambda = 1 - i$ pode ser um valor próprio de uma matriz ortogonal.

e) Se as colunas de uma matriz formam uma base ortogonal de $\mathbb{R}^n$, então a matriz é ortogonal.

f) Se A é uma matriz simétrica e ortogonal, então $A^2 = -I$

g) Uma matriz real anti-simétrica tem valores próprios reais.

h) Toda a matriz unitária tem valores próprios iguais a ± 1.

i) Se $A = PDP^T$, com D diagonal, então A é simétrica.

j) Uma matriz ortogonal é ortogonalmente diagonalizável.

4. Determine condições nos parâmetros reais a, b, c, d, e e f, de modo a que a matriz

$$A = \begin{bmatrix} a & c + id \\ e + if & b \end{bmatrix}$$

seja:

a) hermitiana.

b) unitária.

5. Seja A uma matriz simétrica e $\mathbf{x} \in \mathbb{R}^n$. Mostre que, para o produto interno usual em $\mathbb{R}^n$, se verifica $\langle \mathbf{x}, A\mathbf{x} \rangle = \langle A\mathbf{x}, \mathbf{x} \rangle$.

6. Mostre que se A é simétrica do tipo $n \times n$ e B é uma matriz do tipo $k \times n$, a BAB^T é simétrica.

7. Seja λ um valor próprio de multiplicidade algébrica k da matriz ortogonalmente diagonalizável A. Justifique por que razão a multiplicidade geométrica de λ é também igual a k.

8. Diagonalize ortogonalmente as seguintes matrizes:

a) $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$.

b) $\begin{bmatrix} 2 & 1 & -1 \\ 1 & -1 & 1 \\ -1 & 1 & -1 \end{bmatrix}$.

c) $\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 2 & 1 & -1 \\ 0 & 1 & -1 & 1 \\ 0 & -1 & 1 & -1 \end{bmatrix}$.

9. Diagonalize unitariamente as matrizes seguintes:

a) $\begin{bmatrix} 2 & -i \\ i & 2 \end{bmatrix}$

b) $\begin{bmatrix} 1 & i & 0 \\ -i & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$

10. Para $\mathbf{x} = (x, y) \in \mathbb{R}^2$, determine a matriz simétrica que define as formas quadráticas seguintes e classifique-as.

a) $5x^2 - 3xy + 7y^2$
b) $6xy$.
c) $3x^2 + 5y^2$.

11. Classifique as formas quadráticas seguintes e encontre a mudança de variáveis $\mathbf{x} = P\mathbf{y}$ que transforma a forma quadrática dada noutra sem termos cruzados.

a) $q(\mathbf{x}) - 4x_1x_2 + 3x_2^2$
b) $q(\mathbf{x}) = 3x_1x_2$.
c) $q(\mathbf{x}) = -4x_1^2 + 4x_1x_2 - x_2^2$.
d) $q(\mathbf{x}) = 2x_1^2 + 3x_1x_2 + 2x_2^2$.

12. Seja A uma matriz real, simétrica, e invertível. Mostre que se a forma quadrática $\mathbf{x}^T A\mathbf{x}$ é definida positiva, então a forma quadrática $\mathbf{x}^T A^{-1}\mathbf{x}$ também é definida positiva. [Sugestão: considere uma diagonalização de A].

13. Indique o valor lógico das afirmações seguintes, onde A é uma matriz simétrica real.

a) Uma matriz definida negativa tem pelo menos uma entrada na diagonal principal negativa.
b) Se $\mathbf{x}^T A\mathbf{x} > 0$ para qualquer vector próprio $\mathbf{x}$, então A é definida positiva.
c) Se $\mathbf{x}^T A\mathbf{x} > 0$ para algum vector $\mathbf{x}$, então A é definida positiva.
d) Uma forma quadrática definida positiva pode ser transformada numa forma definida negativa mediante uma mudança de variável $\mathbf{x} = P\mathbf{y}$, onde P é uma matriz ortogonal.
e) Se P é uma matriz ortogonal então a mudança de variável $\mathbf{x} = P\mathbf{y}$ transforma a forma quadrática $\mathbf{x}^T A\mathbf{x}$ numa forma quadrática cuja matriz associada é $P^{-1}AP$.

14. Seja A uma matriz real e simétrica.

a) Mostre que o valor máximo de $\mathbf{x}^T A\mathbf{x}$ na esfera unitária (isto é, para $\|\mathbf{x}\| = 1$), é igual ao maior valor próprio de A e que o valor mínimo é igual ao menor valor próprio.
b) Mostre que o valor máximo M (resp. mínimo m) de $\mathbf{x}^T A\mathbf{x}$, na esfera unitária, é atingido quando $\mathbf{x}$ é um vector próprio de A associado a M (resp. m).
c) Use os resultados das alíneas anteriores para determinar o máximo e o mínimo de $3x_1^2 + 2x_1x_2 - 5x_2^2 + x_3^2$, sob a condição de $x_1^2 + x_2^2 + x_3^2 = 1$. Indique ainda vectores onde estes valores de máximo e mínimo são atingidos.

15. Suponha que a seguinte factorização de A é uma decomposição SVD:

$$A = \begin{bmatrix} \frac{2\sqrt{2}}{3} & 0 & \frac{-1}{3} \\ \frac{-1}{3\sqrt{2}} & \frac{-1}{\sqrt{2}} & \frac{-2}{3} \\ \frac{1}{3\sqrt{2}} & \frac{-1}{\sqrt{2}} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

a) Diga qual o tipo da matriz A e a característica de A.

b) Sem efectuar quaisquer cálculos, indique uma base ortonormada para $N(A)$, $N(A^T)$, $EC(A)$ e $EL(A)$.

16. Determine os valores singulares das matrizes seguintes:

a) $\begin{bmatrix} -2 & 0 \\ 0 & 4 \end{bmatrix}$. b) $\begin{bmatrix} \sqrt{5} & 1 \\ 0 & \sqrt{3} \end{bmatrix}$.

c) $\begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 0 & 1 \end{bmatrix}$. d) $\begin{bmatrix} 2 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$.

17. Seja A uma matriz quadrada invertível. Determine uma decomposição em valores singulares de A^{-1} a partir de uma SVD de A.

18. Seja A uma matriz simétrica real definida positiva. Mostre que a diagonalização ortogonal de A (isto é, $A = PDP^T$) é uma decomposição em valores singulares.

19. Seja A uma matriz quadrada. Mostre que o módulo do determinante de A é igual ao produto dos valores singulares de A.

20. Mostre que toda a matriz quadrada A admite uma *decomposição polar* da forma $A = PQ$, onde P é uma matriz semidefinida positiva com a mesma característica de A, e Q é uma matriz ortogonal.

21. Use uma decomposição SVD da matriz A, do tipo $p \times n$, para mostrar $(A^+)^T = (A^T)^+$.

22. Uma forma quadrática $q_K(\mathbf{x})$ diz-se invariante pela matriz A se, para todo $\mathbf{x}$, é verificada a igualdade $q_K(A\mathbf{x}) = q_K(\mathbf{x})$. Mostre que:

a) O grupo ortogonal $O(n)$ é o grupo das matrizes que deixa invariante a forma quadrática $q(\mathbf{x}) = \mathbf{x}^T\mathbf{x}$, para $\mathbf{x} \in \mathbb{R}^n$.

b) O grupo unitário $U(n)$ é o grupo de matrizes que deixa invariante a forma quadrática $q(\mathbf{x}) = \mathbf{x}^H\mathbf{x}$, para $\mathbf{x} \in \mathbb{C}^n$.

23. Indique o valor lógico das proposições seguintes.

a) A matriz $\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix}$ pertence a $SL(2, \mathbb{C})$.

b) A matriz $\begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix}$ pertence a $SU(2)$.

c) A matriz $\frac{1}{\sqrt{2}}\begin{bmatrix} i & i \\ i & -i \end{bmatrix}$ pertence a $SU(2)$.

d) A matriz $\begin{bmatrix} -1 & 0 & 0 \\ 0 & \cos\theta & \operatorname{sen}\theta \\ 0 & -\operatorname{sen}\theta & \cos\theta \end{bmatrix}$ pertence a $SO(3)$.

24. Identifique um quaternião $\mathfrak{q} = q_0 + xi + yj + zk$ com o par $\mathfrak{q} = (q_0, \mathbf{q}) \in \mathbb{R} \times \mathbb{R}^3$ onde $\mathbf{q} = (x, y, z)$. Mostre que a multiplicação de quaterniões é dada por

$$\begin{aligned} \mathfrak{q}\mathfrak{p} &= (q_0, \mathbf{q})(r_0, \mathbf{r}) \\ &= (q_0 r_0 - \langle \mathbf{q}, \mathbf{r} \rangle \,,\, q_0\mathbf{r} + r_0\mathbf{q} + \mathbf{q} \times \mathbf{r}), \end{aligned}$$

onde $\langle \,,\, \rangle$ é o produto interno usual em $\mathbb{R}^3$, e $\mathbf{q} \times \mathbf{r}$ é o produto vectorial em $\mathbb{R}^3$.

25. Um quaternião $\mathfrak{q} = q_0 + xi + yj + zk$ diz-se um *quaternião imaginário puro* se $q_0 = 0$. Considerando a identificação dos quaterniões com vectores de $\mathbb{R}^4$, indicada no exercício anterior, mostre que

a) O produto de quaterniões imaginários puros, $\mathfrak{q} = (0, \mathbf{q})$ e $\mathfrak{p} = (0, \mathbf{p})$, é um imaginário puro se e só se os vectores $\mathbf{q}$ e $\mathbf{p}$ são ortogonais.

b) Se $\mathfrak{q}$ é um quaternião imaginário puro de módulo igual a 1, então $\mathfrak{q}^2 = -1$.

26. O *grupo simpléctico* $Sp(2n)$ é o grupo de matrizes que deixa invariante a forma bilinear $\varphi_J(\mathbf{x},\mathbf{y}) = \mathbf{x}^T J\mathbf{y}$, com $J = \begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix}$ e I_n a matriz identidade de ordem n. Mostre que:

a) O grupo $Sp(2n)$ é o conjunto das matrizes A, tais que $A^T JA = J$.
b) A álgebra de Lie de $Sp(2n)$ é formada pelas matrizes da forma $\begin{bmatrix} A & B \\ C & -A^T \end{bmatrix}$, onde A, B e C são matrizes $n\times n$ que verificam $B = B^T$ e $C = C^T$.
c) Use os resultados das alíneas anteriores para determinar a dimensão de $Sp(2n)$.

27. Chamam-se *constantes de estrutura* de uma álgebra de Lie $\mathfrak{g}$ às constantes C_{ij}^k definidas por $[E_i, E_j] = \sum_k C_{ij}^k E_k$, onde $\{E_i\}_{i=1,\ldots,p}$ é uma base de $\mathfrak{g}$.

Considere a base $\{E_1, E_2, E_3\}$ da álgebra de Lie $\mathfrak{su}(2)$, onde as matrizes E_i são definidas em (7.54), na página 461, e determine as constantes de estrutura de $\mathfrak{su}(2)$.

28. Relembre que as as matrizes spin de Pauli são definidas por $\sigma_1 = -iE_1, \sigma_2 = iE_2$ e $\sigma_3 = -iE_3$, onde E_i são as matrizes do exercício anterior. Verifique que as matrizes σ_i, para $i = 1, 2, 3$, são matrizes hermitianas. Mostre ainda que estas matrizes geram (sobre $\mathbb{R}$) o subespaço linear das matrizes hermitianas de traço nulo.

29. Determine as exponenciais $e^{\sigma_i t}$, onde σ_i, para $i = 1, 2, 3$, são as matrizes de Pauli. Justifique por que razão as matrizes $U_j = e^{(-1)^j i\sigma_j t}$ pertencem a $SU(2)$.

30. Mostre que:

a) Em $\mathbb{R}^3$, o produto vectorial define um parêntesis de Lie.
b) O espaço linear $\mathbb{R}^3$ munido do produto vectorial e o espaço linear das matrizes anti-simétricas do tipo 3×3, munido do comutador de matrizes, são álgebras de Lie isomorfas.
Sugestão: verifique que a função seguinte é um isomorfismo de álgebras de Lie:

$$\mathbf{x} = (x_1, x_2, x_3) \mapsto$$
$$\begin{bmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{bmatrix} = X.$$

31. Seja $f_A : \mathfrak{su}(2) \to \mathfrak{su}(2)$ a função definida por $f_A(X) = [A, X]$, com

$$A = \begin{bmatrix} i & -1+i \\ 1+i & -i \end{bmatrix}.$$

a) Mostre que f_A é uma função linear.
b) Determine a matriz $M = M(f_A, \mathcal{B}, \mathcal{B})$ que representa f_A na base $\mathcal{B} = (E_1, E_2, E_3)$, onde as matrizes E_i são definidas em (7.54).

32. Seja $\mathfrak{q}$ um quaternião unitário e $\mathfrak{p} \in \mathbb{H}$. Mostre a igualdade

$$\left|\mathfrak{q}\mathfrak{p}\mathfrak{q}^{-1}\right|^2 = |\mathfrak{p}|^2.$$

Capítulo 8

Matrizes quadradas não diagonalizáveis

Este capítulo é dedicado à construção da forma canónica de Jordan[1] de uma matriz quadrada, e sua aplicação na determinação da solução geral do sistema linear homogéneo de equações diferenciais ordinárias, $\mathbf{x}' = A\mathbf{x}$, quando A é uma matriz não diagonalizável.

Embora em termos computacionais a forma canónica de Jordan de uma matriz seja pouco usada devido à sua instabilidade numérica, ela permite a obtenção de resultados importantes quer em álgebra linear quer noutras áreas, como geometria e teoria de grupos. Efectuamos aqui um estudo completo da construção da forma de Jordan para uma matriz quadrada, começando por estudar a forma de Jordan de matrizes nilpotentes.

A demonstração que se apresenta da existência da forma de Jordan de uma matriz, é técnica mas ao mesmo tempo construtiva. São apresentados vários exemplos de determinação desta forma, ilustrando a utilidade do conhecimento detalhado dos passos essenciais da demonstração.

A forma de Jordan de uma matriz é ainda aplicada na determinação de soluções de um sistema linear homogéneo de equações diferenciais, completando-se assim o estudo iniciado no Capítulo 4.

[1]Marie Ennemond Camille Jordan (1838 –1922), matemático francês com contribuições em teoria de grupos, análise matemática e análise complexa.

8.1 Forma canónica de Jordan para matrizes nilpotentes

Quando uma matriz quadrada A não é diagonalizável, o Teorema de Schur (Teorema 7.1, pág. 381) garante que a matriz é semelhante a uma matriz triangular superior. É portanto natural procurar uma matriz triangular J o mais simples possível que seja semelhante à matriz A. A matriz J, designada por forma canónica de Jordan (ou abreviadamente por forma de Jordan) da matriz A, é uma matriz diagonal por blocos em que os blocos J_i na diagonal têm todas as entradas nulas, excepto possivelmente as entradas da diagonal principal e as da diagonal acima desta (supradiagonal) que são iguais a 1. Ou seja, os blocos de J são da forma

$$J_i = \begin{bmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda_i \end{bmatrix}. \tag{8.1}$$

Na base da construção da forma canónica de Jordan de uma matriz quadrada está a forma de Jordan de uma matriz nilpotente. Por isso, iniciamos o nosso estudo estabelecendo a forma de Jordan para matrizes nilpotentes.

Definição 8.1. Uma matriz L diz-se *nilpotente* de índice k, com k um inteiro positivo, se L^k é a matriz nula $\mathbf{O}$ e $L^{k-1} \neq \mathbf{O}$.

Convenciona-se que o índice de nilpotência da matriz nula é igual a 1.

A forma de Jordan para uma matriz nilpotente diz-nos que qualquer matriz nilpotente é semelhante a uma matriz diagonal por blocos N, em que os blocos da diagonal são matrizes triangulares superiores com entradas todas nulas excepto, possivelmente, as entradas da supradiagonal as quais deverão ser iguais a 1. No exemplo que se segue apresentamos algumas matrizes nilpotentes em forma de Jordan.

Exemplo 8.1. Verifiquemos que as matrizes seguintes são nilpotentes e determinemos o seu índice de nilpotência, bem como a dimensão do seu núcleo.

a) A matriz $\mathbf{N} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ é nilpotente, de índice 3, já que $\mathbf{N}^3 = \mathbf{O}$ e

$$\mathbf{N}^2 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \neq \mathbf{O}. \text{ Além disso, } \dim N(\mathbf{N}) = 2.$$

b) A matriz $\mathbf{N} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ é nilpotente, de índice 2, uma vez que $\mathbf{N}^2 = \mathbf{O}$ e $\mathbf{N} \neq \mathbf{O}$. A dimensão do núcleo de $\mathbf{N}$ é $\dim N(\mathbf{N}) = 2$.

c) A matriz $\mathbf{N} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ é nilpotente, de índice 4, visto que $\mathbf{N}^4 = \mathbf{O}$ e $\mathbf{N}^3 \neq \mathbf{O}$. A dimensão do núcleo de $\mathbf{N}$ é $\dim N(\mathbf{N}) = 1$.

d) A matriz nula 4×4 é uma matriz nilpotente de índice 1 e $\dim N(\mathbf{N}) = 4$.

♦

No exemplo anterior apresentámos casos particulares de matrizes nilpotentes, nomeadamente matrizes diagonais por blocos em que cada bloco da diagonal é uma matriz nilpotente com entradas todas iguais a zero, excepto as entradas da supradiagonal que são todas iguais a 1. O número de blocos na diagonal das matrizes $\mathbf{N}$ do Exemplo 8.1 são:

a) 2 blocos: um bloco 3×3 e um bloco 1×1.
b) 2 blocos, ambos do tipo 2×2.
c) 1 bloco do tipo 4×4.
d) 4 blocos do tipo 1×1.

Como veremos, as matrizes $\mathbf{N}$ do exemplo anterior são as possíveis (a menos de uma reordenação dos blocos na diagonal) formas de Jordan de uma matriz nilpotente L do tipo 4×4. Saliente-se o facto de nestas matrizes o número de blocos ser igual à dimensão do núcleo da matriz, e da ordem do(s) maior(es) bloco(s) ser igual ao índice de nilpotência. A relação existente entre o número de blocos e a dimensão do núcleo da matriz, bem como a relação entre a ordem do(s) maior(es) bloco(s) e o índice da matriz, é suficiente, em certos casos, para determinar a forma de Jordan N de uma matriz nilpotente L.

Os valores próprios de uma matriz nilpotente são todos nulos. De facto, se L é nilpotente de índice k e λ é um valor próprio de L, isto é, $L\mathrm{x} = \lambda\mathrm{x}$ para algum

$\mathbf{x} \neq \mathbf{0}$, tem-se

$$L\mathbf{x} = \lambda\mathbf{x} \Longrightarrow L^k\mathbf{x} = \lambda^k\mathbf{x}.$$

Como $L^k = \mathbf{O}$, a última igualdade passa a ser $\mathbf{0} = \lambda^k\mathbf{x}$, a qual é equivalente a $\lambda = 0$ (já que $\mathbf{x} \neq \mathbf{0}$).

Por outro lado, se L é uma matriz de ordem n com um único valor próprio igual a zero, a equação característica de L é $\lambda^n = 0$. Logo, pelo Teorema de Cayley-Hamilton (Teorema 4.3 na página 205), resulta que $L^n = \mathbf{O}$. Se L é uma matriz não nula a igualdade $L^n = \mathbf{O}$ significa que existe $1 < k \leq n$ tal que $L^k = \mathbf{O}$ e $L^{k-1} \neq \mathbf{O}$. Ou seja, que L é nilpotente de índice k.

Assim, uma matriz é nilpotente se e só se zero é o seu único valor próprio.

Recorde-se que sendo L uma matriz diagonalizável, esta matriz é semelhante a uma matriz diagonal com entradas na diagonal principal iguais aos valores próprios de L. Logo, se L é nilpotente e diagonalizável, tem-se $L = P^{-1}\mathbf{O}P = \mathbf{O}$. Tal significa que a única matriz nilpotente diagonalizável é a matriz nula.

Enunciamos a seguir o que acabámos de provar.

Proposição 8.1.
- Uma matriz é nilpotente se e só se zero é único valor próprio da matriz.
- A única matriz nilpotente diagonalizável é a matriz nula.

Como veremos, conhecendo a forma de Jordan para matrizes nilpotentes determina-se a forma de Jordan de uma qualquer matriz quadrada. Apresentamos a seguir um exemplo ilustrativo da relação entre a forma de Jordan de uma matriz nilpotente e a forma de Jordan de uma matriz que não é nilpotente.

Exemplo 8.2. Seja A uma matriz de ordem n com um único valor próprio μ, e suponha-se que se sabe determinar a forma de Jordan N de uma matriz nilpotente.

Pretende-se determinar uma matriz de Jordan J, semelhante a A, tal que J seja diagonal por blocos e cujos blocos J_i na diagonal são da forma (8.1).

A matriz $(A - \mu I)$ é singular, e zero é o único valor próprio desta matriz (uma vez que estamos supondo que μ é o único valor próprio de A). Assim, a matriz $(A - \mu I)$ é nilpotente. Seja N a forma de Jordan de $(A - \mu I)$, isto é, $(A - \mu I) = PNP^{-1}$ com N diagonal por blocos em que as entradas dos blocos são todas nulas excepto as da supradiagonal que são iguais a 1. Logo,

$$(A - \mu I) = PNP^{-1} \Longleftrightarrow A = P(\mu I + N)P^{-1} = PJP^{-1}.$$

A matriz $J = (\mu I + N)$ tem a forma pretendida, sendo portanto a forma de Jordan de A. Ou seja, a forma de Jordan J de A é obtida da forma de Jordan N da matriz nilpotente $(A - \mu I)$. ♦

Pretendemos mostrar que se $L \neq \mathbf{O}$ é uma matriz nilpotente de ordem n e de índice $k > 1$, então existe uma matriz invertível P e uma matriz diagonal por blocos N, tal que

$$P^{-1}LP = N = \begin{bmatrix} \mathbf{N}_1 & \mathbf{O} & \cdots & \mathbf{O} \\ \mathbf{O} & \mathbf{N}_2 & \cdots & \mathbf{O} \\ \vdots & & \ddots & \vdots \\ \mathbf{O} & \mathbf{O} & \cdots & \mathbf{N}_t \end{bmatrix}, \tag{8.2}$$

onde $\mathbf{N}_j$ é uma matriz nilpotente com as entradas na supradiagonal iguais a 1 e zeros nas outras entradas. Ou seja,

$$\mathbf{N}_j = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}.$$

A matriz N é designada por *forma canónica de Jordan*, ou *forma normal de Jordan*, de L.

