Instituto Superior Técnico
Departamento de Matemática

APONTAMENTOS DE ÁLGEBRA LINEAR

0. Introdução

Este texto consiste numa transcrição razoavelmente fiel das apresentações feitas nas aulas teóricas do curso de Álgebra Linear para o mestrado em engenharia Aeroespacial no primeiro semestre de 2018/2019 no IST. O seu objetivo é proporcionar uma referência precisa para a matéria coberta nas aulas teóricas e não substituir os livros de texto indicados na bibliografia na página da cadeira.

A Álgebra Linear é a parte da Matemática que estuda a resolução de equações lineares, ou mais geralmente, que estuda as funções lineares. Os sistemas lineares já vos são familiares. Eis um exemplo:

$$\begin{cases} 2x+3y-z+w=4 \\ -x+2z-w=1 \end{cases}$$

Estamos interessados em saber se um sistema tem solução e, em caso afirmativo, em descrever as soluções de uma forma conveniente. A resolução do sistema pode ser vista como uma questão relativa à função linear

$$f(x,y,z,w) = (2x+3y-z+w, -x+2z-w)$$

De facto a existência de solução é equivalente à afirmação que $(4,1)$ pertence à imagem de f e, quando a solução existe, o conjunto das soluções é a pré-imagem $f^{-1}(\{(4,1)\})$ do ponto $(4,1)$ pela função.

A Álgebra Linear está também fortemente ligada à Geometria. Considerando (x,y,z,w) como coordenadas num espaço euclidiano de dimensão 4 (o espaço-tempo por exemplo), podemos interpretar o sistema acima como descrevendo a interseção de dois (hiper)planos. O sistema terá solução se os hiperplanos se intersetam e nesse caso, o conjunto das soluções do sistema descreve os pontos da interseção.

A Álgebra Linear é ubíqua na Matemática e nas suas aplicações. Por exemplo, o algoritmo de busca de páginas da Google tem por base uma ideia muito simples de álgebra linear como iremos ver mais tarde. A Álgebra Linear que iremos estudar é também usada na compressão de dados e imagens e nas telecomunicações entre muitas outras aplicações. Por outro lado, a Álgebra Linear é também fundamental na Matemática porque as funções lineares servem de modelo (incrivelmente bem sucedido) para funções mais gerais. É essa a ideia do Cálculo, no qual as funções são estudadas recorrendo às suas aproximações lineares (ou derivadas). A Álgebra Linear será assim uma base fundamental para disciplinas de Matemática que estudarão posteriormente como o Cálculo de várias variáveis ou as Equações Diferenciais.

1. O método de Gauss

O método de Gauss é um método para resolver sistemas lineares cuja ideia é a simplificação do sistema através da eliminação sucessiva de variáveis.

Date: December 2, 2019.

Definição 1.1. *Um* sistema linear *de m equações a n incógnitas é uma expressão da forma*

$$
\begin{cases}
a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 \\
a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2 \\
\quad \vdots \\
a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m
\end{cases}
\tag{1}
$$

onde a_{ij}, x_j, b_i para $1 \leq i \leq m$, $1 \leq j \leq n$ denotam números reais (ou complexos). Os números a_{ij} chamam-se os coeficientes do sistema*, os x_i são as* incógnitas *e os b_i os* termos independentes*. Se os termos independentes são nulos (isto é $b_i = 0$ para todo o i) o sistema diz-se* homogéneo.

Estamos interessados em saber se um sistema admite soluções (isto é, se existem números $x_1, \ldots, x_n$ tais que as relações (1) são satisfeitas). Quando isto acontece diz-se que o sistema é *possível*, senão é *impossível*. Quando existem soluções, queremos descrevê-las. Em particular queremos saber se a solução é única (nesse caso diz-se que o sistema é *determinado*) ou não, caso em que o sistema se diz *indeterminado*.

Observe-se que um sistema homogéneo é sempre possível. Tem pelo menos a solução $x_j = 0$ para todo o j, que se chama a *solução trivial*.

Observação 1.2. *Toda a teoria que vamos desenvolver durante o próximo par de meses aplica-se mais geralmente. Os números reais ou complexos podem ser substituídos pelos elementos de qualquer* corpo *(um conjunto com duas operações - soma e multiplicação - que são comutativas, associativas, têm elemento neutro, a multiplicação é distributiva relativamente à soma, todos os elementos têm inverso relativamente à soma e todos os elementos excepto o elemento neutro da soma têm inverso multiplicativo). Um exemplo familiar de corpo além dos conjuntos $\mathbb{R}$ e $\mathbb{C}$ dos números reais e complexos com as suas operações habituais é o conjunto $\mathbb{Q}$ dos números racionais, também com a soma e produto habituais. Um exemplo menos habitual é o conjunto $\{0, 1\}$ com a soma e produto definidas tomando o resto da divisão por 2 da soma e produto usuais.*

O método da eliminação de Gauss é o seguinte algoritmo para simplificar um sistema de equações lineares:

(1) Identificar a primeira variável que ocorre de facto no sistema (isto é, que tem coeficiente não nulo nalguma das equações do sistema).
(2) Se o coeficiente dessa variável na primeira equação for nulo, trocar a primeira equação com outra na qual o coeficiente não é nulo
(3) Subtrair um múltiplo conveniente da primeira equação às restantes de forma a eliminar nelas a variável em questão (isto é tornar o coeficiente dessa variável nulo)
(4) Regressar ao passo (1) considerando apenas o sistema que se obtém esquecendo a primeira equação, a não ser que o sistema fique reduzido a uma única equação, caso em que o algoritmo termina.

Exemplo 1.3. *Considere-se o sistema*

$$\begin{cases} 0x_1 + 0x_2 + 2x_3 - x_4 = 5 \\ 0x_1 + x_2 + 0x_3 + 3x_4 = 1 \\ 0x_1 + 2x_2 + x_3 + x_4 = 2 \end{cases}$$

A primeira variável que ocorre no sistema é x_2. Uma vez que o coeficiente de x_2 na primeira equação é 0, trocamos a primeira equação com a segunda (também poderíamos trocar com a terceira). Obtemos então o sistema

$$\begin{cases} x_2 \qquad + 3x_4 = 1 \\ \qquad 2x_3 - x_4 = 5 \\ 2x_2 + x_3 + x_4 = 2 \end{cases}$$

Subtraímos agora à terceira equação o dobro da primeira para eliminar a variável x_2 obtendo

$$\begin{cases} x_2 \qquad + 3x_4 = 1 \\ \qquad 2x_3 - x_4 = 5 \\ \qquad x_3 - 5x_4 = 0 \end{cases}$$

Voltamos agora ao início mas consideramos apenas as duas últimas equações. A primeira variável é agora x_3 e o seu coeficiente na primeira linha (que é a segunda linha do sistema inicial) é não nulo, pelo que não é necessário trocar a ordem das equações. Subtraindo metade da segunda equação à terceira obtemos o sistema

$$\begin{cases} x_2 \qquad + 3x_4 = 1 \\ \qquad 2x_3 - x_4 = 5 \\ \qquad\quad -\frac{9}{2}x_4 = -\frac{5}{2} \end{cases} \tag{2}$$

O sistema (2) é fácil de resolver começando pela equação debaixo e substituindo repetidamente os resultados obtidos nas equações de cima: da última equação obtemos $x_4 = \frac{5}{9}$ e substituindo na segunda equação obtemos

$$2x_3 = 5 + \frac{5}{9} \Leftrightarrow x_3 = \frac{25}{9}$$

Finalmente substituindo na primeira equação (em geral precisaríamos também do valor de x_3 mas neste sistema isso não acontece) obtemos

$$x_2 = 1 - 3 \cdot \frac{5}{9} = -\frac{2}{3}$$

O conjunto das soluções do sistema é portanto

$$\{(x_1, -\tfrac{2}{3}, \tfrac{25}{9}, \tfrac{5}{9}) \colon x_1 \in \mathbb{R}\} \tag{3}$$

Em particular o sistema é possível e indeterminado.

É um desperdício de tempo escrever as variáveis durante a aplicação dos passos do algoritmo acima. Podemos apenas escrever os coeficientes e termos independentes dos vários sistemas. O procedimento aplicado no exemplo anterior pode então ser abreviado da seguinte forma:

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

$$\xrightarrow{L_3 - \frac{1}{2}L_1} \left[\begin{array}{cccc|c} 0 & 1 & 0 & 3 & 1 \\ 0 & 0 & 2 & -1 & 5 \\ 0 & 0 & 0 & -\frac{9}{2} & -\frac{5}{2} \end{array}\right] \tag{4}$$

As tabelas de números que aparecem acima chamam-se *matrizes* e são objetos fundamentais na álgebra linear. A linha a tracejado antes da última coluna destina-se a lembrar que estamos a resolver um sistema não homogéneo e que a última coluna é formada pelos termos independentes. Quando é claro do contexto a linha a tracejado é por vezes omitida. Quando o sistema é homogéneo a última coluna (formada só por 0s) é omitida.

Exemplo 1.4. *Vamos resolver o sistema*

$$\begin{cases} x + 3y + 2z = 0 \\ 4y + z = 2 \\ -2x - 2y - 3z = 1 \end{cases}$$

Aplicando o método de Gauss obtemos

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

A última equação do sistema descrito pela matriz em que termina o método de Gauss é $0x + 0y + 0z = -1$*, que é impossível. Conclui-se que o sistema inicial é impossível.*

Definição 1.5. *Sejam* m, n *números naturais. Uma* matriz $m \times n$ *de números reais ou complexos é uma função* $\{1, \ldots, m\} \times \{1, \ldots, n\} \to \mathbb{R}$ *(ou* $\mathbb{C}$*). É habitual representar uma tal função por uma tabela de números*

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \qquad \text{onde } a_{ij} \text{ é o valor da função em } (i, j).$$

m *é o* número de linhas *da matriz, enquanto que* n *é o* número de colunas. *Diz-se que uma matriz está* em escada de linhas *se todas as linhas nulas estão em baixo e se a primeira entrada não nula de cada linha, que se denomina por* pivot, *está para a esquerda do pivot da linha abaixo. Isto é,* $[a_{ij}]_{1 \leq i \leq m, 1 \leq j \leq n}$ *está em escada de linhas se quando*

$$a_{ij} = 0 \text{ para } j \leq k \quad (\text{sendo } 0 \leq k \leq n) \quad \Rightarrow \quad a_{i+1\ j} = 0 \text{ para } j \leq k + 1.$$

Note-se que, em termos das matrizes associadas aos sistemas, o que o método de Gauss faz é colocar a matriz do sistema em escada de linhas.

Após a aplicação do método de Gauss temos ainda que resolver iterativamente as equações do sistema, começando pela que está mais abaixo. Este processo pode ser feito de forma muito mais eficiente, efetuando operações semelhantes às do método de Gauss. Este novo algoritmo, uma continuação do método de Gauss, chama-se *Método de Gauss-Jordan* e consiste em, dada uma matriz em escada de linhas,

(1) Multiplicar cada linha não nula pelo inverso do pivot de forma a fazer o pivot igual a 1.
(2) Subtrair múltiplos apropriados das linhas acima de cada linha com pivot até que todas as entradas acima dos pivots fiquem nulas.

Vamos aplicar este algoritmo à matriz em escada de linhas (4) que resultou do Exemplo 1.3.

Exemplo 1.6.

$$\left[\begin{array}{cccc|c} 0 & 1 & 0 & 3 & 1 \\ 0 & 0 & 2 & -1 & 5 \\ 0 & 0 & 0 & -\frac{9}{2} & -\frac{5}{2} \end{array}\right] \xrightarrow[-\frac{2}{9}L_3]{\frac{1}{2}L_2} \left[\begin{array}{cccc|c} 0 & 1 & 0 & 3 & 1 \\ 0 & 0 & 1 & -\frac{1}{2} & \frac{5}{2} \\ 0 & 0 & 0 & 1 & \frac{5}{9} \end{array}\right] \xrightarrow[L_2+\frac{1}{2}L_1]{L_1-3L_3} \left[\begin{array}{cccc|c} 0 & 1 & 0 & 0 & -\frac{2}{3} \\ 0 & 0 & 1 & 0 & \frac{25}{9} \\ 0 & 0 & 0 & 1 & \frac{5}{9} \end{array}\right]$$

Recuperamos assim o conjunto das soluções (3) *obtido acima.*

Quando há muitas equações, o algoritmo de Gauss-Jordan é muito mais eficiente que o processo de substituições sucessivas que usámos antes.

Definição 1.7. *Diz-se que uma matriz está* em escada de linhas reduzida *se está em escada de linhas, os pivots são todos iguais a* 1 *e as entradas acima dos pivots são todas* 0*.*

O algoritmo de Gauss-Jordan coloca portanto uma matriz em escada de linhas numa matriz em escada de linhas reduzida.

Exemplo 1.8. *Vamos resolver o sistema homogéneo*

$$\begin{cases} y + 4w = 0 \\ x - 2y + 3z = 0 \\ 2x - 6y + 16w = 0 \end{cases}$$

Recorde-se que neste caso não incluímos a coluna de 0*s correspondente aos termos dependentes. Obtemos assim*

$$\begin{bmatrix} 0 & 1 & 0 & 4 \\ 1 & -2 & 3 & 0 \\ 2 & -6 & 0 & 16 \end{bmatrix} \xrightarrow{L_1 \leftrightarrow L_2} \begin{bmatrix} 1 & -2 & 3 & 0 \\ 0 & 1 & 0 & 4 \\ 2 & -6 & 0 & 16 \end{bmatrix} \xrightarrow{L_3-2L_1} \begin{bmatrix} 1 & -2 & 3 & 0 \\ 0 & 1 & 0 & 4 \\ 0 & -2 & -6 & 16 \end{bmatrix}$$

$$\xrightarrow{L_3+2L_2} \begin{bmatrix} 1 & -2 & 3 & 0 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & -6 & 24 \end{bmatrix} \xrightarrow{-\frac{1}{6}L_3} \begin{bmatrix} 1 & -2 & 3 & 0 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & -4 \end{bmatrix} \xrightarrow{L_1+2L_2} \begin{bmatrix} 1 & 0 & 3 & 8 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & -4 \end{bmatrix}$$

$$\xrightarrow{L_1-3L_3}\begin{bmatrix} 1 & 0 & 0 & 20 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & -4 \end{bmatrix}$$

Obtemos assim a seguinte solução para o sistema:

$$\begin{cases} x = -20w \\ y = -4w \\ z = 4w \end{cases} \qquad \textit{com } w \in \mathbb{R} \textit{ qualquer.}$$

Exemplo 1.9. *Vamos resolver o sistema linear homogéneo*

$$\begin{cases} x - y + 2z + w - v = 0 \\ 2x - 2y + z - w + 2v = 0 \\ x - y + 5z + 4w - 5v = 0 \end{cases}$$

Aplicando o método de Gauss-Jordan temos

$$\begin{bmatrix} 1 & -1 & 2 & 1 & -1 \\ 2 & -2 & 1 & -1 & 2 \\ 1 & -1 & 5 & 4 & -5 \end{bmatrix} \xrightarrow[L_3-L_1]{L_2-2L_1} \begin{bmatrix} 1 & -1 & 2 & 1 & -1 \\ 0 & 0 & -3 & -3 & 4 \\ 0 & 0 & 3 & 3 & -4 \end{bmatrix} \xrightarrow{L_3+L_2} \begin{bmatrix} 1 & -1 & 2 & 1 & -1 \\ 0 & 0 & -3 & -3 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\xrightarrow{-\frac{1}{3}L_2} \begin{bmatrix} 1 & -1 & 2 & 1 & -1 \\ 0 & 0 & 1 & 1 & -\frac{4}{3} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow{L_1-2L_2} \begin{bmatrix} 1 & -1 & 0 & -1 & \frac{5}{3} \\ 0 & 0 & 1 & 1 & -\frac{4}{3} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Ou seja, o conjunto solução deste sistema é

$$\{(y + w - \tfrac{5}{3}v, y, -w + \tfrac{4}{3}v, w, v) \colon y, w, v \in \mathbb{R}\}$$

Os dois exemplos acima ilustram a seguinte observação relativa à solução de sistemas homogéneos por este método:

- As colunas com pivots correspondem às *variáveis dependentes* do sistema que são expressas em função das restantes.
- As colunas sem pivots correspondem às *variáveis livres* cujo valor pode ser atribuído arbitrariamente numa solução.

Num sistema não homogéneo, o sistema é impossível se houver um pivot na última coluna (como acontece no Exemplo 1.4). Quando o sistema é possível, as colunas com pivot correspondem às variáveis dependentes e as restantes, com excepção da última, às variáveis livres.

Definição 1.10. *A* característica *de uma matriz*[1] *A é o número de pivots que se obtém ao aplicar o método de Gauss (ou Gauss-Jordan).*

Alternativamente a característica é o número de linhas não nulas na matriz que resulta da aplicação do método de Gauss (ou Gauss-Jordan). Ela dá-nos o número mínimo de equações necessárias para descrever a solução do sistema. Note-se que não é imediatamente

[1] Em inglês "rank of a matrix".

claro que a definição de característica faça sentido pois há alguma indeterminação no método de Gauss relativa à escolha das trocas de linha. Podia acontecer que escolhas diferentes durante a aplicação do algoritmo conduzissem a matrizes com números diferentes de pivots no final. Vamos ver que isso não pode acontecer, mas primeiro comecemos por analisar exatamente a razão pela qual os métodos de Gauss e Gauss-Jordan produzem sistemas equivalentes ao inicial.

Suponhamos que temos um sistema linear

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2 \\ \quad\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m \end{cases} \tag{5}$$

Se $(x_1, \ldots, x_n)$ é uma solução do sistema, então para qualquer escolha de $c_1, \ldots, c_m \in \mathbb{R}$ (ou $\mathbb{C}$ consoante os escalares que estejamos a considerar) a seguinte relação será verificada

$$c_1(a_{11}x_1 + \ldots + a_{1n}x_n) + \ldots + c_m(a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n) = c_1b_1 + \ldots + c_mb_m \tag{6}$$

A expressão (6) diz-se uma *combinação linear* das equações do (5). Obtém-se multiplicando a i-ésima equação pela constante c_i e somando as equações resultantes. Concretizando, a combinação linear com coeficientes 2 e -3 das equações

$$x + y = 3 \qquad 2x - 5y = 2$$

é a equação

$$2(x + y) - 3(2x - 5y) = 2 \cdot 3 - 3 \cdot 2 \Leftrightarrow -4x + 17y = 0$$

Observação 1.11. *O conceito de combinação linear é talvez o conceito central da Álgebra Linear. Informalmente, uma combinação linear de coisas é uma expressão que se obtém multiplicando cada coisa por um escalar e somando tudo. Por exemplo, admitindo que se pode multiplicar mamíferos por escalar e somá-los,* 2*morcego*-3*castor é uma combinação linear de mamíferos.*

Quando executamos um passo do algoritmo de Gauss ou Gauss-Jordan, as equações do novo sistema são (por definição do algoritmo) combinações lineares das do sistema anterior. Portanto uma solução do sistema antes da aplicação do passo é ainda uma solução do sistema seguinte. Na realidade, as combinações lineares envolvidas são muito simples. Chamando S ao sistema inicial e S' ao sistema obtido após aplicação de um passo do algoritmo e usando a notação L_i (respetivamente L_i') para a i-ésima equaçao do sistema S (respetivamente S'), temos após um passo do método

$$L_i' = L_j, \quad L_i' = \alpha L_i \text{ com } \alpha \neq 0, \quad \text{ ou } L_i' = L_i - \alpha L_j \text{ com } j \neq i$$

e, no último caso, as linhas distintas da i-ésima permanecem inalteradas e, em particular, $L_j' = L_j$.

Mas as expressões acima permitem também escrever as linhas do sistema S como combinações lineares das linhas de S':

$$L_j = L_i', \quad L_i = \frac{1}{\alpha} L_i' \text{ com } \alpha \neq 0, \quad \text{ ou } L_i = L_i' + \alpha L_j' \text{ com } j \neq i$$

(onde no último caso usámos o facto de L_j e L_j' serem iguais). Conclui-se que as soluções do sistema S' são também soluções do sistema S e portanto que os sistemas S e S' têm exatamente as mesmas soluções. Uma vez que isto acontece durante todas os passos do método conclui-se que *todos os sistemas que ocorrem ao longo da aplicação dos métodos de Gauss e Gauss-Jordan são equivalentes*, isto é, todos têm exatamente o mesmo conjunto de soluções.

Para terminar esta nossa discussão inicial dos sistemas lineares vamos agora provar que a matriz em escada de linhas reduzida no final do método de Gauss-Jordan é independente de quaisquer escolhas, o que mostra que a Definição 1.10 faz sentido (diz-se que a característica está *bem definida*).

A demonstração utilizará um género de argumento que se diz *por redução ao absurdo* e que se baseia no seguinte facto simples da lógica: Se uma afirmação P implica outra afirmação Q e Q é falsa, então P é necessariamente falsa. Em símbolos:

$$((P \Rightarrow Q) \wedge \neg Q) \Rightarrow \neg P$$

Este facto permite-nos provar a validade de uma afirnação A se conseguirmos deduzir uma falsidade a partir da sua negação $\neg A$. Conclui-se então que a afirmação $\neg A$ é falsa, ou seja que A é verdadeira.

Teorema 1.12. *Sejam m, n números naturais e A uma matriz $m \times n$ de números reais ou complexos. Se B e C são matrizes em escada de linhas reduzidas obtidas a partir de A por aplicação dos métodos de Gauss e Gauss-Jordan, então $B = C$.*

Dem. A demonstração é por indução no número n das colunas de A. Para a base da indução precisamos de mostrar que se A é uma matriz com uma única coluna o resultado é verdadeiro. Se A tem apenas uma coluna, ou é nula e então $B = C = 0$ (porque o algoritmo termina imediatamente) ou não é nula e então o algoritmo termina necessariamente com a matriz

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

Em qualquer caso $B = C$.

Para o passo da indução vamos admitir que a afirmação do enunciado é válida se a matriz A tem n colunas e queremos concluir que a afirmação é válida para matrizes com $n+1$ colunas. Vamos admitir por absurdo que isto não é verdade. Então existe uma matriz A com $n+1$ colunas e duas maneiras de aplicar o algoritmo de tal forma que no final obtemos matrizes $B \neq C$.

Quando aplicamos os algoritmos a uma matriz A estamos também a aplicá-los às matrizes que se obtêm de A suprimindo a última coluna (ou qualquer número de colunas à direita). Em particular, escrevendo $A_{\leq n}$ para a matriz que se obtém de A suprimindo a última coluna[2], pela hipótese de indução (uma vez que $A_{\leq n}$ tem n colunas) teremos $B_{\leq n} = C_{\leq n}$. Como estamos a admitir que $B \neq C$ estas matrizes terão que diferir em pelo menos uma das entradas na última coluna. Seja então i tal que $b_{i\ n+1} \neq c_{i\ n+1}$. Recorde-se que os sistemas homogéneos determinados por A, B, e C são equivalentes. Subtraindo as i-ésimas equações dos sistemas correspondentes a B e C obtemos a equação

$$(b_{i\ n+1} - c_{i\ n+1})x_{n+1} = 0$$

(uma vez que $b_{ij} = c_{ij}$ para $j \leq n$). Como o coeficiente de x_{n+1} é não nulo, isto significa que todas as soluções do sistema determinado por A (ou B ou C) satisfazem $x_{n+1} = 0$. Então x_{n+1} não é uma variável livre no sistema de B nem no sistema de C, e portanto tanto B como C têm um pivot na coluna $n + 1$.

Mas observe-se agora que numa matriz em escada de linhas reduzida, um pivot na última coluna ocorre exatamente à direita da primeira linha de 0s na matriz obtida ao suprimir a última coluna. Ou seja, sabendo que B e C têm um pivot na última coluna, a posição do pivot é determinada por $B_{\leq n} = C_{\leq n}$ e portanto é igual para B e C. Ora no final do método de Gauss-Jordan todas as entradas da última coluna são 0 excepto a entrada correspondente ao pivot, que é 1. Conclui-se então que as últimas colunas de B e de C são iguais e portanto $B = C$. Isto contradiz a nossa hipótese que $B \neq C$ e portanto mostra que é impossível obter matrizes distintas ao aplicar o algoritmo de Gauss-Jordan a uma matriz com $(n + 1)$ colunas. Isto conclui o passo de indução e portanto a demonstração. □

Observação 1.13. *A demonstração anterior mostra mais geralmente que se A, B e C são matrizes de sistemas tais que qualquer equação pode ser escrita como combinação linear das equações de cada um dos três sistemas, e B, C estão em escada de linhas reduzidas, então $B = C$. Em particular, se inserirmos trocas de linhas arbitrárias durante a aplicação do método de Gauss (mesmo que isso não seja requerido pelo algoritmo) isso não afectará o resultado do algoritmo de Gauss-Jordan.*

2. O produto de matrizes

Vimos acima que qualquer combinação linear (6) das equações de um sistema linear (5) é satisfeita por uma solução do sistema. Mais geralmente, começando com um sistema linear (5), podemos considerar um novo sistema cujas equações são combinações lineares das equações do sistema inicial. No caso homogéneo (ou seja com $b_i = 0$) um tal sistema

[2]Esta notação *ad hoc* não voltará a ser usada depois desta demonstração.

com k equações tem o aspecto seguinte

$$(7)\quad \begin{cases} c_{11}(a_{11}x_1 + a_{21}x_2 + \ldots + a_{1n}x_n) + \ldots + c_{1m}(a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n) = 0 \\ c_{21}(a_{11}x_1 + a_{21}x_2 + \ldots + a_{1n}x_n) + \ldots + c_{2m}(a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n) = 0 \\ \vdots \\ c_{k1}(a_{11}x_1 + a_{21}x_2 + \ldots + a_{1n}x_n) + \ldots + c_{km}(a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n) = 0 \end{cases}$$

onde $c_{i1}, \ldots, c_{im}$ sáo os coeficientes da combinação linear que produz a i-ésima equação do novo sistema. Estes escalares podem ser dispostos numa matriz $k \times m$.

$$\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1m} \\ c_{21} & c_{22} & \cdots & c_{2m} \\ \vdots & & & \vdots \\ c_{k1} & c_{k2} & \cdots & c_{km} \end{bmatrix}$$

Identificando o sistema inicial com a matriz $[a_{ij}]_{1 \leq i \leq m, 1 \leq j \leq n}$ dos seus coeficientes, podemos pensar neste processo de combinação linear de equações como uma operação que partindo de duas matrizes, $C = [c_{pq}]$ do tipo $k \times m$ e $A = [a_{ij}]$ de tipo $m \times n$ produz uma nova matriz que tem por entradas os coeficientes das equações do sistema (7). Esta nova matriz é de tipo $k \times n$ e tem como entrada ij (correspondente ao coeficiente de x_j na i-ésima equação de (7))

$$(8)\qquad c_{i1}a_{1j} + c_{i2}a_{2j} + \ldots + c_{im}a_{mj} = \sum_{l=1}^{m} c_{il}a_{lj}$$

Definição 2.1. *Sejam k, m, n números naturais, C uma matriz $k \times m$ e A uma matriz $m \times n$ de números reais (ou complexos). O* produto da matriz C pela matriz A *é a matriz $k \times n$, denotada por CA, cuja entrada ij é dada pela expressão* (8).