Antes de passarmos à demonstração deste resultado, observe-se o seguinte:

- Seja P uma matriz invertível, tal que $LP = PN$ com N como em (8.2). Particione-se P em t blocos P_j, $P = [P_1|P_2|\cdots|P_t]$, de tal forma que o número de colunas de cada bloco P_j seja igual à ordem do bloco $\mathbf{N}_j$. Logo, $LP = PN$ implica

$$LP_j = P_j\mathbf{N}_j, \quad \text{para todo } j = 1, \ldots, t. \tag{8.3}$$

  Uma vez que a primeira coluna de cada bloco $\mathbf{N}_j$ é nula, tem-se que a primeira coluna de $P_j\mathbf{N}_j$ é nula, e portanto a primeira coluna de LP_j é nula. Ou seja, a primeira coluna de cada bloco P_j pertence ao núcleo de L.

- Como P é invertível, da observação anterior segue que o núcleo de L tem pelo menos dimensão t, ou seja, que a multiplicidade geométrica do valor próprio zero é pelo menos t.

- Se $\mathbf{u}_j, \mathbf{v}_1, \ldots, \mathbf{v}_k$ são as colunas de P_j, a igualdade $LP_j = P_j\mathbf{N}_j$ em (8.3) é

$$LP_j = \begin{bmatrix} | & | & & | \\ L\mathbf{u}_j & L\mathbf{v}_1 & \cdots & L\mathbf{v}_k \\ | & | & & | \end{bmatrix} = P_j\mathbf{N}_j = \begin{bmatrix} | & | & | & & | \\ \mathbf{0} & \mathbf{u}_j & \mathbf{v}_1 & \cdots & \mathbf{v}_{k-1} \\ | & | & | & & | \end{bmatrix}.$$

  Por conseguinte, as colunas de P_j verificam

$$L\mathbf{u}_j = \mathbf{0}, L\mathbf{v}_1 = \mathbf{u}_j \text{ e } L\mathbf{v}_i = \mathbf{v}_{i-1}, \quad \text{para } i = 2, \ldots, k. \tag{8.4}$$

- Das observações anteriores, conclui-se que a existência da forma de Jordan N para L é equivalente à existência de uma base $B = \{\mathbf{u}_1, \ldots, \mathbf{u}_t\}$ para o núcleo de L, a partir da qual se constroí uma base de $\mathbb{C}^n$ (cujos vectores irão formar as colunas de P) verificando as relações (8.4).

Construção da Forma de Jordan para uma matriz nilpotente

Seja $L \neq \mathbf{O}$ uma matriz nilpotente de ordem n e de índice k.

Vamos começar por construir uma base B para o núcleo de L à custa de bases dos seguintes subespaços

$$M_i = EC(L^i) \cap N(L) \quad \text{para } i = 0, 1, \ldots, k. \tag{8.5}$$

a) Considerem-se os subespaços (8.5). Como $L^0 = I$, e o espaço das colunas de I tem dimensão n, o espaço M_0 é igual a $N(L)$. Por outro lado, como $L^k = \mathbf{O}$, temos $EC(L^k) = \{\mathbf{0}\}$, e portanto $M_k = \{\mathbf{0}\}$.
O espaço das colunas de uma potência positiva X^p, de uma matriz X, está contido (ou é igual) no espaço das colunas de X^{p-1}, uma vez que $\mathbf{y} = X^p\mathbf{x} \Longrightarrow \mathbf{y} = X^{p-1}(X\mathbf{x})$. Tem-se portanto $M_i \subseteq M_{i-1}$, e consequentemente a cadeia de inclusões:

$$\{\mathbf{0}\} = M_k \subseteq M_{k-1} \subseteq \cdots \subseteq M_1 \subseteq M_0 = N(L).$$

b) Construa-se uma base B de $N(L)$ da seguinte forma:
   - Parte-se de uma base S_{k-1} de M_{k-1}, completa-se esta base com um conjunto S_{k-2} de modo a que $S_{k-1} \cup S_{k-2}$ seja uma base de M_{k-2}. De seguida, estende-se $S_{k-1} \cup S_{k-2}$ a uma base de M_{k-3} juntando um conjunto S_{k-3}. Repete-se sucessivamente este processo até obter uma base B de $M_0 = N(L)$:

$$B = S_{k-1} \cup S_{k-2} \cup \cdots \cup S_0 = \{\mathbf{b}_1, \ldots, \mathbf{b}_t\}. \tag{8.6}$$

c) A base B em (8.6) não tem o número de vectores necessários para formar uma base de $\mathbb{C}^n$ já que, sendo $L \neq \mathbf{O}$, o núcleo de L tem dimensão inferior a n (a única matriz nilpotente que é diagonalizável é a matriz nula). Pretendemos por isso completar a base B por forma a obter uma base de $\mathbb{C}^n$ cujos vectores irão constituir as colunas da matriz P. Para tal, determinamos para cada vector de B uma sequência de vectores linearmente independentes do seguinte modo:

   i) Para cada vector $\mathbf{b}_r \in B$ com $\mathbf{b}_r \in S_i$, resolve-se o sistema $L^i\mathbf{x} = \mathbf{b}_r$. O sistema $L^i\mathbf{x} = \mathbf{b}_r$ é sempre possível já que S_i é uma base de $M_i = EC(L^i) \cap N(L)$. Seja $\mathbf{x}_r$ uma solução de $L^i\mathbf{x} = \mathbf{b}_r$, e considerem-se os $(i+1)$ vectores $\mathbf{x}_r, L\mathbf{x}_r, \ldots, L^i\mathbf{x}_r = \mathbf{b}_r$. O conjunto $I_{\mathbf{b}_r}$ formado por estes vectores, ou seja,

$$I_{\mathbf{b}_r} = \left\{L^i\mathbf{x}_r, L^{i-1}\mathbf{x}_r, \ldots, L\mathbf{x}_r, \mathbf{x}_r\right\}, \quad \text{com } \mathbf{b}_r \in S_i \text{ e } L^i\mathbf{x}_r = \mathbf{b}_r\,,$$

   é designado por *cadeia de Jordan* associada a $\mathbf{b}_r$.

   Mostremos que $I_{\mathbf{b}_r}$ é um conjunto linearmente independente. Em primeiro lugar note-se que, por construção, o conjunto S_i estende a base $S_{k-1} \cup S_{k-2} \cup \cdots \cup S_{i+1}$ de M_{i+1} a uma base de M_i. Assim, o vector $\mathbf{b}_r = L^i\mathbf{x}_r$ pertence a S_i mas não pertence a S_{i+1}. Ou seja, $\mathbf{b}_r \in EC(L^i)$ e $\mathbf{b}_r \notin EC(L^{i+1})$. Suponha-se, por absurdo, que existe uma combinação linear nula dos vectores de $I_{\mathbf{b}_r}$ com algum coeficiente não nulo. Seja j o menor índice para o qual $\alpha_j \neq 0$ na combinação linear $\alpha_0\mathbf{x}_r + \alpha_1 L\mathbf{x}_r + \cdots + \alpha_j L^j\mathbf{x}_r + \cdots + \alpha_i L^i\mathbf{x}_r = \mathbf{0}$. Dividindo por α_j, obtém-se

$$\begin{aligned} L^j\mathbf{x}_r &= \beta_{j+1}L^{j+1}\mathbf{x}_r + \cdots + \beta_i L^i\mathbf{x}_r = L^{j+1}\left(\beta_{j+1}\mathbf{x}_r + \cdots + \beta_i L^{i-(j+1)}\mathbf{x}_r\right) \\ &= L^{j+1}\mathbf{y}. \end{aligned}$$

   Logo, o vector $\mathbf{b}_r = L^i\mathbf{x}_r$ poder-se-ia escrever na forma

$$\mathbf{b}_r = L^i\mathbf{x}_r = L^{i-j}(L^j\mathbf{x}_r) = L^{i-j}(L^{j+1}\mathbf{y}) = L^{i+1}\mathbf{y},$$

   o que contraria a hipótese de $\mathbf{b}_r \notin EC(L^{i+1})$. Portanto, os α_i's são todos nulos, o que equivale a dizer que a cadeia $I_{\mathbf{b}_r}$ é linearmente independente.

d) O número de cadeias de Jordan (uma por cada vector da base B do núcleo de L) é igual à dimensão do núcleo de L, ou seja, igual à multiplicidade geométrica do valor próprio zero de L. Além disso, as cadeias de Jordan com o maior número de vectores são aquelas que se constroem a partir dos vectores da base B que pertencem a S_{k-1}. Estas cadeias têm k vectores, onde k é o índice de nilpotência de L.

e) Considere-se agora uma matriz P_r cujas colunas são os vectores da cadeia de Jordan $I_{\mathbf{b}_r} = \{L^i\mathbf{x}_r, L^{i-1}\mathbf{x}_r, \ldots, L\mathbf{x}_r, \mathbf{x}_r\}$, pela ordem pela qual aparecem em $I_{\mathbf{b}_r}$. Isto é,

$$P_r = \begin{bmatrix} | & | & & | \\ L^i\mathbf{x}_r & L^{i-1}\mathbf{x}_r & \cdots & \mathbf{x}_r \\ | & | & & | \end{bmatrix}.$$

Multiplicando a matriz P_r por L, obtém-se

$$LP_r = \begin{bmatrix} | & | & & | \\ L^{i+1}\mathbf{x}_r & L^i\mathbf{x}_r & \cdots & L\mathbf{x}_r \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \mathbf{0} & L^i\mathbf{x}_r & \cdots & L\mathbf{x}_r \\ | & | & & | \end{bmatrix}, \quad (8.7)$$

visto que, sendo $\mathbf{b}_r = L^i\mathbf{x}_r$, resulta $L^{i+1}\mathbf{x}_r = L\mathbf{b}_r = \mathbf{0}$ (o vector $\mathbf{b}_r$ pertence ao núcleo de L). Reescrevendo a igualdade (8.7), temos

$$LP_r = \begin{bmatrix} | & | & & | \\ \mathbf{0} & L^i\mathbf{x}_r & \cdots & L\mathbf{x}_r \\ | & | & & | \end{bmatrix} = P_r \underbrace{\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}}_{\text{bloco de Jordan } \mathbf{N}_r} = P_r\mathbf{N}_r. \quad (8.8)$$

f) Pretendemos agora mostrar que a união $\mathcal{I} = I_{\mathbf{b}_1} \cup \cdots \cup I_{\mathbf{b}_t}$, de todas as cadeias de Jordan construídas a partir dos vectores da base B, é uma base de $\mathbb{C}^n$. Para tal, devemos mostrar que $\mathcal{I}$ tem n vectores linearmente independentes.
Para verificar que a cardinalidade de $\mathcal{I}$ é n precisamos do resultado, que enunciamos a seguir, que relaciona a característica de um produto de matrizes com a característica dos factores.

> Se A é uma matriz do tipo $m \times p$ e B uma matriz é do tipo $p \times n$, então
> $$\mathrm{car}(AB) = \mathrm{car}(B) - \dim EC(B) \cap N(A). \quad (8.9)$$
> Em particular, se A é uma matriz quadrada e $B = A^{j-1}$, com j um inteiro maior que 1, tem-se
> $$\mathrm{car}(A^j) = \mathrm{car}(A^{j-1}) - \dim EC(A^{j-1}) \cap N(A). \quad (8.10)$$

Deixamos como exercício a demonstração do resultado enunciado. Para o efeito, sugere-se que comece com uma base de $EC(B) \cap N(A) \subseteq EC(B)$ e a complete com os vectores $\mathbf{z}_1, \ldots, \mathbf{z}_p$, por forma a obter uma base para $EC(B)$. Mostre em seguida que $\{A\mathbf{z}_1, \ldots, A\mathbf{z}_p\}$ é uma base do espaço das colunas de AB, e portanto que $\dim EC(AB) = \operatorname{car}(AB) = p$.

g) Verifiquemos agora que $\mathcal{I} = I_{\mathbf{b}_1} \cup \cdots \cup I_{\mathbf{b}_t}$ tem n vectores. Usando a igualdade (8.10), a dimensão d_i de $M_i = EC(L^i) \cap N(L)$ é

$$d_i = \dim M_i = \dim EC(L^i) \cap N(L) = \operatorname{car}(L^i) - \operatorname{car}(L^{i+1}) = r_i - r_{i+1}.$$

Por definição de S_i, o número de vectores em cada S_i é igual a

$$\dim M_i - \dim M_{i+1} = d_i - d_{i+1} = \operatorname{car}(L^i) - 2\operatorname{car}(L^{i+1}) + \operatorname{car}(L^{i+2}).$$

Como qualquer cadeia de Jordan construída a partir de um vector de S_i tem $(i+1)$ vectores, o número total de vectores em $\mathcal{I} = I_{\mathbf{b}_1} \cup I_{\mathbf{b}_2} \cup \cdots \cup I_{\mathbf{b}_t}$, é

$$\begin{aligned}
\text{total} &= \sum_{i=0}^{k-1} (i+1)(\dim M_i - \dim M_{i+1}) = \sum_{i=0}^{k-1} (i+1)(d_i - d_{i+1}) \\
&= d_0 - d_1 + 2(d_1 - d_2) + 3(d_2 - d_3) + \cdots + k(d_{k-1} - d_k) \\
&= d_0 + d_1 + \cdots + d_{k-1} = (r_0 - r_1) + (r_1 - r_2) + \cdots + (r_{k-1} - r_k) \\
&= r_0 - r_k = r_0 = \operatorname{car} L^0 = n,
\end{aligned}$$

onde na última igualdade aplicámos o facto de $r_k = \operatorname{car} L^k = \operatorname{car} \mathbf{O} = 0$.

h) Para mostrar que $\mathcal{I} = I_{\mathbf{b}_1} \cup \cdots \cup I_{\mathbf{b}_t}$ é linearmente independente, considerem-se duas quaisquer cadeias de Jordan $I_{\mathbf{b}_r}, I_{\mathbf{b}_s}$

$$\begin{aligned}
I_{\mathbf{b}_r} &= \left\{\mathbf{x}_r, L\mathbf{x}_r, \ldots, L^i\mathbf{x}_r\right\}, \quad \text{com } L^i\mathbf{x}_r = \mathbf{b}_r, \\
I_{\mathbf{b}_s} &= \left\{\mathbf{x}_s, L\mathbf{x}_s, \ldots, L^j\mathbf{x}_s\right\}, \quad \text{com } L^j\mathbf{x}_s = \mathbf{b}_s,
\end{aligned}$$

e $j \geq i$. Como os vectores $\mathbf{b}_r = L^i\mathbf{x}_r$ e $\mathbf{b}_s = L^j\mathbf{x}_s$ pertencem ao núcleo de L, tem-se $L^{i+k}\mathbf{x}_r = L^{j+k}\mathbf{x}_s = \mathbf{0}$ para qualquer inteiro $k \geq 1$. Considere-se a combinação linear nula dos vectores de $I_{\mathbf{b}_i} \cup I_{\mathbf{b}_j}$,

$$\alpha_0\mathbf{x}_r + \alpha_1 L\mathbf{x}_r + \cdots + \alpha_i L^i\mathbf{x}_r + \beta_0\mathbf{x}_s + \beta_1 L\mathbf{x}_s + \cdots + \beta_j L^j\mathbf{x}_s = \mathbf{0}. \quad (8.11)$$

Multiplicando ambos os membros desta igualdade por L^j, e usando o facto de que $L^{i+k}\mathbf{x}_r = L^{j+k}\mathbf{x}_s = \mathbf{0}$ para $k \geq 1$, a igualdade (8.11) reduz-se a

$$\mathbf{0} = \alpha_0 L^j\mathbf{x}_r + \beta_0 L^j\mathbf{x}_s = \begin{cases} \alpha_0\mathbf{b}_r + \beta_0\mathbf{b}_s, & \text{se } i = j \\ \beta_0\mathbf{b}_s, & \text{se } j > i. \end{cases}$$

Como $\mathbf{b}_r$ e $\mathbf{b}_s$ pertencem a uma base de $N(L)$, obtemos: (a) $\alpha_0 = \beta_0 = 0$ se $i = j$; (b) $\beta_0 = 0$ se $j > i$.
No caso (a), multiplicando sucessivamente (8.11) por $L^{j-1}, L^{j-2}, \ldots, L$, conclui-se que todos os α e todos os β são nulos.
No caso (b), multiplicando sucessivamente (8.11) por $L^{j-1}, L^{j-2}, \ldots, L$, conclui-se que todos os β são nulos. De seguida, multiplicando sucessivamente (8.11) por $L^i, L^{i-1}, \ldots, L$, obtemos que todos os α também são nulos. Por conseguinte, $I_{\mathbf{b}_r} \cup I_{\mathbf{b}_s}$ é linearmente independente.

i) Falta verificar que a matriz P, cujas colunas são os vectores da base $\mathcal{I}$, satisfaz a igualdade $LP = PN$. Considere-se a matriz $P = [P_1|\cdots|P_t]$, onde as colunas de cada bloco P_r são os vectores de uma cadeia de Jordan $I_{\mathbf{b}_r}$ assim ordenados : $L^i\mathbf{x}_r, L^{i-1}\mathbf{x}_r, \ldots, L\mathbf{x}_r, \mathbf{x}_r$ (com $L^i\mathbf{x}_r = \mathbf{b}_r \in S_i$).
De (8.7) e (8.8), tem-se $LP_r = P_r\mathbf{N}_r$, e portanto

$$\begin{aligned} LP &= L[P_1|\cdots|P_t] = [LP_1|\cdots|LP_t] \\ &= [P_1\mathbf{N}_1|\cdots|P_t\mathbf{N}_t] = P\begin{bmatrix} \mathbf{N}_1 & \mathbf{O} & \cdots & \mathbf{O} \\ \mathbf{O} & \mathbf{N}_2 & \cdots & \mathbf{O} \\ \vdots & & \ddots & \vdots \\ \mathbf{O} & \mathbf{O} & \cdots & \mathbf{N}_t \end{bmatrix} = PN. \end{aligned}$$

Note-se que cada bloco P_r de P, associado à cadeia de Jordan $I_{\mathbf{b}_r}$ depende da escolha do vector $\mathbf{b}_r \in S_i$ bem como da escolha do vector $\mathbf{x}_r$ que resolve o sistema $L^i\mathbf{x}_i = \mathbf{b}_r$. Portanto, a matriz P não é única.

A demonstração anterior prova o teorema que passamos a enunciar.

Teorema 8.1. Forma canónica de Jordan para uma matriz nilpotente
Se L é uma matriz nilpotente de índice k e de ordem n, então L é semelhante a uma matriz diagonal por blocos N da forma

$$P^{-1}LP = N = \begin{bmatrix} \mathbf{N}_1 & \mathbf{O} & \cdots & \mathbf{O} \\ \mathbf{O} & \mathbf{N}_2 & \cdots & \mathbf{O} \\ \vdots & & \ddots & \vdots \\ \mathbf{O} & \mathbf{O} & \cdots & \mathbf{N}_t \end{bmatrix},$$

onde cada $\mathbf{N}_j$ é uma matriz nilpotente com as entradas na supradiagonal iguais a 1 e zeros nas outras entradas. Além disso,

- O número de blocos em N é $t = \dim N(L)$.
- A maior ordem dos blocos $\mathbf{N}_j$ de N é igual ao índice de nilpotência de L.
- O número de blocos em N do tipo $i \times i$ é igual a

$$\operatorname{car}(L^{i-1}) - 2\operatorname{car}(L^i) + \operatorname{car}(L^{i+1}). \tag{8.12}$$

- Seja $B = S_{k-1} \cup \cdots \cup S_0 = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_t\}$ uma base do núcleo de L, construída a partir dos subespaços $M_i = EC(L^i) \cap N(L)$, para $i = 0, \ldots, k-1$, da forma explicitada em (8.6). Então:
  - A união de todas as cadeias de Jordan, $\mathcal{I} = I_{\mathbf{b}_1} \cup I_{\mathbf{b}_2} \cup \cdots \cup I_{\mathbf{b}_t}$, é uma base de $\mathbb{C}^n$.
  - A matriz invertível $P = [P_1|\cdots|P_t]$, é constituída pelas matrizes P_r que têm por colunas as cadeias de Jordan $I_{\mathbf{b}_r}$. Os vectores coluna de P_r aparecem pela ordem $L^i\mathbf{x}_r, L^{i-1}\mathbf{x}_r, \ldots, L\mathbf{x}_r, \mathbf{x}_r$, sempre que $\mathbf{b}_r \in S_i$, e satisfazem $L^i\mathbf{x}_r = \mathbf{b}_r$.

A matriz N é designada por *forma canónica de Jordan* de L. A menos de uma reordenação de blocos, a matriz N é única. A matriz P não é única.

Como vimos, o resultado anterior assenta na construção de uma base especial para o núcleo da matriz nilpotente L a partir da qual se obtém uma base de $\mathbb{C}^n$ com características apropriadas. Na prática, se o único objectivo for determinar a forma de Jordan N de L, não é necessário determinar uma base de $\mathbb{C}^n$ com as propriedades referidas. Porém, se se pretender determinar uma matriz P que verifica $L = PNP^{-1}$, é conveniente construir a base das colunas de P seguindo a demonstração do Teorema 8.1.

Exemplo 8.3. Determinemos a forma canónica de Jordan para

$$L = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 0 & -1 \\ -1 & 0 & -1 \end{bmatrix}.$$

A matriz L é singular (possui uma coluna nula) e portanto $\lambda = 0$ é um valor próprio da matriz L. Calculando o polinómio característico de L, vemos que não existem valores próprios diferentes de zero, e portanto L é nilpotente (pela Proposição 8.1). O núcleo de L é $N(L) = \mathrm{Span}\{(0, 1, 0)\}$. Como o núcleo de L tem dimensão 1, a respectiva matriz de Jordan N tem apenas um bloco. Logo,

$$N = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Para determinar uma matriz P que satisfaça a igualdade $L = PNP^{-1}$, o Teorema 8.1 diz-nos que devemos construir uma certa base do núcleo de L, e para cada vector dessa base determinar cadeias de Jordan que vão constituir as colunas de P.

Neste caso, como a dimensão de $N(L)$ é 1, podemos usar outra estratégia de cálculo de P baseada na igualdade $L = PNP^{-1}$. Assim, seja P uma matriz que tem nas colunas os vectores $\mathbf{u}, \mathbf{v}_1, \mathbf{v}_2$, por esta ordem, e calcule-se LP e PN.

$$L = PNP^{-1} \iff LP = PN \iff L \begin{bmatrix} | & | & | \\ \mathbf{u} & \mathbf{v}_1 & \mathbf{v}_2 \\ | & | & | \end{bmatrix} = \begin{bmatrix} | & | & | \\ \mathbf{u} & \mathbf{v}_1 & \mathbf{v}_2 \\ | & | & | \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\iff \begin{bmatrix} | & | & | \\ L\mathbf{u} & L\mathbf{v}_1 & L\mathbf{v}_2 \\ | & | & | \end{bmatrix} = \begin{bmatrix} | & | & | \\ \mathbf{0} & \mathbf{u} & \mathbf{v}_1 \\ | & | & | \end{bmatrix}.$$

Ou seja, os vectores coluna de P verificam

$$L\mathbf{u} = \mathbf{0}, \quad L\mathbf{v_1} = \mathbf{u}, \quad L\mathbf{v_2} = \mathbf{v}_1. \tag{8.13}$$

Estas igualdades significam que $\mathbf{u}$ é um vector próprio de L associado ao valor próprio zero, e que os vectores (linearmente independentes) $\mathbf{v}_1, \mathbf{v}_2$ são obtidos a partir de $\mathbf{u}$ de modo a verificarem as igualdades (8.13). Assim, tomando $\mathbf{u} = (0, 1, 0) \in N(L)$, o sistema $L\mathbf{x} = \mathbf{u}$ tem solução geral $S = \{(1/3, t, -1/3) : \ t \in \mathbb{R}\}$. Tomando, por exemplo, $\mathbf{v}_1 = (1/3, 0, -1/3) \in S$ e resolvendo agora $L\mathbf{x} = \mathbf{v}_1$, resulta $\mathbf{v}_2 = (1/9, 0, 2/9)$ (por exemplo). Assim, uma matriz P que satisfaz a igualdade $L = PNP^{-1}$ é

$$P = \begin{bmatrix} 0 & 1/3 & 1/9 \\ 1 & 0 & 0 \\ 0 & -1/3 & 2/9 \end{bmatrix}.$$

♦

No exemplo anterior não foi necessário usar o resultado do Teorema 8.1 sobre a dimensão do(s) maior(es) bloco(s), mas apenas o resultado sobre o número de blocos em N. No entanto, como veremos no próximo exemplo, o resultado enunciado no Teorema 8.1 sobre as dimensões de cada bloco da matriz N não é despiciendo.

Exemplo 8.4. Considere-se a matriz

$$L = \begin{bmatrix} 3 & 3 & 2 & 1 \\ -2 & -1 & -1 & -1 \\ -1 & 1 & 0 & -1 \\ -5 & -4 & -3 & -2 \end{bmatrix}.$$

A matriz L tem determinante nulo uma vez que a quarta linha, L_4, é igual a $L_2 - L_1$. Portanto, zero é valor próprio de L. De facto, L é nilpotente uma vez que L^3 é a matriz nula. Se usarmos o método de eliminação de Gauss, concluímos que L tem característica 2, e portanto o valor próprio zero tem multiplicidade geométrica igual a 2 (já que, $\dim N(L) = 4 - \operatorname{car}(L) = 2$). Do Teorema 8.1, podemos já concluir que a forma de Jordan N de L tem dois blocos. Contrariamente ao que acontecia no exemplo anterior, como agora L é 4×4, o número de blocos não determina a forma de N. De facto, para uma matriz 4×4 podemos ter dois blocos 2×2, ou um bloco 3×3 e um bloco 1×1. Isto significa, que conhecer a multiplicidade geométrica do valor próprio zero de L não é suficiente para concluir qual é a forma de Jordan da matriz.

Para determinar N devemos ter em conta o índice de nilpotência de L. Como

$$L^2 = \begin{bmatrix} -4 & 4 & 0 & -4 \\ 2 & -2 & 0 & 2 \\ 0 & 0 & 0 & 0 \\ 6 & -6 & 0 & 6 \end{bmatrix},$$

e L^3 é a matriz nula, o índice de nilpotência de L é $k = 3$. Do Teorema 8.1, conclui-se que o maior bloco de N é 3×3. Por conseguinte, N tem dois blocos, $\mathbf{N}_1$ do tipo 3×3 e $\mathbf{N}_2$ do tipo 1×1. Ou seja,

$$N = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Pretendemos agora determinar uma matriz P tal que $L = PNP^{-1}$. Para tal, vamos seguir o método explicitado na demonstração do Teorema 8.1, começando por construir uma base B para o núcleo de L. Esta base é obtida completando bases S_i dos espaços $M_i = EC(L^i) \cap N(L)$, que verificam

$$\{\mathbf{0}\} = M_3 \subseteq M_2 \subseteq M_1 \subseteq M_0 = N(L).$$

A matriz L^2 tem característica igual a 1, as suas colunas são dependentes e geram um espaço de dimensão 1, ou seja, $\dim M_2 = 1$. Uma base S_2 para M_2 é $S_2 = \{(-4, 2, 0, 6)\}$. Neste caso, o espaço $M_1 = EC(L) \cap N(L)$ é igual a M_2 uma vez que, usando (8.10), a sua dimensão é

$$\dim M_1 = \operatorname{car}(L) - \operatorname{car}(L^2) = 2 - 1 = 1.$$

É fácil verificar que juntando a S_2 o vector $(-1, -1, 3, 0)$ se obtém uma base para $N(L) = M_0$. Ou seja, uma base do núcleo de L verificando as condições da base B em (8.6) é $B = S_2 \cup S_0 = \{(-4, 2, 0, 6), (-1, -1, 3, 0)\}$.

Construam-se agora as duas cadeias de Jordan associadas a cada vector da base B. Uma vez que o vector $\mathbf{b}_2 = (-4, 2, 0, 6)$ pertence a S_2, uma cadeia de Jordan $I_{\mathbf{b}_2}$ é formada por três vectores, enquanto que a cadeia de Jordan $I_{\mathbf{b}_0}$, associada a $\mathbf{b}_0 = (-1, -1, 3, 0)$, é constituída apenas pelo vector $\mathbf{b}_0$.

Determinando uma solução $\mathbf{x}_2$ do sistema $L^2\mathbf{x} = \mathbf{b}_2$ obtém-se, por exemplo, $\mathbf{x}_2 = (1, 0, 0, 0)$. Uma cadeia de Jordan $I_{\mathbf{b}_2}$ é

$$I_{\mathbf{b}_2} = \left\{L^2\mathbf{x}_2, L\mathbf{x}_2, \mathbf{x}_2\right\} = \{(-4, 2, 0, 6), (3, -2, -1, -5), (1, 0, 0, 0)\}.$$

Uma matriz P tal que $L = PNP^{-1}$, é a matriz que tem por colunas os vectores $L^2\mathbf{x}_2, L\mathbf{x}_2, \mathbf{x}_2, \mathbf{b}_0$, por esta ordem. Ou seja,

$$P = \begin{bmatrix} -4 & 3 & 1 & -1 \\ 2 & -2 & 0 & -1 \\ 0 & -1 & 0 & 3 \\ 6 & -5 & 0 & 0 \end{bmatrix}.$$

♦

Nota 51. *Salientamos que nem todas as bases do núcleo $N(L)$ permitem determinar cadeias de Jordan associadas aos seus vectores. No exemplo anterior, o conjunto $\{(-5,1,3,6),(-1,-1,3,0)\}$ é uma base do núcleo de L, e não existe nenhuma cadeia de Jordan com três vectores associada aos vectores desta base (o sistema $L^2\mathbf{x}=\mathbf{y}$ é impossível para $\mathbf{y}=(-5,1,3,6)$ e $\mathbf{y}=(-1,-1,3,0)$).*

Vimos que é essencial saber determinar bases para os espaços $M_i = EC(L^i)\cap N(L)$. No exercício seguinte, indica-se um método para o cálculo de uma base para a intersecção $EC(B)\cap N(L)$.

Exercício 8.1. Seja L uma matriz do tipo $p\times n$ e B uma matriz $n\times k$. Mostrar que uma base $\mathcal{B}$ para $EC(B)\cap N(L)$ se obtém do seguinte modo:

i) Determinar uma base $\{\mathbf{x}_1,\ldots,\mathbf{x}_s\}$ para $EC(B)$.
ii) Seja X a matriz cujas colunas são os vectores $\{\mathbf{x}_1,\ldots,\mathbf{x}_s\}$.
iii) Determinar uma base $\{\mathbf{u}_1,\ldots,\mathbf{u}_r\}$ para o núcleo de LX.
iv) Uma base para $EC(B)\cap N(L)$ é $\mathcal{B}=\{X\mathbf{u}_1,\ldots,X\mathbf{u}_r\}$.

Resolução: É fácil verificar que os vectores de $\mathcal{B}$ pertencem a $EC(B)\cap N(L)$. Com efeito, $X\mathbf{u}_j$ pertence ao espaço das colunas de X e ao núcleo de L (já que $LX\mathbf{u}_j=\mathbf{0}$). Como $EC(X)=EC(B)$, tem-se portanto que $X\mathbf{u}_j$ pertence $EC(B)\cap N(L)$.

Prove-se agora que $\mathcal{B}$ é uma base de $EC(B)\cap N(L)$. Seja Y a matriz que tem por colunas os vectores $\mathbf{u}_1,\ldots,\mathbf{u}_r$. Comecemos por mostrar que $\mathcal{B}$ é um conjunto linearmente independente. Para tal, basta mostrar que a matriz XY, do tipo $n\times r$, tem característica igual a r. Da expressão (8.9) resulta

$$\operatorname{car}(XY)=\operatorname{car}(Y)-\dim EC(Y)\cap N(X)=\operatorname{car}(Y)=r,$$

onde na última igualdade se aplicou o facto de $N(X)=\{\mathbf{0}\}$ (as colunas de X são linearmente independentes) e portanto $EC(Y)\cap N(X)=\{\mathbf{0}\}$.

Falta mostrar que $\dim EC(B)\cap N(L)=r$. Pelo Teorema da dimensão para matrizes, a igualdade (8.9) é equivalente a

$$\begin{aligned}\operatorname{car}(L_{p\times n}X_{n\times s}) &= \operatorname{car}(X)-\dim EC(X)\cap N(L)\iff\\ &\iff s-\dim N(LX)=s-\dim N(X)-\dim EC(X)\cap N(L).\end{aligned}$$

Por conseguinte,

$$\dim N(LX)=\dim N(X)+\dim EC(X)\cap N(L). \tag{8.14}$$

Como $EC(X)=EC(B)$ e $\dim N(X)=0$, da igualdade anterior resulta $\dim EC(B)\cap N(L)=\dim N(LX)=r$.

Em conclusão, $\mathcal{B}$ é uma base de $EC(B)\cap N(L)$. ▲

Exercício 8.2. Seja L uma matriz nilpotente de índice k.

Mostre que $M_{k-1} = EC(L^{k-1}) \cap N(L)$ é igual a $M_{k-1} = EC(L^{k-1})$.
Sugestão: Usar, por exemplo, a igualdade (8.14). ▲

No exemplo que se segue aplicamos o Exercício 8.1 na determinação da forma de Jordan de uma matriz A com um único valor próprio.

Exemplo 8.5. Considere-se a matriz

$$A = \begin{bmatrix} 4 & 5 & -7 & 6 & 1 \\ 0 & -4 & 12 & -8 & 0 \\ 0 & -1 & 6 & 0 & 0 \\ 0 & 5 & -7 & 10 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{bmatrix}.$$

É óbvio que $\lambda = 4$ é um valor próprio de A, e calculando o polinómio característico de A verifica-se que 4 é o único valor próprio. No Exemplo 8.2 vimos que a forma de Jordan de A se obtém da forma de Jordan da matriz nilpotente $L = A - 4I$. Determine-se a forma de Jordan para L. Esta matriz é

$$L = A - 4I = \begin{bmatrix} 0 & 5 & -7 & 6 & 1 \\ 0 & -8 & 12 & -8 & 0 \\ 0 & -1 & 2 & 0 & 0 \\ 0 & 5 & -7 & 6 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

É fácil ver que a ordem do maior menor principal de L não nulo é 3, pelo que a característica de L é 3 (ver Proposição 3.11, pág. 149). Portanto,

$$\dim N(L) = \dim N(A - 4I) = \dim E(4) = 5 - 3 = 2.$$

O valor próprio $\lambda = 0$ de L possui multiplicidade geométrica igual a 2. Logo, o Teorema 8.1 diz-nos que a forma de Jordan N para L tem dois blocos. Para determinar o tipo dos blocos de N calcule-se o índice de nilpotência de L. Como L^3 é a matriz nula e L^2 é não nula, tem-se que o índice de nilpotência de L é igual a 3. Portanto, a ordem do maior bloco de N é igual a 3. Sendo a matriz L do tipo 5×5, a única maneira de N possuir dois blocos em que um deles é do tipo 3×3 é o outro bloco ser do tipo 2×2. Isto é,

$$N = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Consequentemente, a matriz $(N + 4I) = J$ é a forma de Jordan para A (ver Exemplo 8.2), ou seja,

$$J = \begin{bmatrix} 4 & 1 & 0 & 0 & 0 \\ 0 & 4 & 1 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 4 & 1 \\ 0 & 0 & 0 & 0 & 4 \end{bmatrix}.$$

Determine-se agora uma matriz P tal que $LP = PN$. Para tal vamos seguir o método explicitado na demonstração do Teorema 8.1, começando por construir uma base B do núcleo de L como em (8.6). Sendo $k = 3$ o índice de nilpotência de L, os espaços $M_i = EC(L^i) \cap N(L)$ verificam

$$\{\mathbf{0}\} = M_3 \subseteq M_2 \subseteq M_1 \subseteq M_0 = N(L).$$

Do Exercício 8.2, tem-se $M_2 = EC(L^2)$. A matriz L^2 é

$$L^2 = \begin{bmatrix} 0 & -3 & 4 & -4 & 0 \\ 0 & 12 & -16 & 16 & 0 \\ 0 & 6 & -8 & 8 & 0 \\ 0 & -3 & 4 & -4 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Cada linha da matriz L^2 é múltipla das restantes, pelo que uma base para o espaço das colunas de L^2 é formada por uma qualquer coluna não nula de L^2. Assim, considere-se a base $S_2 = \{(-3, 12, 6, -3, 0)\}$ de $EC(L^2)$.

Uma cadeia de Jordan associada a S_2 tem três vectores. Como já conhecemos a forma de Jordan N, sabemos que existem duas cadeias de Jordan (uma por cada bloco de N) uma com três vectores e outra com dois vectores. A cadeia de Jordan com três vectores é construída a partir de um vector de S_2 e a outra tem de estar associada a um vector de S_1. Pelo que, sem cálculos adicionais, se conclui que $M_1 = M_0 = N(L)$.

Usemos agora o procedimento do Exercício 8.1 para construir uma base $B = S_2 \cup S_1$ do núcleo de L nas condições da base em (8.6). As colunas $3, 4$, e 5 de L formam uma base para o espaço das colunas de L. Seja X a matriz constituída por estas colunas. Então,

$$X = \begin{bmatrix} -7 & 6 & 1 \\ 12 & -8 & 0 \\ 2 & 0 & 0 \\ -7 & 6 & 0 \\ 0 & 0 & 0 \end{bmatrix} \Longrightarrow LX = \begin{bmatrix} 4 & -4 & 0 \\ -16 & 16 & 0 \\ -8 & 8 & 0 \\ 4 & -4 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Uma base para o núcleo de LX é formada pelos vectores $\mathbf{u}_1 = (1,1,0), \mathbf{u}_2 = (0,0,1)$. Logo, pelo Exercício 8.1, os vectores $X\mathbf{u}_1 = (1,0,0,0,0)$ e $X\mathbf{u}_2 = (-1,4,2,-1,0)$ formam uma base para $M_1 = EC(L) \cap N(L)$. Uma vez que $\mathbf{b}_2 = (-3,12,6,-3,0)$ e o vector $\mathbf{b}_1 = X\mathbf{u}_1 = (1,0,0,0,0)$ são linearmente independentes, podemos tomar $S_1 = \{\mathbf{b}_1\}$. Como $\dim N(L) = 2$, a base $B = S_2 \cup S_1$ é uma base para o núcleo de L verificando as condições da base B em (8.6).

Construam-se agora cadeias de Jordan associadas a $\mathbf{b}_2$ e $\mathbf{b}_1$, cadeias estas que têm, respectivamente dois e três vectores. Uma solução do sistema $L^2\mathbf{x} = \mathbf{b}_2$ é $\mathbf{x}_2 = (0,1,0,0,0)$, e portanto uma cadeia de Jordan associada a $\mathbf{b}_2$ é

$$I_{\mathbf{b}_2} = \left\{L^2\mathbf{x}_2, L\mathbf{x}_2, \mathbf{x}_2\right\} = \{(-3,12,6,-3,0),(5,-8,-1,5,0),(0,1,0,0,0)\}.$$

Uma solução de $L\mathbf{x} = \mathbf{b}_1$ é $\mathbf{x}_1 = (1,0,0,0,0)$. Logo,

$$I_{\mathbf{b}_1} = \{L\mathbf{x}_1, \mathbf{x}_1\} = \{(1,0,0,0,0),(0,0,0,0,1)\}.$$

Uma matriz P tal que $P^{-1}LP = N$ tem por colunas os vectores de $I_{\mathbf{b}_2}$ e $I_{\mathbf{b}_1}$ (pela ordem pela qual aparecem nesses conjuntos), isto é,

$$P = \begin{bmatrix} -3 & 5 & 0 & 1 & 0 \\ 12 & -8 & 1 & 0 & 0 \\ 6 & -1 & 0 & 0 & 0 \\ -3 & 5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

♦

8.2 Forma canónica de Jordan para matrizes quadradas

Nos exemplos 8.2 e 8.5, viu-se que a forma de Jordan de uma matriz A com um único valor próprio λ se obtém da forma de Jordan da matriz nilpotente $(A - \lambda I)$. Vamos agora mostrar que no caso geral do espectro $\sigma(A)$ ter cardinalidade superior a 1, a forma canónica de Jordan de A é obtida a partir das formas de Jordan das matrizes $(A - \lambda_i I)$, em que λ_i é um valor próprio de A.

Seja A um matriz do tipo $n \times n$, com espectro $\sigma(A) = \{\lambda_1, \lambda_2, \ldots, \lambda_s\}$. Pretendemos mostrar que A é semelhante a uma matriz diagonal por blocos J, tal que

$$P^{-1}AP = J = \begin{bmatrix} J(\lambda_1) & \mathbf{O} & \cdots & \mathbf{O} \\ \mathbf{O} & J(\lambda_2) & \cdots & \mathbf{O} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{O} & \mathbf{O} & \cdots & J(\lambda_s) \end{bmatrix}, \tag{8.15}$$

onde cada bloco $J(\lambda_i)$ de J é ainda uma matriz diagonal por blocos, cujos blocos $J_k(\lambda_i)$ são da forma

$$J_k(\lambda_i) = \begin{bmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix},$$

onde as entradas omissas são zeros.

Em primeiro lugar, façamos algumas observações sobre o significado de uma matriz ser semelhante a uma matriz diagonal por blocos.

Como se viu no Capítulo 6, qualquer matriz R, de ordem n, pode ser vista como a representação matricial de uma função linear $T : W \to W$, com $\dim W = n$, em relação a uma base $\mathcal{B}$ fixada em W. Seja $\mathcal{B} = (q_1, \dots, q_r, q_{r+1}, \dots, q_n)$ uma base de W e suponha-se $R = M(T, \mathcal{B}, \mathcal{B})$ tem forma

$$R = M(T, \mathcal{B}, \mathcal{B}) = \begin{bmatrix} X_{r\times r} & Y_{r\times(n-r)} \\ \mathbf{O} & Z_{(n-r)\times(n-r)} \end{bmatrix}.$$

Da definição de matriz que representa uma função linear, tem-se

$$\begin{bmatrix} X_{r\times r} & Y_{r\times(n-r)} \\ \mathbf{O} & Z_{(n-r)\times(n-r)} \end{bmatrix} = \begin{bmatrix} | & & | & | & & | \\ (\mathbf{T}(q_1))_{\mathcal{B}} & \cdots & (\mathbf{T}(q_r))_{\mathcal{B}} & (\mathbf{T}(q_{r+1}))_{\mathcal{B}} & \cdots & (\mathbf{T}(q_n))_{\mathcal{B}} \\ | & & | & | & & | \end{bmatrix}.$$

Da igualdade anterior, conclui-se que as imagens por T dos primeiros r vectores da base $\mathcal{B}$ pertencem ao espaço gerado por esses vectores. Ou seja,

$$T(q_i) \in \operatorname{Span}\{q_1, \dots, q_r\}, \quad \text{para } i = 1, \dots, r.$$

Como T é uma função linear e $\{q_1, \dots, q_r\}$ é uma base de $U = \operatorname{Span}\{q_1, \dots, q_r\}$, resulta que U é um subespaço invariante por T, isto é, $T(U) \subseteq U$. No caso da matriz Y ser igual à matriz nula, conclui-se igualmente que os últimos $n - r$ vectores de $\mathcal{B}$ geram um subespaço V invariante por T, e portanto que $W = U \oplus V$, onde U e V são subespaços invariantes por T.

Matrizes que representam uma certa função linear $T : W \to W$ em relação a bases distintas fixadas em W são matrizes semelhantes. Nomeadamente, se $R = M(T, \mathcal{B}, \mathcal{B})$ e $S = M(T, \mathcal{B}', \mathcal{B}')$, então existe uma matriz invertível Q tal que $Q^{-1}SQ = R$. A matriz Q é a matriz que realiza a mudança da base $\mathcal{B}$ para $\mathcal{B}'$, ou seja, $Q = M_{\mathcal{B} \leftarrow \mathcal{B}'}$ (cf. Proposição 6.8, pág 346).

Suponha-se agora que S é semelhante a uma matriz R diagonal por blocos, isto é,

$$Q^{-1}SQ = R = \begin{bmatrix} R_1 & & & \\ & R_2 & & \\ & & \ddots & \\ & & & R_l \end{bmatrix}, \tag{8.16}$$

onde as entradas omissas são matrizes nulas. Considere-se a partição $Q = [Q_1|\cdots|Q_l]$, em que o número de colunas de cada bloco Q_i é igual à ordem de R_i. A relação (8.16) significa que a matriz R representa T em relação a uma base $\mathcal{B} = B_{U_1} \cup \ldots \cup B_{U_l}$, onde B_{U_i} é uma base de um subespaço invariante por T. As colunas de cada bloco Q_i de Q geram o subespaço U_i invariante por T (ou equivalentemente, o subespaço U_i invariante pela matriz R).

Em resumo,

Uma matriz S, de ordem n, é semelhante a uma matriz diagonal por blocos R, como em (8.16), se e só se existe uma matriz invertível $Q = [Q_1|\cdots|Q_l]$, tal que as colunas de cada Q_i geram um subespaço invariante por R.

Exemplos de subespaços invariantes por uma matriz R são o núcleo e o espaço das colunas de uma qualquer potência positiva de R, já que

$$\begin{aligned} \mathbf{y} \in EC(R^p) \Longleftrightarrow \mathbf{y} = R^p\mathbf{x} \quad \text{para algum } \mathbf{x} &\Longrightarrow R\mathbf{y} = R^{p+1}\mathbf{x} = R^p(R\mathbf{x}) \\ &\Longleftrightarrow R\mathbf{y} \in EC(R^p). \\ \mathbf{x} \in N(R^p) \Longleftrightarrow R^p\mathbf{x} = \mathbf{0} &\Longrightarrow R^{p+1}\mathbf{x} = \mathbf{0} \Longleftrightarrow R^p(R\mathbf{x}) = \mathbf{0} \\ &\Longrightarrow R\mathbf{x} \in N(R^p). \end{aligned}$$

Mostrar que uma matriz A admite uma forma de Jordan como em (8.15) passa pela prova de que $\mathbb{C}^n$ se pode escrever como a soma directa de subespaços invariantes por A.