Note-se que a expressão (8) não é mais do que o *produto escalar* da linha i da matriz C com a coluna j da matriz A.

$$\begin{bmatrix} \vdots & & & \vdots \\ c_{i1} & c_{i2} & \cdots & c_{im} \\ \vdots & & & \vdots \end{bmatrix} \begin{bmatrix} \ldots & a_{1j} & \ldots \\ \ldots & a_{2j} & \ldots \\ & \vdots & \\ \ldots & a_{kj} & \ldots \end{bmatrix}$$

Exemplo 2.2.

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

$$= \begin{bmatrix} 2 \cdot 1 + 0 \cdot (-1) + 3 \cdot 0 & 2 \cdot 2 + 0 \cdot 1 + 3 \cdot 3 & 2 \cdot 0 + 0 \cdot (-1) + 3 \cdot 0 & 2 \cdot 0 + 0 \cdot 3 + 3 \cdot 1 \\ 1 \cdot 1 + (-1) \cdot (-1) + 0 \cdot 0 & 1 \cdot 2 + (-1) \cdot 1 + 0 \cdot 3 & 1 \cdot 0 + (-1) \cdot (-1) + 0 \cdot 0 & 1 \cdot 0 - 1 \cdot 3 + 0 \cdot 1 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 13 & 0 & 3 \\ 2 & 1 & 1 & -3 \end{bmatrix}$$

A fórmula (8) para o produto de matrizes admite várias interpretações que facilitam muitas vezes o cálculo e que são já patentes no exemplo anterior:

- A i-ésima linha do produto CA é a combinação linear das linhas de A cujos coeficientes são as entradas da i-ésima linha de C (foi esta aliás a maneira como chegámos à fórmula para o produto de matrizes). Concretamente, no exemplo acima, a primeira linha do produto é igual a
$$2\cdot\begin{bmatrix} 1 & 2 & 0 & 0 \end{bmatrix} + 0\cdot\begin{bmatrix} -1 & 1 & -1 & 3 \end{bmatrix} + 3\cdot\begin{bmatrix} 0 & 3 & 0 & 1 \end{bmatrix}$$
- A j-ésima coluna do produto CA é a combinação linear das colunas de C cujos coeficientes são as entradas da j-ésima coluna de A. No exemplo acima, a primeira coluna do produto é igual a
$$1\cdot\begin{bmatrix} 2 \\ 1 \end{bmatrix} - 1\cdot\begin{bmatrix} 0 \\ -1 \end{bmatrix} + 0\cdot\begin{bmatrix} 3 \\ 0 \end{bmatrix}$$

Em muitos exemplos (como no Exemplo 2.2 acima) o produto calcula-se muito mais rapidamente fazendo as contas por linhas ou colunas do que aplicando a fórmula (8) entrada a entrada.

Usando o produto de matrizes, podemos escrever um sistema (5) usando matrizes para os coeficientes, incógnitas e termos independentes. A expressão (5) é equivalente à igualdade de matrizes

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} \tag{9}$$

que se pode abreviar

$$AX = B$$

Uma vez que entendamos as propriedades do produto de matrizes, poderemos manipular sistemas e resolvê-los de forma análoga à que é já familiar do estudo anterior da resolução de equações numéricas.

Os métodos de Gauss e Gauss-Jordan podem também ser descritos em termos do produto de matrizes. Por exemplo, tendo em conta a descrição do produto de matrizes em termos de combinação linear de linhas, a aplicação da operação $L_2 + 3L_1$ ao sistema (9) consiste na multiplicação em ambos os lados da igualdade pela matriz do tipo $m \times m$

$$\begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 3 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

De forma semelhante, a operação $-2L_2$ corresponde à multiplicação de (9) pela matriz $m \times m$

$$\begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & -2 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

Definição 2.3. *Seja n um número natural. A* matriz identidade *do tipo $n \times n$ é a matriz I_n que tem como entrada ij*

$$\delta_{ij} = \begin{cases} 1 & \text{se } i = j \\ 0 & \text{se } i \neq j \end{cases}$$

ou seja

$$I_n = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

Teorema 2.4 (Propriedades do produto de matrizes)**.** *Sejam k, m, n, p números naturais e A, B, C matrizes do tipo $k \times m$, $m \times n$ e $n \times p$ respetivamente.*

(i) Propriedade associativa do produto: $A(BC) = (AB)C$.
(ii) Elemento neutro para o produto: $I_k A = A$ e $AI_m = A$.

Dem. (i) Temos a verificar que para cada i, j com $1 \leq i \leq k$ e $1 \leq j \leq p$, a entrada ij das matrizes $A(BC)$ e $(AB)C$ são iguais. Escrevendo $(AB)_{ij}$ para a entrada ij do produto das matrizes A e B e aplicando (duas vezes) a fórmula (8) que define o produto de matrizes obtemos

$$\begin{aligned} (A(BC))_{ij} &= \sum_{x=1}^{m} a_{ix}(BC)_{xj} \\ &= \sum_{x=1}^{m} a_{ix} \left(\sum_{y=1}^{n} b_{xy} c_{yj} \right) \\ &= \sum_{x=1}^{m} \sum_{y=1}^{n} a_{ix} b_{xy} c_{yj} \end{aligned}$$

onde na última igualdade aplicámos as propriedades distributiva da soma em relação ao produto (de números) e também as propriedade associativas da soma e multiplicação (de números). De forma inteiramente análoga temos

$$\begin{aligned}((AB)C))_{ij} &= \sum_{z=1}^{n}(AB)_{iz}c_{zj} \\ &= \sum_{z=1}^{n}\left(\sum_{w=1}^{m} a_{iw}b_{wz}\right)c_{zj} \\ &= \sum_{z=1}^{n}\sum_{w=1}^{m} a_{iw}b_{wz}c_{wj}\end{aligned}$$

As expressões obtidas para $(A(BC))_{ij}$ e $((AB)C)_{ij}$ são idênticas[3] (pelas propriedades associativa e comutativa da soma de números) o que conclui a demonstração da igualdade $A(BC) = (AB)C$.

(ii) A demonstração é análoga (mas mais fácil). Exercício.

□

Na proposição anterior vimos propriedades importantes que a multiplicação de matrizes partilha com a multiplicação de números, (embora seja importante notar que a complexidade da multiplicação de matrizes é superior: há matrizes de vários tipos e só quando o número de linhas do fator da esquerda é igual ao número de colunas do fator da direita se pode efetuar a multiplicação). Há também diferenças importantes:

Exemplo 2.5 (A multiplicação de matrizes não é comutativa)**.** *Note-se que os produtos AB e BA só poderão ser matrizes do mesmo tipo se A e B forem matrizes quadradas com igual número de linhas. Se escolhermos duas destas matrizes ao acaso (com mais de uma linha!), a probabilidade de os produtos serem diferentes é 100%. Por exemplo,*

$$\begin{bmatrix} 1 & 2 \\ 3 & -1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ -2 & 1 \end{bmatrix} = \begin{bmatrix} -3 & 3 \\ 5 & 2 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 \\ -2 & 1 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 3 & -1 \end{bmatrix} = \begin{bmatrix} 4 & 1 \\ 1 & -5 \end{bmatrix}$$

Uma das propriedades da multiplicação de números que é muito útil é a chamada *lei do corte*:

Se $a \neq 0$ e $ab = ac$ então $b = c$.

Definição 2.6. *A matriz $m \times n$* nula *é a matriz que tem todas as entradas iguais a* 0*. É denotada por* 0 *(deixando implícitas as dimensões).*

É imediato da definição do produto que (sempre que os produtos façam sentido) temos

$$A \cdot 0 = 0 \qquad 0 \cdot A = 0$$

[3]Os índices dos somatórios são *variáveis mudas*. Obtém-se uma expressão da outra substituindo o índice x por w e y por z.

Exemplo 2.7 (A lei do corte não é válida para o produto de matrizes)**.** *Seja A a matriz* $\begin{bmatrix} 2 & -1 \\ 4 & -2 \end{bmatrix}$. *Entâo*

$$A^2 \overset{def}{=} AA = \begin{bmatrix} 2 & -1 \\ 4 & -2 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 4 & -2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

portanto, apesar de $A \neq 0$ temos

$$AA = A \cdot 0.$$

Definição 2.8. *Uma matriz $n \times n$, A diz-se* invertível *se existe uma matriz B (necessariamente também $n \times n$) tal que*

$$AB = BA = I_n$$

Uma tal matriz B diz-se uma inversa *de A.*

Proposição 2.9. *Seja A uma matriz $n \times n$* invertível, *B, C matrizes $n \times m$ e E, F matrizes $m \times n$. Então*

$$AC = AD \Rightarrow C = D \qquad e \qquad EA = FA \Rightarrow E = F$$

Dem. Provamos apenas a primeira implicação deixando a segunda como exercício. Seja B uma inversa de A. Então

$$AC = AD \Rightarrow B(AC) = B(AD) \Leftrightarrow (BA)C = (BA)D \Leftrightarrow I_nC = I_nD \Leftrightarrow C = D$$

□

Vamos também necessitar de outras operações com matrizes que têm uma natureza muito mais elementar do que o produto.

Definição 2.10. *Sejam A, B matrizes $m \times n$. A* soma *das matrizes A e B é a matriz do mesmo tipo $A + B$ que tem como entrada ij*

$$(A + B)_{ij} = a_{ij} + b_{ij}$$

O produto de uma matriz A $m \times n$ pelo escalar $\lambda \in \mathbb{R}$ (ou $\mathbb{C}$) *é a matriz λA também do tipo $m \times n$ cuja entrada ij é*

$$(\lambda A)_{ij} = \lambda a_{ij}$$

Por exemplo

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

e

$$\sqrt{2} \begin{bmatrix} 1 & 1 \\ -1 & 2 \\ 4 & 0 \end{bmatrix} = \begin{bmatrix} \sqrt{2} & \sqrt{2} \\ -\sqrt{2} & 2\sqrt{2} \\ 4\sqrt{2} & 0 \end{bmatrix}$$

Vejamos algumas propriedades fundamentais destas operações cujas demonstrações são imediatas e ficam como exercício.

Proposição 2.11 (Propriedades da soma de matrizes)**.** *Sejam A, B, C matrizes $m \times n$. Então*

(i) (Associatividade) $A + (B + C) = (A + B) + C$
(ii) (Comutatividade) $A + B = B + A$
(iii) (Existência de elemento neutro) $A + 0 = A$
(iv) (Existência de inversos/simétricos) Existe D *tal que* $A + D = 0$

É fácil verificar (exercício) que o simétrico de uma matriz é único. Usa-se a notação $-A$ para o simétrico de uma matriz e claramente a componente ij da matriz $-A$ é dada por $-a_{ij}$.

Proposição 2.12 (Propriedades do produto por escalar)**.** *Sejam A, B matrizes $m \times n$ e λ, μ escalares reais (ou complexos). Então*

(i) $1 \cdot A = A$
(ii) $\lambda(\mu A) = (\lambda\mu)A$
(iii) $\lambda(A + B) = \lambda A + \lambda B$
(iv) $(\lambda + \mu)A = \lambda A + \mu A$

Outras propriedades do produto por escalar que são muitas vezes utilizadas são as seguintes

$$0 \cdot A = 0, \qquad (-1) \cdot A = -A$$

Estas propriedades são de verificação imediata a partir da definição do produto por escalar mas podem também ser deduzidas das propriedades indicadas nas Proposições acima (sem usar a definição). Fica como exercício a realização dessas deduções.

Vejamos agora algumas relações entre a soma e o produto por escalar com o produto de matrizes.

Proposição 2.13 (Distributividade)**.** *Sejam A uma matrix $m \times n$, B e C matrizes $n \times p$ e D uma matriz $p \times q$. Então*

$$A(B + C) = AB + AC \qquad (B + C)D = BD + CD$$

Dem. Verificamos apenas a primeira igualdade dado que a demonstração da segunda é inteiramente análoga. Temos que ver que para cada i, j com $1 \leq i \leq m$ e $1 \leq j \leq p$, as entradas ij das matrizes $A(B + C)$ e $AB + AC$ são iguais. De acordo com (8) a entrada ij de $A(B + C)$ é dada pela expressão

$$\begin{aligned}
\sum_{k=1}^{n} a_{ik}(B + C)_{kj} &= \sum_{k=1}^{n} a_{ik}(b_{kj} + c_{kj}) \\
&= \sum_{k=1}^{n} a_{ik}b_{kj} + a_{ik}c_{kj} \\
&= (AB)_{ij} + (AC)_{ij}
\end{aligned}$$

o que mostra a igualdade pretendida. $\square$

Podemos usar as propriedades acima para desenvolver e simplificar expressões como estamos habituados a fazer com os números mas devido às diferenças indicadas acima, isto requer algum cuidado. Por exemplo, se A e B são matrizes $n \times n$ temos

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

Esta expressão é (pela lei do corte para a soma de matrizes) igual à expressão habitual

$$A^2 + 2AB + B^2$$

se e só se for satisfeita a seguinte igualdade pelas matrizes A, B

$$AB = BA$$

o que, como já indicámos acima, quase nunca se verifica.

Definição 2.14. *Sejam A, B matrizes $n \times n$. Diz-se que A e B* comutam *se $AB = BA$.*

É imediato verificar que a matriz λI_n comuta com qualquer outra matriz $n \times n$, uma vez que, pela interpretação do produto de matrizes em termos de combinações lineares de linhas e colunas, multiplicar A à esquerda por λI_n consiste em multiplicar cada linha de A por λ, enquanto que multiplicar por λI_n à direita consiste em multiplicar por λ cada coluna de A. Portanto

$$(\lambda I_n)A = \lambda A = A(\lambda I_n)$$

Um dos exercícios da ficha para as aulas práticas da próxima semana pede-vos que verifiquem que estas matrizes - os múltiplos escalares da matriz identidade - são na realidade as únicas matrizes que têm esta propriedade de comutar com todas as outras. A igualdade acima é um caso particular da seguinte propriedade que relaciona o produto de matrizes com o produto por escalar. A demonstração (muito fácil) é deixada como exercício.

Proposição 2.15. *Sejam A uma matriz $m \times n$, B uma matriz $n \times p$ e λ um escalar real (ou complexo). Então*

$$\lambda(AB) = A(\lambda B) = (\lambda A)B$$

Exemplo 2.16. *Seja A uma matriz $n \times n$. Então (uma vez que $3I_n$ comuta com A)*

$$(A + 3I_n)^2 = A^2 + 2(3I_n)A + (3I_n)^2 = A^2 + 6A + 9I_n$$

Já vimos que a invertibilidade de uma matriz é uma propriedade útil, permitindo-nos por exemplo a aplicação da lei do corte.

Proposição 2.17 (Unicidade da inversa)**.** *Seja A uma matriz $n \times n$. Se B e C são inversas de A então $B = C$.*

Dem. Temos

$$B = BI_n = B(AC) = (BA)C = I_nC = C$$

□

A partir de agora escrevemos

$$A^{-1} \quad \text{para } a \text{ inversa da matriz } A.$$

Notemos as seguintes consequências da unicidade da inversa.

Proposição 2.18. *Sejam A, B matrizes $n \times n$ invertíveis. Então*

(i) AB é invertível e $(AB)^{-1} = B^{-1}A^{-1}$
(ii) A^{-1} é invertível e $(A^{-1})^{-1} = A$.

Dem. Mostramos apenas a primeira afirmação deixando a segunda como exercício. Uma vez que a inversa é única, tudo o que é necessário fazer é verificar que as relações na Definição 2.8 são satisfeitas:

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

e, analogamente,

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AI_nA^{-1} = AA^{-1} = I_n$$

□

Põe-se agora a questão de como saber se uma matriz é invertível e nesse caso calcular a matriz inversa. Na realidade já aprendemos a calcular a inversa! Se B é a inversa de A então

$$AB = I_n$$

Tendo em conta a interpretação do produto AB como um cálculo de combinações lineares de colunas de A, isto diz-nos que as entradas da i-ésima coluna de A são os coeficientes da combinação linear das colunas de A que produz a i-ésima coluna da matriz identidade. Se denotarmos a i-ésima coluna de B por X_i, isto diz-nos que a seguinte relação é satisfeita

$$AX_i = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{10}$$

(onde a entrada não nula da matriz à direita está na i-ésima linha). Assim *podemos calcular a i-ésima coluna da inversa resolvendo o sistema linear* (10) para o que podemos usar os métodos de Gauss e Gauss-Jordan. Para calcular a inversa temos que resolver n sistemas lineares mas não há qualquer razão para o fazer separadamente. Como os coeficientes do sistema são os mesmos para todos os sistemas podemos resolver todos ao mesmo tempo:

Exemplo 2.19. *Vamos calcular A^{-1} para a matriz $A = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 3 & 0 \\ 4 & 0 & 5 \end{bmatrix}$*

Aplicamos o método de Gauss-Jordan aos sistemas com termos independentes $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ *e* $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ *simultaneamente:*

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

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

As colunas da matriz à direita são as soluções de cada um dos sistemas e portanto as colunas da matriz inversa. Assim, se a matriz A for invertível então teremos necessariamente

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

Exemplo 2.20. *Vamos calcular A^{-1} para a matriz* $A = \begin{bmatrix} 1 & 3 & 1 \\ 0 & -1 & 0 \\ 2 & 0 & 1 \end{bmatrix}$

Temos

$$\left[\begin{array}{ccc|ccc} 1 & 3 & 1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 \\ 2 & 0 & 1 & 0 & 0 & 1 \end{array}\right] \xrightarrow{L_3-2L_1} \left[\begin{array}{ccc|ccc} 1 & 3 & 1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 \\ 0 & -6 & -1 & -2 & 0 & 1 \end{array}\right]$$

$$\xrightarrow{L_3-6L_2} \left[\begin{array}{ccc|ccc} 1 & 3 & 1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & -2 & -6 & 1 \end{array}\right] \xrightarrow[-L_3]{-L_2} \left[\begin{array}{ccc|ccc} 1 & 3 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 2 & 6 & -1 \end{array}\right]$$

$$\xrightarrow{L_1-L_3} \left[\begin{array}{ccc|ccc} 1 & 3 & 0 & -1 & -6 & 1 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 2 & 6 & -1 \end{array}\right] \xrightarrow{L_1-3L_2} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -1 & -3 & 1 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 2 & 6 & -1 \end{array}\right]$$

Assim, se a matriz A for invertível então teremos necessariamente

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

Resta perceber porque é que a matriz B calculada nos exemplos anteriores é de facto uma inversa de A. A maneira como foi determinada torna claro que $AB = I_n$, mas para que B seja a inversa é ainda necessário que $BA = I_n$. Isto está longe de ser óbvio (embora seja fácil de verificar nos exemplos acima ou em qualquer exemplo concreto).

Antes de explicar a razão pela qual o método anterior pode ser sempre usado para achar a inversa (ou ver que uma matriz não é invertível) vamos primeiro responder à seguinte pergunta natural: Porque não achar a inversa por linhas resolvendo o sistema determinado pela equação $BA = I_n$ linha a linha? De facto podemos fazê-lo, mas a matriz dos coeficientes do sistema não será A, e dado que o método de Gauss-Jordan (tal como nós o apresentámos) se aplica imediatamente apenas à solução de sistemas $Ax = b$ com x e b matrizes coluna, é mais prático fazer as contas como fizemos acima.

Esta questão aponta no entanto para um aspeto básico do cálculo matricial que diz respeito à simetria entre linhas e colunas. A atribuição do primeiro índice às linhas e do segundo às colunas é claramente apenas uma convenção pelo que é natural considerar a seguinte *simetria* das matrizes.

Definição 2.21. *Seja A uma matriz $m \times n$. A* matriz transposta de A *é a matriz A^T, do tipo $n \times m$ cuja entrada ij é*

$$(A^T)_{ij} = a_{ji}$$

Por exemplo

$$\begin{bmatrix} 1 & -1 & 2 \\ 0 & 3 & 2 \end{bmatrix}^T = \begin{bmatrix} 1 & 0 \\ -1 & 3 \\ 2 & 2 \end{bmatrix} \qquad \text{e} \qquad \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}^T = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$$

Proposição 2.22 (Propriedades da transposição)**.** *(i)* $(A^T)^T = A$
(ii) $(\alpha A)^T = \alpha A^T$
(iii) $(A + B)^T = A^T + B^T$
(iv) $(AB)^T = B^T A^T$.

Dem. As primeiras três propriedades são muito fáceis de demonstrar e ficam como exercício. Quanto à última, suponhamos que A é uma matriz $m \times n$ e B é uma matriz $n \times p$, de forma a que $(AB)^T$ é uma matriz $p \times m$. Dados i, j com $1 \leq i \leq p$ e $1 \leq j \leq m$ temos então que a entrada ij da matriz $(AB)^T$ é

$$\left((AB)^T\right)_{ij} = (AB)_{ji} = \sum_{k=1}^{n} a_{jk} b_{ki} = \sum_{k=1}^{n} (A^T)_{kj} (B^T)_{ik} = \sum_{k=1}^{n} (B^T)_{ik} (A^T)_{kj} = (B^T A^T)_{ij}$$

conforme queríamos demonstrar. □

Usando esta simetria e a propriedade (iv) acima, é imediato verificar que a solução do sistema para uma linha da matriz inversa mencionado anteriormente não é mais do que a solução do sistema

$$A^T x = b$$

com b a coluna correspondente da matriz identidade. Isto sugere uma relação entre a transposição e a inversão... Qual?

Justifiquemos então finalmente o nosso método de cálculo de inversas:

Teorema 2.23. *Seja A uma matriz $n \times n$ de números reais ou complexos. As seguintes afirmações são equivalentes:*

(i) A é invertível
(ii) Para cada matriz $n \times 1$, B, o sistema $AX = B$ tem solução e esta é única.
(iii) A tem característica n

Dem. Vamos ver que (i)⇒(ii)⇒(iii)⇒(i).

(i)⇒(ii): Multiplicando o sistema dos dois lados por A^{-1} temos

$$A^{-1}AX = A^{-1}B \Rightarrow I_n X = A^{-1}B \Rightarrow X = A^{-1}B$$

Logo se a solução existe, ela é única e é dada por $X = A^{-1}B$. Mas é fácil verificar que $A^{-1}B$ é de facto uma solução

$$A(A^{-1}B) = I_n B = B$$

o que conclui a prova desta implicação

(ii)⇒(iii): Esta implicação é equivalente à implicação $\neg$(iii) $\Rightarrow$ $\neg$(ii) que passamos a demonstrar. Se a característica de A não é igual a n, então no final do método de Gauss-Jordan, alguma das colunas não tem pivot. A variável correspondente é então livre na solução do sistema homogéneo $AX = 0$, que tem portanto infinitas soluções. Conclui-se que a solução do sistema $AX = 0$ não é única e portanto a afirmação (ii) é falsa.

(iii)⇒(i): Se A tem característica n, então aplicando o método de Gauss-Jordan a matriz A é transformada na matriz I_n (uma vez que esta é a única matriz $n \times n$ em escada de linhas reduzida com característica n). Mas, como já observámos, cada passo do método de Gauss-Jordan consiste na multiplicação à esquerda por uma matriz. Nomeadamente:

- A operação $L_i \leftrightarrow L_j$, com $i \neq j$ corresponde à multiplicação à esquerda pela matriz

$$S_{ij} = \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 0 & & 1 & & \\ & & & \ddots & & & \\ & & 1 & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}$$

em que os $\ddots$ indicam 1s, todas as entradas não indicadas são 0 e os 0s na diagonal ocorrem nas linhas i e j.

- A operação αL_i com $\alpha \neq 0$ corresponde à multiplicação pela matriz

$$D_{i,\alpha} \begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & \alpha & & \\ & & & \ddots & \\ & & & & 1 \end{bmatrix}$$

com todas as entradas fora da diagonal 0 e todas as entradas na diagonal 1 exceto a i-ésima que é α.

- A operação $L_i + \alpha L_j$ com $i \neq j$ e $\alpha \neq 0$ corresponde à multiplicação pela matriz

$$I_n + \alpha E_{ij} = \begin{bmatrix} 1 & & & & \\ & \ddots & & \alpha & \\ & & & \ddots & \\ & & & & \ddots & \\ & & & & & 1 \end{bmatrix}$$

em que todas as entradas da diagonal são 1 e todas as entradas fora da diagonal são 0 exceto a entrada ij, que é igual a α. O esquema acima corresponde ao caso em que $i < j$ e portanto à fase final do método de Gauss-Jordan. A fase inicial do método de Gauss consiste na multiplicação por estas matrizes com $i > j$, caso em que a entrada não nula fora da diagonal está abaixo da diagonal.

Em termos do produto de matrizes, a observação que o método de Gauss-Jordan termina na matriz I_n expressa a igualdade

$$E_k \cdots E_2 E_1 A = I_n \tag{11}$$

em que k é o número de passos do método de Gauss-Jordan e cada uma das matrizes E_i, correspondente ao passo i do método, é alguma das matrizes referidas acima. Ora cada matriz E_i é invertível! De facto, é imediato verificar que

- $S_{ij}^{-1} = S_{ij}$
- $D_{i,\alpha}^{-1} = D_{i,\frac{1}{\alpha}}$
- $(I_n + \alpha E_{ij})^{-1} = I_n - \alpha E_{ij}$

Multiplicando a igualdade (11) pelas inversas das matrizes $E_k, E_{k-1}, \ldots$ obtemos

$$A = E_1^{-1} E_2^{-1} \cdots E_k^{-1}$$

Uma vez que A é um produto de matrizes invertíveis, pela Proposição 2.18, A é invertível.

□

Vemos assim que, quando aplicamos o método de Gauss-Jordan para resolver simultaneamente os n sistemas lineares correspondentes à equação $AB = I_n$, só há duas possibilidades: ou a aplicação do método mostra que a característica de A é menor do que n e então A não é invertível ou, a característica de A é n e então a matriz A é invertível. Neste último caso, uma vez que a matriz B calculada pelo método de Gauss-Jordan satisfaz $AB = I_n$, temos

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

3. Espaços vetoriais

Um espaço vetorial é um "sítio onde se podem fazer combinações lineares". Para isto tudo o que é necessário é saber como somar e como multiplicar por escalar os objetos do espaço vetorial. Para que estas combinações lineares se comportem como estamos habituados nos exemplos que vimos até agora é necessário que satisfaçam certas propriedades que são especificadas na definição de espaço vetorial.

O arquétipo de um espaço vetorial é $\mathbb{R}^n = \{(x_1, \ldots, x_n) \colon x_i \in \mathbb{R}\}$ em que a multiplicação por escalar é definida por

$$\alpha \cdot (x_1, \ldots, x_n) = (\alpha x_1, \ldots, \alpha x_n)$$

e a soma por

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n)$$

Nos casos em que $n = 1, 2$ ou 3, estamos habituados a identificar $\mathbb{R}^n$ geometricamente com o conjunto dos vetores com origem em $(0, \ldots, 0)$, e sabemos interpretar geometricamente o produto por escalar e a soma.

Por exemplo, o conjunto de todas as combinações lineares de dois vetores em $\mathbb{R}^3$ formam um plano que passa pela origem e contém os dois vetores.

A definição de espaço vetorial vai-nos permitir transferir a nossa intuição geométrica sobre o comportamento de vetores no espaço para um sem-fim de novas situações!

Definição 3.1. *Um* espaço vetorial real *é um conjunto não vazio V, cujos elementos se designam por* vetores, *juntamente com duas funções*

- *Multiplicação por escalar: $\mathbb{R} \times V \to V$ que a um par (α, v) associa um vetor αv.*
- *Soma de vetores: $V \times V \xrightarrow{+} V$ que a um par de vetores (v, w) associa um vetor $v + w$*

satisfazendo as seguintes relações:

(i) Para todos os $u, v, w \in V$, $u + (v + w) = (u + v) + w$.
(ii) Para todos os $u, v \in V$, $u + v = v + u$.
(iii) Existe um elemento $0 \in V$ tal que, para todo o $v \in V$ se tem $v + 0 = v$.
(iv) Para todo o $v \in V$ existe um elemento $w \in V$ tal que $v + w = 0$.
(v) Para todo o $v \in V$, tem-se $1v = v$.
(vi) Para todos os $\alpha, \beta \in \mathbb{R}$, e $v \in V$ tem-se $\alpha(\beta v) = (\alpha\beta)v$.
(vii) Para todos os $\alpha \in \mathbb{R}$ e $v, w \in V$ tem-se $\alpha(v + w) = \alpha v + \alpha w$.
(viii) Para todos os $\alpha, \beta \in \mathbb{R}$ e $v \in V$ tem-se $(\alpha + \beta)v = \alpha v + \beta v$.

Não é difícil mostrar que o elemento w tal que $v + w = 0$ é único: se $v + w = v + w' = 0$ então

$$w' = w' + 0 = w' + (v + w) = (w' + v) + w = 0 + w = w + 0 = w$$

O único w tal que $w + v = 0$ chama-se o simétrico de v e denota-se por $-v$.

Observação 3.2. *(i) Substituindo na definição acima $\mathbb{R}$ por $\mathbb{C}$ obtemos a definição de um* espaço vetorial complexo. *Mais geralmente se $\mathbb{K}$ é um corpo (ver Observação 1.2) e substituirmos $\mathbb{R}$ por $\mathbb{K}$ obtemos a noção de espaço vetorial sobre o corpo $\mathbb{K}$.*
(ii) É também comum usar a terminologia espaço linear *em vez de* espaço vetorial.

Definição 3.3. *Seja V um espaço vetorial e $v_1, \ldots, v_k$ elementos de V. Diz-se que $v \in V$ é uma combinação linear dos vetores $v_1, \ldots, v_k$ se existem $\alpha_1, \ldots, \alpha_k \in \mathbb{R}$ tais que*

$$v = \alpha_1 v_1 + \ldots + \alpha_k v_k$$

Os escalares $\alpha_1, \ldots, \alpha_k$ chamam-se os coeficientes *da combinação linear.*

Exemplo 3.4. *(1) $\mathbb{R}^n$ com a soma e produto por escalar definidos coordenada a coordenada é um espaço vetorial real. A validade dos axiomas na Definição 3.1 é uma consequência imediata das propriedades das operações de soma e produto de números reais. Por exemplo a propriedade associativa da soma de vetores segue imediatamente da propriedade associativa da soma de números reais. Analogamente $\mathbb{C}^n = \{(z_1, \ldots, z_n) \colon z_i \in \mathbb{C}\}$ é um espaço vetorial complexo, com as operações de soma e produto por escalar definidas componente a componente.*

(2) O conjunto $M_{m\times n}(\mathbb{R})$ das matrizes $m \times n$ reais é um espaço vetorial real. É esse o conteúdo das Proposições 2.11 e 2.12. Analogamente, o conjunto das matrizes $M_{m\times n}(\mathbb{C})$ é um espaço vetorial complexo.

(3) Seja S um conjunto não vazio. O conjunto $F(S;\mathbb{R}) = \{f \colon S \to \mathbb{R}\}$ das funções de S para $\mathbb{R}$ munido das operações

$$(f+g)(x) \stackrel{def}{=} f(x) + g(x) \qquad (\alpha f)(x) \stackrel{def}{=} \alpha f(x)$$

é um espaço vetorial real. Analogamente o conjunto das funções com valores complexos é um espaço vetorial complexo. Note-se que este exemplo contém os dois exemplos anteriores. De facto $\mathbb{R}^n$ é basicamente o caso em que o conjunto S é $\{1, \ldots, n\}$ e $M_{m\times n}(\mathbb{R})$ é, por definição, o caso em que $S = \{1, \ldots, m\} \times \{1, \ldots, n\}$.

Observação 3.5. *É habitual referirmo-nos a um espaço vetorial apenas pelo conjunto subjacente deixando implícitas a estrutura de soma de vetores e multiplicação por escalares quando estas são claras do contexto. Por exemplo, quando falamos do espaço vetorial $M_{m\times n}(\mathbb{R})$ referimo-nos a este conjunto com as operações habituais de soma e multiplicação por escalar.*

Exemplo 3.6. *Sejam $v, w \in \mathbb{R}^3$ dois vetores não colineares. Pelo significado geométrico da soma de vetores e produto por escalar, o conjunto das combinações lineares de v e w é o plano que passa pela origem e contém v e w. Dado um ponto u desse plano, o significado dos coeficientes α, β na combinação linear $u = \alpha v + \beta w$ é o seguinte (familiar da noção de coordenadas cartesianas)*

- *αv é o ponto de interseção da reta paralela a w que passa por u, com a reta determinada por v e pela origem (que é o conjunto $\{\lambda v \colon \lambda \in \mathbb{R}\}$).*
- *βw é o ponto de interseção da reta paralela a v que passas por u, com a reta $\{\lambda w \colon \lambda \in \mathbb{R}\}$*

Vejamos mais alguns exemplos e não-exemplos de espaços vetoriais.

Exemplo 3.7. *(i) O conjunto V de todos os polinómios reais com as operações de soma e produto por escalar habituais é um espaço vetorial. Note-se que V está contido no* *conjunto das funções reais $F(S,\mathbb{R})$ e que as operações de soma e produto por escalar são a restrição aos polinómios das operações definidas para as funções. Isso torna a verificação da maioria dos axiomas na Definição 3.1 automáticas. De facto, uma vez que se observe que a soma de polinómios e a multiplicação de um escalar por um polinómio são polinómios, a validade das propriedades (i)-(ii) e (v)-(viii) é imediata e resta apenas observar que a função nula é um polinómio logo (iii) é satisfeito e que a função simétrica de um polinómio é um polinómio logo (iv) é também satisfeito.*

(ii) Seja $V = \{(x,y) \in \mathbb{R}^2 \colon x \geq 0, y \geq 0\}$ com a soma habitual de vetores em $\mathbb{R}^2$ e com o produto por escalar definido por

$$\alpha(x,y) \overset{def}{=} (|\alpha|x, |\alpha|y)$$

Com estas operações V não é um espaço vetorial porque os axiomas (iv) e (vii) não são verificados. Por exemplo o vetor $(1,0)$ não tem simétrico e $(0,0) = 0(1,0) = (1+(-1))(1,0) \neq 1(1,0) + (-1)(1,0) = (2,0)$. Em geral, se α e β têm sinais contrários e $v \neq 0$, a igualdade $(\alpha+\beta)v = \alpha v + \beta v$ não se verifica.

Definição 3.8. *Seja V um espaço vetorial. Um subconjunto $W \subset V$ diz-se um* subespaço vetorial de V *se munido das operações de V é um espaço vetorial.*

Implícito na definição anterior está que W é *fechado* para as operações de V, isto é que se $w_1, w_2 \in W$ e $\alpha \in \mathbb{R}$ então a soma de w_1 e w_2 em V pertence a W e o produto por escalar em V, αw_1, pertence a W.

Exemplo 3.9. *O Exemplo 3.7 (i) verifica que o conjunto dos polinómios é um subespaço vetorial de $F(S;\mathbb{R})$.*

Como observámos no Exemplo 3.7 (i) quando $W \subset V$ é um subconjunto de um espaço vetorial fechado para a soma e multiplicação por escalar, a verificação de que W é um espaço vetorial pode reduzir-se à verificação que o elemento neutro da soma e os simétricos (em V) de elementos de W pertencem a V. A próxima proposição mostra que mesmo estas verificações não são necessárias.

Proposição 3.10. *Seja V um espaço vetorial. Se W é um subconjunto não vazio de V fechado para a soma e multiplicação por escalar, então W é um subespaço vetorial de V.*

Proof. Como já observámos, a verificação dos axiomas (i)-(ii) e (v)-(viii) é imediata. É um exercício para as aulas práticas verificar que, para qualquer $v \in V$, o produto por escalar $0v$ é o elemento neutro para a soma. Como W é não vazio e fechado para o produto por escalar conclui-se que $0 \in W$ e portanto o axioma (iii) é verificado. É também um exercício para as aulas práticas verificar que o simétrico de $v \in V$ é o produto por escalar $(-1)v$. Uma vez que W é fechado para o produto por escalar conclui-se que o axioma (iv) é verificado em W. □

Exemplo 3.11. *(i) Seja V o espaço vetorial de todos os polinómios reais. O subconjunto $W \subset V$ formado pelos polinómios de grau menor ou igual a 3 é um subespaço vetorial. De facto, de acordo com a proposição anterior basta observar que a soma de* *polinómios de grau* ≤ 3 *tem grau* ≤ 3 *e que o produto de um polinómio de grau* ≤ 3 *por um escalar tem ainda grau* ≤ 3.

(ii) O plano $W = \{(x,y,z) \in \mathbb{R}^3 \colon x+y+z=0\}$ *é um subespaço vetorial de* $\mathbb{R}^3$. *De acordo com a Proposição acima basta notar que se* $(x,y,z), (x',y',z') \in W$ *e* $\alpha \in \mathbb{R}$ *então* $(x+x')+(y+y')+(z+z')=0$ *e* $(\alpha x)+(\alpha y)+(\alpha z)=0$ *logo* $(x+x', y+y', z+z') \in W$ *e* $(\alpha x, \alpha y, \alpha z) \in W$.

(iii) Seja A *uma matriz* $m \times n$. *O* núcleo de A *é o conjunto*

$$N(A) = \{x \in \mathbb{R}^n \colon A \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = 0\}$$

Este conjunto é um subespaço vetorial de $\mathbb{R}^n$ *(o argumento é exatamente o mesmo que no exemplo anterior).*

Intuitivamente devemos pensar nos espaços vetoriais como sendo objetos que se comportam de forma semelhante ao espaço euclidiano usual - $\mathbb{R}^3$ - e nos subespaços vetoriais como sendo subconjuntos com comportamento semelhante ao das retas e planos em $\mathbb{R}^3$ que passam pela origem.

Definição 3.12. *Seja* V *um espaço vetorial e* $S \subset V$ *um subconjunto. A* expansão linear *de* S *em* V *é o conjunto* $L(S)$ *das combinações lineares de elementos de* S, *isto é*

$$L(S) = \{\alpha_1 v_1 + \ldots + \alpha_n v_n \colon \alpha_1, \ldots, \alpha_n \in \mathbb{R}, v_1, \ldots, v_n \in S, n \in \mathbb{N}\}$$

Por convenção $L(\emptyset) = \{0\}$.

Exemplo 3.13. *(i) Seja* V *o espaço vetorial dos polinómios reais. Vamos determinar se* $x + 2x^3 \in L(S)$ *onde* $S = \{1-x, x+x^2+x^3, x^2\}$. *Por definição, a pergunta é se existem escalares* $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}$ *tais que*

$$x + 2x^3 = \alpha_1(1-x) + \alpha_2(x+x^2+x^3) + \alpha_3 x^2$$

Como dois polinómios são iguais se têm os mesmos coeficientes, a igualdade anterior é equivalente ao sistema

$$\begin{cases} \alpha_1 = 0 \\ -\alpha_1 + \alpha_2 = 1 \\ \alpha_2 + \alpha_3 = 0 \\ \alpha_2 = 2 \end{cases} \Leftrightarrow \begin{cases} \alpha_1 = 0 \\ \alpha_2 = 1 \\ \alpha_3 = -1 \\ \alpha_2 = 2 \end{cases}$$

Uma vez que o sistema é impossível, conclui-se que $x + 2x^3 \notin L(S)$. *Neste caso não se justificava a utilização do método de Gauss para a resolução do sistema. Mas note-se que se tivéssemos escrito o sistema acima da forma habitual, a matriz à qual iríamos* *aplicar o método de Gauss seria*

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

Os coeficientes dos elementos de S aparecem nas primeiras três colunas, enquanto que a última coluna contém os coeficientes do polinómio $x + 2x^3$.

(ii) Sendo $S = \{(1,3,2),(0,1,4),(1,4,6)\} \subset \mathbb{R}^3$, vamos determinar equações cartesianas que definam $L(S)$. Os elementos de $L(S)$ são os vetores $(a,b,c) \in \mathbb{R}^3$ para os quais é possível achar $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}$ tais que

$$(a,b,c) = \alpha_1(1,3,2) + \alpha_2(0,1,4) + \alpha_3(1,4,6) = (a,b,c)$$

Ou seja, são os vetores (a,b,c) tais que o seguinte sistema é possível

$$\left[\begin{array}{ccc|c} 1 & 0 & 1 & a \\ 3 & 1 & 4 & b \\ 2 & 4 & 6 & c \end{array}\right] \underset{L_3-2L_1}{\overset{L_2-3L_1}{\longrightarrow}} \left[\begin{array}{ccc|c} 1 & 0 & 1 & a \\ 0 & 1 & 1 & b-3a \\ 0 & 4 & 4 & c-2a \end{array}\right] \overset{L_3-4L_2}{\longrightarrow} \left[\begin{array}{ccc|c} 1 & 0 & 1 & a \\ 0 & 1 & 1 & b-3a \\ 0 & 0 & 0 & c-4b+10a \end{array}\right]$$

Conclui-se que $(a,b,c) \in L(S) \Leftrightarrow c - 4b + 10a = 0$. Geometricamente, $L(S)$ é um plano que passa pela origem. Normalmente, esperaríamos que três vetores em $\mathbb{R}^3$ formassem um referencial e que qualquer outro vetor se pudesse escrever como combinação linear deles mas neste caso $(1,3,2)+(0,1,4) = (1,4,6)$ e portanto podemos escrever qualquer combinação linear dos três vetores de S usando apenas os dois primeiros. A expansão linear destes dois vetores é um plano que tem equação paramétrica

$$(x,y,z) = \alpha_1(1,3,2) + \alpha_2(0,1,4), \qquad \text{com } \alpha_1, \alpha_2 \in \mathbb{R}$$

e, como vimos acima, equação cartesiana

$$10x - 4y + z = 0.$$

Proposição 3.14. *Seja V um espaço vetorial e $S \subset V$ um subconjunto. Então $L(S)$ é o mais pequeno subespaço vetorial de V que contém S. Mais precisamente*

- *$L(S)$ é um subespaço vetorial de V e $S \subset L(S)$.*
- *Se $W \subset V$ é um subespaço vetorial de V que contém S, então $L(S) \subset W$.*

Dem. Se S é vazio então as condições são claramente verificadas. Suponhamos que S é não vazio. $L(S)$ contém S porque dado $v \in S$ temos que $1 \cdot v = v$ é uma combinação linear de elementos de S e portanto pertence a $L(S)$. Para ver que $L(S)$ é um subespaço vetorial precisamos de ver que $L(S)$ é fechado para a soma e para o produto por escalar. Seja $\lambda \in \mathbb{R}$ um escalar e $\alpha_1 v_1 + \ldots + \alpha_n v_n$ um elemento de S. Então

$$\lambda(\alpha_1 v_1 + \ldots + \alpha_n v_n) = (\lambda\alpha_1)v_1 + \ldots + (\lambda\alpha_n)v_n$$

é também uma combinação linear de elementos de S e portanto pertence a $L(S)$. Conclui-se que $L(S)$ é fechado para o produto por escalar. Por outro lado, dados dois elementos

$\alpha_1 v_1 + \ldots + \alpha_n v_n$ e $\beta_1 w_1 + \ldots + \beta_m w_m$ em $L(S)$ a sua soma é

$$\alpha_1 v_1 + \ldots + \alpha_n v_n + \beta_1 w_1 + \ldots + \beta_m w_m$$

que é ainda uma combinação linear de elementos de S. Conclui-se que $L(S)$ também é fechado para a soma de vetores e portanto é um subespaço vetorial de V.

Finalmente, seja W um qualquer subespaço vetorial de V que contém S. Então dados $v_1, \ldots, v_n \in S$ e $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ temos que $\alpha_i v_i \in W$ (pois W é fechado para o produto por escalar) e portanto

$$\alpha_1 v_1 + \ldots + \alpha_n v_n \in W$$

(porque W é fechado para a soma). Conclui-se que W contém qualquer combinação linear de elementos de S, ou seja, que W contém $L(S)$. ☐

Devido ao resultado enunciado na Proposição anterior, chamamos a $L(S)$ o *subespaço gerado por* S e se $W = L(S)$ dizemos que W *é gerado por* S e que S é um conjunto de geradores para W.

Exemplo 3.15. *(i) Vamos achar um conjunto de geradores para o subespaço*

$$W = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} : a+b-2c = 0, d-c+a = 0 \right\} \subset M_{2\times 2}(\mathbb{R})$$

(é imediato verificar que W é de facto um subespaço vetorial de $M_{2\times 2}(\mathbb{R})$).

Podemos resolver o sistema dado pelas condições que definem W (aqui não se justifica a aplicação do método de Gauss)

$$\begin{cases} a+b-2c = 0 \\ d-c+a = 0 \end{cases} \Leftrightarrow \begin{cases} c = \frac{1}{2}a + \frac{1}{2}b \\ d = -\frac{1}{2}a + \frac{1}{2}b \end{cases}$$

O elemento típico de W pode portanto escrever-se na forma

$$\begin{bmatrix} a & b \\ \frac{1}{2}a + \frac{1}{2}b & -\frac{1}{2}a + \frac{1}{2}b \end{bmatrix} = a \begin{bmatrix} 1 & 0 \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix} + b \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \quad \text{com } a, b \in \mathbb{R}$$

logo

$$S = \left\{ \begin{bmatrix} 1 & 0 \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \right\}$$

é um conjunto de geradores para W.

Seja A uma matriz $m \times n$. Chama-se *espaço das linhas de* A, e denota-se por $EL(A)$ ao subespaço de $\mathbb{R}^n$ gerado pelas linhas de A. Por exemplo, para

$$\begin{bmatrix} 2 & 0 & 1 & 4 \\ 0 & 3 & 1 & 2 \end{bmatrix} \tag{12}$$

temos

$$EL(A) = L(\{(2,0,1,4), (0,3,1,2)\}) \subset \mathbb{R}^4$$

Quando aplicamos o método de Gauss(-Jordan) a uma matriz, o espaço das linhas não muda. De facto suponhamos que

$$A = A_1 \to A_2 \to \cdots \to A_k$$

é uma sucessão de matrizes obtida por aplicação o método de Gauss-Jordan à matriz A. Uma vez que as linhas de A_{i+1} são combinações lineares das linhas da matriz A_i temos que

$$\{\text{linhas de } A_{i+1}\} \subset EL(A_i)$$

e portanto, pela Proposição 3.14 temos $EL(A_{i+1}) \subset EL(A_i)$. Mas, as linhas de A_i também são combinações lineares das linhas de A_{i+1}, logo $EL(A_i) \subset EL(A_{i+1})$ e conclui-se que $EL(A_i) = EL(A_{i+1})$. O método de Gauss-Jordan dá-nos portanto um método para determinar um conjunto de geradores particularmente simples para o espaço das linhas de uma matriz: as linhas não nulas da matriz em escada de linhas reduzida obtida como output do algoritmo.

Analogamente definimos o *espaço das colunas* de uma matriz A do tipo $m \times n$ como o subespaço de $\mathbb{R}^m$ gerado pelas colunas de A. Por exemplo, para a matriz (12) temos

$$EC(A) = L(\{(2,0),(0,3),(1,1),(4,2)\}) = \mathbb{R}^2.$$

Note-se que *não é verdade* que o espaço das colunas permaneça inalterado ao longo da aplicação do método de Gauss.

Um espaço vetorial V diz-se *finitamente gerado* se existe um conjunto finito $S \subset V$ tal que $V = L(S)$.

Exemplo 3.16. *O espaço vetorial V formado por todos os polinómios reais não é finitamente gerado. De facto, sendo $S = \{p_1, \ldots, p_k\} \subset V$. um conjunto finito de polinómios e n_i o grau do polinómio p_i podemos tomar*

$$N = \max\{n_1, \ldots, n_k\}$$

e claramente x^{N+1} não pode ser escrito como combinação linear de elementos de S. Isto mostra que não existe um conjunto finito de geradores para V.

4. Dependência linear, bases e dimensão

Chegamos agora a alguns dos conceitos fundamentais da Álgebra Linear.

Definição 4.1. *Seja V um espaço vetorial. Um conjunto $S \subset V$ diz-se* linearmente dependente *se existem $v_1, \ldots, v_n \in S$ distintos e escalares $\alpha_1, \ldots, \alpha_n$ não todos nulos tais que*

$$\alpha_1 v_1 + \ldots + \alpha_n v_n = 0$$

Caso contrário, S diz-se linearmente independente. *Um conjunto $B \subset V$ diz-se uma* base *de V se é linearmente independente e gera V.*

Note-se que a negação da condição de dependência linear é logicamente equivalente à seguinte condição, que utilizamos normalmente para testar independência linear: S é linearmente independente se e só se dados $v_1, \ldots, v_n$ elementos distintos de S e escalares $\alpha_1, \ldots, \alpha_n$ tais que $\alpha_1 v_1 + \ldots + \alpha_n v_n = 0$ temos necessariamente $\alpha_1 = \cdots = \alpha_n = 0$.