Começamos por mostrar que uma matriz quadrada A não invertível é semelhante a uma matriz com dois blocos na diagonal, em que um desses blocos é nilpotente e o outro é invertível (ver Proposição 8.2 abaixo).

Seja R uma matriz de ordem n e considerem-se os núcleos e os espaços das colunas das potências não negativas de R. Como,

$$N(R^p) \subseteq N(R^{p+1}), \qquad EC(R^p) \supseteq EC(R^{p+1}),$$

tem-se

$$N(R) \subseteq N(R^2) \subseteq \cdots \subseteq N(R^{p-1}) \subseteq N(R^p) \subseteq N(R^{p+1}) \subseteq \cdots$$

e

$$EC(R) \supseteq EC(R^2) \supseteq \cdots \supseteq EC(R^{p-1}) \supseteq EC(R^p) \supseteq EC(R^{p+1}) \supseteq \cdots . \tag{8.17}$$

Estas inclusões dizem-nos que a dimensão de $N(R^p)$ aumenta com p e que a dimensão de $EC(R^p)$ diminui com p. No entanto, tem de existir necessariamente uma potência de R a partir da qual $N(R^p) = N(R^{p+1})$, caso contrário existiria uma potência de R para a qual a dimensão do seu núcleo seria maior do que a ordem de R, o que é impossível. Seja k o menor inteiro positivo tal que $N(R^k) = N(R^{k+1})$. Como $\dim EC(R^k) = \operatorname{car}(R^k) = n - \dim N(R^k)$, temos $EC(R^k) = EC(R^p)$ para $p \geq k$.

Para referência futura, passamos a definir índice de uma matriz.

Definição 8.2. Índice de uma matriz

O índice de uma matriz quadrada A é o menor inteiro não negativo k tal que

$$N(A^k) = N(A^{k+1}), \quad \text{ou equivalentemente, tal que} \quad EC(A^k) = EC(A^{k+1}).$$

Se A é invertível o seu índice é igual zero.

Mostramos de seguida que matrizes singulares de índice k são semelhantes a matrizes diagonais por blocos, com um bloco nipotente e um bloco invertível.

Proposição 8.2. Seja A é uma matriz de ordem n singular, de índice k, e tal que $\operatorname{car}(A^k) = s$.

Existe uma matriz invertível Q tal que

$$Q^{-1}AQ = \begin{bmatrix} L & \mathbf{O} \\ \mathbf{O} & C_{s\times s} \end{bmatrix}, \tag{8.18}$$

onde L é uma matriz nilpotente de índice k, e C é uma matriz invertível do tipo $s \times s$. As primeiras $(n-s)$ colunas de Q formam uma base do núcleo $N(A^k)$ e as últimas s colunas de Q uma base para o espaço das colunas $EC(A^k)$.

Passamos a designar a matriz L em (8.18) por *parte nilpotente* de A.

Demonstração. Já vimos que $N(A^k)$ e $EC(A^k)$ são subespaços invariantes por A, e pelo Teorema da dimensão para matrizes, a soma das suas dimensões é igual a n. Para provar que estes subespaços são complementares, ou seja, que $\mathbb{C}^n = N(A^k) \oplus EC(A^k)$, falta mostrar que $N(A^k) \cap EC(A^k) = \{\mathbf{0}\}$. Seja $\mathbf{x} \in N(A^k) \cap EC(A^k)$, ou seja,

$$A^k\mathbf{x} = \mathbf{0}, \quad \text{e} \quad \mathbf{x} = A^k\mathbf{y} \quad \text{para algum } \mathbf{y} \in \mathbb{C}^n.$$

Então,

$$A^k\mathbf{x} = A^{2k}\mathbf{y} = \mathbf{0} \Longrightarrow \mathbf{y} \in N(A^{2k}) = N(A^k) \Longrightarrow \mathbf{x} = \mathbf{0}.$$

Considere-se a matriz em blocos $Q = [X|Y]$ tal que as colunas do bloco X e do bloco Y constituem, respectivamente, uma base de $N(A^k)$ e de $EC(A^k)$. Como $N(A^k)$ e $EC(A^k)$ são subespaços invariantes por A, a matriz $Q^{-1}AQ$ é uma matriz diagonal por blocos, com dois blocos L e C na diagonal de ordens, respectivamente $(n-s)$ e s, uma vez que $\dim EC(A^k) = \operatorname{car}(A^k) = s$.

Mostremos agora que a matriz L é nilpotente e que C é invertível. Para tal considere-se que Q^{-1} está particionada na forma $Q^{-1} = \left[\begin{array}{c} Z \\ \hline W \end{array}\right]$ em que o número de linhas de Z é $(n-s)$. Como $(Q^{-1}AQ)^k = Q^{-1}A^kQ$, tem-se

$$\begin{aligned} \begin{bmatrix} L^k & \mathbf{O} \\ \mathbf{O} & C^k \end{bmatrix} &= Q^{-1}A^kQ = \left[\begin{array}{c} Z \\ \hline W \end{array}\right] A^k \left[X|Y\right] = \left[\begin{array}{c} Z \\ \hline W \end{array}\right] \left[A^kX|A^kY\right] \\ &= \begin{bmatrix} ZA^kX & ZA^kY \\ WA^kX & WA^kY \end{bmatrix} = \begin{bmatrix} \mathbf{O} & ZA^kY \\ \mathbf{O} & WA^kY \end{bmatrix}, \end{aligned}$$

onde a última igualdade resulta do facto da colunas de X formarem uma base do núcleo de A^k. Da igualdade anterior, obtém-se $L^k = \mathbf{O}$. Além disso, como matrizes semelhantes têm a mesma característica, tem-se

$$s = \operatorname{car}(A^k) = \operatorname{car}(Q^{-1}A^kQ) = \operatorname{car}(C^k).$$

Sendo C^k uma matriz de ordem s e $\operatorname{car}(C^k) = s$, conclui-se que C^k é invertível e portanto C também o é.

Para verificar que L é nilpotente de índice k, falta mostrar que $L^{k-1} \neq \mathbf{O}$. Suponha-se, por absurdo, que o índice de L não é k, ou seja, que $L^{k-1} = \mathbf{O}$.

Então,

$$\begin{aligned}\operatorname{car}(A^{k-1}) &= \operatorname{car}(Q^{-1}A^{k-1}Q) = \operatorname{car}\begin{bmatrix} L^{k-1} & \mathbf{O} \\ \mathbf{O} & C^{k-1}\end{bmatrix} = \operatorname{car}\begin{bmatrix} \mathbf{O} & \mathbf{O} \\ \mathbf{O} & C^{k-1}\end{bmatrix} \\ &= \operatorname{car}(C^{k-1}) = s = \operatorname{car}(A^k),\end{aligned}$$

o que é impossível, já que por definição de índice de A se tem $\operatorname{car}(A^{k-1}) < \operatorname{car}(A^k)$. Portanto, $L^{k-1} \neq \mathbf{O}$. □

Aplicamos agora a proposição anterior e o Teorema da forma de Jordan para matrizes nilpotentes, para obter a forma de Jordan de uma matriz quadrada A.

Teorema 8.2. Forma canónica de Jordan

Para toda a matriz A de ordem n, com espectro $\sigma(A) = \{\lambda_1, \lambda_2, \ldots, \lambda_s\}$, existe uma matriz invertível P tal que

$$P^{-1}AP = J = \begin{bmatrix} J(\lambda_1) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & J(\lambda_2) & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & J(\lambda_s) \end{bmatrix}. \tag{8.19}$$

- J possui um bloco $J(\lambda_i)$ para cada valor próprio $\lambda_i \in \sigma(A)$.
- Cada bloco $J(\lambda_i)$ tem $t_i = \dim N(A - \lambda_i I)$ blocos de Jordan

$$J(\lambda_i) = \begin{bmatrix} J_1(\lambda_i) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & J_2(\lambda_i) & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & J_{t_i}(\lambda_i) \end{bmatrix},$$

  com

$$J_k(\lambda_i) = \begin{bmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix}$$

- A ordem dos maiores blocos em $J(\lambda_i)$ é k_i, onde k_i é o índice de $(A - \lambda_i I)$.
- O número de blocos de Jordan em $J(\lambda_i)$ do tipo $j \times j$, é dado por

$$r_{j-1}(\lambda_i) - 2r_j(\lambda_i) + r_{j+1}(\lambda_i),$$

  onde $r_j(\lambda_i) = \operatorname{car}(A - \lambda_i I)^j$.

A matriz J é designada por *forma canónica de Jordan* para A. A menos de uma reordenação de blocos a matriz J é única.

Demonstração. Se λ_i é um valor próprio de A, a matriz $(A - \lambda_i I)$ é singular. A Proposição 8.2 garante que $(A - \lambda_i I)$ é semelhante a uma matriz diagonal por blocos, sendo um dos blocos na diagonal uma matriz nilpotente L_i e o outro uma matriz invertível C_i. Isto é, existe uma matriz invertível X_i tal que

$$X_i^{-1}(A - \lambda_i I)X_i = \begin{bmatrix} L_i & \mathbf{O} \\ \mathbf{O} & C_i \end{bmatrix}. \tag{8.20}$$

Uma vez que L_i é nilpotente, pelo Teorema 8.1, existe uma matriz invertível Y_i tal que

$$Y_i^{-1} L_i Y_i = N(\lambda_i) = \begin{bmatrix} \mathbf{N}_1(\lambda_i) & & & \\ & \mathbf{N}_2(\lambda_i) & & \\ & & \ddots & \\ & & & \mathbf{N}_{t_i}(\lambda_i) \end{bmatrix}, \tag{8.21}$$

onde:

- Cada bloco $\mathbf{N}_j(\lambda_i)$ é uma matriz triangular superior com todas as entradas nulas excepto as da supradiagonal que são todas iguais a 1.
- Existem t_i blocos em $N(\lambda_i)$ onde $t_i = \dim N(L_i) = \dim N(A - \lambda_i I)$.
- O número de blocos do tipo $j \times j$ em $N(\lambda_i)$ é dado por (8.12):

$$\operatorname{car}(A - \lambda_i I)^{j-1} - 2\operatorname{car}(A - \lambda_i I)^{j} + \operatorname{car}(A - \lambda_i I)^{j+1}.$$

Mostramos agora que a matriz A é semelhante a uma matriz diagonal por blocos, com dois blocos na diagonal, um bloco que é $J(\lambda_i)$ e o outro bloco não tem λ_i no seu espectro.

Para tal, considere-se a matriz $Q_i = X_i \begin{bmatrix} Y_i & \mathbf{O} \\ \mathbf{O} & I \end{bmatrix}$. Esta matriz é invertível uma vez que X_i e Y_i são invertíveis. Portanto,

$$\begin{aligned} Q_i^{-1}(A - \lambda_i I)Q_i &= \begin{bmatrix} Y_i^{-1} & \mathbf{O} \\ \mathbf{O} & I \end{bmatrix} X_i^{-1}(A - \lambda_i I)X_i \begin{bmatrix} Y_i & \mathbf{O} \\ \mathbf{O} & I \end{bmatrix} \\ &= \begin{bmatrix} Y_i^{-1} & \mathbf{O} \\ \mathbf{O} & I \end{bmatrix} \begin{bmatrix} L_i & \mathbf{O} \\ \mathbf{O} & C_i \end{bmatrix} \begin{bmatrix} Y_i & \mathbf{O} \\ \mathbf{O} & I \end{bmatrix} \qquad \text{(por (8.20))} \\ &= \begin{bmatrix} N_i & \mathbf{O} \\ \mathbf{O} & C_i \end{bmatrix}. \end{aligned}$$

Equivalentemente,

$$Q_i^{-1} A Q_i = \begin{bmatrix} N_i + \lambda_i I & \mathbf{O} \\ \mathbf{O} & C_i + \lambda_i I \end{bmatrix} = \begin{bmatrix} J(\lambda_i) & \mathbf{O} \\ \mathbf{O} & \mathbf{A}_i \end{bmatrix}, \tag{8.22}$$

com $\mathbf{A}_i = C_i + \lambda_i I$. De (8.21), segue que o bloco $J(\lambda_i)$ é da forma

$$J(\lambda_i) = N(\lambda_i) + \lambda_i I = \begin{bmatrix} \mathbf{N}_1(\lambda_i) + \lambda_i I & \mathbf{O} & \cdots & \mathbf{O} \\ \mathbf{O} & \mathbf{N}_2(\lambda_i) + \lambda_i I & \cdots & \mathbf{O} \\ \vdots & & \ddots & \vdots \\ \mathbf{O} & \mathbf{O} & \cdots & \mathbf{N}_{t_i}(\lambda_i) + \lambda_i I \end{bmatrix}.$$

Determinemos agora o espectro de $C_i + \lambda_i I = \mathbf{A}_i$. Para tal, note-se que se $\sigma(X) = \{\mu_1, \ldots, \mu_r\}$ é o espectro de uma qualquer matriz X, então o espectro da matriz $(X + \theta I)$ é $\sigma(X + \theta I) = \{\mu_1 + \theta, \ldots, \mu_r + \theta\}$, uma vez que $\det((X + \theta I) - \lambda I) = \det(X - (\lambda - \theta)I)$.

Como matrizes semelhantes têm os mesmos valores próprios e a matriz C_i em (8.20) é invertível, temos que λ_i não pertence ao espectro de C_i (já que $0 \notin \sigma(C_i)$). Da igualdade (8.20) segue ainda que o espectro de C_i é

$$\sigma(C_i) = \{(\lambda_1 - \lambda_i), \ldots, (\lambda_{i-1} - \lambda_i), (\lambda_{i+1} - \lambda_i), \ldots, (\lambda_s - \lambda_i)\}.$$

Portanto $\sigma(C_i + \lambda_i I) = \sigma(\mathbf{A}_i) = \{\lambda_1, \ldots, \lambda_{i-1}, \lambda_{i+1}, \ldots, \lambda_s\} = \sigma(A) \setminus \{\lambda_i\}$.

Escolha-se outro valor próprio λ_j do espectro de A, distinto de λ_i. Repetindo o procedimento anterior agora para a matriz $(\mathbf{A}_i - \lambda_j I)$ em (8.22), obtemos uma matriz invertível Q_j tal que

$$Q_j^{-1}(\mathbf{A}_i - \lambda_j I)Q_j = \begin{bmatrix} J(\lambda_j) & \mathbf{O} \\ \mathbf{O} & C_j + \lambda_j I \end{bmatrix} = \begin{bmatrix} J(\lambda_j) & \mathbf{O} \\ \mathbf{O} & \mathbf{A}_j \end{bmatrix},$$

onde o espectro de $\mathbf{A}_j$ é $\sigma(\mathbf{A}_j) = \sigma(A) \setminus \{\lambda_i, \lambda_j\}$, e o bloco $J(\lambda_j)$ tem as propriedades acima indicadas para o bloco $J(\lambda_i)$ com λ_i substituído por λ_j.

Tomando a matriz invertível $S_j = Q_i \begin{bmatrix} I & \mathbf{O} \\ \mathbf{O} & Q_j \end{bmatrix}$, de (8.22) resulta

$$S_j^{-1} A S_j = \begin{bmatrix} J(\lambda_i) & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & J(\lambda_j) & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \mathbf{A}_j \end{bmatrix} \quad \text{com } \sigma(\mathbf{A}_j) = \sigma(A) \setminus \{\lambda_i, \lambda_j\}.$$

Continuando este processo até esgotar os valores próprios de A, conclui-se que A é semelhante a uma matriz diagonal por blocos, em que cada bloco é da forma $J(\lambda_i) = (N_i + \lambda_i I)$, onde N_i é forma de Jordan para a parte nilpotente de $(A - \lambda_i I)$.

A unicidade de J (a menos de um reordenação dos blocos) segue como consequência da unicidade da forma de Jordan de cada uma das matrizes nilpotentes $N(\lambda_i)$. $\square$

Para referência futura, definimos *índice de um valor próprio* λ da matriz A como sendo o índice da matriz $(A - \lambda I)$.

Definição 8.3. Seja A uma matriz quadrada e λ um valor próprio de A.
O *índice* de λ é o índice da matriz $(A - \lambda I)$. Ou seja, o índice de λ é o menor inteiro positivo k tal que $N(A - \lambda I)^k = N(A - \lambda I)^{k+1}$.

Como veremos no próximo exemplo, a forma canónica de Jordan de uma matriz A possui toda a informação sobre os valores próprios de A.

Exemplo 8.6. Suponhamos que A é uma matriz com a seguinte forma de Jordan:

$$J = \begin{bmatrix} \begin{array}{ccc|} 5 & 1 & 0 \\ 0 & 5 & 1 \\ 0 & 0 & 5 \\ \hline \end{array} & & & & \\ & \boxed{\begin{matrix} 5 & 1 \\ 0 & 5 \end{matrix}} & & & \\ & & \boxed{\begin{matrix} 2 & 1 \\ 0 & 2 \end{matrix}} & & \\ & & & \boxed{3} & \\ & & & & \boxed{3} \end{bmatrix} = \begin{bmatrix} J(5) & & \\ & J(2) & \\ & & J(3) \end{bmatrix}.$$

A matriz A é do tipo 9×9 e o seu espectro é $\sigma(A) = \{5, 2, 3\}$. As multiplicidades algébricas ($mult\ alg$) e geométricas ($mult\ geom$) destes valores próprios, bem como o índice (ind) de cada valor próprio, obtêm-se da forma de J. Assim,

- A multiplicidade algébrica do valor próprio λ é dada pelo tamanho do bloco $J(\lambda)$ em J,

$$mult\,alg(5) = 5, \quad mult\,alg(2) = 2 \quad \text{e} \quad mult\,alg(3) = 2.$$

- A multiplicidade geométrica do valor próprio λ, ou seja, $\dim N(A - \lambda I)$, é dada pelo número de blocos de $J(\lambda)$,

$$mult\,geom(5) = 2, \quad mult\,geom(2) = 1 \quad \text{e} \quad mult\,geom(3) = 2.$$

- O índice do valor próprio λ é a maior ordem dos blocos na diagonal de $J(\lambda)$,

$$ind(5) = 3, \quad ind(2) = 2 \quad \text{e} \quad ind(3) = 1.$$

- O valor próprio 3 tem índice 1, ou seja, 3 é um valor próprio semi-simples de A. Recorde-se que um valor próprio λ é semi-simples se as suas multiplicidades algébrica e geométrica são iguais. Neste caso, a restrição de A ao núcleo de $(A - \lambda I)$ é diagonalizável embora a matriz A possa não o ser.

♦

Uma matriz invertível P tal que $P^{-1}AP = J$, onde $J = \operatorname{diag}(J(\lambda_1), \ldots, J(\lambda_s))$ é a forma de Jordan de A, é obtida de forma análoga ao caso de matrizes nilpotentes, construindo-se agora cadeias de Jordan associadas a vectores próprios de A. Mais explicitamente, para cada valor próprio λ_i de A de índice k_i e de multiplicidade geométrica t_i, constrói-se uma base $B_i = S_{k_i-1} \cup \cdots \cup S_1 \cup S_0$ para o núcleo $(A - \lambda_i I)$ da seguinte forma:

Determinam-se conjuntos S_i tais que

- S_{k_i-1} é uma base para $\mathcal{M}_{k_i-1} = EC(A - \lambda_i I)^{k_i-1} \cap N(A - \lambda_i I)$.
- $S_{k_i-1} \cup S_{k_i-2}$ é uma base para $\mathcal{M}_{k_i-2} = EC(A - \lambda_i I)^{k_i-2} \cap N(A - \lambda_i I)$.
- $\vdots$
- $B_i = S_{k_i-1} \cup \cdots \cup S_1 \cup S_0 = \{\mathbf{b}_1, \ldots, \mathbf{b}_{t_i}\}$ é uma base para $N(A - \lambda_i I)$.

Seguidamente, para cada vector $\mathbf{b}_r \in B_i$, constrói-se uma cadeia de Jordan $I_{\mathbf{b}_r}$ da seguinte forma:

- Se $\mathbf{b}_r \in S_j$, determina-se uma solução $\mathbf{x}_r$ do sistema $(A - \lambda_i I)^j \mathbf{x} = \mathbf{b}_r$. A cadeia $I_{\mathbf{b}_r}$ é constituída pelos vectores $(A - \lambda_i I)^j \mathbf{x}_r, (A - \lambda_i I)^{j-1} \mathbf{x}_r, \ldots, \mathbf{x}_r$. Os vectores de $I_{\mathbf{b}_r}$ são: um vector próprio de A, $\mathbf{b}_r = (A - \lambda_i I)^j \mathbf{x}_r$, e os restantes vectores (associados ao vector próprio $\mathbf{b}_r$) são designados por *vectores próprios generalizados*, de ordem j.

- Os vectores $\mathbf{b}_r = (A - \lambda_i I)^j \mathbf{x}_r$ $\mathbf{y}_1 = (A - \lambda_i I)^{j-1} \mathbf{x}_r$, ..., $\mathbf{y}_j = \mathbf{x}_r$, da cadeia $I_{\mathbf{b}_r}$, verificam:

$$\begin{array}{ll}
(A - \lambda_i I)\mathbf{b}_r = \mathbf{0} & \Longleftrightarrow \mathbf{b}_r \text{ é um valor próprio de } A \\
(A - \lambda_i I)\mathbf{y}_1 = \mathbf{b}_r & \Longrightarrow (A - \lambda_i I)^2 \mathbf{y}_1 = \mathbf{0} \\
(A - \lambda_i I)\mathbf{y}_2 = \mathbf{y}_1 & \Longrightarrow (A - \lambda_i I)^3 \mathbf{y}_2 = \mathbf{0} \\
\quad\vdots & \quad\vdots \\
(A - \lambda_i I)\mathbf{y}_j = \mathbf{y}_{j-1} & \Longrightarrow (A - \lambda_i I)^{j+1} \mathbf{y}_j = \mathbf{0}.
\end{array}$$

  Ou seja,

$$\mathbf{y}_j \in N(A - \lambda_i I)^{j+1} \quad \text{e} \quad \mathbf{y}_j \notin N(A - \lambda_i I)^j. \tag{8.23}$$

- Sendo $\tilde{P}_r$ a matriz cujas colunas são os vectores $\mathbf{b}_r, \mathbf{y}_1, \ldots, \mathbf{y}_j$, por esta ordem, tem-se $A\tilde{P}_r = \tilde{P}_r J_r(\lambda_i)$, com

$$J_j(\lambda_i) = \begin{bmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix}.$$

- O conjunto $\mathcal{I}_i = I_{\mathbf{b}_1} \cup \cdots \cup I_{\mathbf{b}_{t_i}}$ é uma base de um subespaço de $\mathbb{C}^n$ de dimensão p, onde p é a multiplicidade algébrica de λ_i. Além disso, a matriz P_i que tem para colunas os vectores de $\mathcal{I}_i$, ordenados da forma acima indicada, é tal que

$$AP_i = P_i J(\lambda_i) = P_i \begin{bmatrix} J_1(\lambda_i) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & J_2(\lambda_i) & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & J_{t_i}(\lambda_i) \end{bmatrix}.$$

- A matriz P é uma matriz em blocos $P = [P_1|P_2|\cdots|P_s]$ em que cada P_i é a matriz definida no item anterior.