Exemplo 4.2. *(i) Seja $S = \{v\}$ um conjunto com um único elemento. Se $v = 0$ então S é linearmente dependente uma vez que $1 \cdot 0$ é uma combinação linear com coeficientes não nulos de elementos de S que produz o vetor 0. Se $v \neq 0$, então S é linearmente* *independente. De facto, uma combinação linear de elementos de S com coeficientes não nulos é da forma αv com $\alpha \neq 0$ e é uma consequência dos axiomas de espaço vetorial que sendo $\alpha \neq 0$ e $v \neq 0$ então $\alpha v \neq 0$ (ver o último exercício da Ficha 4).*

(ii) *Se S contém o vetor nulo então S é linearmente independente (pois $1 \cdot 0 = 0$).*

(iii) *Mais geralmente, se $S \subset S'$ e S é linearmente dependente, o mesmo é verdade para S' (pois a combinação linear com coeficientes não todos nulos que certifica a dependência linear de S, certifica também a dependência linear de S'). Equivalentemente, se S' é um conjunto linearmente independente e $S \subset S'$ então S é também linearmente independente.*

(iv) *Seja $S = \{v, w\}$ um conjunto com dois elementos (distintos). Então S é linearmente dependente se e só se v e w são colineares, isto é se um deles é um múltiplo escalar do outro. De facto, se existem α_1, α_2 não ambos nulos tais que*

$$\alpha_1 v + \alpha_2 w = 0$$

ou $\alpha_1 \neq 0$ e então $v = -\frac{\alpha_2}{\alpha_1} w$, ou $\alpha_2 \neq 0$ e $w = -\frac{\alpha_1}{\alpha_2} v$.

(v) *Generalizando o exemplo anterior vemos que um conjunto $S \subset V$ é linearmente dependente se e só se um dos elementos de S pode ser expresso como uma combinação linear dos restantes elementos de S. De facto uma das implicações é imediata e para ver a outra, se S é linearmente dependente podemos escolher $v_1, \ldots, v_n \in S$ e escalares $\alpha_1, \ldots, \alpha_n$ não todos nulos de tal forma que*

$$\alpha_1 v_1 + \ldots + \alpha_n v_n = 0$$

Assumindo, por exemplo, que $\alpha_i \neq 0$ temos que

$$v_i = -\frac{\alpha_1}{\alpha_i} v_1 - \ldots - \frac{\alpha_{i-1}}{\alpha_i} v_{i-1} - \frac{\alpha_{i+1}}{\alpha_i} v_{i+1} - \ldots - \frac{\alpha_n}{\alpha_i} v_n$$

é uma combinação linear de $v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n$.

(vi) *O subconjunto $\{(1,2), (0,3), (1,0)\} \subset \mathbb{R}^2$ é linearmente dependente uma vez que*

$$(1,2) - (1,0) - \tfrac{2}{3}(0,3) = (0,0)$$

Como nenhum par de vetores do conjunto é colinear, se retirarmos qualquer dos vetores ao conjunto obtemos um conjunto linearmente independente, que claramente gera $\mathbb{R}^2$ e constitui portanto uma base para $\mathbb{R}^2$.

(vii) *O conjunto $B = \{e_1 = (1, 0, \ldots, 0), e_2 = (0, 1, 0, \ldots, 0), \ldots, e_n = (0, \ldots, 0, 1)\}$ é uma base de $\mathbb{R}^n$ chamada a* base canónica. *De facto, dado $(x_1, \ldots, x_n) \in \mathbb{R}^n$ temos*

$$(x_1, \ldots, x_n) = x_1 e_1 + \ldots + x_n e_n$$

logo $L(B) = \mathbb{R}^n$ e se $\alpha_1, \ldots, \alpha_n$ são números reais e $\alpha_1 e_1 + \ldots + \alpha_n e_n = 0$ então dado que

$$\alpha_1 e_1 + \ldots + \alpha_n e_n = (\alpha_1, \ldots, \alpha_n)$$

temos $\alpha_1 = \cdots = \alpha_n = 0$ o que mostra que B é linearmente independente.

(viii) *Se A é uma matriz $m \times n$ em escada de linhas, então as linhas não nulas constituem uma base para $EL(A)$. De facto já vimos acima que as linhas não nulas geram $EL(A)$ e se uma combinação linear das linhas se anular, o sistema para os coeficientes da combinação linear que se obtém considerando apenas as componentes correspondentes* *às colunas que contêm pivots implica imediatamente que os coeficientes da combinação linear são todos nulos. Por exemplo, para*

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

olhando apenas para a primeira e terceira componente dos vetores na equação

$$\alpha_1(2,1,1,4) + \alpha_2(0,0,1,2) = (0,0,0,0)$$

vemos que

$$2\alpha_1 = 0 \qquad e \qquad \alpha_1 + \alpha_2 = 0$$

pelo que $\alpha_1 = \alpha_2 = 0$.

O método de Gauss dá-nos portanto uma maneira prática de determinar uma base para o espaço das linhas de uma matriz (e, na prática, para qualquer subespaço de um espaço vetorial finitamente gerado).

(ix) É um exercício simples verificar que $\{1, x, x^2, \ldots, x^n, \ldots\}$ *é uma base para o espaço vetorial dos polinómios reais.*

Intuitivamente, uma base para um espaço vetorial é um "referencial". De facto, se B é uma base de V, os coeficientes da combinação linear que exprime um vetor $v \in V$ em termos dos elementos de B são *únicos*: Admitindo que $B = \{v_1, \ldots, v_n\}$, qualquer vetor v pode ser escrito na forma

$$v = \alpha_1 v_1 + \ldots + \alpha_n v_n$$

(porque B gera V) mas se tivermos também

$$v = \beta_1 v_1 + \ldots + \beta_n v_n$$

então subtraindo as duas igualdades temos

$$0 = (\alpha_1 - \beta_1)v_1 + \ldots + (\alpha_n - \beta_n)v_n$$

e, uma vez que, B é um conjunto linearmente independente, isto implica que $\alpha_1 - \beta_1 = 0, \ldots, \alpha_n - \beta_n = 0$. Os coeficientes dos elementos da base chamam-se as *coordenadas* de v na base B. Uma base permite assim identificar os vetores de V com listas de escalares (ou seja com $\mathbb{R}^n$ onde $n = \dim V$).

Vejamos agora algumas propriedades importantes relativas à dependência linear. Sugerimos que ao ler os enunciados que se seguem se tenha em mente o exemplo de $\mathbb{R}^3$ e a interpretação geométrica usual da combinação linear de vetores no espaço assim como dos subespaços lineares de $\mathbb{R}^3$ - retas, planos, etc.

Proposição 4.3. *Seja* V *um espaço vetorial e* $S \subset V$ *um conjunto linearmente independente. Se* $v \notin L(S)$ *então* $S \cup \{v\}$ *é linearmente independente.*

Dem. Sejam $v_1, \ldots, v_n$ vetores distintos de S e $\alpha_1, \ldots, \alpha_n, \alpha_{n+1}$ escalares. Temos a verificar que se

$$\alpha_1 v_1 + \ldots + \alpha_n v_n + \alpha_{n+1} v = 0$$

então $\alpha_1 = \cdots = \alpha_{n+1} = 0$. Notamos primeiro que α_{n+1} é necessariamente 0 porque senão

$$v = -\frac{\alpha_1}{\alpha_{n+1}}v_1 - \cdots - -\frac{\alpha_n}{\alpha_{n+1}}v_n$$

é uma combinação linear de elementos de S, contrariando a hipótese da Proposição. Mas então

$$\alpha_1 v_1 + \ldots + \alpha_n v_n = 0$$

Como S é linearmente independente segue que $\alpha_1 = \cdots = \alpha_n = 0$. □

Proposição 4.4. *Seja V um espaço vetorial e $S \subset V$ um subconjunto.*

(i) Se S é finito, existe um subconjunto $T \subset S$ tal que $L(T) = L(S)$ e T é linearmente independente.
(ii) Se S é uma base de V com n elementos, então qualquer subconjunto de V com mais de n elementos é linearmente dependente.
(iii) Se V é um espaço vetorial finitamente gerado, qualquer subconjunto linearmente independente $T \subset V$ está contido numa base de V.

Dem. (i) A demonstração é por indução no número de elementos de S. Se $S = \{v\}$ tem apenas um elemento então, ou $v = 0$ e então podemos tomar $T = \emptyset$ ou $v \neq 0$ e então S é linearmente independente e podemos tomar $T = S$. Suponhamos agora que a afirmação é válida para conjuntos com n elementos e suponhamos que S tem $n+1$ elementos. Se S é linearmente independente então podemos tomar $T = S$. Senão podemos escolher um vetor $v \in S$ que se pode escrever como combinação linear dos elementos de $S \setminus \{v\}$, e então $L(S) = L(S \setminus \{v\})$. Como $S \setminus \{v\}$ tem n elementos, por hipótese de indução existe $T \subset S \setminus \{v\}$ linearmente independente tal que $L(T) = L(S \setminus \{v\}) = L(S)$, o que conclui a demonstração.

(ii) Seja $S = \{v_1, \ldots, v_n\}$ uma base de V. Vamos mostrar que qualquer conjunto $\{w_1, \ldots, w_{n+1}\}$ formado por $n+1$ elementos distintos de V é linearmente dependente. Uma vez que S é uma base, existem escalares a_{ij} tais que

$$\begin{aligned} w_1 &= a_{11}v_1 + \ldots + a_{1n}v_n \\ w_2 &= a_{21}v_1 + \ldots + a_{2n}v_n \\ &\vdots \\ w_{n+1} &= a_{(n+1)1}v_1 + \ldots + a_{(n+1)n}v_n \end{aligned}$$

Escrevendo uma combinação linear

$$\alpha_1 w_1 + \ldots + \alpha_{n+1} w_{n+1}$$

na base S obtemos

$$(\alpha_1 a_{11} + \alpha_2 a_{21} + \ldots + \alpha_{n+1} a_{(n+1)1})v_1 + \ldots + (\alpha_1 a_{1n} + \alpha_2 a_{2n} + \ldots + \alpha_{n+1} a_{(n+1)n})v_n \tag{13}$$

Pretendemos mostrar que existem α_i's não todos nulos tais que a expressão (13) é nula. Mas para que (13) seja nula basta que se verifiquem as condições

$$\begin{aligned} \alpha_1 a_{11} + \alpha_2 a_{21} + \ldots + \alpha_{n+1} a_{(n+1)1} &= 0 \\ &\vdots \\ \alpha_1 a_{1n} + \alpha_2 a_{2n} + \ldots + \alpha_{n+1} a_{(n+1)n} &= 0 \end{aligned}$$

Estas condições dizem que $(\alpha_1, \ldots, \alpha_n)$ é uma solução do sistema linear homogéneo cuja matriz dos coeficientes é a_{ij}. Trata-se de um sistema de n equações com $(n+1)$ incógnitas logo tem sempre (infinitas) soluções não nulas (pois há pelo menos uma variável livre). Isto conclui a demonstração.

(iii) Seja T um conjunto linearmente independente e $S = \{v_1, \ldots, v_n\}$ um conjunto de geradores para V. Claramente $L(T \cup S) = V$ (mas infelizmente $T \cup S$ pode não ser linearmente independente). Seja v_i o primeiro vetor de S que não pertence a $L(T)$ (se não existir então $L(T) = L(T \cup S) = V$ e T é a base requerida). Temos, por um lado, que $L(T \cup \{v_i, \ldots, v_n\}) = L(T \cup S) = V$ e por outro, pela Proposição 4.3, que $T \cup \{v_i\}$ é linearmente independente. Seja agora v_j o primeiro dos vetores $v_{i+1}, \ldots, v_n$ que não pertence a $L(T \cup \{v_i\})$ (se não existir então $L(T \cup \{v_i\}) = V$ e $T \cup \{v_i\}$ é a base requerida). Temos agora que

$$L(T \cup \{v_i, v_j, v_{j+1}, \ldots v_n\}) = V$$

e, pela Proposição 4.3, que $(T \cup \{v_i\}) \cup \{v_j\}$ é linearmente independente. Prosseguindo desta maneira, obtemos após um número finito de passos uma base para V contendo T. Este procedimento é na realidade um algoritmo para achar esta base.

□

Observação 4.5. *A demonstração da Proposição 4.4(iii) pode ser formalizada usando indução no número de vetores de V que é necessário acrescentar ao conjunto T para obter um conjunto de geradores para V. Se este número é 0 então T é já uma base de V. O argumento na demonstração acima pode facilmente ser adaptado para demonstrar o passo da indução: Se quando basta acrescentar n vetores a T para gerar V, o conjunto T pode ser completado de forma a obter uma base, então o mesmo se verifica quando basta acrescentar $n+1$ vetores a T para gerar V.*

Podemos agora facilmente demonstrar o seguinte resultado fundamental.

Teorema 4.6. *Seja V um espaço vetorial finitamente gerado. Então V tem uma base finita e todas as bases de V têm o mesmo número de elementos.*

Dem. Seja S um conjunto finito tal que $V = L(S)$. Pela Proposição 4.4(i), o conjunto S contém um subconjunto T que é linearmente independente e tal que $L(T) = L(S) = V$. T é portanto uma base de V e trata-se de um conjunto finito porque S é finito.

Suponhamos que B é uma outra base de V. A Proposição 4.4(ii) garante que $\#B \leq \#T$ (senão B seria linearmente dependente). Mas o mesmo raciocínio diz-nos que $\#T \leq \#B$ e portanto B e T têm o mesmo número de elementos. □

Definição 4.7. *O número de elementos de qualquer base de um espaço finitamente gerado chama-se a* dimensão de V *e denota-se por* $\dim V$. *Se um espaço vetorial* V *não tem uma base finita, diz-se que tem* dimensão infinita.

É imediato da Proposição 4.4(i) que um espaço tem dimensão infinita se e só se não é finitamente gerado.

Exemplo 4.8. *À luz do Exemplo 4.2(vii),(viii) e (ix) temos*

(i) $\dim \mathbb{R}^n = n$.
(ii) Se A *é uma matriz, então* $\dim EL(A)$ *é igual à característica da matriz* A.
(iii) O espaço dos polinómios tem dimensão infinita.

Intuitivamente, a dimensão de um conjunto é o número de parâmetros reais (ou coordenadas) que necessitamos para descrever os pontos do conjunto. Por exemplo a superfície da Terra tem dimensão 2 pois um ponto à superfície da terra é descrito por dois números reais - a latitude e a longitude. Estas questões serão discutidas mais tarde na disciplina de Cálculo 2. O Teorema 4.6 encoraja esta nossa intuição ao afirmar que numa gama restrita de exemplos - aqueles em que o conjunto em questão tem a estrutura de um espaço vetorial finitamente gerado - não há qualquer ambiguidade quanto ao número de parâmetros necessários para descrever o conjunto.

Exemplo 4.9. *A dimensão do espaço* $M_{2\times 4}(\mathbb{R})$ *é* 8. *De facto é imediato verificar que as oito matrizes*

$$E_{11} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, E_{12} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \ldots, E_{42} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

constituem uma base. Mais geralmente $\dim M_{m\times n}(\mathbb{R}) = mn$. *Uma base é dada pelas matrizes* $\{E_{ij}\}_{1\leq i\leq m, 1\leq j\leq n}$ *onde* E_{ij} *designa a matriz que tem* 1 *como entrada* ij *e todas as restantes entradas iguais a* 0.

Corolário 4.10. *Seja* V *um espaço vetorial de dimensão* n.

(i) Qualquer conjunto linearmente independente com n *vetores é uma base de* V.
(ii) Qualquer conjunto de geradores tem pelo menos n *elementos.*

Dem. (i) Seja S um conjunto linearmente independente com n vetores. Se $L(S) \neq V$ então podemos escolher $v \notin L(S)$ e, pela Proposição 4.3, o conjunto $S \cup \{v\}$ é um conjunto linearmente independente com $n+1$ elementos. Isto não é possível pela Proposição 4.4 (ii)

(ii) Pela Proposição 4.4(i), qualquer conjunto de geradores contém uma base e portanto, pelo Teorema 4.6 tem pelo menos n elementos.

□

Observação 4.11. *Todos os resultados demonstrados acima que assumem que o espaço vetorial em questão é finitamente gerado admitem versões para espaços vetoriais arbitrários. Por exemplo em qualquer espaço vetorial é verdade que duas bases têm o mesmo número de* *elementos, no sentido em que é possível definir uma correspondência bijetiva entre os elementos de uma base e da outra. A demonstração destas versões mais gerais requer alguns conhecimentos de Teoria dos Conjuntos pelo que não discutiremos estes resultados.*

Vejamos como as propriedades dos conjuntos linearmente independentes e bases demonstrados acima podem auxiliar o cálculo de bases e a determinação se um conjunto é ou não linearmente dependente.

Exemplo 4.12. *Vamos verificar que o conjunto* $B = \{(1,0,1),(1,1,0),(0,0,3)\}$ *é uma base para* $\mathbb{R}^3$ *e determinar as componentes de* $(1,2,1)$ *nesta base.*

Uma vez que $\dim \mathbb{R}^3 = 3$*, de acordo com o Corolário 4.10(i) para ver que* B *é uma base basta-nos verificar que* B *é um subconjunto linearmente independente de* $\mathbb{R}^3$*. Podemos fazer isto (pelo menos) de duas formas:*

- *Usando a definição:* B *é linearmente independente se e só se*
$$\alpha(1,0,1)+\beta(1,1,0)+\gamma(0,0,3) = (0,0,0) \Rightarrow \alpha=\beta=\gamma=0$$
A equação à esquerda da implicação é um sistema linear homogéneo cujas incógnitas são os coeficientes α,β,γ*. Resolvendo o sistema vemos se o conjunto é ou não linearmente independente:*
$$\begin{cases} \alpha+\beta=0 \\ \beta=0 \\ \alpha+3\gamma=0 \end{cases} \Leftrightarrow \begin{cases} \alpha=0 \\ \beta=0 \\ \gamma=0 \end{cases}$$
o que mostra que B *é linearmente independente. Neste caso não se justificava utilizar o método de Gauss para resolver o sistema, mas vale a pena notar (para quando as contas sejam mais complicadas) que o sistema em questão tem como coeficientes a matriz cujas colunas são os elementos do conjunto* B*. No exemplo acima:*
$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 3 \end{bmatrix}$$
- *Alternativamente podemos usar a observação feita no Exemplo 4.2(viii) acima. Se escrevermos os elementos de* B *nas linhas de uma matriz e aplicarmos o método de Gauss à matriz obteremos, no final, uma base para* $L(B)$ *e, em particular, calcularemos a dimensão da expansão linear de* B*.* B *será linearmente independente se e só se* $\dim L(B)$ *for igual ao número de elementos de* B*. De facto, se* $\dim L(B) < \#B$ *então pela Proposição 4.4 (ii)* B *será linearmente dependente. Por outro lado, se* $\dim L(B) = \#B$*,* B *não pode ser linearmente dependente porque, se assim fosse, a Proposição 4.4 (i) garantiria a existência de uma base para* $L(B)$ *com menos elementos que* B *o que contradiria o Teorema 4.6.*

Finalmente, a determinação das componentes de um vetor numa dada base consiste na solução de um sistema linear:
$$\alpha(1,0,1)+\beta(1,1,0)+\gamma(0,0,3) = (1,2,1)$$

que podemos escrever na forma de uma matriz aumentada

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

donde obtemos os coeficientes $\alpha = -1, \beta = 2, \gamma = \frac{2}{3}$.

Exemplo 4.13. *Consideremos o conjunto* $S = \left\{ \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 1 & 2 \end{bmatrix}, \begin{bmatrix} 0 & 2 \\ 1 & 3 \end{bmatrix} \right\} \subset M_{2\times 2}(\mathbb{R})$. *Vamos determinar uma base para o espaço* $L(S) \subset M_{2\times 2}(\mathbb{R})$ *e completá-la de forma a obter uma base para* $M_{2\times 2}(\mathbb{R})$.

A observação básica para realizar estes cálculos é que estas matrizes se identificam naturalmente com vetores de $\mathbb{R}^4$ *através da correspondência*

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \Leftrightarrow (a, b, c, d)$$

De facto tanto a soma como o produto por escalar são, em ambos os casos, efetuados coordenada a coordenada. Para determinar uma base para $L(S)$ *podemos portanto (conforme o Exemplo 4.4(viii)) aplicar o método de Gauss a uma matriz cujas linhas são os vetores de* $\mathbb{R}^4$ *correspondentes aos elementos de* S*:*

$$\begin{bmatrix} 1 & 2 & 0 & 1 \\ -1 & 0 & 1 & 2 \\ 0 & 2 & 1 & 3 \end{bmatrix} \xrightarrow{L_2+L_1} \begin{bmatrix} 1 & 2 & 0 & 1 \\ 0 & 2 & 1 & 3 \\ 0 & 2 & 1 & 3 \end{bmatrix} \xrightarrow{L_3-L_2} \begin{bmatrix} 1 & 2 & 0 & 1 \\ 0 & 2 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Conclui-se que uma base para $L(S)$ *é*

$$\left\{ \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 2 \\ 1 & 3 \end{bmatrix} \right\}$$

(e portanto $L(S)$ *tem dimensão 2). Para completar este conjunto de forma a obter uma base de* $M_{2\times 2}(\mathbb{R})$ *precisamos de juntar dois vetores ao conjunto acima de forma a que o conjunto resultante seja ainda linearmente independente. Isto porque* $\dim M_{2\times 2}(\mathbb{R}) = 4$ *e portanto, pelo Corolário 4.10, qualquer subconjunto linearmente independente de* $M_{2\times 2}(\mathbb{R})$ *com quatro elementos constitui uma base para* $M_{2\times 2}(\mathbb{R})$.

Podemos novamente apoiar-nos na correspondência entre $M_{2\times 2}(\mathbb{R})$ *e* $\mathbb{R}^4$ *e no facto de as linhas de uma matriz em escada de linhas serem linearmente independentes. Uma vez que*

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

está em escada de linhas, o conjunto

$$\left\{ \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 2 \\ 1 & 3 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

é uma base de $M_{2\times 2}(\mathbb{R})$ *contendo a base de* $L(S)$.

5. Mudanças de coordenadas

Definição 5.1. *Uma* base ordenada B *de um espaço vetorial de dimensão finita* V *é uma sequência finita* $B = (v_1, \ldots, v_n)$ *de vetores distintos* $v_i \in V$ *tais que o conjunto* $\{v_1, \ldots, v_n\}$ *é linearmente independente e gera* V

Como o nome indica, a diferença entre base e base ordenada é que numa base ordenada escolhemos explicitamente uma ordem para os vetores da base. Há um primeiro vetor da base, um segundo, etc... Na realidade até agora, quando fizemos cálculos escolhemos implicitamente uma ordem para os vetores das bases envolvidas de forma a poder identificar o espaço vetorial em questão com $\mathbb{R}^n$.

Uma base ordenada $B = (v_1, \ldots, v_n)$ determina uma bijeção natural

$$V \longleftrightarrow \mathbb{R}^n$$

que faz corresponder a um vetor $v \in V$ os seus coeficientes na base B, na ordem indicada,

$$v = \alpha_1 v_1 + \ldots + \alpha_n v_n \longleftrightarrow (\alpha_1, \ldots, \alpha_n)$$

O escalar α_i diz-se *a i-ésima coordenada* de v na base ordenada B.

Exemplo 5.2. *(i) A base ordenada canónica de* $\mathbb{R}^n$ *é* $B = (e_1, \ldots, e_n)$*, onde* $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ *(com o* 1 *na posição* i*). Uma vez que*

$$(x_1, \ldots, x_n) = x_1 e_1 + \ldots + x_n e_n$$

as coordenadas de $(x_1, \ldots, x_n)$ *na base canónica são* $(x_1, \ldots, x_n)$.

(ii) Para $0 < \alpha < \frac{\pi}{2}$*, seja* $B = ((\cos\alpha, \operatorname{sen}\alpha), (-\operatorname{sen}\alpha, \cos\alpha))$ *a base ordenada de* $\mathbb{R}^2$ *que se obtém rodando os vetores da base canónica um ângulo* α *no sentido anti-horário. Vamos achar as coordenadas do vetor* $(1, 0)$ *na base* B.

Podemos fazê-lo usando a interpretação geométrica das coordenadas (conforme o Exemplo 3.6) e trigonometria elementar obtendo $(\cos\alpha, -\operatorname{sen}\alpha)$ *ou, alternativamente, resolvendo o sistema*

$$(1,0) = c_1(\cos\alpha, \operatorname{sen}\alpha) + c_2(-\operatorname{sen}\alpha, \cos\alpha) \Leftrightarrow \begin{cases} c_1 \cos\alpha - c_2 \operatorname{sen}\alpha = 1 \\ c_1 \operatorname{sen}\alpha + c_2 \cos\alpha = 0 \end{cases}$$

A combinação linear $\cos\alpha L_1 + \operatorname{sen}\alpha L_2$ *das duas equações do sistema produz* $c_1 = \cos\alpha$*, e substituindo na segunda equação temos*

$$\cos\alpha \operatorname{sen}\alpha + c_2 \cos\alpha = 0 \Leftrightarrow c_2 = -\operatorname{sen}\alpha$$

(uma vez que $\cos\alpha > 0$*). Em geral, podemos ver geometricamente qual é a relação entre as coordenadas* (a, b) *de um vetor na base canónica e as suas coordenadas na base* B*. As coordenadas na base* B *obtêm-se de* (a, b) *rodando este vetor um ângulo* α no sentido horário.

Vimos no exemplo anterior que as coordenadas na nova base B podiam ser obtidas a partir das coordenadas noutra base (a base canónica) através de uma certa transformação. É natural perguntar em geral qual é a relação entre as coordenadas de um vetor $v \in V$ em duas bases ordenadas $B_1 = (v_1, \ldots, v_n)$ e $B_2 = (w_1, \ldots, w_n)$ de V dadas.

Seja
$$v = \alpha_1 v_1 + \ldots + \alpha_n v_n$$
Para achar as coordenadas de v na base B_2 podemos escrever os vetores v_i na base B_2:
$$\begin{aligned} v_1 &= a_{11}w_1 + a_{21}w_2 + \ldots + a_{n1}w_n \\ v_2 &= a_{12}w_1 + a_{22}w_2 + \ldots + a_{n2}w_n \\ &\vdots \\ v_n &= a_{1n}w_1 + a_{2n}w_2 + \ldots + a_{nn}w_n \end{aligned}$$
Substituindo na fórmula para v obtemos
$$\begin{aligned} v &= \alpha_1(a_{11}w_1 + a_{21}w_2 + \ldots + a_{n1}w_n) + \alpha_2(a_{12}w_1 + a_{22}w_2 + \ldots + a_{n2}w_n) + \\ &\quad \ldots + \alpha_n(a_{1n}w_1 + a_{2n}w_2 + \ldots + a_{nn}w_n) \\ &= (a_{11}\alpha_1 + a_{12}\alpha_2 + \ldots + a_{1n}\alpha_n)w_1 + (a_{21}\alpha_1 + a_{22}\alpha_2 + \ldots + a_{2n}\alpha_n)w_2 + \\ &\quad \ldots + (a_{n1}\alpha_1 + a_{n2}\alpha_2 + \ldots + a_{nn}\alpha_n)w_n \end{aligned}$$
Escrevendo $(\beta_1, \ldots, \beta_n)$ para as coordenadas do vetor v na base B_2 temos portanto
$$\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}$$
onde na coluna j da matriz $[a_{ij}]$ aparecem as coordenadas do vetor v_j na base B_2.

Proposição 5.3. *Seja V um espaço vetorial de dimensão n e B_1 e B_2 bases ordenadas para V. Existe uma única matriz $n \times n$, denotada por $S_{B_1 \to B_2}$, tal que para todo o vetor $v \in V$, as coordenadas $(\beta_1, \ldots, \beta_n)$ de v na base B_2 e as coordenadas $(\alpha_1, \ldots, \alpha_n)$ de v na base B_1 estão relacionadas da seguinte forma*
$$\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} = S_{B_1 \to B_2} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}$$
A esta matriz chama-se a matriz de mudança de coordenadas da base B_1 para a base B_2.

Dem. Já observámos acima que é possível relacionar as coordenadas através de uma matriz. Para ver que a matriz é única note-se que se existir uma tal matriz S então a j-ésima coluna da matriz terá necessariamente de consistir nas coordenadas do j-ésimo vetor da base B_1 na base B_2. De facto, as coordenadas desse vetor (chamemos-lhe v_j) na base B_1 são $(0, \ldots, 0, 1, 0, \ldots, 0)$ com o 1 na j-ésima posição, e ao multiplicarmos a matriz S por este vetor de coordenadas obtemos a j-ésima coluna de S que tem então que conter as coordenadas de v_j na base B_2. □

Exemplo 5.4. *A matriz de mudança de base da base canónica B_{can} de $\mathbb{R}^2$ para a base B do Exemplo 5.2 é dada por*

$$S_{B_{can}\to B} = \begin{bmatrix} \cos\alpha & \operatorname{sen}\alpha \\ -\operatorname{sen}\alpha & \cos\alpha \end{bmatrix}$$

De fato, a primeira coluna contém as componentes do primeiro vetor da base canónica na base B como vimos no Exemplo 5.2 e da mesma forma podemos verificar que a segunda coluna contém as coordenadas do vetor $(0,1)$ na base B. Note-se que o efeito que tem a multiplicação desta matriz por um vetor coluna é a rotação do vetor um ângulo α no sentido horário conforme tínhamos previsto geometricamente.

Proposição 5.5. *Seja V um espaço vetorial de dimensão finita e B_1, B_2, B_3 bases ordenadas para V. Temos as seguintes relações entre as matrizes de mudança de coordenadas:*

(i) $S_{B_1\to B_3} = S_{B_2\to B_3}S_{B_1\to B_2}$
(ii) $S_{B_2\to B_1} = (S_{B_1\to B_2})^{-1}$

Dem. (i) Sejam X_1, X_2 e X_3 os vetores coluna contendo as coordenadas de um dado vetor $v \in V$. Por definição das matrizes de mudança de coordenadas temos

$$X_2 = S_{B_1\to B_2}X_1, \qquad X_3 = S_{B_2\to B_3}X_2$$

Substituindo a primeira equação na segunda obtemos

$$X_3 = S_{B_2\to B_3}\left(S_{B_1\to B_2}X_1\right) = \left(S_{B_2\to B_3}S_{B_1\to B_2}\right)X_1$$

Uma vez que a equação anterior é válida para qualquer vetor $v \in V$ e a matriz de mudança de coordenadas é única conclui-se que

$$S_{B_1\to B_3} = S_{B_2\to B_3}S_{B_1\to B_2}$$

(ii) Claramente, para qualquer base ordenada B com n elementos, temos que a matriz de mudança de coordenadas da base B para ela própria é a matriz identidade I_n. Aplicando o ponto (i) com $B_3 = B_1$ obtemos

$$I_n = S_{B_2\to B_1}S_{B_1\to B_2}$$

e da mesma forma, trocando B_1 com B_2

$$I_n = S_{B_1\to B_2}S_{B_2\to B_1}$$

o que mostra que $S_{B_1\to B_2}$ e $S_{B_2\to B_1}$ são matrizes inversas.

□

Observação 5.6. *Note-se que o ponto (ii) da Proposição anterior diz, em particular, que uma matriz de mudança de base é sempre invertível. Reciprocamente, é um exercício da ficha para as aulas práticas que qualquer matriz invertível é uma matriz de mudança de base (a partir de qualquer base dada).*

6. Transformações lineares

Na Matemática estuda-se certos objetos cuja natureza depende da área da Matemática. Por exemplo, na Álgebra Linear estuda-se espaços vetoriais, enquanto que em Geometria se pode estudar, por exemplo, curvas e superfícies. Normalmente estes objetos consistem em conjuntos munidos de certa estrutura adicional. No caso dos espaços vetoriais esta estrutura adicional toma a forma das operações de soma de vetores e o produto de vetores por escalares. Para estudar os objetos em questão é sempre necessário pensar em como se relacionam entre eles. As relações entre os objetos manifestam-se através de funções entre os conjuntos subjacentes que preservam a estrutura adicional. No caso que nos interessa agora isso leva-nos à seguinte definição.

Definição 6.1. *Sejam V e W espaços vetoriais. Uma função $f\colon V \to W$ diz-se uma* transformação linear *de V para W se*

(i) $f(v_1 + v_2) = f(v_1) + f(v_2)$ para todos os $v_1, v_2 \in V$.
(ii) $f(\alpha v) = \alpha f(v)$ para todo o $v \in V$ e escalar α.

As transformações lineares são portanto as funções entre os conjuntos subjacentes aos espaços vetoriais que preservam a soma e o produto por escalar. Note-se que na definição acima aparecem duas somas (em geral) distintas no axioma (i): do lado esquerdo do sinal de igual, a soma é a soma de vetores em V, enquanto que do lado direito se trata da soma em W. Analogamente para os dois produtos por escalar que aparecem no axioma (ii).

Chamamos a atenção para as seguintes consequências imediatas dos axiomas acima: uma transformação linear leva necessariamente o vetor $0 \in V$ no vetor $0 \in W$. De facto, sendo $v \in V$ um vetor qualquer sabemos que $0 \cdot v = 0$. Como f preserva o produto por escalar temos então

$$f(0) = f(0 \cdot v) = 0 \cdot f(v) = 0 \in W$$

A outra observação importante é que uma transformação linear leva combinações lineares em V para combinações lineares em W: dados escalares $\alpha_1, \ldots, \alpha_n$ e vetores $v_1, \ldots, v_n$ temos

$$\begin{aligned} f(\alpha_1 v_1 + \ldots + \alpha_n v_n) &= f(\alpha_1 v_1) + f(\alpha_2 v_2) + \ldots + f(\alpha_n v_n) \\ &= \alpha_1 f(v_1) + \ldots + \alpha_n f(v_n) \end{aligned}$$

Vejamos alguns exemplos de transformações lineares $f\colon V \to W$.

Exemplo 6.2. *(1) Sejam $V = W = \mathbb{R} = \mathbb{R}^1$. A função $f\colon \mathbb{R} \to \mathbb{R}$ definida pela expressão $f(x) = 2x$ é uma transformação linear. De facto temos*

$$f(x_1 + x_2) = 2(x_1 + x_2) = 2x_1 + 2x_2 = f(x_1) + f(x_2)$$

$$f(\alpha x) = 2(\alpha x) = \alpha(2x) = \alpha f(x)$$

O gráfico de f é uma linha reta que passa pela origem. Mais geralmente, é fácil ver (exercício) que uma função $f\colon \mathbb{R} \to \mathbb{R}$ é uma transformação linear se e só se f é uma função linear, isto é, da forma $f(x) = ax$ para algum número real $a \in \mathbb{R}$. Assim, as transformações lineares são as funções reais de variável real cujos gráficos são retas que passam pela origem.

Por exemplo, a expressão $f(x) = 3x + 1$ *não define uma transformação linear de* $\mathbb{R}$ *para* $\mathbb{R}$*. De facto* $f(0+0) = 1$ *é diferente de* $f(0) + f(0) = 1 + 1 = 2$*. Alternativamente,* $f(0) = 1 \neq 0$ *e vimos acima que uma transformação linear leva sempre o vetor nulo do conjunto de partida no vetor nulo do conjunto de chegada.*

(2) *Sejam* $V = W = \mathbb{R}^2$ *e identifiquemos como habitualmente* $\mathbb{R}^2$ *com o plano. Considere-se a função* $f : \mathbb{R}^2 \to \mathbb{R}^2$ *definida geometricamente como "rotação de 90 graus em torno da origem no sentido anti-horário". Apelando ao significado geométrico da soma de vetores e produto por escalar é imediato verificar que esta transformação preserva a soma de vetores e o produto por escalar pelo que é uma transformação linear.*

Podemos verificar a afirmação anterior obtendo uma expressão analítica para a função f*. Sendo* (a, b) *um vetor no primeiro quadrante é imediato verificar que após a rotação o vetor fica com coordenadas* $(-b, a)$*. É fácil verificar que o mesmo sucede para qualquer vetor pelo que a expressão analítica para a rotação é*

$$f(a, b) = (-b, a)$$

Podemos agora ver que f *é uma transformação linear:*

$$\begin{aligned} f((a_1, b_1) + (a_2, b_2)) &= f(a_1 + a_2, b_1 + b_2) \\ &= (-b_1 - b_1, a_1 + a_2) = (-b_1, a_1) + (-b_2, a_2) \\ &= f(a_1, b_1) + f(a_2, b_2) \end{aligned}$$

e

$$f(\alpha(a, b)) = f(\alpha a, \alpha b) = (-\alpha b, \alpha a) = \alpha(-b, a) = f(a, b)$$

Note-se que identificando os vetores de $\mathbb{R}^2$ *com matrizes coluna* 2×1*, podemos escrever* f *da seguinte forma*

$$f\left(\begin{bmatrix} a \\ b \end{bmatrix}\right) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$$

(3) *Seja* $V = \mathbb{R}^n, W = \mathbb{R}^m$ *e* A *uma matriz* $m \times n$*. Identificando como habitualmente vetores de* $\mathbb{R}^n$ *com matrizes coluna podemos definir uma transformação linear* $f : \mathbb{R}^n \to \mathbb{R}^m$ *através da fórmula*

$$f(x) = Ax$$

O exemplo anterior é um caso particular deste. De facto, o primeiro exemplo também é. Nesse caso, $A = [a]$ *é uma matriz* 1×1*.*

(4) *Seja* $W = F(\mathbb{R}, \mathbb{R})$ *o espaço vetorial das funções reais de variável real e*

$$V = \{f \in W : f \text{ é diferenciável}\}$$

o subespaço vetorial formado pelas funções diferenciáveis. Então a aplicação $T : V \to W$ *definida por*

$$T(f) = f'$$

ou seja a operação de derivação, é uma transformação linear. De facto temos

$$T(f + g) = (f + g)' = f' + g' = T(f) + T(g)$$

e

$$T(\alpha f) = (\alpha f)' = \alpha f'$$

pelas regras de derivação para a soma e para o produto por escalar. Estas regras dizem precisamente que a operação de derivação é uma transformação linear. Este exemplo é, pelo menos aparentemente, muito diferente dos anteriores. O conceito de transformação linear estabelece assim uma relação entre operações tão diferentes como uma rotação do plano e a operação de derivação de uma função.

(5) *Seja* $V = M_{m\times n}(\mathbb{R})$ *e* $W = M_{p\times q}(\mathbb{R})$ *e sejam* B *uma matriz* $p \times m$ *e* C *uma matriz* $n \times q$. *Então a aplicação* $T\colon V \to W$ *definida pela fórmula*

$$T(A) = BAC$$

é uma transformação linear:

$$\begin{aligned} T(A_1 + A_2) &= B(A_1 + A_2)C = (BA_1 + BA_2)C \\ &= BA_1C + BA_2C = T(A_1) + T(A_2) \end{aligned}$$

(pela distributividade do produto de matrizes em relação à soma, e associatividade da multiplicação de matrizes) e

$$T(\alpha A) = B(\alpha A)C = (\alpha BA)C = \alpha BAC$$

pela relação entre o produto de matrizes e o produto por escalar. Um exemplo concreto é por exemplo a transformação $T\colon M_{2\times 2}(\mathbb{R}) \to M_{4\times 3}(\mathbb{R})$ *determinada pelas matrizes*

$$B = \begin{bmatrix} 1 & 3 \\ -2 & 0 \\ -1 & 1 \\ 2 & 0 \end{bmatrix}, \qquad C = \begin{bmatrix} 0 & 1 & 2 \\ -1 & 1 & 0 \end{bmatrix}$$

que é dada pela fórmula

$$T\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{bmatrix} 1 & 3 \\ -2 & 0 \\ -1 & 1 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 0 & 1 & 2 \\ -1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} -b-3d & a+b+3c+3d & 2a+6c \\ 2b & -2a-2b & -4a \\ b-d & -a-b+c+d & -2a+2c \\ -2b & 2a+2b & 4a \end{bmatrix}$$

(6) *Seja* V *o espaço vetorial dos polinómios e* $W = \mathbb{R}^2$. *Então a função* $f\colon V \to \mathbb{R}^2$ *definida por*

$$f(p) = (p(1), p''(2))$$

é uma transformação linear:

$$\begin{aligned} f(p+q) &= ((p+q)(1), (p+q)''(2)) = (p(1) + q(1), p''(2) + q''(2)) \\ &= (p(1), p''(2)) + (q(1), q''(2)) = f(p) + f(q) \end{aligned}$$

$$f(\alpha p) = ((\alpha p)(1), (\alpha p)''(2)) = (\alpha p(1), \alpha p''(2)) = \alpha(p(1), p''(2))$$

porque a soma de funções e a multiplicação de uma função por escalar são calculadas ponto a ponto e pelas regras de derivação. Note-se que este exemplo é, pelo menos aparentemente, de uma natureza bastante diferente dos exemplos (1)-(5) acima.

Proposição 6.3. *Sejam V, W espaços vetoriais, $B = \{v_1, \ldots, v_n\}$ uma base para V e $w_1, \ldots, w_n$ vetores quaisquer de W. Então existe uma única transformação linear $f \colon V \to W$ tal que $f(v_i) = w_i$.*

Dem. Começamos por mostrar a unicidade. Suponhamos que $f \colon V \to W$ é uma transformação linear tal que $f(v_i) = w_i$. Dado um vetor $v \in V$ qualquer, existem escalares $\alpha_1, \ldots, \alpha_n$ únicos tais que

$$v = \alpha_1 v_1 + \ldots + \alpha_n v_n$$

Uma vez que uma transformação linear preserva combinações lineares, teremos necessariamente

$$f(v) = f(\alpha_1 v_1 + \ldots + \alpha_n v_n) = \alpha_1 f(v_1) + \ldots + \alpha_n f(v_n) = \alpha_1 w_1 + \ldots + \alpha_n w_n \tag{14}$$

Obtemos assim uma fórmula para f que mostra a unicidade da transformação linear (caso exista). Para verificar que existe e completar a demonstração resta ver que a expressão (14) define efetivamente uma transformação linear com as propriedades requeridas. Seja então $f \colon V \to W$ a função definida pela expressão (14).

- f envia o vetor $v_i \in B$ em w_i: Temos $v_i = 0 \cdot v_1 + \ldots + 0 \cdot v_{i-1} + 1 \cdot v_i + 0 \cdot v_{i+1} + \ldots 0 \cdot v_n$ logo $f(v_i) = 0 \cdot w_1 + \ldots + 0 \cdot w_{i-1} + 1 \cdot w_i + 0 \cdot w_{i+1} + \ldots 0 \cdot w_n = w_i$.
- f é uma transformação linear: Sejam $v = \alpha_1 v_1 + \ldots + \alpha_n v_n$ e $w = \beta_1 v_1 + \ldots + \beta_n v_n$ dois vetores quaisquer de V. Então $v + w = (\alpha_1 + \beta_1) v_1 + \ldots + (\alpha_n + \beta_n) v_n$ pelo que

$$\begin{aligned} f(v+w) &= (\alpha_1 + \beta_1) w_1 + \ldots + (\alpha_n + \beta_n) w_n \\ &= (\alpha_1 v_1 + \ldots + \alpha_n v_n) + (\beta_1 v_1 + \ldots + \beta_n v_n) = f(v) + f(w) \end{aligned}$$

  e, dado um escalar α temos $\alpha v = \alpha\alpha_1 v_1 + \ldots + \alpha\alpha_n v_n$ e portanto

$$f(\alpha v) = \alpha\alpha_1 w_1 + \ldots + \alpha\alpha_n w_n = \alpha(\alpha_1 w_1 + \ldots + \alpha_n w_n) = \alpha f(v)$$

  o que conclui a demonstração.

$\square$

O resultado anterior pode ser visto (pelo menos) de duas maneiras diferentes. Por um lado, dá-nos um método para construir transformações lineares: basta escolher uma base para o espaço de partida e decidir qual o valor que irá tomar em cada vetor da base. Além disso a demonstração acima dá-nos uma fórmula ((14)) para a transformação linear assim obtida. Por outro lado, a Proposição diz-nos que as transformações lineares são funções excepcionalmente simples. Para definir uma função de V para W é normalmente necessário decidir o seu valor individualmente para cada vetor de V. A Proposição anterior diz que quando f é linear, todo o comportamento da função é completamente determinado pelos valores que toma num número finito de elementos do domínio (os vetores constituintes de uma base).

Observação 6.4. *A Proposição 6.3 é ainda válida quando a base de V é um conjunto infinito, sendo a demonstração essencialmente a mesma. Deixamos esta verificação como exercício às leitoras interessadas.*

Exemplo 6.5. *A transformação linear* $T\colon \mathbb{R}^2 \to \mathbb{R}^3$ *tal que* $T(1,0) = (2,1,-3)$ *e* $T(0,1) = (4,1,5)$ *é a função definida pela expressão*

$$T(a,b) = a(2,1,-3) + b(4,1,5) = (2a+4b, a+b, -3a+5b)$$

que pode ser representada matricialmente por

$$T\left(\begin{bmatrix} a \\ b \end{bmatrix}\right) = \begin{bmatrix} 2 & 4 \\ 1 & 1 \\ -3 & 5 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$$

Claramente o exemplo anterior pode ser generalizado a qualquer transformação linear de $\mathbb{R}^m$ para $\mathbb{R}^n$ e vemos assim que o Exemplo 6.2 (3) é na realidade exaustivo. Vamos agora ver que em completa generalidade, desde que os espaços vetoriais envolvidos tenham dimensão finita, uma transformação linear é determinada por uma matriz. Antes disso aproveitamos para introduzir notação para as coordenadas de um vetor numa base ordenada.

Definição 6.6. *Seja* V *um espaço vetorial,* $B = (v_1, \ldots, v_n)$ *uma base ordenada para* V *e* $v = \alpha_1 v_1 + \ldots + \alpha_n v_n$ *um vetor de* V*. Escrevemos* $[v]_B$ *para a matriz coluna* $n \times 1$ *cujas componentes são as coordenadas de* v *(por ordem):*

$$[v]_B = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}$$

Uma base finita B com n elementos determina uma função $f\colon V \to M_{n\times 1}(\mathbb{R})$ definida por

$$f(v) = [v]_B$$

que é uma bijeção (pela unicidade das coordenadas). Aliás é esta identificação que temos usado, informalmente, para efetuar cálculos em espaços vetoriais de polinómios e matrizes.

Exercício 6.7. *Dado um espaço vetorial* V *e uma base* $B = (v_1, \ldots, v_n)$ *para* V*, verifique que a função* $f\colon V \to M_{n\times 1}(\mathbb{R})$ *definida por* $f(v) = [v]_B$ *é uma transformação linear.*

Proposição 6.8. *Sejam* V, W *espaços vetoriais e* $B_1 = (v_1, \ldots, v_m)$ *e* $B_2 = (w_1, \ldots, w_n)$ *bases ordenadas para* V *e* W *respetivamente. Seja* $f\colon V \to W$ *uma transformação linear. Então existe uma única matriz* $A_{f,B_1,B_2} \in M_{m\times n}(\mathbb{R})$ *tal que, para todo o vetor* $v \in V$ *se tem*

$$[f(v)]_{B_2} = A_{f,B_1,B_2}[v]_{B_1}$$

A matriz A_{f,B_1,B_2} diz-se a matriz que *representa a transformação linear* f *com respeito às bases* B_1 *e* B_2.

Exemplo 6.9. *(i) Seja* V *um espaço vetorial com bases* $B_1 = (v_1, \ldots, v_n)$ *e* $B_2 = (w_1, \ldots, w_n)$ *e* $\mathrm{Id}\colon V \to V$ *a função identidade (definida por* $\mathrm{Id}(v) = v$*). É imediato verificar que* Id *é uma transformação linear. Temos então, por definição de matriz mudança de base*

$$A_{\mathrm{Id},B_1,B_2} = S_{B_1\to B_2}$$

De facto, a identidade

$$[\mathrm{Id}(v)]_{B_2} = A_{\mathrm{Id},B_1,B_2}[v]_{B_1} \Leftrightarrow [v]_{B_2} = A_{\mathrm{Id},B_1,B_2}[v]_{B_1}$$

mostra que A_{Id,B_1,B_2} satisfaz a relação que caracteriza a matriz de mudança de coordenadas, e como tal (por unicidade), é a matriz de mudança de coordenadas $S_{B_1\to B_2}$.

(ii) Seja V o espaço vetorial dos polinómios de grau ≤ 3 e considere-se a transformação linear $T\colon V\to V$ definida por $T(p)=p'$. Uma vez que

$$T(a+bx+cx^2+dx^3) = b+2cx+3dx^2,$$

sendo $B=(1,x,x^2,x^3)$ a base canónica, a equação $[T(p)]_B = A_{T,B,B}[p]_B$ para a matriz $A_{T,B,B}$ fica

$$\begin{bmatrix} b \\ 2c \\ 3d \\ 0 \end{bmatrix} = A_{T,B,B}\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$$

e conclui-se então que

$$A_{T,B,B} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Vale a pena refletir durante um momento no facto de a matriz acima representar a operação de derivação (embora no contexto restrito dos polinómios de grau menor ou igual a 3).

Dem. da Proposição 6.8. Vejamos primeiro ver que se a matriz A_{f,B_1,B_2} existir, ela é única. Para o i-ésimo vetor da base B_1, $v=v_i$, a equação que caracteriza a matriz A_{f,B_1,B_2} é

$$[f(v_i)] = A_{f,B_1,B_2}[v_i]_{B_1}$$

mas, uma vez que $[v_i]_{B_1}$ tem todas as entradas iguais a 0 exceto a i-ésima que é igual a 1, o produto no termo direito da equação acima é a i-ésima coluna da matriz A_{f,B_1,B_2}. Isto mostra que a matriz A_{f,B_1,B_2} fica univocamente determinada: se existir, a sua i-ésima coluna é necessariamente igual a $[f(v_i)]_{B_2}$.

Para completar a demonstração basta agora verificar que a matriz $n\times n$ cuja i-ésima coluna é $[f(v_i)]_{B_2}$ satisfaz a equação do enunciado. Seja $v=\alpha_1v_1+\ldots+\alpha_nv_n$ um vetor de V. Então

$$\begin{aligned}
[f(v)]_{B_2} &= [f(\alpha_1v_1+\ldots+\alpha_nv_n)]_{B_2} \\
&= [\alpha_1f(v_1)+\ldots+\alpha_nf(v_n)]_{B_2} \\
&= \alpha_1[f(v_1)]_{B_2}+\ldots+\alpha_n[f(v_n)]_{B_2}
\end{aligned}$$

onde na segunda igualdade usámos o facto de f ser uma transformação linear e na terceira o Exercício 6.7. Pela definição do produto de matrizes a expressão

$$\alpha_1[f(v_1)]_{B_2}+\ldots+\alpha_n[f(v_n)]_{B_2}$$

é exatamente o produto da matriz que tem por i-ésima coluna $[f(v_i)]_{B_2}$ pelo vetor coluna com componentes $(\alpha_1, \ldots, \alpha_n)$, que por sua vez, é exatamente $[v]_{B_1}$. Isto conclui a demonstração. $\square$

A Proposição 6.8 permite identificar uma transformação linear entre espaços vetoriais de dimensão finita com uma matriz mediante a escolha de bases para o espaço vetorial de partida e de chegada. Além disso explica como obter a matriz em questão: é a matriz cuja i-ésima coluna contém as coordenadas do i-ésimo vetor da base do espaço de partida na base do espaço de chegada.