No exemplo seguinte determinamos a forma de Jordan J de uma matriz A bem como uma matriz P tal que $P^{-1}AP = J$.

Exemplo 8.7. Determinemos a forma canónica de Jordan da matriz

$$A = \begin{bmatrix} 5 & 2 & -3 & 2 & 0 & 0 \\ -1 & -6 & 16 & -11 & 0 & 0 \\ 0 & -4 & 11 & -4 & 0 & 0 \\ 1 & 5 & -7 & 10 & 0 & 0 \\ 1 & 3 & -4 & 3 & 5 & 4 \\ 0 & 3 & -4 & 4 & -1 & 1 \end{bmatrix}.$$

O espectro de A é $\sigma(A) = \{3, 5\}$. O valores próprios 3 e 5 têm, respectivamente multiplicidade algébrica 2 e 4. A forma de Jordan de A possui dois blocos $J(3)$ e $J(5)$, ou seja, A é semelhante à matriz

$$J = \begin{bmatrix} J(5) & \mathbf{O} \\ \mathbf{O} & J(3) \end{bmatrix}, \tag{8.24}$$

onde $J(5)$ é 4×4 e $J(3)$ é 2×2. Para determinar $J(5)$ e $J(3)$, começamos por determinar a multiplicidade geométrica dos valores próprios de A. Resolvendo os sistemas $(A - \lambda_i I)\mathbf{x} = \mathbf{0}$, conclui-se que

$$E(5) = \left\{(a,b,c,d,e,f) \in \mathbb{R}^5 : a = d - e, b = -4d + 3e, c = -2d + 2e, f = 0\right\}.$$

$$E(3) = \left\{(a,b,c,d,e) \in \mathbb{R}^4 : a = b = c = d = 0, e = -2f\right\}.$$

Ou seja, $\dim E(5) = 2$ e $\dim E(3) = 1$. Portanto, $J(5)$ é formado por dois blocos e $J(3)$ por um bloco. Dado que $J(5)$ é 4×4, a multiplicidade geométrica do valor próprio 5 não permite concluir se serão dois blocos 2×2, ou um bloco 3×3 e um bloco 1×1. No entanto podemos já concluir que $J(3)$ é a matriz

$$J(3) = \begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}.$$

Calculando as potências $(A - 5I)^2$ e $(A - 5I)^3$ verificamos que o índice de $\lambda = 5$ é igual a dois, uma vez que $N(A - 5I)^2 = N(A - 5I)^3$. Por conseguinte, como $J(5)$ é formado por dois blocos e o maior bloco é de ordem 2, tem-se

$$J(5) = \begin{bmatrix} 5 & 1 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 0 & 0 & 5 & 1 \\ 0 & 0 & 0 & 5 \end{bmatrix}.$$

Determinemos agora uma matriz P tal que $P^{-1}AP = J$, em que J é da forma (8.24). A quinta e sexta colunas de P são, respectivamente, um vector próprio $\mathbf{u}$ associado ao valor próprio 3, e um vector próprio generalizado determinado a partir de $\mathbf{u}$, resolvendo o sistema $(A-3I)\mathbf{x} = \mathbf{u}$. Tome-se para $\mathbf{u} \in E(3)$ o vector $\mathbf{u} = (0,0,0,0,-2,1)$. Um vector próprio generalizado é $\mathbf{v} = (0,0,0,0,-1,0)$, como facilmente se verifica calculando $(A - 3I)\mathbf{v}$.

Para determinar as quatro primeiras colunas de P, vamos começar por calcular uma base para $N(A - 5I)$ da forma $S_{k_i-1} \cup \cdots \cup S_0$, onde k_i é o índice de $\lambda = 5$. Como o índice de $\lambda = 5$ é igual a 2, os subespaços $\mathcal{M}_i = EC(A-5I)^i \cap N(A-5I)$ verificam:

$$\{\mathbf{0}\} = \mathcal{M}_2 \subseteq \mathcal{M}_1 \subseteq \mathcal{M}_0 = N(A - 5I).$$

Sendo $\dim N(A-5I) = 2$, o espaço das colunas de $(A-5I)$ tem dimensão 4. Para determinar uma base S_1 de $\mathcal{M}_1$ usamos agora o procedimento do Exercício 8.1.

Para tal, considere-se uma matriz X cujas colunas formam uma base de $EC(A - 5I)$,

$$A - 5I = \begin{bmatrix} 0 & 2 & -3 & 2 & 0 & 0 \\ -1 & -11 & 16 & -11 & 0 & 0 \\ 0 & -4 & 6 & -4 & 0 & 0 \\ 1 & 5 & -7 & 5 & 0 & 0 \\ 1 & 3 & -4 & 3 & 0 & 4 \\ 0 & 3 & -4 & 4 & -1 & -4 \end{bmatrix} \Longrightarrow X = \begin{bmatrix} -3 & 2 & 0 & 0 \\ 16 & -11 & 0 & 0 \\ 6 & -4 & 0 & 0 \\ -7 & 5 & 0 & 0 \\ -4 & 3 & 0 & 4 \\ -4 & 4 & -1 & -4 \end{bmatrix}.$$

O núcleo de $(A - 5I)X$ é

$$N((A-5I)X) = \{(a,b,c,d) \in \mathbb{R}^4 : \ a = b - c/4, d = 0\}.$$

Uma base de $N((A-5I)X)$ é formada pelos vectores $\mathbf{w}_1 = (0,1,4,0)$ e $\mathbf{w}_2 = (1,1,0,0)$. Por conseguinte, do Exercício 8.1, uma base S_1 para $\mathcal{M}_1$ é formada pelos vectores $X\mathbf{w}_1 = (2,-11,-4,5,3,0)$ e $X\mathbf{w}_2 = (-1,5,2,-2,-1,0)$. Temos portanto, $\mathcal{M}_1 = \mathcal{M}_0 = N(A-5I)$, e uma base S_1 é

$$S_1 = \{\mathbf{w}_1, \mathbf{w}_2\} = \{(2,-11,-4,5,3,0), (-1,5,2,-2,-1,0)\}.$$

Resolvendo o sistema $(A-5I)\mathbf{x} = \mathbf{b}$ para $\mathbf{b} = \mathbf{w}_1$ e $\mathbf{b} = \mathbf{w}_2$, obtemos vectores próprios generalizados associados, respectivamente a $\mathbf{w}_1$ e a $\mathbf{w}_2$. Nomeadamente, os vectores:

$$\mathbf{v}_1 = (4,-12,-8,1,0,0), \qquad \mathbf{v}_2 = (0,1,1,0,-1,0).$$

Uma matriz P, tal que $P^{-1}AP = J$, tem por colunas os vectores $\mathbf{w}_1, \mathbf{y}_1, \mathbf{w}_2, \mathbf{y}_2, \mathbf{u}$ e $\mathbf{v}$, por esta ordem, ou seja,

$$P = \begin{bmatrix} 2 & 4 & -1 & 0 & 0 & 0 \\ -11 & -12 & 5 & 1 & 0 & 0 \\ -4 & -8 & 2 & 1 & 0 & 0 \\ 5 & 1 & -2 & 0 & 0 & 0 \\ 3 & 0 & -1 & -1 & -2 & -1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

Note-se que sabendo que $J(5)$ era constituído por dois blocos 2×2 e que cada cadeia de Jordan associada a um vector de S_j tem $(j+1)$ vectores, poderíamos ter concluído imediatamente que $\mathcal{M}_1 = \mathcal{M}_0 = N(A-5I)$. ♦

8.3 Solução geral do sistema de EDOs $\mathbf{x}' = A\mathbf{x}$

No Capítulo 4 vimos que o conjunto solução geral de um sistema de equações diferenciais $\mathbf{x}' = A\mathbf{x}$ é um espaço linear de dimensão n, onde n é a ordem da matriz A (ver Exercício 4.4, pág. 222). Assim, para determinar a solução geral de $\mathbf{x}' = A\mathbf{x}$, onde A é uma matriz real de ordem n, precisamos de n soluções (reais) linearmente independentes do sistema. Quando a matriz A é diagonalizável, podemos obter n soluções linearmente independentes do sistema de EDOs a partir de uma base constituída por vectores próprios. De facto, da Proposição 4.13 (pág. 225) sabemos que $e^{\lambda t}\mathbf{u}$ é solução de $\mathbf{x}' = A\mathbf{x}$ se e só se $\mathbf{u}$ é um vector próprio de A associado ao valor próprio λ. Além disso, se a matriz A tem valores próprios complexos, para cada par $\lambda, \overline{\lambda}$ de valores próprios (complexos) conjugados, são soluções reais (linearmente independentes) do sistema: $\mathrm{Re}(e^{\lambda t}\mathbf{u})$ e $\mathrm{Im}(e^{\lambda t}\mathbf{u})$, com $\mathbf{u}$ um vector próprio associado a λ. Por conseguinte, se $B = \{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n\}$ é uma base de vectores próprios de A, com $\mathbf{u}_i$ associado ao valor próprio λ_i, uma base para o conjunto solução geral do sistema de EDOs é dada pelas seguintes soluções:

$e^{\lambda_i t}\mathbf{u}_i$	para $\lambda_i \in \mathbb{R}$,
$\mathrm{Re}(e^{\lambda_i t}\mathbf{u}_i)$ e $\mathrm{Im}(e^{\lambda_i t}\mathbf{u}_i)$	para cada par $\lambda_i, \overline{\lambda_i}$ de valores próprios (complexos) conjugados.

No caso em que A não é diagonalizável, não existe um número suficiente de vectores próprios que possibilite a construção de uma base do conjunto solução geral da equação diferencial $\mathbf{x}' = A\mathbf{x}$. No entanto, o Teorema 8.2 garante a existência de uma base de $\mathbb{C}^n$ formada por vectores próprios e vectores próprios generalizados. Pretendemos agora obter soluções da equação diferencial $\mathbf{x}' = A\mathbf{x}$ a partir de vectores próprios generalizados.

Seja λ um valor próprio de A e $\mathbf{u}$ um vector próprio associado a partir do qual se constrói a cadeia de Jordan $I_{\mathbf{u}} = \{\mathbf{u}, \mathbf{y}_1, \ldots, \mathbf{y}_p\}$. Como vimos anteriormente, $I_{\mathbf{u}}$ é um conjunto linearmente independente e seus vectores satisfazem:

$$\begin{aligned} (A - \lambda I)\mathbf{u} &= \mathbf{0} \\ (A - \lambda I)\mathbf{y}_1 &= \mathbf{u} \\ (A - \lambda I)\mathbf{y}_2 &= \mathbf{y}_1 \\ &\vdots \\ (A - \lambda I)\mathbf{y}_p &= \mathbf{y}_{p-1}. \end{aligned} \tag{8.25}$$

Na proposição que enunciamos a seguir estabelecem-se soluções (linearmente independentes) da equação diferencial $\mathbf{x}' = A\mathbf{x}$ associadas aos vectores de uma cadeia de Jordan $I_{\mathbf{u}}$.

Proposição 8.3. Seja $I_{\mathbf{u}} = \{\mathbf{u}, \mathbf{y}_1, \ldots, \mathbf{y}_p\}$ uma cadeia de Jordan contruída a partir do vector próprio $\mathbf{u}$ de A, verificando as relações (8.25).

As funções

$$\mathbf{x}_k(t) = e^{\lambda t}\left[\mathbf{y}_{k-1} + t\mathbf{y}_{k-2} + \frac{t^2}{2!}\mathbf{y}_{k-3} + \cdots + \frac{t^{k-2}}{(k-2)!}\mathbf{y}_1 + \frac{t^{k-1}}{(k-1)!}\mathbf{u}\right],$$

para $k = 1, \ldots, p+1$, são soluções linearmente independentes do sistema de EDOs, $\mathbf{x}' = A\mathbf{x}$.

Demonstração. Mostremos que $\mathbf{x}_k(t)$ é solução de $\mathbf{x}' = A\mathbf{x}$. Derivando a expressão de $\mathbf{x}_k$ em ordem a t, temos

$$\begin{aligned} \mathbf{x}_k'(t) &= \lambda e^{\lambda t}\left[\mathbf{y}_{k-1} + t\mathbf{y}_{k-2} + \tfrac{t^2}{2!}\mathbf{y}_{k-3} + \cdots + \tfrac{t^{k-2}}{(k-2)!}\mathbf{y}_1 + \tfrac{t^{k-1}}{(k-1)!}\mathbf{u}\right] \\ &+ e^{\lambda t}\left[\mathbf{y}_{k-2} + t\mathbf{y}_{k-3} + \cdots + \tfrac{t^{k-3}}{(k-3)!}\mathbf{y}_1 + \tfrac{t^{k-2}}{(k-2)!}\mathbf{u}\right]. \end{aligned} \tag{8.26}$$

Substituindo as igualdades (8.25) na segunda parcela da soma anterior, obtemos

$$\begin{aligned} \mathbf{x}_k'(t) = {}& \lambda e^{\lambda t}\left[\mathbf{y}_{k-1} + t\mathbf{y}_{k-2} + \frac{t^2}{2!}\mathbf{y}_{k-3} + \cdots + \frac{t^{k-2}}{(k-2)!}\mathbf{y}_1 + \frac{t^{k-1}}{(k-1)!}\mathbf{u}\right] \\ &+ e^{\lambda t}\left[(A-\lambda I)\mathbf{y}_{k-1} + t(A-\lambda I)\mathbf{y}_{k-2} + \cdots + \frac{t^{k-2}}{(k-2)!}(A-\lambda I)\mathbf{y}_1\right] \\ &= e^{\lambda t}\left[A\mathbf{y}_{k-1} + tA\mathbf{y}_{k-2} + \cdots + \frac{t^{k-2}}{(k-2)!}A\mathbf{y}_1 + \lambda\frac{t^{k-1}}{(k-1)!}\mathbf{u}\right]. \end{aligned}$$

Como $\mathbf{u}$ é um vector próprio de A associado a λ, verifica-se $\lambda\mathbf{u} = A\mathbf{u}$. Substituindo $\lambda\mathbf{u} = A\mathbf{u}$ na última parcela da expressão anterior, tem-se

$$\mathbf{x}_k'(t) = Ae^{\lambda t}\left[\mathbf{y}_{k-1} + t\mathbf{y}_{k-2} + \frac{t^2}{2!}\mathbf{y}_{k-3} + \cdots + \frac{t^{k-1}}{(k-1)!}\mathbf{u}\right] = A\mathbf{x}_k(t).$$

Ou seja, $\mathbf{x}_k$ é solução de $\mathbf{x}' = A\mathbf{x}$.

Falta mostrar a independência linear das soluções

$$\begin{aligned} \mathbf{x}_1(t) &= e^{\lambda t}\mathbf{u} \\ \mathbf{x}_2(t) &= e^{\lambda t}(\mathbf{y}_1 + t\mathbf{u}) \\ &\vdots \\ \mathbf{x}_{p+1}(t) &= e^{\lambda t}\left(\mathbf{y}_p + t\mathbf{y}_{p-1} + \cdots + \tfrac{t^p}{p!}\mathbf{u}\right). \end{aligned}$$

De acordo com a Proposição 4.12 (pág. 222) é suficiente mostrar que para algum $t_0 \in \mathbb{R}$, o conjunto de vectores $\{\mathbf{x}_1(t_0), \ldots, \mathbf{x}_{p+1}(t_0)\}$ é linearmente independente. Calculando as soluções anteriores em $t_0 = 0$ obtém-se $\{\mathbf{u}, \mathbf{y}_1, \ldots, \mathbf{y}_p\}$. Este conjunto é linearmente independente uma vez que é uma cadeia de Jordan. Logo, as $(p+1)$ soluções $\mathbf{x}_k(t)$ são linearmente independentes. $\square$

É fácil concluir, da proposição anterior e do Teorema 8.2, que tomando as soluções $\mathbf{x}_k(t)$ associadas a cada cadeia de Jordan para cada bloco de Jordan na decomposição (8.19), se obtém uma base para o conjunto solução geral. No exemplo a seguir ilustramos este procedimento.

Exemplo 8.8. Determinar a solução geral da equação diferencial $\mathbf{x}' = A\mathbf{x}$ onde A é a matriz do exemplo 8.7. Nesse exemplo, determinámos a forma de Jordan J de A e uma matriz P, tal que $P^{-1}AP = J$. As matrizes obtidas foram:

$$J = \begin{bmatrix} 5 & 1 & 0 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 5 & 1 & 0 & 0 \\ 0 & 0 & 0 & 5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 3 \end{bmatrix} \text{ e } P = \begin{bmatrix} 2 & 4 & -1 & 0 & 0 & 0 \\ -11 & -12 & 5 & 1 & 0 & 0 \\ -4 & -8 & 2 & 1 & 0 & 0 \\ 5 & 1 & -2 & 0 & 0 & 0 \\ 3 & 0 & -1 & -1 & -2 & -1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

A primeira, terceira e quinta colunas de P são vectores próprios de A, ou seja,

$$\mathbf{u} = (2, -11, -4, 5, 3, 0), \mathbf{v} = (-1, 5, 2, -2, -1, 0) \text{ e } \mathbf{w} = (0, 0, 0, 0, -2, 1).$$

Além disso, da matriz P obtêm-se as seguintes cadeias de Jordan associadas a estes vectores:

$$\begin{aligned} I_{\mathbf{u}}(5) &= \{(2, -11, -4, 5, 3, 0), (4, -12, -8, 1, 0, 0)\} = \{\mathbf{u}, \mathbf{u}_1\}, \\ I_{\mathbf{v}}(5) &= \{(-1, 5, 2, -2, -1, 0), (0, 1, 1, 0, -1, 0)\} = \{\mathbf{v}, \mathbf{v}_1\}, \\ I_{\mathbf{w}}(3) &= \{(0, 0, 0, 0, -2, 1), (0, 0, 0, 0, -1, 0)\} = \{\mathbf{w}, \mathbf{w}_1\}. \end{aligned}$$

Uma base de $\mathbb{R}^n$ constituída por vectores próprios e vectores próprios generalizados é $I_{\mathbf{u}}(5) \cup I_{\mathbf{v}}(5) \cup I_{\mathbf{w}}(3)$. Pela Proposição 8.3, uma base para o conjunto solução geral da equação diferencial é constituída pelas funções:

$$\mathbf{x}_1(t) = e^{5t}\mathbf{u}, \quad \mathbf{x}_2(t) = e^{5t}(\mathbf{u}_1 + t\mathbf{u}), \quad \mathbf{x}_3(t) = e^{5t}\mathbf{v}, \quad \mathbf{x}_4(t) = e^{5t}(\mathbf{v}_1 + t\mathbf{v}),$$
$$\mathbf{x}_5(t) = e^{3t}\mathbf{w}, \quad \mathbf{x}_6(t) = e^{3t}(\mathbf{w}_1 + t\mathbf{w}).$$

Assim, a solução geral da equação $\mathbf{x}' = A\mathbf{x}$, é

$$\mathbf{x}(t) = c_1\mathbf{x}_1(t) + c_2\mathbf{x}_2(t) + c_3\mathbf{x}_3(t) + c_4\mathbf{x}_4(t) + c_5\mathbf{x}_5(t) + c_6\mathbf{x}_6(t),$$

com c_1, c_2, c_3, c_4, c_5 e c_6 constantes reais arbitrárias. ♦

Exemplo 8.9. Determinemos e^{At} para a seguinte matriz real A de ordem p.

$$A = \begin{bmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda \end{bmatrix}.$$

Sabemos da Proposição 4.14 (pág. 229), que e^{At} é dada por $e^{At} = X(t)X(0)^{-1}$, onde $X(t)$ é uma qualquer matriz solução fundamental da equação $\mathbf{x}' = A\mathbf{x}$. Isto é, $X(t)$ é uma matriz cujas colunas são p soluções (reais) linearmente independentes da referida equação diferencial.

A matriz A é um bloco de Jordan, sendo λ o único valor próprio de A. É fácil verificar que o espaço próprio $E(\lambda)$ é gerado pelo vector $\mathbf{e}_1$ da base canónica de $\mathbb{R}^p$ e que a base canónica de $\mathbb{R}^p$ é uma cadeia de Jordan associada a $\mathbf{e}_1$. Da Proposição 8.3, obtemos a seguinte base do conjunto solução geral da equação $\mathbf{x}' = A\mathbf{x}$:

$$\begin{aligned} \mathbf{x}_1 &= e^{\lambda t}\mathbf{e}_1 \\ \mathbf{x}_2 &= e^{\lambda t}[\mathbf{e}_2 + t\mathbf{e}_1] \\ &\vdots \\ \mathbf{x}_p &= e^{\lambda t}[\mathbf{e}_p + t\mathbf{e}_{p-1} + \cdots + \tfrac{t^{p-1}}{(p-1)!}\mathbf{e}_1]. \end{aligned}$$

Uma matriz solução fundamental para o sistema de EDOs é uma matriz que possui estas soluções nas colunas. Por exemplo,

$$X(t) = e^{\lambda t}\begin{bmatrix} | & | & & | \\ \mathbf{e}_1 & \mathbf{e}_2 + t\mathbf{e}_1 & \cdots & \mathbf{e}_p + t\mathbf{e}_{p-1} + \cdots + \frac{t^{p-1}}{(p-1)!}\mathbf{e}_1 \\ | & | & & | \end{bmatrix}.$$

A matriz $X(0)$ é a matriz identidade. Logo, $e^{At} = X(t)X(0)^{-1} = X(t)$. Ou seja,

$$e^{At} = X(t) = e^{\lambda t} \begin{bmatrix} 1 & t & \frac{t^2}{2} & \frac{t^3}{3!} & \cdots & \frac{t^{p-1}}{(p-1)!} \\ 0 & 1 & t & \frac{t^2}{2} & \cdots & \frac{t^{p-2}}{(p-2)!} \\ 0 & 0 & 1 & t & \cdots & \frac{t^{p-3}}{(p-3)!} \\ & & & \ddots & \ddots & \vdots \\ & & & & \ddots & t \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix}. \tag{8.27}$$

♦

Exemplo 8.10. Determinemos a solução geral do sistema $\mathbf{x}' = A\mathbf{x}$ onde A é a matriz

$$A = \begin{bmatrix} 5 & -1 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 4 \end{bmatrix}.$$

A matriz A tem um único valor próprio $\lambda = 4$ de multiplicidade geométrica 2. Como a matriz A é do tipo 3×3 e a sua forma de Jordan J tem dois blocos, podemos já concluir que J é:

$$J = \begin{bmatrix} 4 & 1 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix}. \tag{8.28}$$

O espaço das colunas e o núcleo de $(A - 4I)$ são:

$$A - 4I = \begin{bmatrix} 1 & -1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \Longrightarrow EC(A - 4I) = \mathrm{Span}\{(1,1,0)\}$$
$$N(A - 4I) = \mathrm{Span}\{(1,1,0), (0,0,1\}.$$

Tomando $\mathbf{u} = (1,1,0) \in EC(A-4I) \cap N(A-4I)$ e $\mathbf{v} = (0,0,1) \in N(A-4I)$, as cadeias de Jordan associadas a estes vectores possuem, respectivamente, 2 e 1 vectores. Resolvendo o sistema $(A - 4I)\mathbf{x} = \mathbf{u}$, um vector próprio generalizado associado a $\mathbf{u}$ é $\mathbf{w} = (1,0,0)$. Tomando as cadeias de Jordan

$$I_{\mathbf{u}} = \{\mathbf{u}, \mathbf{w}\}, \qquad I_{\mathbf{v}} = \{\mathbf{v}\},$$

do Teorema 8.3 obtemos as seguinte soluções linearmente independentes da equação diferencial:

$$\mathbf{x}_1(t) = e^{4t}\mathbf{u}, \quad \mathbf{x}_2(t) = e^{4t}(\mathbf{w} + t\mathbf{u}), \quad \text{e} \quad \mathbf{x}_3(t) = e^{4t}\mathbf{v}.$$

Logo, uma matriz solução fundamental para a equação $\mathbf{x}' = A\mathbf{x}$ é

$$X(t) = e^{4t}\begin{bmatrix} 1 & 1+t & 0 \\ 1 & t & 0 \\ 0 & 0 & 1 \end{bmatrix} = e^{4t}\begin{bmatrix} | & | & | \\ \mathbf{u} & \mathbf{w}+t\mathbf{u} & \mathbf{v} \\ | & | & | \end{bmatrix} =$$

A solução geral do sistema de EDOs é dada por $\mathbf{x}(t) = X(t)\mathbf{c}$, onde $\mathbf{c}$ é um vector coluna constante arbitrário.