Isto é extremamente útil para fazer contas com transformações lineares como iremos ver em seguida. Convém no entanto notar que a Proposição não se aplica a todos os exemplos de transformação linear que queremos considerar - por exemplo, à operação de derivação. Por outro lado, o objeto em que normalmente estamos interessados é a transformação linear ela própria e não uma (das muitas possíveis) representações matriciais que usamos para calcular. Uma analogia que pode ser útil é que uma transformação linear é como uma ideia, que se pode exprimir em várias línguas, as bases nos espaços de partida e de chegada são como uma escolha de língua, e a matriz que representa a transformação linear é a palavra que representa a ideia na língua escolhida.

Vamos agora continuar o estudo das transformações lineares que se vai prolongar até ao final do semestre.

Definição 6.10. *Sejam V e W espaços vetoriais. Escrevemos $L(V, W)$ para o conjunto das transformações lineares de V para W. Dadas $f, g \in L(V, W)$ e um escalar α definimos a* soma *de f e g como sendo a função $f + g \colon V \to W$ definida pela expressão*

$$(f + g)(v) = f(v) + g(v)$$

e definimos o produto de uma transformação linear f pelo escalar α *como sendo a função $\alpha f \colon V \to W$ definida pela expressão*

$$(\alpha f)(v) = \alpha \cdot f(v).$$

Proposição 6.11. *Sejam V e W espaços vetoriais. Com as operações de soma e produto por escalar definidas acima, o conjunto $L(V, W)$ é um espaço vetorial.*

Dem. Temos a verificar que as operações de soma e produto por escalar estão bem definidas, isto é, que dadas $f, g \in L(V, W)$ e um escalar f, as funções $f + g$ e αf estão ainda em $L(V, W)$ e depois os oito axiomas que estas operações devem satisfazer num espaço vetorial.

Vemos primeiro que $f + g$ é uma transformação linear: dados $v_1, v_2 \in V$ temos

$$\begin{aligned} (f+g)(v_1+v_2) &= f(v_1+v_2) + g(v_1+v_2) = f(v_1) + f(v_2) + g(v_1) + g(v_2) \\ &= f(v_1) + g(v_1) + f(v_2) + g(v_2) = (f+g)(v_1) + (f+g)(v_2) \end{aligned}$$

e dado um escalar α e $v \in V$ temos

$$(f+g)(\alpha v) = f(\alpha v) + g(\alpha v) = \alpha f(v) + \alpha g(v) = \alpha(f(v) + g(v)) = \alpha((f+g)(v))$$

A verificação que $(\alpha f) \in L(V, W)$ é análoga e fica como exercício. A verificação dos axiomas de espaço vetorial é também deixada como exercício. Notamos apenas que o vetor

$0 \in L(V, W)$ é a transformação linear identicamente nula que envia todos os vetores $v \in V$ para $0 \in W$. $\square$

Proposição 6.12. *Sejam V, W, U espaços vetoriais e $f: V \to W$, e $g: W \to U$ transformações lineares. Então a função composta*

$$g \circ f: V \to U$$

é uma transformação linear.

Dem. Temos a verificar que $g \circ f$ preserva a soma e o produto por escalar.

- Dados $v_1, v_2 \in V$ temos

$$(g \circ f)(v_1 + v_2) = g(f(v_1 + v_2)) = g(f(v_1) + f(v_2)) = g(f(v_1)) + g(f(v_2)) = (g \circ f)(v_1) + (g \circ f)(v_2)$$

  onde na segunda igualdade usámos o facto de f ser uma transformação linear, e na terceira, o facto de g ser uma transformação linear.
- Dados um escalar α e um vetor $v \in V$ temos

$$(g \circ f)(\alpha v) = g(f(\alpha v)) = g(\alpha f(v)) = \alpha g(f(v)) = \alpha(g \circ f)(v)$$

  onde, tal como acima, na segunda igualdade usámos o facto de f ser uma transformação linear, e na terceira, o facto de g ser uma transformação linear.

$\square$

Proposição 6.13. *Sejam V, W espaços vetoriais e $f: V \to W$ uma transformação linear. Se a função f é invertível (isto é, se é bijetiva) então a função inversa $f^{-1}: W \to V$ é uma transformação linear.*

Proof. Temos a verificar que a função inversa f^{-1} preserva a soma e a multiplicação por escalar. Sejam w_1, w_2 vetores de W. Como f é sobrejetiva existem vetores v_1 e v_2 de V tais que $f(v_1) = w_1$ e $f(v_2) = w_2$. Então

$$f^{-1}(w_1 + w_2) = f^{-1}(f(v_1) + f(v_2)) = f^{-1}(f(v_1 + v_2)) = (f^{-1} \circ f)(v_1 + v_2) = v_1 + v_2$$

onde na segunda igualdade usámos o facto de f ser uma transformação linear. Por definição de função inversa temos que $v_1 = f^{-1}(w_1)$ e $v_2 = f^{-1}(w_2)$. Substituindo na igualdade acima concluímos que $f^{-1}: W \to V$ preserva a soma de vetores. A verificação que f^{-1} preserva o produto por escalar é análoga e fica como exercício. $\square$

Observação 6.14. *Alternativamente, na demonstração anterior poderíamos ter aplicado a função injetiva (por hipótese) f às expressões $f^{-1}(w_1 + w_2)$ e $f^{-1}(w_1) + f^{-1}(w_2)$ e verificado que essas contas produziam o mesmo resultado. A injetividade de f garante então que $f^{-1}(w_1 + w_2) = f^{-1}(w_1) + f^{-1}(w_2)$.*

Definição 6.15. *Sejam V, W espaços vetoriais. Uma transformação linear invertível $f: V \to W$ diz-se um* isomorfismo *de espaços vetoriais.*

A palavra isomorfismo vem de "iso" - igual - e "morphos" - forma. Um isomorfismo entre dois espaços vetoriais é uma equivalência entre eles. O isomorfismo estabelece uma correspondência bijetiva entre os conjuntos subjacentes (um "dicionário" entre os vetores de um dos espaços e os vetores do outro). Uma vez que a função e a sua inversa preservam as operações dos espaços vetoriais ou, equivalentemente, as combinações lineares, qualquer propriedade ou afirmação acerca de um dos espaços (que se possa expressar usando combinações lineares) será verdadeira se e só se for verdadeira no outro. Por exemplo um conjunto será linearmente (in)dependente num espaço se e só se a sua imagem através do isomorfismo for linearmente (in)dependente no outro. A verificação da afirmação anterior assim como de outras do mesmo género ficará como exercício na próxima Ficha para as aulas práticas.

Exemplo 6.16. *(i) As funções $M_{n\times 1}(\mathbb{R}) \to \mathbb{R}^n$ e $M_{1\times n}(\mathbb{R}) \to \mathbb{R}^n$ definidas por*

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \mapsto (x_1, \ldots, x_n) \qquad e \qquad \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \mapsto (x_1, \ldots, x_n)$$

são isomorfismos de espaços vetoriais. De facto as funções descritas acima são claramente bijetivas e também transformações lineares (pela definição de soma e produto por escalar nos vários espaços envolvidos).

(ii) Seja V um espaço vetorial com base ordenada $B = (v_1, \ldots, v_n)$. A função $f\colon V \to M_{n\times 1}(\mathbb{R})$ definida por

$$f(v) = [v]_B$$

que calcula a matriz coluna das coordenadas na base ordenada B é um isomorfismo. Que f é uma transformação linear é o conteúdo do Exercício 6.7. A função f é também bijetiva: a sobrejetividade de f traduz o facto que qualquer n-tuplo $(\alpha_1, \ldots, \alpha_n)$ de escalares formar as coordenadas de um vetor de V (nomeadamente de $v = \alpha_1 v_1 + \ldots \alpha_n v_n$), enquanto que a injetividade de f é uma consequència da unicidade das coordenadas de um vetor (que por sua vez é uma consequência de B ser um conjunto linearmente independente).

(iii) Sejam V, W espaços vetoriais e $B_1 = (v_1, \ldots, v_n), B_2 = (w_1, \ldots, w_m)$ bases ordenadas para V e W respetivamente. A função

$$\Phi\colon L(V, W) \to M_{m\times n}(\mathbb{R})$$

definida por (ver Proposição 6.8 para o significado da notação)

$$\Phi(f) = A_{f,B_1,B_2}$$

é um isomorfismo de espaços vetoriais. Portanto uma transformação linear entre espaços vetoriais finitamente gerados pode ser identificada com uma matriz, uma vez escolhidas bases ordenadas para o domínio e conjunto de chegada da transformação linear.

Temos que verificar que Φ é uma transformação linear e que é invertível (ou bijetiva) enquanto função.

- *Sejam $f, g\colon V \to W$ transformações lineares. Dados $v_1 v_2 \in V$ temos*

$$[(f+g)(v)]_{B_2} = [f(v) + g(v)]_{B_2} = [f(v)]_{B_2} + [g(v)]_{B_2} \tag{15}$$

onde na primeira igualdade usámos a definição de soma de transformações lineares e na segunda o facto que a operação de calcular as coordenadas é linear (algo que usámos também no ponto (ii) acima). Por definição das matrizes que representam f, g, e pela distributividade em relação à soma do produto de matrizes obtemos

$$[f(v)]_{B_2} + [g(v)]_{B_2} = A_{f,B_1,B_2}[v]_{B_1} + A_{g,B_1,B_2}[v]_{B_1} = (A_{f,B_1,B_2} + A_{g,B_1,B_2})[v]_{B_1} \tag{16}$$

Das igualdades (15) *e* (16) *obtemos, novamente por definição da matriz que representa* $(f + g)$,

$$A_{f+g,B_1,B_2} = A_{f,B_1,B_2} + A_{g,B_1,B_2}$$

ou seja

$$\Phi(f + g) = \Phi(f) + \Phi(g)$$

A demonstração que $\Phi(\alpha f) = \alpha\Phi(f)$ é análoga e fica como exercício. Concluímos que Φ é uma transformação linear.

- *Recorde-se da demonstração da Proposição 6.8 que a matriz $\Phi(f)$ tem como i-ésima coluna $[f(v_i)]_{B_2}$. Dada uma matriz A, pela Proposição 6.3 e o exemplo (ii) acima existe uma transformação linear f tal que $[f(v_i)]_{B_2}$ é a i-ésima coluna de A. Temos então $\Phi(f) = A$, o que mostra que Φ é sobrejetiva. Por outro lado, suponhamos que f e g são transformações lineares tais que $\Phi(f) = \Phi(g)$ então, para cada $i = 1, \ldots, n$, as coordenadas de $f(v_i)$ e $g(v_i)$ são iguais. Mas isto significa que $f(v_i) = g(v_i)$ para cada i, e então pela Proposição 6.3 temos que $f = g$. Isto mostra que Φ é uma função injetiva e portanto, dado que também é sobrejetiva, invertível.*

Conclui-se que Φ é um isomorfismo de espaços vetoriais.

Os exemplos anteriores dizem-nos que qualquer espaço vetorial real finitamente gerado é equivalente a $\mathbb{R}^n$ e que uma transformação linear entre tais espaços pode ser identificada com uma matriz. Estes factos são muito úteis para fazer contas. Já foram usados muitas vezes e continuarão a ser usados até ao final do semestre para esse efeito. No entanto não seria uma boa ideia concluir daqui que nos podemos concentrar exclusivamente em $\mathbb{R}^n$ e nas matrizes. Apesar de ser possível identificar um espaço finitamente gerado com algum $\mathbb{R}^n$ *não há em geral nenhuma maneira canónica de o fazer.* A identificação é feita através de uma *escolha* de base e há muitas escolhas possíveis. Um espaço vetorial geral não possui coordenadas especiais (ao contrário do que acontece em $\mathbb{R}^n$ e em vários outros exemplos que temos vindo a considerar como os espaços de matrizes) e esta é uma diferença muito importante. Veremos em breve que as soluções de certas equações diferenciais formam espaços vetoriais nos quais não há habitualmente qualquer "base canónica".

Proposição 6.17. *Sejam V, W, U espaços vetoriais, B_1, B_2, B_3 bases ordenadas para V, W, U respetivamente, e $f : V \to W$, $g : W \to U$ transformações lineares. Então a matriz que representa a transformação linear $g \circ f$ nas bases dadas é o produto da matriz que representa g pela matriz que representa f. Isto é,*

$$A_{g\circ f,B_1,B_3} = A_{g,B_2,B_3}A_{f,B_1,B_2}$$

Dem. Dado $v \in V$ temos pela definição das matrizes que representam f e g

$$\begin{aligned}[(g \circ f)(v)]_{B_3} &= [g(f(v))]_{B_3} = A_{g,B_2,B_3}[f(v)]_{B_2} \\ &= A_{g,B_2,B_3}(A_{f,B_1,B_2}[v]_{B_1}) = (A_{g,B_2,B_3}A_{f,B_1,B_2})[v]_{B_1}\end{aligned}$$

donde, pela unicidade da matriz que representa $g \circ f$ conclui-se que

$$A_{g\circ f,B_1,B_3} = A_{g,B_2,B_3}A_{f,B_1,B_2}$$

conforme pretendido. □

Esta proposição explica a associatividade do produto de matrizes: o produto de matrizes é a tradução através dos isomorfismos do Exemplo 6.16(iii) da composição de funções, que é uma operação associativa.

Observação 6.18. *É possível pensar visualmente na correspondência entre transformações lineares e matrizes, e em particular na Proposição anterior da seguinte forma. Considere-se o diagrama*

$$\begin{tikzcd} V \arrow[r, "f"] \arrow[d, "{[\cdot]_{B_1}}"', "\cong"] & W \arrow[d, "{[\cdot]_{B_2}}"', "\cong"] \\ M_{n\times 1}(\mathbb{R}) \arrow[r, "A_{f,B_1,B_2}"] & M_{m\times 1}(\mathbb{R}) \end{tikzcd} \tag{17}$$

onde as setas representam transformações lineares com domínio a origem da seta e conjunto de chegada o término da seta. As setas pretendem representar visualmente que os vetores do espaço da origem são "transportados" pela transformação linear do seu domínio até ao espaço vetorial de chegada. O símbolo $\cong$ designa isomorfismo e os isomorfismos no diagrama acima são os do Exemplo 6.16(ii) que calculam a matriz coluna das coordenadas, ou seja, $v \mapsto [v]_{B_1}$ para a seta da esquerda e $w \mapsto [w]_{B_2}$ para a seta da direita. A equação

$$[f(v)]_{B_2} = A_{f,B_1,B_2}[v]_{B_1} \tag{18}$$

diz que se obtém o mesmo resultado quando se faz um vetor $v \in V$ seguir os dois possíveis trajetos do canto superior esquerdo até ao canto inferior direito em (17)*: do lado esquerdo de* (18) *temos o efeito de seguir primeiro a seta de cima e depois a seta da direita; do lado direito de* (18) *segue-se primeiro a seta da esquerda e depois a de baixo.*

Quando independentemente do caminho seguido entre dois nós do diagrama se obtém sempre o mesmo resultado diz-se que o diagrama é comutativo. *Portanto a equação* (18) *traduz a comutatividade de* (17).

Nestes termos, a Proposição 6.17 traduz a comutatividade do retângulo exterior no seguinte diagrama

$$\begin{tikzcd} V \arrow[r, "f"] \arrow[d, "{[\cdot]_{B_1}}"', "\cong"] & W \arrow[r, "g"] \arrow[d, "{[\cdot]_{B_2}}"', "\cong"] & U \arrow[d, "{[\cdot]_{B_3}}"', "\cong"] \\ M_{n\times 1}(\mathbb{R}) \arrow[r, "A_{f,B_1,B_2}"] & M_{m\times 1}(\mathbb{R}) \arrow[r, "A_{g,B_2,B_3}"] & M_{p\times 1}(\mathbb{R}) \end{tikzcd}$$

que é claramente uma consequência da comutatividade dos dois quadrados.

Corolário 6.19. *Sejam V, W espaços vetoriais, $f: V \to W$ uma transformação linear invertível e B_1, B_2 bases para V e W respetivamente. Então $A_{f^{-1},B_2,B_1} = (A_{f,B_1,B_2})^{-1}$.*

Dem. Uma vez que $f \circ f^{-1} = \mathrm{Id}_W$ e $f^{-1} \circ f = \mathrm{Id}_V$, e que a matriz que representa a transformação linear identidade com respeito a uma mesma base num espaço vetorial é a matriz identidade, pela Proposição anterior temos

$$A_{f,B_1,B_2}A_{f^{-1},B_2,B_1} = I \qquad A_{f^{-1},B_2,B_1}A_{f,B_1,B_2} = I$$

(onde I designa a matriz identidade). □

7. Subespaços associados a uma transformação linear

Definição 7.1. *Seja $f: V \to W$ uma transformação linear. O* núcleo de f *é o conjunto*

$$N(f) = \{v \in V : f(v) = 0\}$$

e a imagem de f *é o conjunto*

$$f(V) = \{f(v) : v \in V\} \subset W$$

Proposição 7.2. *Seja $f: V \to W$ uma transformação linear. Então $N(f)$ é um subespaço vetorial de V e $f(V)$ é um subespaço vetorial de W.*

Dem. Uma vez que $f(0) = 0$ temos que $0 \in N(f)$ e $0 \in f(V)$ pelo que estes conjuntos são não vazios. Vejamos que $N(f)$ é um subespaço vetorial:

- Sendo $v_1, v_2 \in N(f)$ temos $f(v_1+v_2) = f(v_1)+f(v_2) = 0+0 = 0$ logo $v_1+v_2 \in N(f)$.
- Sendo α um escalar e $v \in N(f)$ temos $f(\alpha v) = \alpha f(v) = \alpha 0 = 0$ logo $\alpha v \in N(f)$.

Quanto a $f(V)$:

- Dados $w_1, w_2 \in f(V)$, existem $v_1, v_2 \in V$ tais que $f(v_1) = w_1$ e $f(v_2) = w_2$. Então $f(v_1 + v_2) = w_1 + w_2$ logo $w_1 + w_2 \in f(V)$.
- Dado um escalar α e $w = f(v) \in f(V)$ temos $\alpha w = f(\alpha v) \in f(V)$.

□

Por definição de sobrejetividade, uma transformação linear é sobrejetiva se e só se $f(V) = W$. A injetividade de f pode ser determinada em termos do núcleo como explica o seguinte resultado.

Proposição 7.3. *Uma transformação linear $f: V \to W$ é injetiva se e só se $N(f) = \{0\}$.*

Dem. Suponhamos que f é injetiva. Se $v \in N(f)$ então $f(v) = 0 = f(0)$. Uma vez que f é injetiva conclui-se que $v = 0$, logo $N(f) = \{0\}$.

Suponhamos agora que $N(f) = \{0\}$. Então se $f(v_1) = f(v_2)$ temos $f(v_1 - v_2) = 0$ e portanto $v_1 - v_2 \in N(f) = \{0\}$, ou seja, $v_1 = v_2$. □

A Proposição anterior pode ser vista como mais uma manifestação do "bom comportamento" das transformações lineares. A condição $N(f) = \{0\}$ é equivalente (uma vez que $f(0) = 0$) à proposição

$$f(x) = f(0) \Rightarrow x = 0$$

que é um caso particular da condição geral de injetividade

$$f(x) = f(y) \Rightarrow x = y.$$

A Proposição 7.3 diz que, quando uma função é linear, para verificar a condição de injetividade podemos assumir que um dos elementos do domínio é 0. Se for verdade nesse caso particular então é verdade em geral.

É natural perguntar a que correspondem o núcleo e a imagem de uma transformação linear em termos de coordenadas, ou seja através do "dicionário" descrito no diagrama (17). Quanto ao núcleo, temos

$$v \in N(f) \Leftrightarrow f(v) = 0 \Leftrightarrow [f(v)]_{B_2} = 0$$

uma vez que um vetor é nulo se e só se as suas coordenadas numa base são todas nulas. Por (18) isto acontece se e só se

$$A_{f,B_1,B_2}[v]_{B_1} = 0$$

ou seja, se o vetor de $\mathbb{R}^n$ formado pelas coordenadas de v pertence ao núcleo da matriz A_{f,B_1,B_2} que representa a transformação linear f. Assim, não muito surpreendentemente, em coordenadas, o núcleo de uma transformação linear corresponde ao núcleo da matriz que representa a transformação linear.

Quanto à imagem de f, a sua tradução em coordenadas é o conjunto

$$\{[f(v)]_{B_2} \colon v \in V\} \subset M_{m\times 1}(\mathbb{R})$$

Novamente por (18) temos que este conjunto é igual a

$$\{A_{f,B_1,B_2}[v]_{B_1} \colon v \in V\}$$

Mas sendo v um vector arbitrário de V, a sua matriz coluna de coordenadas é uma matriz arbitrária em $M_{n\times 1}(\mathbb{R})$ e portanto este conjunto não é mais do que o espaço das colunas da matriz A_{f,B_1,B_2}. Ou seja, em coordenadas, *a imagem de uma transformação linear f é o espaço das colunas da matriz que representa f.*

Chegamos agora a um dos resultados básicos da Álgebra Linear, cuja importância se irá tornando clara com o desenrolar do semestre.

Teorema 7.4. *Seja V um espaço vetorial finitamente gerado, W um espaço vetorial e $f \colon V \to W$ uma transformação linear. Então*

$$\dim N(f) + \dim f(V) = \dim V$$

Dem. Seja $\{v_1, \ldots, v_k\}$ uma base para o subespaço $N(f) \subset V$ (que é finitamente gerado porque V é). Pela Proposição 4.4(iii) podemos completar este conjunto com um número finito de vetores distintos $\{v_{k+1} \ldots, v_n\}$ de tal forma que $\{v_1, \ldots, v_n\}$ seja uma base para V. Vamos verificar que $\{f(v_{k+1}), \ldots, f(v_n)\}$ é uma base de $f(V)$. Teremos então

$$\dim N(f) = k, \quad \dim f(V) = n - k, \quad \dim V = n$$

o que verifica a afirmação do enunciado.

- $\{f(v_{k+1}), \ldots, f(v_n)\}$ gera V: Seja w um vetor em $f(V)$. Então existe $v \in V$ tal que $f(v) = w$. Uma vez que $\{v_1, \ldots, v_n\}$ é uma base, existem escalares $\alpha_1, \ldots, \alpha_n$ tais que $v = \alpha_1 v_1 + \ldots + \alpha_n v_n$. Então

$$\begin{aligned} f(\alpha_1 v_1 + \ldots + \alpha_k v_k + \alpha_{k+1} v_{k+1} + \ldots + \alpha_n v_n) &= f(\alpha_1 v_1 + \ldots + \alpha_k v_k) + \alpha_{k+1} f(v_k) + \ldots + \alpha_n f(v_n) \\ &= 0 + \alpha_{k+1} f(v_k) + \ldots + \alpha_n f(v_n) \end{aligned}$$

  onde na segunda igualdade usámos o facto de o vetor $\alpha_1 v_1 + \ldots + \alpha_k v_k$ pertencer ao núcleo de f. A expressão acima mostra que w é uma combinação linear de $f(v_{k+1}), \ldots, f(v_n)$ pelo que estes vetores geram $f(V)$.
- $\{f(v_{k+1}), \ldots, f(v_n)\}$ é linearmente independente: Suponhamos que $\beta_1, \ldots, \beta_{n-k}$ são escalares tais que

$$\beta_1 f(v_{k+1}) + \ldots + \beta_{n-k} f(v_n) = 0$$

  Então $f(\beta_1 v_{k+1} + \ldots + \beta_{n-k} v_n) = 0$, logo $\beta_1 v_{k+1} + \ldots + \beta_{n-k} v_n \in N(f)$. Portanto existem escalares $\alpha_1, \ldots, \alpha_k$ tais que $\alpha_1 v_1 + \ldots + \alpha_k v_k = \beta_1 v_{k+1} + \ldots + \beta_{n-k} v_n$ ou seja tais que

$$\alpha_1 v_1 + \ldots + \alpha_k v_k - \beta_1 v_{k+1} - \ldots - \beta_{n-k} v_n = 0$$

  Uma vez que $\{v_1, \ldots, v_n\}$ é uma base de V tal só pode acontecer se $\alpha_1 = \ldots = \alpha_k = -\beta_1 = \ldots = -\beta_{n-k} = 0$. Conclui-se que $\beta_1 = \cdots = \beta_{n-k} = 0$ e portanto que $\{f(v_{k+1}), \ldots, f(v_n)\}$ é linearmente independente.

$\square$

Definição 7.5. *Sendo V um espaço finitamente gerado, W um espaço vetorial e $f \colon V \to W$ uma transformação linear, o número* $\dim f(V)$ *chama-se a* característica *da transformação linear f (rank em inglês) e o número* $\dim N(f)$ *chama-se a* nulidade *de f (nullity em inglês).*

O Teorema 7.4 é conhecido em inglês por “the rank-nullity Theorem”. Tem o seguinte corolário extremamente útil:

Corolário 7.6. *Sejam V e W espaços vetoriais finitamente gerados* com a mesma dimensão *e seja $f \colon V \to W$ uma transformação linear. Então as seguintes afirmações são equivalentes*

(i) f é invertível (isto é, f é bijetiva).
(ii) f é injetiva (equivalentemente, $N(f) = \{0\}$).
(iii) f é sobrejetiva (isto é, $f(V) = W$).

Dem. É claro que a afirmação (i) implica as afirmações (ii) e (iii), e, por definição (ii) juntamente com (iii) implicam (i). Para demonstrar a equivalência das afirmações basta assim ver que quando (ii) se verifica, (iii) também se verifica e vice-versa.