Calculemos agora a exponencial $e^{At} = X(t)X(0)^{-1}$. Como

$$X(0) = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

tem-se

$$e^{At} = X(t)X(0)^{-1} = X(t)\begin{bmatrix} 0 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = e^{4t}\begin{bmatrix} 1+t & -t & 0 \\ t & 1-t & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Refira-se que a solução geral do sistema também é da forma $\mathbf{x}(t) = e^{At}\mathbf{c}$, onde $\mathbf{c}$ é um vector coluna constante arbitrário. ♦

Para finalizar, refira-se um método de cálculo de e^{At} distinto do que foi usado nos exemplos anteriores. Suponha-se que $A = PJP^{-1}$, onde J é a forma de Jordan de A. Então,

$$\mathbf{x}' = A\mathbf{x} \Longleftrightarrow \mathbf{x}' = PJP^{-1}\mathbf{x} \Longleftrightarrow \mathbf{y}' = J\mathbf{y}, \quad \text{com } \mathbf{y} = P^{-1}\mathbf{x}.$$

É claro da equivalência anterior que $X(t)$ é uma matriz solução fundamental do sistema $\mathbf{x}' = A\mathbf{x}$ se e só se $Y(t) = P^{-1}X(t)$ é uma matriz solução fundamental do sistema $\mathbf{y}' = J\mathbf{y}$. Por conseguinte, se $Y(t) = P^{-1}X(t)$ é uma matriz solução fundamental do sistema $\mathbf{y}' = J\mathbf{y}$, tem-se $e^{Jt} = Y(t)Y^{-1}(0)$ e

$$e^{Jt} = Y(t)Y^{-1}(0) = P^{-1}X(t)X^{-1}(0)P = P^{-1}e^{At}P \Longleftrightarrow e^{At} = Pe^{Jt}P^{-1}.$$

Deixamos como exercício verificar que se uma matriz K é diagonal por blocos $K = \operatorname{diag}(K_1, \ldots, K_p)$, então $e^{Kt} = \operatorname{diag}\left(e^{K_1 t}, \ldots, e^{K_p t}\right)$.

A exponencial e^{Jt}, onde J é uma matriz de Jordan na forma (8.19), é portanto da forma

$$e^{Jt} = \begin{bmatrix} e^{J(\lambda_1)t} & & & \\ & e^{J(\lambda_2)t} & & \\ & & \ddots & \\ & & & e^{J(\lambda_s)t} \end{bmatrix}. \tag{8.29}$$

Além disso, cada bloco $J(\lambda_i)$ de J é ainda uma matriz diagonal por blocos cujos blocos são matrizes da forma (8.27) (ver Exemplo 8.9). Portanto, cada bloco $e^{J(\lambda_i)t}$ em (8.29) é uma matriz em blocos cujos blocos são da forma (8.27).

Exemplo 8.11. Considere-se de novo a matriz A do Exemplo 8.7 e a correspondente matriz de Jordan J calculada nesse exemplo. A matriz J obtida é constituída por dois blocos $J(5)$ e $J(3)$, respectivamente, do tipo 4×4 e do tipo 2×2. Isto é,

$$J = \begin{bmatrix} J(5) & \mathbf{O} \\ \mathbf{O} & J(3) \end{bmatrix} \quad \text{com } J(5)_{4\times 4} = \begin{bmatrix} J_1(5)_{2\times 2} & \mathbf{O} \\ \mathbf{O} & J_2(5)_{2\times 2} \end{bmatrix}, \; J(3)_{2\times 2} = \begin{bmatrix} J_1(3) \end{bmatrix}.$$

Além disso, $J_1(5) = J_2(5) = \begin{bmatrix} 5 & 1 \\ 0 & 5 \end{bmatrix}$ e $J_1(3) = \begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}$.

Usando a expressão (8.27) do Exemplo 8.9 , temos

$$e^{J_1(5)t} = e^{J_2(5)t} = e^{5t} \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}, \quad e^{J(3)t} = e^{3t} \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}.$$

Por conseguinte,

$$e^{Jt} = \begin{bmatrix} e^{5t} & te^{5t} & 0 & 0 & 0 & 0 \\ 0 & e^{5t} & 0 & 0 & 0 & 0 \\ 0 & 0 & e^{5t} & te^{5t} & 0 & 0 \\ 0 & 0 & 0 & e^{5t} & 0 & 0 \\ 0 & 0 & 0 & 0 & e^{3t} & te^{3t} \\ 0 & 0 & 0 & 0 & 0 & e^{3t} \end{bmatrix}.$$

Sendo P uma matriz tal que $A = PJP^{-1}$, a exponencial e^{At} é igual a $e^{At} = Pe^{Jt}P^{-1}$. ♦

Exercício 8.3. Seja A a matriz (diagonal por blocos) do Exemplo 8.10. Calcule e^{At} a partir de e^{Jt} onde J é a matriz de Jordan em (8.28). Confirme o seu resultado com a exponencial obtida no referido exemplo. ▲

No caso da matriz A do sistema $\mathbf{x}' = A\mathbf{x}$ não ser diagonalizável e ter valores próprios complexos, determina-se uma base constituída por soluções reais, tomando as partes reais e imaginárias das soluções complexas obtidas a partir de uma base de $\mathbb{C}^n$ formada por vectores próprios e vectores próprios generalizados. No exemplo a seguir ilustramos este procedimento.

Exemplo 8.12. Determinemos a solução geral do sistema $\mathbf{x}' = A\mathbf{x}$, em que A é a matriz

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

O polinómio característico de A é $p(\lambda) = (\lambda^2+1)^2$, e portanto $\lambda = i$ e $\overline{\lambda} = -i$ são valores próprios de A com multiplicidades algébricas iguais a dois. Os espaços próprios $E(i)$ e $E(-i)$ são:

$$E(i) = \operatorname{Span}\{(i,1,0,0)\} \quad E(-i) = \operatorname{Span}\{(-i,1,0,0)\}.$$

A matriz A não é portanto diagonalizável (a multiplicidade geométrica de cada valor próprio é igual a 1). Como a dimensão de cada espaço próprio é igual a 1, e a matriz A é do tipo 4×4, conclui-se que a forma de Jordan J de A é

$$J = \begin{bmatrix} i & 1 & 0 & 0 \\ 0 & i & 0 & 0 \\ 0 & 0 & -i & 1 \\ 0 & 0 & 0 & -i \end{bmatrix}.$$

Determinemos uma cadeia de Jordan associada ao vector próprio $\mathbf{u} = (i,1,0,0) \in E(i)$. A cadeia $I_{\mathbf{u}}$ é formada por $\mathbf{u}$ e por um vector próprio generalizado $\mathbf{w}$ verificando $(A - iI)\mathbf{w} = \mathbf{u}$. Resolvendo este sistema, obtemos para $\mathbf{w}$, por exemplo, $\mathbf{w} = (0,0,i,1)$. Logo,

$$I_{\mathbf{u}} = \{\mathbf{u}, \mathbf{w}\} = \{(i,1,0,0), (0,0,i,1)\}.$$

Usando a Proposição 8.3, da cadeia $I_{\mathbf{u}}$ obtêm-se duas soluções (complexas) linearmente independentes. Mais precisamente, as soluções

$$\mathbf{y}_1(t) = e^{it}\mathbf{u} = (\cos t + i \operatorname{sen} t) \begin{bmatrix} i \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -\operatorname{sen} t + i\cos t \\ \cos t + i \operatorname{sen} t \\ 0 \\ 0 \end{bmatrix},$$

e

$$\mathbf{y}_2(t) = e^{it}(\mathbf{w} + t\mathbf{u}) = (\cos t + i \operatorname{sen} t)\begin{bmatrix} it \\ t \\ i \\ 1 \end{bmatrix} = \begin{bmatrix} -t \operatorname{sen} t + it \cos t \\ t \cos t + it \operatorname{sen} t \\ i \cos t - \operatorname{sen} t \\ \cos t + i \operatorname{sen} t \end{bmatrix}.$$

Aplicando a Proposição 4.13, temos que os conjuntos $S_1 = \{\operatorname{Re} \mathbf{y}_1(t), \operatorname{Im} \mathbf{y}_1(t)\}$ e $S_2 = \{\operatorname{Re} \mathbf{y}_2(t), \operatorname{Im} \mathbf{y}_2(t)\}$ são linearmente independentes. Deixamos como exercício verificar que $S = S_1 \cup S_2$ é uma base do conjunto solução geral. Os elementos de S são:

$$\mathbf{x}_1(t) = \operatorname{Re} \mathbf{y}_1(t) = \begin{bmatrix} -\operatorname{sen} t \\ \cos t \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{x}_2(t) = \operatorname{Im} \mathbf{y}_1(t) = \begin{bmatrix} \cos t \\ \operatorname{sen} t \\ 0 \\ 0 \end{bmatrix},$$

$$\mathbf{x}_3(t) = \operatorname{Re} \mathbf{y}_2(t) = \begin{bmatrix} -t \operatorname{sen} t \\ t \cos t \\ -\operatorname{sen} t \\ \cos t \end{bmatrix}, \quad \mathbf{x}_4(t) = \operatorname{Im} \mathbf{y}_2(t) = \begin{bmatrix} t \cos t \\ t \operatorname{sen} t \\ \cos t \\ \operatorname{sen} t \end{bmatrix}.$$

Determinemos agora a exponencial e^{At}. Uma matriz solução fundamental $X(t)$ tem por colunas as soluções $\mathbf{x}_i(t)$, e a exponencial e^{At} é dada por $e^{At} = X(t)X^{-1}(0)$. Tomando para $X(t)$ a matriz cujas colunas são as soluções $\mathbf{x}_2(t), \mathbf{x}_1(t), \mathbf{x}_4(t), \mathbf{x}_3(t)$, por esta ordem, a matriz $X(0)$ é igual à identidade. Portanto, $e^{At} = X(t)$ é a matriz

$$e^{At} = \begin{bmatrix} \cos t & -\operatorname{sen} t & t \cos t & -t \operatorname{sen} t \\ \operatorname{sen} t & \cos t & t \operatorname{sen} t & t \cos t \\ 0 & 0 & \cos t & -\operatorname{sen} t \\ 0 & 0 & \operatorname{sen} t & \cos t \end{bmatrix}.$$

A solução geral do sistema é $\mathbf{x}(t) = e^{At}\mathbf{c}$, com $\mathbf{c}$ um vector constante arbitrário. ♦

Exercícios

1. Seja L uma matriz nilpotente 4×4, de índice k, e $\dim N(L) = t$. Indique o valor lógico das afirmações seguintes.

a) Se $k = 1$, então $t = 2$.
b) Se $k = 3$, então $t = 2$.
c) Se $t = 2$, então $k = 3$.
d) Se $t = 3$, então $k = 2$.

2. Verifique que as matrizes seguintes são nilpotentes e determine a sua forma de Jordan.

$$a) \begin{bmatrix} -2 & -1 & 0 \\ -2 & 0 & 1 \\ -16 & -6 & 2 \end{bmatrix} \quad b) \begin{bmatrix} -2 & -1 & 0 \\ 4 & 2 & 0 \\ -4 & -2 & 0 \end{bmatrix}.$$

3. Para cada uma das matrizes A seguintes, determine a forma de Jordan J e uma matriz P tal que $A = PJP^{-1}$.

$$a)\ A = \begin{bmatrix} 0 & -1 & 0 \\ -2 & 2 & 1 \\ -16 & -6 & 4 \end{bmatrix}.$$

$$b)\ A = \begin{bmatrix} 0 & -1 & 0 \\ 4 & 4 & 0 \\ -10 & -4 & 3 \end{bmatrix}.$$

$$c)\ A = \begin{bmatrix} 4 & 0 & -2 & 1 \\ 0 & 3 & 0 & 0 \\ 0 & -1 & 3 & 0 \\ 0 & 0 & -1 & 4 \end{bmatrix}.$$

$$d)\ A = \begin{bmatrix} 3 & -1 & 0 & 0 \\ -1 & 3 & 1 & 0 \\ 0 & -1 & 3 & 0 \\ 1 & 0 & -2 & 4 \end{bmatrix}.$$

4. Seja $\sigma(A) = \{2, 4\}$ o espectro de uma matriz A de ordem 7, tal que $mult.alg.(4) = 4$, $mult.geo.(4) = 2$ e $mult.geo.(2) = 1$.

Considere ainda que a forma de Jordan de A é $J = \begin{bmatrix} J(4) & \\ & J(2) \end{bmatrix}$.

Indique o valor lógico das afirmações seguintes.

a) Os blocos $J(4)$ e $J(2)$ são do mesmo tipo.
b) A restrição de A ao núcleo $N(A - 2I)$ é diagonalizável.
c) O bloco $J(2)$ é a matriz diagonal $J(2) = \mathrm{diag}(2, 2, 2)$.
d) Os dados do problema permitem determinar $J(4)$.
e) Os dados do problema permitem determinar $J(2)$.
f) A matriz $(A - 4I)$ é nilpotente.
g) A parte nilpotente (ver Proposição 8.2) de $(A - 2I)$ é semelhante à matriz nilpotente

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

5. Determine a exponencial e^{At} para cada uma das matrizes do Exercício 3, bem como a solução geral do sistema de equações diferencias $\mathbf{x}' = A\mathbf{x}$.

6. Seja $A = \begin{bmatrix} -1 & -2 & 1 \\ 1 & 1 & -1 \\ 0 & -1 & 0 \end{bmatrix}$.

a) Mostre que A é uma matriz nilpotente.
b) Calcule a exponencial e^{At} e verifique que essa exponencial satisfaz a igualdade

$$e^{At} = I + At + \frac{1}{2}At^2,$$

onde I designa a matriz identidade de terceira ordem.

Apêndice A

Números complexos

Sendo a e b números reais, um número complexo z tem a forma

$$z = a + ib = a + bi \quad \text{com } \ i^2 = -1 \text{ e } \ a, b \in \mathbb{R}.$$

A expressão $i = \sqrt{-1}$ é designada por a unidade imaginária dos números complexos. Os números reais a e b, respectivamente, a parte real e a parte imaginária de z, são designados por $\mathrm{Re}(z) = a$ e $\mathrm{Im}(z) = b$. O conjunto dos números complexos é designado por $\mathbb{C}$. Note-se que um número real a é um número complexo da forma $z = a + 0i$. Ou seja, $\mathbb{R} \subset \mathbb{C}$.

Os números complexos foram inventados para permitir calcular soluções de equações como $x^2 + 1 = 0$, equação esta que, apesar de não ter soluções reais, possui duas soluções complexas $x = \pm i$.

Dois números complexos são iguais se e só se as respectivas parte real e imaginária são iguais. Isto é, $z = a + ib$ e $w = c + id$ são iguais se e só se

$$\mathrm{Re}(z) = a = \mathrm{Re}(w) = c \quad \text{e} \quad \mathrm{Im}(z) = b = \mathrm{Im}(w) = d.$$

Podemos operar com números complexos, definindo a adição e a multiplicação do modo seguinte:

$$\begin{aligned} (a + ib) + (c + id) &= (a + c) + i(b + d) \\ (a + ib)(c + id) &= (ac - bd) + i(bc + ad). \end{aligned}$$

Estas operações gozam das propriedades comutativa, associativa e distributiva, como facilmente se deduz usando as propriedades das operações de adição e multiplicação de números reais.

Exemplo A.1.

$$(2+3i)(5-10i) = 10 - 20i + 15i - 30i^2 = 40 - 5i.$$

♦

Define-se o *conjugado* de $z = a + ib$ como sendo o número complexo $\overline{z} = a - ib$ (o conjugado de z obtém-se de z substituindo a sua parte imaginária pelo respectivo simétrico). Por exemplo, o conjugado de $z = -2 - 3i$ é $\overline{z} = -2 + 3i$. Dado um complexo $z = a + ib$, temos

$$\begin{aligned} z\overline{z} &= (a+ib)(a-ib) = a^2 - iab + iab - i^2 b \\ &= a^2 + b^2. \end{aligned}$$

Define-se *módulo* de z como sendo o número real não negativo

$$|z| = \sqrt{z\overline{z}} = \sqrt{a^2 + b^2}.$$

Note-se que se z é real (isto é, $b = 0$) então $|z| = \sqrt{a^2} = |a|$ coincide com o módulo de um número real.

Se $z \neq 0$, o inverso de z é

$$\frac{1}{z} = \frac{\overline{z}}{z\overline{z}} = \frac{\overline{z}}{|z|^2}.$$

Por exemplo,

$$\frac{1}{1+2i} = \frac{1-2i}{5} \qquad \text{e} \qquad \frac{1}{i} = \frac{-i}{-i^2} = -i.$$

Exercício A.1. Mostre que se z e w são números complexos, são satisfeitas as seguintes relações:

a) $z + \overline{z} = 2\,\mathrm{Re}(z)$.
b) $z - \overline{z} = 2i\,\mathrm{Im}(z)$.
c) $|\mathrm{Re}(z)| \leq |z|$.
d) $\overline{z} = z$ se e só se z é real.
e) $\overline{z + w} = \overline{z} + \overline{w}$.
f) $\overline{zw} = \overline{z}\,\overline{w}$.
g) $z\overline{z} = |z|^2 \geq 0$
h) $|zw| = |z||w|$.
i) $|z + w| \leq |z| + |w|$.

▲

A um número complexo $z = a + ib$ podemos fazer corresponder (de forma biunívoca) o ponto do plano de coordenadas (a, b). O eixo horizontal é designado por *eixo real* e o vertical por *eixo imaginário*. Geometricamente, o conjugado de um número complexo obtém-se por reflexão em relação ao eixo real. Além disso, o módulo de um complexo $z = a + ib$ representa a distância de (a, b) à origem (ver Figura A.1).

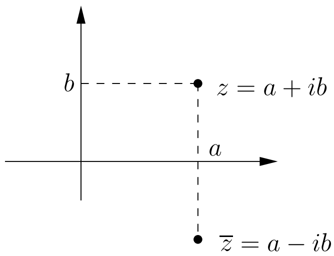


Figura A.1: Representação geométrica de um complexo $z = a + ib$ e do seu conjugado.

É fácil mostrar que a soma de $z = a + ib$ com $w = c + id$ corresponde à soma dos vectores (a, b) e (c, d) associados respectivamente a z e a w.

Qualquer número complexo admite dois tipos de representação, a *representação cartesiana* $z = a + ib$ e a representação dita polar. A *representação polar* de um complexo $z = a + ib$ é dada em termos do ângulo θ que o vector (a, b) faz com a parte positiva do eixo real, e de $|z| = \sqrt{a^2 + b^2}$, ou seja da distância de z à origem. Na Figura A.2 ilustra-se a representação polar de um complexo. Tendo em conta essa ilustração, dado o complexo $z = a + ib$ tem-se $a = |z| \cos\theta$ e $b = |z| \operatorname{sen}\theta$. Logo, a *representação polar* de z é

$$z = |z| \cos\theta + i|z| \operatorname{sen}\theta = |z| \left(\cos\theta + i \operatorname{sen}\theta\right).$$

Ao ângulo θ chamamos *argumento* de z e denotamos por $\theta = \arg(z)$. O argumento de z não é único já que podemos somar um múltiplo de 2π e obtemos o mesmo número complexo. Por exemplo,

$$1 + i = \sqrt{2} \left(\cos\frac{\pi}{4} + i \operatorname{sen}\frac{\pi}{4}\right) = \sqrt{2} \left(\cos\frac{9\pi}{4} + i \operatorname{sen}\frac{9\pi}{4}\right).$$

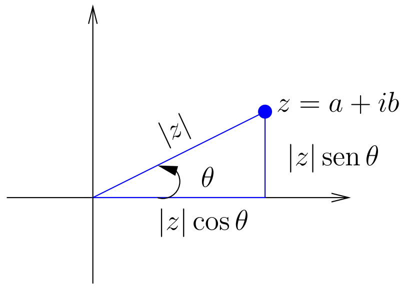


Figura A.2: Representação polar de um complexo z.

Porém, o argumento θ de um complexo é único se considerarmos $-\pi < \theta \leq \pi$. A este valor chamamos o *argumento principal* de z e designamo-lo por $\theta = \operatorname{Arg}(z)$.