Suponhamos que f é injetiva. Então $\dim N(f) = 0$ e portanto pelo Teorema 7.4 e a hipótese sobre a dimensão dos espaços V e W temos

$$\dim f(V) = \dim V = \dim W$$

Ou seja $f(V)$ é um subespaço de W com a mesma dimensão que W. Então temos necessariamente $f(V) = W$ (por exemplo, pelo Corolário 4.10(i)) e portanto f é também sobrejetiva.

Suponhamos agora que f é sobrejetiva, ou seja que $\dim f(V) = \dim W$. Aplicando o Teorema 7.4 e a hipótese $\dim V = \dim W$ temos

$$\dim f(V) + \dim N(f) = \dim V \Leftrightarrow \dim V + \dim N(f) = \dim V \Leftrightarrow \dim N(f) = 0$$

logo $N(f) = \{0\}$ e portanto, pela Proposição 7.3, f é injetiva. □

Em vista da interpretação da imagem de uma transformação linear f como o espaço das colunas da matriz que a representa, o Teorema 7.4 tem a seguinte consequência importante (que está longe de ser óbvia!).[4]

Proposição 7.7. *Seja A uma matriz $m \times n$. Então o espaço das linhas e o espaço das colunas de A têm a mesma dimensão (que é a característica de A). Isto é,*

$$\dim EC(A) = \dim EL(A) = \textit{característica de } A$$

Proof. A dimensão do espaço das linhas é o número de pivots da matriz A após aplicação do método de Gauss, enquanto que a dimensão do núcleo de A é o número de variáveis livres no sistema homogéneo associado a A, ou seja, o número de colunas de A sem pivot. Isto significa que

$$\dim EL(A) = n - \dim N(A)$$

Por outro lado, no caso da transformação linear $f\colon M_{n\times 1}(\mathbb{R}) \to M_{m\times 1}(\mathbb{R})$ definida por $f(x) = Ax$, o Teorema 7.4 diz que

$$\dim EC(A) + \dim N(A) = n \Leftrightarrow \dim EC(A) = n - \dim N(A)$$

Conclui-se portanto que $\dim EC(A) = \dim EL(A)$ e este número é a característica de A. □

A Proposição anterior justifica também a atribuição do nome "característica" de f à dimensão de $f(V)$. Podemos agora aproveitar para atualizar os nossos critérios para a invertibilidade de uma matriz (comparem com o Teorema 2.23)

Proposição 7.8. *Seja A uma matriz $n \times n$. Então as seguintes afirmações são equivalentes*

(i) *A é invertível.*
(ii) *A característica de A é n (equivalentemente $\dim EL(A) = n$).*
(iii) *Para cada matriz $b \in M_{n\times 1}(\mathbb{R})$ a equação $Ax = b$ tem solução única (equivalentemente, a função $x \mapsto Ax$ é bijetiva).*
(iv) $N(A) = 0$
(v) $EC(A) = \mathbb{R}^n$
(vi) *Existe $B \in M_{n\times n}(\mathbb{R})$ tal que $AB = I_n$*
(vii) *Existe $B \in M_{n\times n}(\mathbb{R})$ tal que $BA = I_n$*

[4]Para uma explicação conceptual desta igualdade que é independente da nossa discussão inicial dos sistemas lineares e do método de Gauss ver o último exercício da Ficha 9.

Dem. A equivalência das primeiras três afirmações foi já vista no Teorema 2.23 embora a equivalência de (i) com (iii) possa agora ser interpretada conceptualmente como uma consequência da Proposição 6.13 e Corolário 6.19. A equivalência de (iii), (iv) e (v) é uma consequência do Corolário 7.6 e da interpretação do núcleo e espaço das colunas da matriz como núcleo e imagem da transformação linear associada.

É claro da definição de invertibilidade que $(i) \Rightarrow (vi)$ e (vii). Reciprocamente se existe B tal que $AB = I_n$ então o espaço das colunas de A contém as colunas da matriz identidade, e portanto $EC(A) = \mathbb{R}^n$, que é a condição (v). Por outro lado se existe B tal que $BA = I_n$ então dado $x \in N(A)$ temos $x = I_n x = BAx = B0 = 0$ pelo que $N(A) = \{0\}$ que é a condição (iv). Vemos assim que (vi) e (vii) são também equivalentes às restantes condições. □

Exemplo 7.9 (Significado geométrico dos pivots)**.** *Suponhamos que A é uma matriz $m \times n$. Associada a esta matriz está um plano em $\mathbb{R}^n$ contendo a origem - o espaço das linhas $EL(A)$ - que tem uma certa dimensão $k \leq n$ a que se chama a característica da matriz A. Quando aplicamos o método de eliminação de Gauss à matriz, o plano determinado pelas linhas das sucessivas matrizes permanece sempre o mesmo e no final do método obtemos uma matriz em escada de linhas cujas linhas formam uma base para $EL(A)$.*

Apesar de a matriz obtida no final do método de Gauss não ser única (há alguma indeterminação na aplicação do método), há certas características comuns a todas estas matrizes como por exemplo a localização dos pivots (isto é parte do conteúdo do Teorema 1.12). Vamos ver qual é a explicação geométrica para este facto começando por alguns exemplos simples.

Consideremos primeiro o caso em que $m = 1$ e $n = 3$. Temos portanto que

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

e que $EL(A)$ é uma linha em $\mathbb{R}^3$ (vamos ignorar o caso trivial em que $a = b = c = 0$). Neste caso a matriz A já está em escada de linhas e há três possibilidades: o pivot pode estar na primeira, segunda ou terceira entrada da matriz.

1º caso: *Se o pivot está na primeira posição isto significa que a projeção da linha no eixo dos xx produz todo o eixo dos xx.*

2º caso: *Se o pivot está na segunda posição isto significa que a projeção da linha no eixo dos xx é nula, ou seja que a linha está contida no plano yz, e que a projeção desta linha no eixo dos yy não é nula (i.e. a linha não coincide com o eixo dos zz).*

3º caso: *Se o pivot está na terceira posição então a linha é o eixo dos zz.*

Em seguida consideremos o caso em que $m = 2$ e $n = 3$. Então

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

Se a característica de A for menor ou igual a 1 o espaço $EL(A)$ é uma linha ou um ponto e já vimos o que acontece. Suponhamos portanto que a característica de A é 2. Há novamente três casos a considerar para a matriz em escada de linhas resultante do método de Gauss: os dois pivots estão ou nas entradas 11 e 22 da matriz, ou 11 e 23 ou 21 e 32.

1º caso: *Os pivots estão nas entradas* 11 *e* 22*: Isto significa que a projecão do plano* $EL(A)$ *no eixo dos* xx *é todo o eixo dos* xx *e que a projeção de* $EL(A)$ *no plano* xy *é todo o plano* xy *(ou seja que o plano é o gráfico de uma função sobre o plano* xy*).*

2º caso: *Os pivots estão nas entradas* 11 *e* 32*. Isto significa que a projeção de* $EL(A)$ *no eixo* xx *é todo o eixo* xx *mas a projeção no plano* xy *é apenas uma linha (gerada pela projeção em* xy *do vetor que aparece na primeira linha da matriz, projeção esta que não pertence ao eixo dos* yy*). Assim,* $EL(A)$ *é um plano vertical contendo o eixo dos* zz*, que não é o plano* yz*.*

3º caso: *Os pivots estão nas entradas* 21 *e* 32*. Neste caso, a projeção do plano* $EL(A)$ *no eixo dos* xx *é apenas a origem, o que significa que* $EL(A)$ *é o plano* yz*.*

Em geral, a posição dos pivots dá-nos informação sobre a posição do plano $EL(A)$ *relativamente aos planos coordenados. Sendo* $j \leq n$ *e*

$$\pi_j \colon \mathbb{R}^n \to \mathbb{R}^j$$

a projeção no plano $(x_1, \ldots, x_j)$ *definida por* $\pi_j(x_1, \ldots, x_n) = (x_1, \ldots, x_j)$*, podemos associar a* A *uma função crescente*

$$\{1, \ldots, n\} \xrightarrow{d} \{0, 1, \ldots, k\}$$

(onde k *é a caraterística de* A*) definida por*

$$d(j) = \dim \pi_j(EL(A))$$

Ou seja, $d(j)$ *é a dimensão da projeção de* $EL(A)$ *no plano* $(x_1, \ldots, x_j)$ *ou, alternativamente, a dimensão do espaço das linhas da matriz* $m \times j$ *que se obtém esquecendo as últimas* $n - j$ *colunas de* A*. Temos assim* $d(1) = 0$ *ou* 1 *e* $d(n) = k$*.*

Os pivots ocorrem exatamente nas colunas em que a função $d(j)$ *"salta", isto é, na primeira coluna ou nas colunas* $j > 1$ *para as quais* $d(j) > d(j-1)$*.*

Finalmente, note-se que a projeção do plano $EL(A)$ *no plano* $\mathbb{R}^k$ *determinado pelas variáveis correspondentes às colunas em que ocorrem pivots é um isomorfismo de espaços vetoriais. Isto significa que* $EL(A)$ *é o gráfico de uma função sobre esse plano. Fica como exercício para as alunas interessadas a verificação que a transformação linear de* $\mathbb{R}^k$ *para* $\mathbb{R}^{n-k}$ *cujo gráfico é* $EL(A)$ *é determinada (nas bases canónicas) pela matriz* $m \times (n-k)$ *que se obtém da matriz em escada de linhas* reduzida *obtida de* A *no final do método do Gauss-Jordan considerando apenas as colunas sem pivot.*

Vejamos agora alguns exemplos práticos de aplicação dos últimos resultados.

Exemplo 7.10 (Exercício 7 da Ficha 8)**.** *Vamos achar uma base para o núcleo e a imagem da transformação linear*

$$f \colon M_{2\times 2}(\mathbb{R}) \to M_{2\times 2}(\mathbb{R})$$

definida por

$$f(A) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} A - A \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

Escrevendo

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

obtemos

$$f(A) = \begin{bmatrix} a+c & b+d \\ c & d \end{bmatrix} - \begin{bmatrix} a & a+b \\ c & c+d \end{bmatrix} = \begin{bmatrix} c & d-a \\ 0 & -c \end{bmatrix}$$

Conclui-se que $N(f)$ é definido pelas equações $c = 0, d - a = 0$ ou seja que

$$N(f) = \left\{ \begin{bmatrix} a & b \\ 0 & a \end{bmatrix} : a, b \in \mathbb{R} \right\}$$

Uma vez que

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

vemos que uma base para $N(f)$ é

$$\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \right\}$$

O Teorema 7.4 diz-nos já que a imagem de f terá dimensão

$$\dim M_{2\times 2}(\mathbb{R}) - \dim N(f) = 4 - 2 = 2$$

pelo que para achar uma base para a imagem só precisamos de dois vetores não colineares em $f(M_{2\times 2}(\mathbb{R}))$.

Neste exemplo é também fácil achar a base diretamente, sem recorrer a este atalho. A expressão acima para $f(A)$ diz-nos que a imagem de f é formada pelas matrizes da forma

$$\begin{bmatrix} \alpha & \beta \\ 0 & -\alpha \end{bmatrix} = \alpha \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} + \beta \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

com α, β in $\mathbb{R}$. Assim uma base para a imagem de f é

$$\left\{ \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \right\}$$

Exemplo 7.11 (Exercício 8 da Ficha 8)**.** *Seja V um espaço vetorial de dimensão finita com base ordenada B_1 e $f\colon V \to V$ uma transformação linear representada na base B_1 pela matriz A. Sendo $S = S_{B_1 \to B_2}$ a matriz de mudança de coordenadas da base B_1 para outra base B_2 vejamos qual é a fórmula para a matriz que representa f na base B_2 em termos das matrizes A e S.*

Queremos a matriz $B = A_{f,B_2,B_2}$ tal que para todo o vetor $v \in V$ se tem

$$[f(v)]_{B_2} = B[v]_{B_2}$$

e sabemos que

$$[f(v)]_{B_1} = A[v]_{B_1} \quad e \quad [v]_{B_2} = S[v]_{B_1} \text{ para quaisquer } v$$

Então

$$[f(v)]_{B_2} = S[f(v)]_{B_1} = SA[v]_{B_1} = SAS^{-1}[v]_{B_2}$$

A unicidade da matriz B diz-nos então que

$$B = SAS^{-1} \tag{19}$$

Vamos aproveitar a fórmula acima para calcular a matriz que representa a transformação linear $f\colon \mathbb{R}^3 \to \mathbb{R}^3$ definida por

$$f(x, y, z) = (x - 2y, x + z, y + z)$$

relativamente à base ordenada $B = ((1, 0, 1), (0, 1, 1), (-1, 0, 1))$ de $\mathbb{R}^3$. Na base canónica temos

$$\begin{bmatrix} x-2y \\ x+z \\ y+z \end{bmatrix} = \begin{bmatrix} 1 & -2 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

A matriz de mudança de coordenadas que se pode obter imediatamente é $S_{B\to B_{can}}$ uma vez que esta tem como colunas as coordenadas dos vetores de B na base canónica:

$$S_{B\to B_{can}} = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

Conclui-se assim que a matriz que representa f com respeito à base B é

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

Alternativamente poderíamos ter calculado esta matriz diretamente achando as coordenadas da imagem por f dos elementos da base B na própria base B (são estas as colunas da matriz que acabámos de obter).

Finalmente vejamos se é possível que as matrizes $\begin{bmatrix} 2 & -1 \\ -1 & 0 \end{bmatrix}$ e $\begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$ representem a mesma transformação linear $f\colon V \to W$ (com respeito a algumas bases de V e W). Note-se que simplesmente pelo facto de se tratar de matrizes 2×2 sabemos que V e W têm dimensão 2 (os vetores de V e W têm duas coordenadas).

Se duas matrizes representam a transformação linear elas são "a mesma coisa com nomes diferentes" (formalmente uma obtém-se da outra através da fórmula (19) - diz-se que são matrizes semelhantes*). Ora a primeira matriz tem caraterística 2 enquanto que a segunda tem caraterística 1. Isto significa que qualquer transformação linear representada pela primeira matriz é um isomorfismo, enquanto que qualquer transformação representada pela segunda não é injetiva nem sobrejetiva (tanto o núcleo como a imagem têm dimensão 1). As matrizes não podem portanto representar a mesma transformação linear.*

Exemplo 7.12 (Exercício 5(a) da Ficha 8)**.** *Seja $f\colon \mathbb{R}^7 \to \mathbb{R}^4$ uma transformação linear. Quais são os possíveis valores para $(\dim N(f), \dim f(\mathbb{R}^7))$? Como $f(\mathbb{R}^7)$ é um subespaço de $\mathbb{R}^4$, a sua dimensão está entre 0 e 4. Uma vez que*

$$\dim N(f) + \dim f(\mathbb{R}^7) = 7$$

vemos que $N(f)$ tem dimensão entre 3 e 7. Assim que

$$(\dim N(f), \dim f(\mathbb{R}^7)) \in \{(7,0),(6,1),(5,2),(4,3),(3,4)\}$$

Todos estes valores são possíveis. O primeiro par é o valor correspondente à transformação linear nula enquanto que os outros correspondem, por exemplo, às projeções $\pi_j(x_1,\ldots,x_7) = (x_1,\ldots x_j)$ com $j = 1,\ldots,4$.

Exemplo 7.13 (Exercício 12(b) da Ficha 8). *O* dual *de um espaço vetorial real é o espaço vetorial $V^* = L(V,\mathbb{R})$ das transformações lineares de V para $\mathbb{R}$ que se chamam normalmente* funcionais lineares. *Já vimos muitos exemplos de elementos do dual de um espaço. Por exemplo, se V é o espaço vetorial dos polinómios reais as funções ϕ e ψ definidas por*

$$\phi(p) = p(0), \quad \psi(p) = p'(1)$$

são elementos de V^*. Muito em breve definirão em Cálculo 1 o integral de uma função contínua e então sendo V o espaço vetorial das funções contínuas $[0,1] \to \mathbb{R}$ temos que

$$f \mapsto \int_0^1 f(x)dx$$

é um elemento do dual de V.

Outros exemplos importantes são as funções coordenadas. *Se $B = (v_1,\ldots,v_n)$ é uma base para V então a função*

$$\varphi_i\colon V \to \mathbb{R}$$

definida por

$$\varphi_i(\alpha_1 v_1 + \ldots + \alpha_n v_n) = \alpha_i$$

é, como é imediato verificar, um elemento de V^*. Note-se que φ_i não é mais do que a função que calcula a i-ésima coordenada de V na base B, isto é a i-ésima função coordenada. Vamos verificar que $B^* = (\varphi_1,\ldots,\varphi_n)$ são uma base de V^* que se chama a base dual de B.

Temos a verificar que os elementos de B^* são linearmente independentes e geram V^*. Sejam $\alpha_1,\ldots,\alpha_n$ escalares tais que

$$\alpha_1\varphi_1 + \ldots + \alpha_n\varphi_n = 0$$

Avaliando o termo esquerdo da expressão anterior no i-ésimo elemento v_i da base B, e dado que

$$\varphi_i(v_j) = \begin{cases} 1 & \text{se } i = j \\ 0 & \text{se } i \neq j \end{cases}$$

obtemos

$$\alpha_1\varphi_1(v_i) + \ldots + \alpha_n\varphi_n(v_i) = 0 + \ldots + 0 + \alpha_i + 0 + \ldots + 0 = \alpha_i$$

pelo que $\alpha_i = 0$. Como isto é verdade para todo o i conclui-se que $\alpha_1 = \cdots = \alpha_n = 0$, ou seja que os elementos de B^* são linearmente independentes.

Par ver que B^* gera V^*, seja $\varphi \in V^*$ um elemento qualquer. Queremos resolver a equação

$$\alpha_1\varphi_1 + \ldots + \alpha_n\varphi_n = \varphi$$

Avaliando o termo esquerdo da equação anterior no elemento v_i da base obtemos

$$0 + \ldots + 0 + \alpha_i \varphi_i(v_i) + 0 + \ldots + 0 = \varphi(v_i)$$

logo α_i é necessariamente igual a $\varphi(v_i)$. Para esta escolha de coeficientes, os elementos $\alpha_1\varphi_1 + \ldots + \alpha_n\varphi_n$ e φ de V^* tomam os mesmos valores na base B e portanto coincidem. Conclui-se assim que B^* é um conjunto gerador e portanto uma base.

Note-se a expressão engraçada para exprimir um funcional φ qualquer em função desta base:

$$\varphi = \varphi(v_1)\varphi_1 + \ldots + \varphi(v_n)\varphi_n$$

8. Equações lineares

Definição 8.1. *Uma* equação linear *é uma equação da forma*

$$f(x) = w$$

onde $f\colon V \to W$ é uma transformação linear, w é um vetor de W e a incógnita x é um vetor de V a determinar. A equação diz-se homogénea *quando $w = 0$.*

É claro que uma equação linear tem solução se e só e $w \in f(V)$. O conjunto das soluções é controlado pelo núcleo de f no seguinte sentido.

Proposição 8.2 (Princípio da sobreposição)**.** *Seja $f\colon V \to W$ uma transformação linear. Se v é uma solução da equação linear $f(v) = w$, o conjunto de todas as soluções é*

$$v + N(f) = \{v + z \colon z \in N(f)\} \subset V$$

Dem. Se v é uma solução e $z \in N(f)$ temos que $f(v+z) = f(v) + f(z) = w + 0 = w$ logo $v + z$ é uma solução. Assim

$$v + N(f) \subset \{u \in V \colon f(u) = w\}$$

Reciprocamente, seja u uma solução qualquer da equação. Então $u = v + (u - v)$ e $f(u - v) = f(u) - f(v) = w - w = 0$ pelo que $u - v \in N(f)$ e portanto $u \in v + N(f)$. Conclui-se que

$$\{u \in V \colon f(u) = w\} \subset v + N(f)$$

o que termina a demonstração. □

Geometricamente, o resultado anterior diz que o conjunto das soluções é o "plano" paralelo a $N(f)$ (que é um "plano" em V contendo a origem) que passa por uma solução particular qualquer da equação.

É costume enunciar o resultado da Proposição 8.2 da seguinte forma;

> A solução geral de uma equação linear é dada por uma solução particular da equação mais a solução geral da equação homogénea.

Por uma solução particular entende-se uma qualquer solução v fixada para a equação. Por solução geral entende-se o conjunto das soluções. Assim a afirmação acima diz apenas que o conjunto das soluções de uma equação linear é obtido somando todas as soluções da equação homogénea a uma qualquer solução da equação que consigamos determinar.

Exemplo 8.3 (O oscilador harmónico). *Seja $x\colon \mathbb{R} \to \mathbb{R}$ uma função que descreve a posição de uma partícula presa a uma mola em função do tempo. A partícula é atuada unicamente pela força exercida pela extensão ou contração da mola, que é proporcional ao deslocamento da mola em relação à sua posição de repouso. Assumindo que 0 é a coordenada da posição de repouso, a equação de Newton diz-nos que*

$$x''(t) + kx(t) = 0 \tag{20}$$

onde k é uma constante positiva determinada pelas características físicas da mola e a massa da partícula (recorde que x'' é a aceleração e note que a força exercida pela mola, mx'' tem o sentido contrário ao deslocamento x). Para simplificar as contas vamos assumir a partir de agora que $k = 1$.

Sendo $V \subset F(\mathbb{R}, \mathbb{R})$ o subespaço vetorial formado pelas funções duas vezes diferenciáveis e T a transformação linear

$$T\colon V \to F(\mathbb{R}, \mathbb{R})$$

definida pela expressão

$$T(x) = x'' + x$$

vemos que o núcleo de T é exatamente o conjunto das soluções de (20) *(com $k = 1$) que formam portanto um subespaço vetorial de V.*

É fácil adivinhar duas soluções para a equação

$$x'' + x = 0 \tag{21}$$

pois claramente $x(t) = \cos t$ e $x(t) = \operatorname{sen} t$ são soluções. Como o conjunto das soluções é um espaço vetorial temos mais geralmente que

$$x(t) = \alpha_1 \cos t + \alpha_2 \operatorname{sen} t, \qquad \text{com } \alpha_1, \alpha_2 \in \mathbb{R} \tag{22}$$

são soluções.

Para o ano que vem irão aprender que uma solução de uma equação diferencial como (20) *é completamente determinada por $x(0)$ e $x'(0)$ (fisicamente isto diz que a evolução da posição da partícula é completamente determinada pela sua posição e velocidade iniciais). Assim o conjunto das soluções é um espaço vetorial de dimensão 2 (um vetor é determinado por dois números reais) e portanto a fórmula* (22) *descreve a solução geral da equação* (21)*.*

No caso da equação (21) *podemos verificar a afirmação anterior diretamente recorrendo à* conservação da energia. *Definindo a quantidade*

$$E(t) = (x')^2 + x^2$$

(correspondendo à soma das energia cinética e potencial) temos

$$\frac{dE}{dt} = 2x'x'' + 2xx' = 2x'(-x) + 2xx' = 0$$

logo a quantidade $(x')^2 + x^2$ é conservada ao longo do tempo para qualquer solução da equação diferencial (21)*. Em particular se $x(t)$ for uma solução com $x(0) = x'(0) = 0$ teremos $(x'(t))^2 + x(t)^2 = 0$ para todo o t e portanto $x(t) = 0$.*

Isto permite-nos concluir que os valores de $x(0)$ e $x'(0)$ determinam completamente a solução $x(t)$ para todo o t: se $x(t)$ e $y(t)$ forem soluções de (21) *com $x(0) = y(0)$ e*

$x'(0) = y'(0)$ então $u(t) = x(t) - y(t)$ é também uma solução de (21) (porque se trata de uma equação linear!) que satisfaz $u(0) = u'(0) = 0$. Mas então $u(t) = 0$ e portanto $x(t) = y(t)$.

É agora imediato verificar que as soluções (22) permitem atribuir valores arbitrários a $x(0)$ e $x'(0)$ mediante variação dos coeficientes α_1 e α_2 (na realidade $\alpha_1 = x(0)$ e $\alpha_2 = x'(0)$) e portanto descrevem todas as soluções de (21).

Suponhamos agora que queremos resolver a equação[5]

$$x'' + x = t^3 \tag{23}$$

Trata-se agora de uma equação linear não homogénea. Não é no entanto difícil descobrir uma solução particular desta equação tentando encontrar um polinómio que a satisfaça. Se o fizer irá ver que o único polinómio que satisfaz esta equação é

$$x(t) = t^3 - 6t$$

A Proposição 8.2 diz-nos então que a solução geral da equação (23) é

$$x(t) = t^3 - 6t + \alpha_1 \cos t + \alpha_2 \operatorname{sen} t, \qquad \text{com } \alpha_1, \alpha_2 \in \mathbb{R}.$$

9. Valores e vetores próprios

O nosso objetivo para uma boa parte do que resta do semestre é estudar com mais detalhe transformações lineares de um espaço vetorial nele próprio, isto é transformações lineares $f\colon V \to V$ onde V é um espaço vetorial. Neste caso há certas questões que podemos colocar que não fariam sentido para uma transformação linear arbitrária.

Por exemplo podemos perguntar se há alguma direção que seja preservada pela transformação f. Se isso acontecer teremos pelo menos uma descrição muito simples do efeito de f ao longo dessas direções. É isto o que acontece por exemplo com as rotações do espaço que preservam o eixo da rotação e das reflexões num plano que preservam tanto as direções do plano de reflexão como a direção ortogonal ao plano (na qual o sentido é trocado).