Exemplo A.2. Determinemos a representação polar dos seguintes números complexos usando para argumento o seu argumento principal.

$$(a)\ z = \sqrt{3} + i \qquad (b)\ z = -1 - i$$

(a) O valor do módulo de z é

$$|z| = \sqrt{3+1} = \sqrt{4} = 2.$$

Para $a = \sqrt{3}$ e $b = 1$ tem-se

$$\sqrt{3} = 2\cos\theta$$
$$1 = 2\operatorname{sen}\theta,$$

ou seja, $\cos\theta = \frac{\sqrt{3}}{2}$ e $\operatorname{sen}\theta = \frac{1}{2}$. O único valor de θ em $]-\pi, \pi]$ é $\theta = \frac{\pi}{6}$ (= 30^o). Consequentemente, a representação polar de z é

$$z = 2\left(\cos\frac{\pi}{6} + i\operatorname{sen}\frac{\pi}{6}\right).$$

(b) O valor do módulo de z é

$$|z| = \sqrt{1+1} = \sqrt{2}.$$

Para $a = -1$ e $b = -1$, vem

$$-1 = \sqrt{2}\cos\theta$$
$$-1 = \sqrt{2}\operatorname{sen}\theta,$$

ou seja, $\cos\theta = \frac{-1}{\sqrt{2}} = \frac{-\sqrt{2}}{2}$ e $\operatorname{sen}\theta = \frac{-1}{\sqrt{2}} = \frac{-\sqrt{2}}{2}$. O único valor de θ em $]-\pi,\pi]$ é $\theta = \frac{-3\pi}{4}$. Assim, a forma polar de z é

$$z = \sqrt{2}\left(\cos\frac{-3\pi}{4} + i\operatorname{sen}\frac{-3\pi}{4}\right).$$

♦

É possível estender a função exponencial aos números complexos e provar (ver Marsden & Hoffman [8]) que $\cos\theta + i\operatorname{sen}\theta$ é igual à exponencial complexa $e^{i\theta}$, onde e é o número irracional, base dos logaritmos neperianos, aproximadamente igual a $e \approx 2.71828$. Ou seja,

$$\cos\theta + i\operatorname{sen}\theta = e^{i\theta}. \tag{A.1}$$

Logo, a representação polar de um número complexo pode escrever-se

$$z = |z|e^{i\theta}.$$

Observando a Figura A.1 verificamos que se θ é o argumento de z, então $-\theta$ é o argumento do conjugado de z. Por exemplo, o conjugado de $z = e^{i\theta}$ é $\overline{z} = e^{-i\theta}$.

A interpretação geométrica da multiplicação de dois números complexos é facilitada se usarmos a sua representação polar. Para tal, consideremos

$$z_1 = r_1(\cos\theta_1 + i\operatorname{sen}\theta_1) \quad \text{e} \quad z_2 = r_2(\cos\theta_2 + i\operatorname{sen}\theta_2).$$

O produto z_1z_2 é dado por

$$z_1z_2 = r_1r_2\left[(\cos\theta_1\cos\theta_2 - \operatorname{sen}\theta_1\operatorname{sen}\theta_2) + i\left(\operatorname{sen}\theta_1\cos\theta_2 + \cos\theta_1\operatorname{sen}\theta_2\right)\right].$$

Tendo em conta as igualdades trigonométricas

$$\begin{aligned}\cos(\theta_1+\theta_2) &= \cos\theta_1\cos\theta_2 - \operatorname{sen}\theta_1\operatorname{sen}\theta_2\\ \operatorname{sen}(\theta_1+\theta_2) &= \operatorname{sen}\theta_1\cos\theta_2 + \cos\theta_1\operatorname{sen}\theta_2,\end{aligned}$$

obtém-se

$$z_1z_2 = |z_1||z_2|\left(\cos(\theta_1+\theta_2) + i\operatorname{sen}(\theta_1+\theta_2)\right).$$

Consequentemente, o módulo do produto z_1z_2 é o produto dos módulos, e um argumento de z_1z_2 é a soma de argumentos de z_1 e z_2. Isto é,

$$|z_1z_2| = |z_1||z_2| \quad \text{e} \quad \arg(z_1z_2) = \arg(z_1) + \arg(z_2).$$

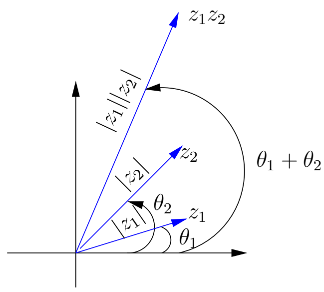


Figura A.3: O produto de dois complexos.

Note que, usando a representação polar de complexos em termos de exponenciais complexas, são válidas as igualdades

$$z_1z_2 = r_1e^{i\theta_1}r_2e^{i\theta_2} = r_1r_2e^{i(\theta_1+\theta_2)}. \tag{A.2}$$

Na Figura A.3 ilustramos a geometria associada ao produto de dois complexos.

Exercício A.2. Mostre que para $z_2 \neq 0$ são verdadeiras as seguintes igualdades:

$$\left|\frac{z_1}{z_2}\right| = \frac{|z_1|}{|z_2|} \quad \text{e} \quad \arg\left(\frac{z_1}{z_2}\right) = \arg(z_1) - \arg(z_2).$$

▲

Fórmula de *De Moivre*

Se n é um inteiro positivo e $z = r(\cos\theta + i\,\mathrm{sen}\,\theta)$, da equação (A.2), tem-se

$$z^n = r^n\left(\cos(n\theta) + i\,\mathrm{sen}(n\theta)\right).$$

No caso particular do complexo ter módulo igual a 1 ($r = 1$ na fórmula anterior) resulta a expressão seguinte conhecida por fórmula de *De Moivre*[1]:

$$(\cos\theta + i\,\mathrm{sen}\,\theta)^n = \cos(n\theta) + i\,\mathrm{sen}(n\theta). \tag{A.3}$$

A fórmula de Moivre em termos de exponenciais complexas escreve-se:

$$(e^{i\theta})^n = e^{in\theta}.$$

[1] Abraham de Moivre (1667–1754), matemático francês.

Raízes índice n de um complexo

Dado um número complexo z, pretendemos determinar números complexos w tais que

$$w^n = z.$$

Ou seja, pretende-se calcular w tal que $w = \sqrt[n]{z}$. Para o efeito é conveniente usarmos a representação polar de w e de z. Seja $w = \rho(\cos\alpha + i\,\mathrm{sen}\,\alpha)$ e $z = r(\cos\theta + i\,\mathrm{sen}\,\theta)$. Usando a igualdade (A.2), a expressão $w^n = z$ é equivalente a

$$w^n = \rho^n\left(\cos(n\alpha) + i\,\mathrm{sen}(n\alpha)\right) = r(\cos\theta + i\,\mathrm{sen}\,\theta).$$

Assim, $w^n = z$ se e só se

$$\rho^n = r \quad \text{e} \quad n\alpha = \theta + 2k\pi, \quad \text{para } k \in \mathbb{Z}.$$

Logo, os números complexos w pretendidos são

$$w = \sqrt[n]{r}\left(\cos\frac{\theta + 2k\pi}{n} + i\,\mathrm{sen}\,\frac{\theta + 2k\pi}{n}\right), \quad \text{para } k \in \mathbb{Z},$$

onde $\sqrt[n]{r}$ designa a raiz índice n do número real positivo r. Da expressão anterior conclui-se que $w = \sqrt[n]{z}$ são números complexos com módulo igual $\sqrt[n]{r}$, isto é, são números complexos que se situam numa circunferência de centro na origem e raio $\sqrt[n]{r}$. Além disso, existem n números complexos w distintos dados por:

$$w_k = \sqrt[n]{r}\left(\cos\frac{\theta + 2k\pi}{n} + i\,\mathrm{sen}\,\frac{\theta + 2k\pi}{n}\right), \quad \text{para } k = 0, 1, 2, \ldots, (n-1), \tag{A.4}$$

cujos argumentos são $\dfrac{\theta + 2k\pi}{n}$ para $k = 0, 1, 2, \ldots, (n-1)$. Assim, as n raízes índice n de um complexo situam-se numa circunferência de centro na origem e raio $\sqrt[n]{r}$, e dividem esta circunferência em n partes iguais.

Por exemplo, determinemos as raízes cúbicas e quartas da unidade, isto é, $\sqrt[3]{1}$ e $\sqrt[4]{1}$. Como um argumento de 1 é zero e o módulo de 1 é 1, estas raízes situam-se numa circunferência de centro na origem e raio 1. O número real 1 é uma raiz e portanto dividindo esta circunferência, respectivamente, em 3 e 4 partes iguais, tem-se

$$w_0 = 1, \quad w_1 = \frac{-1}{2} + i\frac{\sqrt{3}}{2}, \quad w_2 = \frac{-1}{2} - i\frac{\sqrt{3}}{2}$$

e

$$u_0 = 1, \quad u_1 = i, \quad u_2 = -1, \quad u_3 = -i.$$

Na Figura A.4 encontram-se representadas estas raízes.

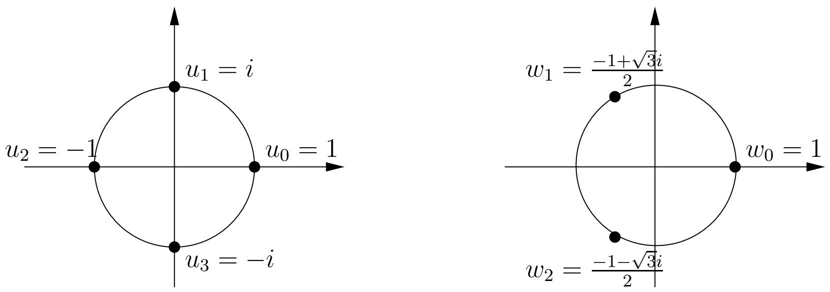


Figura A.4: À esquerda as raízes quartas da unidade e à direita as raízes cúbicas.

Exemplo A.3. Determinemos todas as raízes quartas de -16. Como -16 se encontra na parte negativa do eixo real podemos tomar para argumento o valor $\theta = \pi$. Logo,

$$\begin{aligned}\sqrt[4]{-16} &= \sqrt[4]{16}\left(\cos\frac{\pi+2k\pi}{4} + i\operatorname{sen}\frac{\pi+2k\pi}{4}\right)\\ &= 2\left(\cos\left(\frac{\pi}{4}+k\frac{\pi}{2}\right) + i\operatorname{sen}\left(\frac{\pi}{4}+k\frac{\pi}{2}\right)\right), \quad \text{para } k = 0,1,2,3,\end{aligned}$$

são os 4 números complexos (distintos) iguais a $\sqrt[4]{-16}$. Designando por w_k estas raízes, temos

$$w_0 = \sqrt{2}+i\sqrt{2}, \quad w_1 = -\sqrt{2}+i\sqrt{2}, \quad w_2 = -\sqrt{2}-i\sqrt{2}, \quad w_3 = \sqrt{2}-i\sqrt{2}.$$

♦

Exemplo A.4. Determinemos todas as raízes quartas de $z = 1+i$. Um argumento de $1+i$ é $\pi/4$ e o módulo de z é $\sqrt{2}$. Logo, quatro números complexos distintos iguais a $\sqrt[4]{1+i}$ são

$$\begin{aligned}\sqrt[4]{1+i} &= \sqrt[4]{2}\left(\cos\frac{\pi/4+2k\pi}{4} + i\operatorname{sen}\frac{\pi/4+2k\pi}{4}\right)\\ &= \sqrt[4]{2}\left(\cos\left(\frac{\pi}{16}+k\frac{\pi}{2}\right) + i\operatorname{sen}\left(\frac{\pi}{16}+k\frac{\pi}{2}\right)\right), \quad \text{para } k = 0,1,2,3,\end{aligned}$$

Na Figura A.5 encontram-se representadas estas raízes. ♦

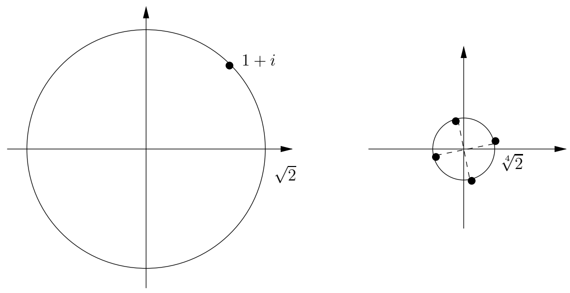


Figura A.5: À esquerda o complexo $1 + i$ e à direita as raízes $\sqrt[4]{1 + i}$.

Bibliografia

[1] F. R. Dias Agudo [1992], Introdução à Algebra Linear e Geometria Analítica, 4. edição, Escolar Editora.

[2] Martin Braun [1992], Differential Equations and Their Applications, Texts in Applied Mathematics, vol. 11, 4th edition, Springer.

[3] Kurt Bryan and Tanya Leise [2006], *The \$ 25,000,000,000 eigenvector. The linear Algebra behind Google*, SIAM Review, Vol. 48, No. 3, 569-581.

[4] P.M. Cohn [2005], Basic Algebra, Springer.

[5] R. Howe [1983], *Very basic Lie Theory*, American Mathematical Monthly, **90**, 600-623; Correction, Amer. Math. Monthly **91** (1984), 247.

[6] Dan Kalman [1996], *The singularly valuable Decomposition: The SVD of a matrix*, The college mathematics Journal, Vol. 27, No 1, 2-23.

[7] Amy Langville, Carl Meyer[2006], Google's PageRank and Beyond: The Science of Search Engine Rankings. Princeton University Press.

[8] Jerrold E. Marsden and Michael J. Hoffman[1998], Basic complex analysis, Third edition, W. H. Freeman.

[9] Carl D. Meyer [2000], Matrix Analysis and Applied Linear Algebra, SIAM.

[10] P. Olver e C. Shakiban [2006], Applied Linear Algebra, Pearson, Prentice Hall.

[11] Gilbert Strang [2003], Introduction to Linear Algebra, Third edition, Wellesley-Cambridge Press.

[12] Loyd N. Trefethen and David Bau III [1997], Numerical Linear Algebra, SIAM.

[13] Thomas Yuster [1984], *The reduced row form of a matrix is unique: a simple proof*, Mathematics Magazine, Vol. 57, No. 2, 93-94.

Soluções (prob. ímpares)

Capítulo 1

1. a) e c).

3. $p(x) = -1/3x^2 + 2x + 4/3$

5. a)F. b) V. c) F.

7. a) $\begin{bmatrix} -7 \\ -10 \end{bmatrix}$. b)$A = \begin{bmatrix} -2 & 7 \\ -4 & 10 \end{bmatrix}$,

$$f\left(\begin{bmatrix} 3 \\ 2 \end{bmatrix}\right) = \begin{bmatrix} 8 \\ 8 \end{bmatrix} \text{ e } f\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} -2 \\ -4 \end{bmatrix}.$$

9. a) Sim; $\operatorname{car}(A) = 3$. b) c) d) f) e j): Sim; $\operatorname{car}(A) = 2$. g) e e) Não; $\operatorname{car}(A) = 2$. h) Não; $\operatorname{car}(A) = 1$. i) Não; $\operatorname{car}(A) = 3$.

11. a) $\operatorname{car}(A) = \begin{cases} 1 & \text{se } \alpha = 2 \text{ ou } \alpha = -2 \\ 2 & \text{se } \alpha \neq \pm 2 \end{cases}$

e $\operatorname{car}(A|\mathbf{b}) = \begin{cases} 1 & \text{se } \alpha = -2 \\ 2 & \text{se } \alpha \neq -2 \end{cases}$.

b) Se $\alpha = 2$ sistema impossível; se $\alpha = -2$ sistema indeterminado com grau de indeterminação 2; se $\alpha \neq \pm 2$ o sistema é indeterminado com grau de indeterminação 1.

Solução geral:
para $\alpha = -2$, $\{(y - z, y, z) : y, z \in \mathbb{R}\}$ e para $\alpha \neq \pm 2$, $\{(y - \frac{1}{\alpha-2}, y, \frac{1}{\alpha-2}) : y \in \mathbb{R}\}$.

13. a) $\{(3, 1, 2)\}$

b) $\{(w - 1, 2z, z, w) : z, w \in \mathbb{R}\}$

15. a) $A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix}$

b) $A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 4 & 4 & 4 & 4 \\ 9 & 9 & 9 & 9 \\ 16 & 16 & 16 & 16 \end{bmatrix}$

c) $A = \begin{bmatrix} 0 & 2 & 3 & 4 \\ -2 & 0 & 3 & 4 \\ -3 & -3 & 0 & 4 \\ -4 & -4 & -4 & 0 \end{bmatrix}$

17.
a) $A\mathbf{b} = \begin{bmatrix} 5a + 3c + 4 & a + c & 3a - c + 4 \end{bmatrix}^T$
b) $a, 2$ e c.

19. $\begin{bmatrix} -10 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ e $m = 1/6$, $b = 8/3$.

21.
a) $AB =$
$$= \begin{bmatrix} 1 + 4\sqrt{3} & -2 + \sqrt{2} & \pi + 8 - \sqrt{2} \\ -2 + \sqrt{3} & -5 + \sqrt{3} & -2\pi + 2 - \sqrt{3} \end{bmatrix}$$
BA não faz sentido.

b) $AB = [-1 - 2i]$, $BA = \begin{bmatrix} -2 & -(i+1) \\ 5 - i & -3 - 2i \end{bmatrix}$.

23. a) V. b) F. c) V. d) V.

25. a) $A^n = I_2$

b) e c) $A^n = \begin{cases} A & \text{se } n \equiv 1 \pmod 4 \\ -I & \text{se } n \equiv 2 \pmod 4 \\ -A & \text{se } n \equiv 3 \pmod 4 \\ I & \text{se } n \equiv 0 \pmod 4 \end{cases}$

d) $A^n = \begin{bmatrix} \cos(n\alpha) & -\operatorname{sen}(n\alpha) \\ \operatorname{sen}(n\alpha) & \cos(n\alpha) \end{bmatrix}$

27. a) $\begin{bmatrix} 1 & -2 \\ -2 & 3 \end{bmatrix}$ b) $\begin{bmatrix} 5 & 5/2 \\ 5/2 & 4 \end{bmatrix}$

29. Todas falsas.

31. b)

33. a) $E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 5 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$

e $E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1/2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ b) $A^{-1} = E_3 E_2 E_1$

c) $A = E_1^{-1} E_2^{-1} E_3^{-1}$.

35. a) $A^{-1} = \begin{bmatrix} \frac{3}{2} & \frac{-11}{10} & \frac{-6}{5} \\ -1 & 1 & 1 \\ \frac{-1}{2} & \frac{7}{10} & \frac{-4}{5} \end{bmatrix}$.

b) $A^{-1} = \begin{bmatrix} \frac{-18}{5} & \frac{19}{10} & \frac{8}{5} \\ \frac{9}{5} & \frac{-7}{10} & \frac{-4}{5} \\ 2 & -1 & -1 \end{bmatrix}$.

c) $A^{-1} = \frac{1}{2}\begin{bmatrix} 1 & -1 & 1 \\ -1 & 1 & 1 \\ 1 & 1 & -1 \end{bmatrix}$.

d) $A^{-1} = \begin{bmatrix} \frac{7}{2} & 0 & -3 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$.

e) $A^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{-1}{3} & \frac{1}{3} & 0 & 0 \\ 0 & \frac{-1}{5} & \frac{1}{5} & 0 \\ 0 & 0 & \frac{-1}{7} & \frac{1}{7} \end{bmatrix}$.

37. $BA^2 - A = 0$.

39. Todas falsas.

41. b).

43. Todas falsas.

45. $\begin{bmatrix} B^{-1} & 0 \\ -D^{-1}CB^{-1} & D^{-1} \end{bmatrix}$.

Capítulo 2

1. a) $\operatorname{sign}(3,5,1,2,4) = -1$.
b) $\operatorname{sign}(2,4,1,5,3) = +1$.
c) $\operatorname{sign}(4,3,2,5,1) = -1$

3. a) -189 b) $\frac{-8}{7}$ c) $\frac{-1}{56}$ d) 7.

5. $\begin{vmatrix} b+c & c+a & b+a \\ a & b & c \\ 1 & 1 & 1 \end{vmatrix} = \overset{P8}{=} \begin{vmatrix} a+b+c & a+b+c & a+b+c \\ a & b & c \\ 1 & 1 & 1 \end{vmatrix}$ $= 0$ (a primeira linha é igual à terceira linha multiplicada por $(a+b+c)$).

7. Falso. Um contra-exemplo é: $A = I$ e $B = \begin{bmatrix} -1 & 0 \\ 0 & 0 \end{bmatrix}$.

9. a) $k = -1$ b) $k = \frac{5\pm\sqrt{17}}{2}$

11. a) $\frac{-4}{3}$ b) -48 c) $\frac{1}{288}$.

13. a) $\det(A) = 3$ ou $\det(A) = -3$. b) Se $\det(A) = 3$, então $b_{32} = 1/3$; se $\det(A) = -3$, então $b_{3,2} = -1/3$.

15. a) $x = 1$ b) $x = 0, x = 1$.

17. b).

19. a) F b) V c) V d) V.

Capítulo 3

1. Exemplos de conjuntos geradores:
a) $\{(-1,2,0,0),(1,0,1,0),(0,0,0,1)\}$.
b) $\{(6,-2,1,0),(-1,0,0,1)\}$.
c) $\{(1,0,0,0)\}$.

3. a) Sim. b) Sim. c) Não. d) Sim.

5. A-Não. B-Sim. C- Sim. D-Sim. E-Sim.

7. Por exemplo, $A = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}$ e $\mathbf{b} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$.

9. Por exemplo, $\begin{bmatrix} 4 & 2 & 6 \\ 2 & 1 & 3 \\ 6 & 3 & 3 \end{bmatrix}$.

11. a) F. b) F. c) V. d) F.

13. a) Sim, b) Não c) Sim d) Não.

15. Por exemplo,
$EC(A) : \{(1,0,0),(-3,1,0)\}$;
$EL(A) : \{(1,-3,0,1),(0,1,0,-1)\}$;
$N(A) : \{(2,1,0,1),(0,0,1,0)\}$;
$\dim EC(A) = \dim EL(A) = \dim N(A) = 2$.
$EC(B) : \{(2,6),(5,-8)\}$;
$EL(B) : \{(2,1,5),(6,3,-8)\}$;
$N(B) : \{(1,-2,0)\}$;
$\dim EC(B) = \dim EL(B) = 2$ e $\dim N(B) = 1$.
$EC(C) : \{(2,3,5),(-1,1,0)\}$;
$EL(C) : \{(2,-1,0,1),(3,1,-1,0)\}$;
$N(C) : \{(1,2,5,0),(-1,3,0,5)\}$;
$\dim EC(C) = \dim EL(C) = 2$ e $\dim N(C) = 2$.

17. c)

19. a) F. b) V. c) F. d) F. e) V.

21. a) F. b) V. c) V. d) F.

23. a) $(14,-2)$ b) $(4,-2,10)$.

25. $\begin{bmatrix} 2/7 & -3/14 \\ 1/7 & 1/7 \end{bmatrix}$ e $\mathbf{w}_B = (\frac{1}{7}, \frac{4}{7})$.

27. $3\begin{bmatrix} 2 & 1 \\ 0 & 4 \end{bmatrix} - 4\begin{bmatrix} 1 & -1 \\ 0 & 3 \end{bmatrix} + \begin{bmatrix} 3 & 2 \\ 0 & 5 \end{bmatrix}$. Sim, formam uma base.

29. c)

31. d)

33. a) Por exemplo, $BC = (1, t, t^2, t^3)$.
$\mathbf{p}_B = (1, 0, -1, 0)$
b) Não é base já que, $\dim V = 4$ e S tem 5 vectores.
c) $\dim S = 3$ e uma base é:
$\{1 - 2t, 1 + t^2, t\}$.
d) Não é subespaço de V (o zero de V não pertence a W).

35. a) $B_{U+V} = \{(1, 0, 0, 0), (1, 0, 2, 0)\}$,
$U \cap V = \{(0, 0, 0, 0)\}$;
$\dim U + V = 2, \quad \dim U \cap V = 0$.
b) $B_{U+V} = \{(-2, 0, 0, 1), (1, 1, 1, 0), (1, 0, -1, 0)\}$,
$B_{U \cap V} = \{(-1, 1, 1, 1)\}$;
$\dim U + V = 3, \quad \dim U \cap V = 1$.

Capítulo 4

1. a) Sim; $\lambda = 3$. b) Sim; $\lambda = 2$. c) Não. d) Não. e) Sim; $\lambda = 1$.

3. a) F. b) V. c) F.

5. $\lambda = 6$ e $\lambda = 0$.
$B = \{(1, 1, 1), (-1, 1, 0), (-1, 0, 1)\}$.

7. a) $\operatorname{car}(A) = 2$. b) $\det(A^T A) = 0$. c) $1, 1/3$ e $1/4$.

9. a) Não. b) Sim (e.g. $P = I$). c) Não. d) Sim; $P = \begin{bmatrix} 1 & -2 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ e $D = \operatorname{diag}(2, 3, 3)$.
e) Sim; $P = \begin{bmatrix} 1 & 1 & 0 \\ i & -i & 0 \\ 0 & 0 & 1 \end{bmatrix}$ e $D = \operatorname{diag}(i, -i, 2)$.

11. a) V. b) F. c) F. d) F. e) V. f) F. g) F. h) V.

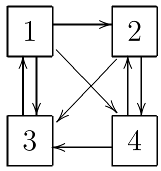
13. a) V. b) V. c) F. d) F. e) V. f) V.

15. b)

17. a) A soma por colunas é igual a 1 logo $\lambda = 1$ é valor próprio. Como o traço é 1.92, então o outro valor próprio é 0.92. b) $A\mathbf{v} = \mathbf{v}$ e $A\mathbf{u} = 0.92\mathbf{u}$, logo $\mathbf{u}$ e $\mathbf{v}$ são vectores próprios associados respectivamente a 0.92 e 1. c) Ver (4.15) d) $\frac{1}{8}\begin{bmatrix} 5 \\ 3 \end{bmatrix}$.

19. $(0, 0, 1)$.

21. a)



b) $(0.30, 0.20, 0.30, 0.20)$

23. 34.012 m

25. a) $c_1 e^{(1+\sqrt{5})t}\begin{bmatrix} 1 \\ \sqrt{5} \end{bmatrix} + c_2 e^{(1-\sqrt{5})t}\begin{bmatrix} 1 \\ -\sqrt{5} \end{bmatrix}$.
b) $c_1 e^{(2+\sqrt{3})t}\begin{bmatrix} 2 \\ \sqrt{3} - 1 \end{bmatrix} +$
$+ c_2 e^{(2-\sqrt{3})t}\begin{bmatrix} 2 \\ -1 - \sqrt{3} \end{bmatrix}$.
c) $c_1 e^{2t}\begin{bmatrix} 2\cos t - 4\,\mathrm{sen}\, t \\ 3\cos t - \mathrm{sen}\, t \\ \cos t \end{bmatrix} +$
$+ c_2 e^{2t}\begin{bmatrix} 2\,\mathrm{sen}\, t + 4\cos t \\ \cos t + 3\,\mathrm{sen}\, t \\ \mathrm{sen}\, t \end{bmatrix} + c_3 e^{-t}\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.

27. a) $e^t\begin{bmatrix} \cos t + \mathrm{sen}\, t & -2\,\mathrm{sen}\, t \\ \mathrm{sen}\, t & \cos t - \sin t \end{bmatrix}$
b) $\begin{bmatrix} e^t & -e^t + e^{3t} & 0 \\ 0 & e^{3t} & 0 \\ 0 & 0 & e^t \end{bmatrix}$

c) $\begin{bmatrix} e^{2t} & 0 & 0 \\ e^{3t}-e^{2t} & e^{3t} & 0 \\ 0 & 0 & e^{t} \end{bmatrix}$

29. $t^2 + c_1e^{3t} + c_2e^{-2t}$.

Capítulo 5

1. a) $3\sqrt{11}$ b) $3+\sqrt{58}$ c) $9-3\sqrt{58}$
d) $\frac{1}{3}(1,2,0,2)$ e) 1 f) arccos $\left(\frac{16}{3\sqrt{58}}\right)$.

3. a) $\frac{-7}{2}$ b) $\frac{-3}{8}$.

5. $\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + 2\langle \mathbf{u},\mathbf{v}\rangle$
$\|\mathbf{u}-\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\langle \mathbf{u},\mathbf{v}\rangle$.
Somando estas igualdades obtém-se o resultado.

A soma dos quadrados dos comprimentos das diagonais de um paralelogramo é igual ao dobro da soma dos quadrados dos comprimentos das suas arestas.

7. $k=8$.

9. a) i)F. ii) F. b) $\{(1,2,-1,2),(0,-1,3,-2)\}$
c) $\{(1,3,1),(2,5,1)\}$.

11. a) $\{(5,-4,1)\}$
b) $\{(-2,1,0,0),(-3,0,1,0),(-4,0,0,1)\}$.

13. $\frac{\sqrt{6}}{6}$ e $\frac{1}{6}(11,2,7)$.

15. $2x-y+4z=0$.

17. a) Não b) $\frac{5+i}{5}(1+2i,1-2i)$.

19. b).

21. $\{\frac{1}{\sqrt{3}}(1,1,1), \frac{\sqrt{3}}{3\sqrt{2}}(1,1,-2), \frac{\sqrt{2}}{2}(1,-1,0)\}$.

23. b) $\dim U = 1$ e $\dim U^{\perp} = 3$.
c) $B_U = \left\{\frac{1}{\sqrt{2}}\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\right\}$
$B_{U^\perp} = \left\{\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\right\}$
d) $\operatorname{proj}_U A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ e $\operatorname{proj}_{U^\perp} A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$ e) $\operatorname{proj}_U A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. f) $\sqrt{5}$

25. a) Não b) $k=0$ c) Não

27. $\sqrt{\frac{2}{3}}$ e $\frac{1}{3}(2,0,4,1)$.

29. a) $\sqrt{3}$ b) $\frac{2}{\sqrt{3}}$ c) 2

31. Não

33. $\{a,b) \in \mathbb{R}^2 : \ b = 3/2 - 2a, \ a \in \mathbb{R}\}$

35. $18y - 7x = 90$

37. a) V. b) F. c) V. d) F. e) V.

Capítulo 6

1. a) Sim b) Não c) Sim d) Não e) Não f) Sim.

3. a) Não. b) Não. c) Não. d) Sim. e) Não.

5. $(-27,4,24)$

7. a) $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$; $(-1,-2,0),(0,0,1)$ e $(1,0,1)$.
b) $A = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$; $(-2,-1,0),(0,0,1)$ e $(0,1,1)$.

9. a) $(\mathbf{T}(\mathbf{v_1}))_B = (1,-2)_B$ e $(\mathbf{T}(\mathbf{v_2}))_B = (3,-2)_B$
b) $T(\mathbf{v}_1) = (3,-5)$ e $T(\mathbf{v}_2) = (5,1)$.
c) $\frac{1}{7}\begin{bmatrix} -3 & 8 \\ -23 & -4 \end{bmatrix}$ e
$T(x_1,x_2) = \frac{1}{7}(-3x_1+8x_2,-23x_1-4x_2)$.
d) $T(1,1) = \frac{1}{7}(5,-27)$.

11. a) $(S\circ T)(x,y) = (2x,-x+3y)$
b) $A = \begin{bmatrix} 2 & 0 \\ -1 & 3 \end{bmatrix}$. $(S\circ T)$ é um isomorfismo visto que $\det(A) \neq 0$.
c) $C = M^{-1}AM = \begin{bmatrix} -1 & 2 \\ -6 & 6 \end{bmatrix}$.

13. a) F b) V c) V d) F.

15. a)

17. a) Sim b) Sim.

19. a) F b) F c) V d) F.

21. a) $T(A+B) = (A+B)+(A+B)^T = (A+A^T)+(B+B^T) = T(A)+T(B)$

$T(\alpha A) = (\alpha A) + (\alpha A)^T = \alpha(A + A^T) = \alpha T(A)$

b) $\begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$.

c) $B_{Im(T)} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$

$B_{N(T)} = \left\{ \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \right\}$

d) Não é injectiva nem sobrejectiva.

e) $\left\{ \begin{bmatrix} 1 & 2-x \\ x & 1 \end{bmatrix} : x \in \mathbb{R} \right\}$.

23. a) $T(2,1,2) = (7,-1,-2)$

b) $A = \begin{bmatrix} 2 & -3 & 3 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$.

c) $B = \begin{bmatrix} 2 & -2 & 2 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix}$.

d) $C = \begin{bmatrix} 2 & -1 & 0 \\ 0 & -1 & -1 \\ 0 & 0 & -1 \end{bmatrix}$.

e) $D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$.

d) $A = M^{-1}B$; $A = CM$;
$A = M^{-1}DM$.

25. a) $A = M(T, BC, BC) = \begin{bmatrix} 5 & -3 & 0 \\ 1 & 0 & -1 \\ 1 & 0 & 0 \end{bmatrix}$.

b) $C = M(T, B, B') = \begin{bmatrix} 2 & 4 & 1 \\ 1 & 1 & 1 \\ -1 & 0 & 0 \end{bmatrix}$.

c)
$$\begin{tikzcd} \mathbb{R}^3_{BC} \arrow[r, "A"] \arrow[d, "M_{B\leftarrow BC}=P"'] & \mathbb{R}^3_{BC} \arrow[d, "S=M_{B'\leftarrow BC}"] \\ \mathbb{R}^3_{B} \arrow[r, "C"'] & \mathbb{R}^3_{B'} \end{tikzcd}$$

$S = M_{B'\leftarrow BC} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix}$

$P = M_{B\leftarrow BC} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix}$.

27. a) DPA. b) $DPAK^{-1}$. c) $QDPA$.

29. a) $M = \begin{bmatrix} 2 & 1 & 2 & -1 \\ 0 & 0 & 1 & -1 \\ -3 & -1 & 0 & -1 \\ 0 & 2 & 1 & 1 \end{bmatrix}$,

$\mathrm{car}(M) = 3$.

b) $\dim N(\phi^*) = 1$, e uma base para $N(\phi^*)$ é constituída por: $-6v^1 + 11v^2 - 4v^3 + v^4$.

Capítulo 7

1. a) Unitária e simétrica. b) Hermitiana. c) Se $b = 0$ é unitária e hermitiana. d) Nenhuma. e) Simétrica.

3. a) F. b) F. c) V. d) F. e) F. f) F. g) F. h) F. i) V. j) F.

5. $\langle \mathbf{x}, A\mathbf{x} \rangle = \mathbf{x}^T A\mathbf{x} = (A^T\mathbf{x})^T\mathbf{x} = \langle A^T\mathbf{x}, \mathbf{x} \rangle = \langle A\mathbf{x}, \mathbf{x} \rangle$.

7. Sendo a matriz diagonalizável, todos os seus valores próprios têm multiplicidades algébrica e geométrica iguais.

9. a) $U = \frac{1}{\sqrt{2}} \begin{bmatrix} i & -i \\ 1 & 1 \end{bmatrix}$, $D = \mathrm{diag}(1,3)$.

b) $U = \frac{1}{\sqrt{2}} \begin{bmatrix} i & 0 & -i \\ 1 & 0 & 1 \\ 0 & \sqrt{2} & 0 \end{bmatrix}$, $D = \mathrm{diag}(2,2,0)$.

11. a) Indefinida; $P = \frac{1}{\sqrt{5}} \begin{bmatrix} -1 & 2 \\ 2 & 1 \end{bmatrix}$.

b) Indefinida; $P = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}$.

c) Semidefinida negativa; $P = \frac{1}{\sqrt{5}} \begin{bmatrix} -2 & 1 \\ 1 & 2 \end{bmatrix}$.

d) Definida positiva; $P = \frac{1}{\sqrt{2}}\begin{bmatrix}1 & -1\\ 1 & 1\end{bmatrix}$.

13. a)V. b) V. c) F. d) F. e) V.

15. a) A é 3×2 e $\mathrm{car}(A) = 2$.

b) $N(A) = \{(0,0)\}$;

$B_{EL(A)}\left\{(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}), (\frac{-\sqrt{2}}{2}, \frac{\sqrt{2}}{2})\right\}$;

$B_{N(A^T)} = \{(\frac{-1}{3}, \frac{-2}{3}, \frac{2}{3})\}$;

$B_{EC(A)} = \left\{(\frac{2\sqrt{2}}{3}, \frac{-\sqrt{2}}{6}, \frac{\sqrt{2}}{6}), (0, \frac{-\sqrt{2}}{2}, \frac{\sqrt{2}}{2})\right\}$.

17. $A^{-1} = V\Sigma^{-1}U^T$ para $A = U\Sigma V^T$.

19. $|\det A| = |\det(U\Sigma V^T)| = |\pm\det(\Sigma)| = |\det(\Sigma)| = \sigma_1\sigma_2\cdots\sigma_n$.

21. Sendo $A = U\Sigma V^T$ uma SVD de A tem-se

$A^T = V\Sigma^T U^T,\ A^+ = V\Sigma^+ U^T$

$(A^T)^+ = U(\Sigma^T)^+ V^T$

$(A^+)^T = U(\Sigma^+)^T V^T$.

Dada a forma de Σ e de Σ^+ verifica-se $(\Sigma^T)^+ = (\Sigma^+)^T$.

23. a)V. b) F. c)V. d) F.

25. a) Do exercício anterior, para $\mathfrak{q} = (0, \mathbf{q})$ e $\mathfrak{p} = (0, \mathbf{q})$ tem-se

$\mathfrak{q}\mathfrak{p} = (-\langle \mathbf{q}, \mathbf{r}\rangle\ ,\ \mathbf{q} \times \mathbf{r})$

Logo, $\mathfrak{q}\mathfrak{p}$ é um quaternião imaginário puro se e só se $\langle \mathbf{q}, \mathbf{r}\rangle = 0$.

b) Se $\mathfrak{q} = (0, \mathbf{q})$ e $\|\mathbf{q}\| = 1$ tem-se

$\mathfrak{q}^2 = (-\langle \mathbf{q}, \mathbf{q}\rangle\ ,\ \mathbf{q} \times \mathbf{q}) = (-\|\mathbf{q}\|^2, 0)$, ou seja $\mathfrak{q}^2 = -1$.

27. $C_{12}^3 = -C_{21}^3 = 2, C_{13}^2 = -C_{31}^2 = -2$, $C_{23}^1 = -C_{32}^1 = 2$ e $C_{ij}^k = 0$ para $i = k$ ou $j = k$ ou $i = j$.

29. $e^{\sigma_1 t} = \frac{1}{2}\begin{bmatrix}e^t + e^{-t} & e^t - e^{-t}\\ e^t - e^{-t} & e^t + e^{-t}\end{bmatrix}$,

$e^{\sigma_2 t} = \frac{1}{2}\begin{bmatrix}e^t + e^{-t} & -i(e^t - e^{-t})\\ i(e^t - e^{-t}) & e^t + e^{-t}\end{bmatrix}$,

$e^{\sigma_3 t} = \begin{bmatrix}e^t & 0\\ 0 & e^{-t}\end{bmatrix}$.

Pertencem a $SU(2)$ porque $(-1)^j i\sigma_j = E_j \in \mathfrak{su}(2)$.

31. $M = 2\begin{bmatrix}0 & -1 & 1\\ 1 & 0 & -1\\ -1 & 1 & 0\end{bmatrix}$

Capítulo 8

1. a) F. b) V. c) F. d) V.

3. a) $J = \begin{bmatrix}2 & 1 & 0\\ 0 & 2 & 1\\ 0 & 0 & 2\end{bmatrix}$ e $P = \begin{bmatrix}1 & 0 & 0\\ -2 & -1 & 0\\ 2 & -2 & -1\end{bmatrix}$

b) $J = \begin{bmatrix}2 & 1 & 0\\ 0 & 2 & 0\\ 0 & 0 & 3\end{bmatrix}$ e $P = \begin{bmatrix}1 & 0 & 0\\ -2 & -1 & 0\\ 2 & -2 & -1\end{bmatrix}$

c) $J = \begin{bmatrix}3 & 1 & 0 & 0\\ 0 & 3 & 0 & 0\\ 0 & 0 & 4 & 1\\ 0 & 0 & 0 & 4\end{bmatrix}$ e $P = \begin{bmatrix}1 & 0 & 1 & 0\\ 0 & -1 & 0 & 0\\ 1 & 0 & 0 & 0\\ 1 & 1 & 0 & 1\end{bmatrix}$.

d) $J = \begin{bmatrix}3 & 1 & 0 & 0\\ 0 & 3 & 1 & 0\\ 0 & 0 & 3 & 0\\ 0 & 0 & 0 & 4\end{bmatrix}$ e $P = \begin{bmatrix}1 & 0 & 1 & 0\\ 0 & -1 & 0 & 0\\ 1 & 0 & 0 & 0\\ 1 & 1 & 0 & 1\end{bmatrix}$.

5. a) $e^{2t}\begin{bmatrix}1 - 2t + 3t^2 & t^2 - t & -t^2/2\\ -2(t + 3t^2) & 1 + 2t^2 & t + t^2\\ -16t + 6t^2 & -6t + 2t^2 & 1 + 2t - t^2\end{bmatrix}$.

b) $\begin{bmatrix}e^{2t}(1 - 2t) & -te^{2t} & 0\\ 4te^{2t} & (1 - 2t)e^{2t} & 0\\ e^{2t}(6 - 6e^t - 4t) & 2e^{2t}(1 - e^t - t) & e^{3t}\end{bmatrix}$.

c) $\begin{bmatrix}e^{4t} & e^{3t}(te^t - t) & e^{3t}(1 - e^t - te^t) & te^{4t}\\ 0 & e^{3t} & 0 & 0\\ 0 & -te^{3t} & e^{3t} & 0\\ 0 & e^{3t}(-1 + e^t - t) & e^{3t}(1 - e^t) & e^{4t}\end{bmatrix}$.

d) $\frac{1}{2}\begin{bmatrix}e^{3t}(2 + t^2) & -2te^{3t} & -t^2e^{3t}\\ -2te^{3t} & 2e^{3t} & 2te^{3t}\\ t^2e^{3t} & -2te^{3t} & -e^{3t}(t^2 - 2)\\ te^{3t}(2 + t) & 2e^{3t}(-1 + e^t - t) & e^{3t}(2 - 2e^t - 2t -\end{bmatrix}$

Em cada alínea a solução geral é $\mathbf{x}(t) = e^{At}\mathbf{c}$, com $\mathbf{c}$ um vector constante arbitrário.

Índice

Lista de símbolos

$\square$	Fim de demonstração
$L_i \leftrightarrow L_j$	Troca da linha i com a linha j.
αL_i	Multiplicar a linha i pelo escalar $\alpha \neq 0$.
$L_i + \alpha L_j$	Substituir a linha i pela soma da linha i com a linha j multiplicada pelo escalar α.
$A\mathbf{x} = \mathbf{b}$	Sistema de equações lineares com matriz dos coeficientes A e termo independente $\mathbf{b}$.
$[A\|\mathbf{b}]$	Matriz aumentada do sistema $A\mathbf{x} = \mathbf{b}$.
a_{ij}	Entrada na linha i e na coluna j da matriz $A = [a_{ij}]$.
$\mathrm{car}(A)$	Característica da matriz A.
I_n ou I	Matriz identidade de ordem n.
E_i	Matriz elementar.
A^T	Matriz transposta de A.
A^{-1}	Inversa da matriz A.
A^k	Potência de A.
$\mathrm{tr}(A)$	Traço da matriz A.
LU	factorização LU.
$\det(A)$	Determinante da matriz A.
$\mathrm{sign}\,\sigma$	Sinal da permutação σ.
C_{ij}	cofactor da entrada ij.
$cof(A)$	Matriz dos cofactores da matriz A.
$\mathrm{ad}(A)$	Matriz adjunta da matriz A.
$N(A)$	Núcleo de A.
$EC(A)$	Espaço das colunas de A.
$EL(A)$	Espaço das linhas de A.
Span	Espaço gerado por, ou expansão linear de.
dim	Dimensão.
$\mathbf{x}_B$	Vector das coordenadas de x na base ordenada B.
$U + V$	Soma de subespaços.
$U \oplus V$	Soma directa de subespaços.
$U \cap V$	Intersecção de subespaços.
$M_{B' \leftarrow B}$	Matriz de mudança da base B para a base B'.
$E(\lambda)$	Espaço próprio do valor próprio λ.
$mult\,alg(\lambda)$	Multiplicidade algébrica do valor próprio λ.
$mult\,geom(\lambda)$	Multiplicidade geométrica do valor próprio λ.
$\langle x, y\rangle$	Produto interno de x por y.
A^H	Transposta da matriz conjugada de A.
dist	Distância.
$\\|\cdot\\|$	Norma.
$\\|\cdot\\|_2$	Norma-2.
$\mathrm{proj}_x\, u$	Projecção ortogonal de u sobre x.

$S^{\perp}$	Complemento ortogonal do subespaço S.
u_S, ou $\mathrm{proj}_S\, u$	Projecção ortogonal do vector u sobre o subespaço S.
$\mathbf{u} \times \mathbf{v}$	Produto externo (ou vectorial) do vector $\mathbf{u}$ pelo vector $\mathbf{v}$.
$\mathbf{w} \cdot (\mathbf{u} \times \mathbf{v})$	Produto misto dos vectores $\mathbf{w}, \mathbf{u}$ e $\mathbf{v}$.
$Im(f)$	Contradomínio de f, ou imagem de f.
$N(f)$	Núcleo da função linear f.
f^{-1}	Inversa da função f.
$L(V, W)$	Espaço das funções lineares entre os espaços vectoriais V e W.
$M(T, B, B')$	Matriz que representa a função linear na base B do domínio e na base B' do conjunto de chegada.
V^*	Espaço dual.
T^*	Função dual da função linear T.
$\langle\langle \xi, v \rangle\rangle$	Imagem de v por $\xi \in V^*$.
$q_A(\mathbf{x})$	Forma quadrática com matriz associada A.
A^+	Pseudoinversa de A.
$\mathbb{H}$	Quaterniões.
$\mathfrak{g}$	Álgebra de Lie do grupo G.
$[A, B]$	Comutador de matrizes.