Definição 9.1. *Seja $f\colon V \to V$ uma transformação linear. Um vetor $v \in V \setminus \{0\}$* não nulo *diz-se um* vetor próprio *de f se existe um escalar λ tal que $f(v) = \lambda v$. Nesse caso λ diz-se um* valor próprio *de f e diz-se que v é um* vetor próprio associado *ao valor próprio λ.*

Exemplo 9.2. *Considere-se a transformação linear $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ definida por*

$$f(x,y) = (x + 2y, 2x + y)$$

ou seja, em notação matricial,

$$\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

[5]Fisicamente esta equação corresponde a adicionar ao sistema mecânico considerado anteriormente uma força exterior dependente do tempo que actua com intensidade t^3/m (onde m é a massa da partícula).

Claramente

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

e

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

Pelo que $v_1 = (1,1)$ *e* $v_2 = (1,-1)$ *são vetores próprios de* f *com valores próprios* 3 *e* -1 *respetivamente. Note-se que quaisquer múltiplos não nulos destes vetores são ainda vetores próprios (com os mesmos valores próprios).*

Os vetores v_1 *e* v_2 *formam uma base de* $\mathbb{R}^2$ *em termos da qual é extremamente simples compreender o efeito que a transformação linear* f *tem sobre os vetores de* $\mathbb{R}^2$*: Ao longo da direção de* v_1 *(a diagonal do primeiro quadrante)* f *expande por um fator de* 3*, enquanto que na direção ortogonal, (a diagonal do quarto quadrante),* f *reflete. Com base nisto é fácil descrever o efeito que* f *teria num desenho qualquer no plano.*

Note-se ainda que, uma vez que $f(v_1) = 3v_1$ *e* $f(v_2) = -v_2$ *temos que a representação de* f *com respeito à base* $B = (v_1, v_2)$ *é*

$$A_{f,B,B} = \begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix}$$

Vimos no exemplo anterior como é fácil descrever uma transformação linear $f\colon V \to V$ quando existe uma base para V formada por vetores próprios. Isto sugere a seguinte definição.

Definição 9.3. *Uma transformação linear* $f\colon V \to V$ *diz-se* diagonalizável *se existe uma base para* V *constituída por vetores próprios de* f*. Uma matriz* $A \in M_{n\times n}(\mathbb{R})$ *diz-se diagonalizável, se a transformação linear de* $\mathbb{R}^n$ *representada por* A *(com respeito à base canónica) é diagonalizável.*

A razão da palavra diagonalizável é, claro, que a representação de uma transformação linear diagonalizável numa base $B = (v_1, \dots, v_n)$ de vetores próprios é uma matriz diagonal

$$\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix}$$

onde λ_i é o valor próprio associado a v_i.

Pôe-se agora a questão de como encontrar os valores próprios e vetores próprios de uma transformação linear. Note-se que não existem sempre, por exemplo uma rotação do plano (por um ângulo que não seja 0 ou π) não fixa nenhuma direção.

Se A for uma matriz $n \times n$ que representa f numa dada base temos que tentar achar um vetor $v \neq 0$ tal que

$$Av = \lambda v \Leftrightarrow Av = \lambda I_n v \Leftrightarrow (A - \lambda I_n)v = 0$$

Assim, *um valor próprio de* f *é um escalar* λ *para o qual a matriz* $A - \lambda I_n$ *tenha núcleo não trivial, ou equivalentemente, tal que* $A - \lambda I_n$ *não seja invertível.* Uma vez achados os valores próprios, os vetores próprios são fáceis de determinar - são elementos não triviais do núcleo.

Embora seja possível achar os valores próprios usando o método de Gauss (tratando λ como um parâmetro), é útil ter um critério geral para que λ seja um valor próprio. É esse o nosso próximo objetivo. Iremos ver que há um polinómio nas entradas de uma matriz quadrada - o determinante da matriz - que se anula precisamente quando a matriz não é invertível. Quando calculamos esse polinómio para a matriz $(A - \lambda I_n)$ e igualamos a 0 obtemos uma equação para achar os valores próprios λ.

Esse polinómio pode não ter raízes reais, mas o Teorema Fundamental da Álgebra garante-nos que existem sempre raízes complexas. Assim, desde que estejamos dispostos a encarar as matrizes como matrizes complexas, existem sempre vetores próprios. Além disso, uma matriz genérica (isto é escolhida "ao calhas") é diagonalizável, até com valores próprios todos distintos.

10. Determinantes

Queremos procurar um critério para que uma matriz $n \times n$ tenha núcleo não trivial ou, equivalentemente, para que tenha característica menor que n. Geometricamente, esta condição traduz-se no espaço das linhas da matriz ter dimensão menor ou igual a $(n - 1)$.

Pensando primeiro nos casos mais familiares em que $n = 2$ ou 3, esta condição geométrica pode ainda traduzir-se no seguinte. Quando $n = 2$, podemos considerar o paralelogramo

$$P(v_1, v_2) = \{\alpha v_1 + \beta v_2 \colon 0 \leq \alpha, \beta \leq 1\}$$

que tem um vértice na origem e duas das arestas dadas pelas linhas da matriz, que identificamos como vetores v_1 e v_2 de $\mathbb{R}^2$,

A matriz tem característica menor ou igual a 1 se e só se este paralelogramo degenera num segmento de reta ou num ponto, ou seja, se e só se *a área do paralelogramo* $P(v_1, v_2)$ *é nula.*

Analogamente, uma matriz 3×3 terá característica menor que 3 se e só se o paralelipípedo

$$P(v_1, v_2, v_3) = \{\alpha v_1 + \beta v_2 + \gamma v_3 \colon 0 \leq \alpha, \beta, \gamma \leq 1\}$$

(em que v_1, v_2, v_3 designam as linhas da matriz identificadas com vetores de $\mathbb{R}^3$) tiver volume nulo.

Mais geralmente pode definir-se uma noção de volume n-dimensional para um subconjunto de $\mathbb{R}^n$ como irão ver em Cálculo 2 e então a condição geral que procuramos é equivalente ao volume do paralelipípedo n-dimensional $P(v_1, \ldots, v_n)$ ter volume n-dimensional nulo.

O nosso objetivo é portanto encontrar uma fórmula para a área/volume de um paralelogramo/paralelipípedo a partir das coordenadas dos vetores que formam as arestas. A observação básica que nos permite obter esta fórmula é a seguinte (faça um desenho):

> Ao deslizar o ponto final da aresta de um paralelogramo ao longo de uma linha paralela à outra aresta, a área não se altera

ou seja

$$\text{área}(P(v_1, v_2)) = \text{área}(P(v_1 + \alpha v_2, v_2)) \tag{24}$$

(e claro que o mesmo se verifica se deslizarmos o ponto final de v_2 ao longo da direção v_1). Esta fórmula diz-nos por exemplo que as áreas dos paralelogramos correspondentes às matrizes

$$\begin{bmatrix} a & 0 \\ c & d \end{bmatrix} \qquad \text{e} \qquad \begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix}$$

são iguais, pois $(0, d)$ pode obter-se de (c, d) deslizando ao longo de $(a, 0)$ (a não ser que $a = 0$, mas nesse caso as áreas são nulas e a afirmação permanece verdadeira). Assim, a área do paralelogramo com arestas $(a, 0)$ e (c, d) é a área do retângulo com arestas $(a, 0)$ e $(0, d)$, ou seja $|ad|$ (mesmo que a ou d sejam 0). Mas a fórmula (24) diz-nos mais geralmente que quando aplicamos o método de Gauss a uma matriz 2×2, a área do paralelogramo associado não muda! Supondo que $a \neq 0$ temos

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \xrightarrow{L_2 - \frac{c}{a} L_1} \begin{bmatrix} a & b \\ 0 & d - \frac{bc}{a} \end{bmatrix}$$

logo concluímos que a área de um paralelogramo com arestas (a, b) e (c, d) é

$$\text{área } (P((a, b), (c, d))) = |a| \cdot \left| d - \frac{bc}{a} \right| = |ad - bc|$$

(a leitora interessada poderá verificar que esta fórmula permanece válida mesmo quando $a = 0$). E obtemos assim a condição desejada nas entradas da matriz:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \text{ tem núcleo não trivial sse } ad - bc = 0$$

ou, equivalentemente,

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \text{ é invertível sse } ad - bc \neq 0$$

Podemos fazer um raciocínio análogo para matrizes 3×3 mas a fórmula obtida será agora mais complicada. Novamente o volume de um paralelipípedo $P(v_1, v_2, v_3)$ em $\mathbb{R}^3$ não se alterará se deslizarmos o ponto final de uma das arestas paralelamente ao plano determinado pelas outras duas, ou seja, por exemplo

$$\text{volume } P(v_1 + \alpha v_2, v_2, v_3) = \text{volume } P(v_1, v_2, v_3)$$

Portanto o volume de um paralelipípedo com arestas as linhas da matriz

$$\begin{bmatrix} a & b & c \\ 0 & e & f \\ 0 & 0 & i \end{bmatrix}$$

será o volume do paralelipípedo reto com arestas de comprimento $|a|, |e|$ e $|i|$, e podemos reduzir a este caso usando eliminação de Gauss:

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \underset{L_3-\frac{g}{a}L_1}{\xrightarrow{L_2-\frac{d}{a}L_1}} \begin{bmatrix} a & b & c \\ 0 & e-\frac{db}{a} & f-\frac{dc}{a} \\ 0 & h-\frac{gb}{a} & i-\frac{gc}{a} \end{bmatrix} \xrightarrow{L_3-\frac{h-\frac{gb}{a}}{e-\frac{db}{a}}L_1} \begin{bmatrix} a & b & c \\ 0 & e-\frac{db}{a} & f-\frac{dc}{a} \\ 0 & 0 & i-\frac{gc}{a}-\frac{h-\frac{gb}{a}}{e-\frac{db}{a}}(f-\frac{dc}{a}) \end{bmatrix}$$

Obtemos assim a fórmula

$$\text{volume } (P((a,b,c),(d,e,f),(g,h,i))) = |a| \left| e - \frac{db}{a} \right| \left| i - \frac{gc}{a} - \frac{h-\frac{gb}{a}}{e-\frac{db}{a}}(f-\frac{dc}{a}) \right|$$

que, reduzindo ao mesmo denominador e cancelando alguns termos se transforma em:

$$\text{volume } (P((a,b,c),(d,e,f),(g,h,i))) = |aei + bfg + cdh - ceg - bdi - afh|$$

Fica como exercício verificar que esta fórmula é válida mesmo nos casos em que $a = 0$, ou $a \neq 0$ mas $e - \frac{db}{a} = 0$, nos quais a eliminação de Gauss feita acima tem de ser modificada.

O cálculo anterior sugere que não será prático obter e manipular diretamente uma expressão para o volume de um paralelipípedo n-dimensional. Com efeito, para $n = 4$ veremos que a fórmula análoga tem 24 termos, para $n = 5$, 120 termos, e em geral o número de termos é $n!$. Uma expressão de tal complexidade só pode ser manipulada conceptualmente.

Abstraindo as propriedades, não do volume, mas da expressão mais fundamental que obtivemos acima para $n = 2, 3$ cujo módulo é o volume, obtemos a seguinte definição.

Definição 10.1. *Uma função* determinante *para as matrizes* $n \times n$ *é uma função*

$$\det\colon M_{n\times n}(\mathbb{R}) \to \mathbb{R}$$

que se denota por

$$\det \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \qquad \textit{ou} \qquad \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix}$$

que satisfaz as seguintes propriedades.

(i) **Multilinearidade:** *Para cada* $1 \leq i \leq n$ *temos*

$$\begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{i1}+b_{i1} & & a_{in}+b_{in} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix} = \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{i1} & & a_{in} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix} + \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ b_{i1} & & b_{in} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix}$$

e, para α um escalar qualquer,

$$\begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ \alpha a_{i1} & & \alpha a_{in} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix} = \alpha \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{i1} & & a_{in} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix}$$

(ii) **Alternância:** $\det A = 0$ *se duas linhas da matriz A forem iguais.*
(iii) **Normalização:** $\det I_n = 1$.

Em concreto, no caso das matrizes 2×2, a primeira propriedade diz por exemplo que

$$\begin{vmatrix} 2 & 1 \\ 1+3 & 2+4 \end{vmatrix} = \begin{vmatrix} 2 & 1 \\ 1 & 2 \end{vmatrix} + \begin{vmatrix} 2 & 1 \\ 3 & 4 \end{vmatrix} \quad \text{e} \quad \begin{vmatrix} 2 & 1 \\ -2\cdot 3 & -2 \cdot 4 \end{vmatrix} = -2 \begin{vmatrix} 2 & 1 \\ 3 & 4 \end{vmatrix}$$

Observação 10.2. *Se quiséssemos abstrair as propriedades do volume que usámos para chegar às fórmulas para a área de um paralelogramo e o volume de um paralelipípedo, a única alteração que teríamos de fazer nas propriedades (i) a (iii) acima seria substituir α por $|\alpha|$ do lado direito da segunda igualdade na propriedade (i), e insistir que a função volume tomasse valores em $\mathbb{R}_0^+$.*

Identificando as linhas de uma matriz $n \times n$ com vetores de $\mathbb{R}^n$, podemos pensar na função determinante como uma função $D\colon \mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \mathbb{R}$ que associa um escalar a um n-tuplo $(v_1, \ldots, v_n)$ de vetores de $\mathbb{R}^n$ (v_i é a i-ésima linha da matriz). Deste ponto de vista, a propriedade de multilinearidade escreve-se

$$D(v_1, \ldots, \alpha v_i + \beta v_i', \ldots, v_n) = \alpha D(v_1, \ldots, v_i, \ldots, v_n) + \beta D(v_1, \ldots, v_i', \ldots, v_n) \tag{25}$$

onde $v_1, \ldots, v_n \in \mathbb{R}^n$ são vetores arbitrários e α, β escalares arbitrários. A equação (25) diz que, para cada i entre 1 e n, a função $D_i\colon \mathbb{R}^n \to \mathbb{R}$ que se obtém quando fixamos todos os vectores excepto o i-ésimo,

$$D_i(v) = D(v_1, \ldots, v_{i-1}, v, v_{i+1}, \ldots, v_n)$$

é linear (ou seja, um elemento do dual de $\mathbb{R}^n$).

Em geral, uma função $D\colon V \times \cdots \times V \to \mathbb{R}$ satisfazendo (25) diz-se uma *função multilinear*[6] (é linear em cada variável independentemente).

A razão para o nome da segunda propriedade na definição de determinante é a seguinte.

Proposição 10.3. *Seja $D\colon V \times \cdots \times V \to \mathbb{R}$ uma função multilinear. Então as seguintes condições são equivalentes*

(i) $D(v_1, \ldots, v_n) = 0$ *se* $v_i = v_j$ *para algum* $i \neq j$.
(ii) Se $i \neq j$, então $D(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n) = -D(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_n)$ para todos os $v_1, \ldots, v_n$ (isto é, a troca de dois argumentos tem como efeito a troca de sinal do valor da função).

Dem.

[6] Também se chama um tensor-n covariante em V.

$(i) \Rightarrow (ii)$ Supondo que $i < j$, e aplicando a linearidade primeiro na i-ésima variável e depois na j-ésima obtemos

$$\begin{aligned} D(v_1, \ldots, v_i + v_j, \ldots, v_i + v_j, \ldots, v_n) &= D(v_1, \ldots, v_i, \ldots, v_i + v_j, \ldots, v_n) + \\ &\quad D(v_1, \ldots, v_j, \ldots, v_i + v_j, \ldots, v_n) = \\ &= D(v_1, \ldots, v_i, \ldots, v_i, \ldots, v_n) + D(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n) \\ &\quad + D(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_n) + D(v_1, \ldots, v_j, \ldots, v_j, \ldots, v_n) \end{aligned}$$

Substituindo os termos com argumentos repetidos por 0 obtém-se

$$0 = 0 + D(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n) + D(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_n) + 0$$

que é equivalente à condição (ii).

$(ii) \Rightarrow (i)$ Se $v_i = v_j$, então a troca do i-ésimo argumento com o j-ésimo não tem nenhum efeito. Portanto

$$D(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n) = -D(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_n) = -D(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n)$$

e portanto $D(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n) = 0$.

□

É um resultado importante que as propriedades (i) a (iii) na definição de determinante especificam completamente essa função:

Teorema 10.4. *Existe uma única função determinante* $\det M_{n\times n}(\mathbb{R}) \to \mathbb{R}$

A demonstração deste teorema segue o padrão usual: iremos ver que só há uma possibilidade para uma tal função (obtendo no processo uma fórmula para o determinante) e depois verificar que essa única possibilidade satisfaz de facto os axiomas da definição. Começamos por ilustrar este processo usando os axiomas para ver que a única função determinante nas matrizes 2×2 é

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$$

Sendo $a, b, c, d \in \mathbb{R}$ quaisquer e aplicando a linearidade do determinante na primeira linha da matriz temos

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = a \begin{vmatrix} 1 & 0 \\ c & d \end{vmatrix} + b \begin{vmatrix} 0 & 1 \\ c & d \end{vmatrix}$$

e aplicando agora a linearidade na segunda linha obtemos

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = a\left(c \begin{vmatrix} 1 & 0 \\ 1 & 0 \end{vmatrix} + d \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}\right) + b\left(c \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} + d \begin{vmatrix} 0 & 1 \\ 0 & 1 \end{vmatrix}\right)$$

Os primeiro e último termos do lado direito do sinal de igual na expressão acima são nulos porque as linhas das matrizes em questão estão repetidas. Pelas propriedades (iii) e (ii) respetivamente temos

$$\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1 \qquad \text{e} \qquad \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = -1$$

portanto

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

é a única função real das matrizes 2×2 que satisfaz as condições da Definição 10.1.

Façamos agora o caso mais realista de uma matriz 3×3. Assumindo que existe a função determinante e usando linearidade na primeira linha obtemos

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a \begin{vmatrix} 1 & 0 & 0 \\ d & e & f \\ g & h & i \end{vmatrix} + b \begin{vmatrix} 0 & 1 & 0 \\ d & e & f \\ g & h & i \end{vmatrix} + c \begin{vmatrix} 0 & 0 & 1 \\ d & e & f \\ g & h & i \end{vmatrix} \tag{26}$$

Desenvolvendo o primeiro termo do lado direito do sinal de igual usando linearidade na segunda linha obtemos

$$a \begin{vmatrix} 1 & 0 & 0 \\ d & e & f \\ g & h & i \end{vmatrix} = a \left(d \begin{vmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ g & h & i \end{vmatrix} + e \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ g & h & i \end{vmatrix} + f \begin{vmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ g & h & i \end{vmatrix} \right)$$

O primeiro termo na soma do lado direito é nulo porque a primeira linha está repetida. Da mesma forma, cada parcela do lado direito em (26) vai dar origem a dois termos não nulos quando aplicarmos linearidade ao longo da segunda linha da matriz. Podemos agora aplicar linearidade ao longo da terceira linha a cada um destes 6 termos. Por exemplo, para o primeiro dos seis resultaria

$$ae \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ g & h & i \end{vmatrix} = ae \left(g \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{vmatrix} + h \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{vmatrix} + i \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} \right) = aei$$

uma vez que os dois primeiros termos da soma anterior têm linhas repetidas e o determinante da matriz identidade é 1. Aplicando o mesmo raciocínio para os restantes termos não nulos na expansão até à segunda linha obtemos a seguinte expressão para o determinante:

$$aei + afh \begin{vmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{vmatrix} + bdi \begin{vmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{vmatrix} + bfg \begin{vmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{vmatrix} + cdh \begin{vmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{vmatrix} + ceg \begin{vmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{vmatrix}$$

Os determinantes das matrizes com 0s e $1s$ são ± 1 consoante o número de vezes qye temos que trocar um par de linhas para transformar a matriz na identidade é par ou ímpar. Recuperamos assim a expressão para o determinante de uma matriz 3×3:

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = aei - afh - bdi + bfg + cdh - ceg$$

Procedendo desta forma para uma matriz $n \times n$ é agora claro que vamos obter uma expressão para o determinante. Haverá um termo não nulo na expressão para cada matriz de $1s$ e 0s que tenha exatamente um 1 em cada linha, e tal que os 1s nunca apareçam repetidos numa coluna. Para descrever estes termos por meio de uma expressão necessitamos de alguma terminologia.

Definição 10.5. *Uma* permutação *do conjunto* $\{1,\ldots,n\}$ *é uma função bijetiva*

$$\sigma\colon \{1,\ldots,n\} \to \{1,\ldots,n\}$$

Designamos por Σ_n *o conjunto de todas estas permutações.*

Uma permutação descreve uma troca de ordem. Deve ser familiar do ensino secundário que o número de elementos de Σ_n é $n!$. Os termos na expansão do determinante vão corresponder precisamente às permutações: se chamarmos $\sigma(i)$ à coluna em que aparece o 1 na linha i, a condição que não apareçam dois 1s na mesma coluna é $\sigma(i) \neq \sigma(j)$ para $i \neq j$, ou seja é a injetividade da função σ. Como uma função injetiva de um conjunto com n elementos para ele próprio é necessariamente uma bijeção, conclui-se que a função determinada por uma matriz de 0s e 1s satisfazendo as condições indicadas é uma bijeção.

O termo do determinante de A correspondente a uma permutação σ será dado pelo produto das entradas de A que ocorriam nas posições onde estão os 1s, ou seja o produto dos $a_{i\sigma(i)}$ com $i = 1,\ldots,n$. O termo terá um sinal que será $\pm$ consoante o número de vezes que temos que trocar pares de linhas para transformar a matriz de 0s e 1s na identidade é par ou impar. Chamando a este sinal $\operatorname{sgn}(\sigma)$ - *o sinal da permutação* σ - obtemos a seguinte expressão para o determinante:

$$\det(A) = \sum_{\sigma\in\Sigma_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)} \tag{27}$$

O argumento anterior torna claro que se existir uma função determinante, ela é única (tem que ser dada pela fórmula (27)!). Mas neste momento não é ainda claro que uma tal função exista. Há muitas maneiras de trocar pares de linhas de forma a obter a matriz identidade a partir de uma matriz de 0s e 1s. Se para uma das maneiras o número de trocas fosse par e para outra maneira fosse ímpar concluir-se-ia que a função determinante não podia existir.

Não é fácil verificar diretamente que o sinal de uma permutação está bem definido. Em vez disso vamos dar uma construção indutiva do determinante. Uma vez que isto esteja feito teremos implicitamente provado que o sinal de uma permutação está bem definido! Será necessariamente

$$\operatorname{sgn}(\sigma) = \det A(\sigma) \qquad \text{com } A(\sigma) \text{ a matriz com entradas } a_{ij} = \begin{cases} 1 & \text{se } j = \sigma(i) \\ 0 & \text{caso contrário.} \end{cases} \tag{28}$$

A matriz $A(\sigma)$ diz-se uma *matriz de permutação*. O efeito que tem nas coordenadas de um vetor linha ou coluna é uma permutação das coordenadas. Por exemplo,

$$A(\sigma) \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} x_{\sigma(1)} \\ x_{\sigma(2)} \\ \vdots \\ x_{\sigma(n)} \end{bmatrix}$$

É um bom exercício ver o que acontece quando se multiplica à esquerda por um vetor linha.

Dem. do Teorema 10.4. Já vimos que se existir uma função determinante ela é única (e dada pela fórmula (27)). Vamos ver por indução em n que existe uma função determinante para matrizes $n \times n$. Quando $n = 1$, é imediato que

$$\det([a_{11}]) = a_{11}$$

Suponhamos que já definimos uma função determinante nas matrizes $n \times n$. Dada uma matriz A do tipo $(n+1) \times (n+1)$, seja A_{1i} a matriz $n \times n$ que se obtém de A suprimindo a primeira linha e a i-ésima coluna. Vamos definir

$$\det(A) = a_{11} \det(A_{11}) - a_{12} \det(A_{12}) + \ldots + (-1)^n a_{1(n+1)} \det A_{1(n+1)} \tag{29}$$

fórmula esta que é motivada pela relação entre os determinantes para matrizes 3×3 e 2×2 que obtivemos anteriormente.

Temos a verificar que $\det A$ verifica as condições $(i) - (iii)$ da Definição 10.1. A condição (i) é verificada porque a expressão (29) é claramente linear na primeira linha da matriz A e, por hipótese de indução, nas restantes, uma vez que as funções $\det(A_{1i})$ são multilineares.

A condição (iii) também é verificada porque as entradas na primeira linha da matriz identidade I_{n+1} com excepção da primeira são todas nulas. Uma vez que $(I_{(n+1)})_{11} = I_n$ obtemos

$$\det(I_{n+1}) = 1 \cdot \det(I_n) = 1.$$

Resta-nos verificar que se uma das linhas de A está repetida então $\det A = 0$. Se a repetição ocorrer nas linhas i e j com $i, j \geq 2$ então todos os termos $\det(A_{1i})$ em (29) se anulam (por hipótese de indução) e portanto $\det A = 0$. Se $i = 1$, podemos assumir que $j = 2$ uma vez que, por hipótese de indução, o termo direito da equação (29) troca de sinal quando trocamos a linha j de A com a segunda linha.

Suponhamos assim que A tem a primeira e segunda linha iguais. Se A é uma matriz 2×2 a expressão (29) é

$$\det(A) = a_{11}a_{22} - a_{12}a_{21} = a_{11}a_{12} - a_{12}a_{11} = 0$$

Se $n > 1$, podemos, por hipótese de indução aplicar a expressão (29) às matrizes $n \times n$ A_{1i}. A entrada $1j$ na primeira linha de A_{1i} é igual a

$$\begin{cases} a_{2j} & \text{se } j < i \\ a_{2(j+1)} & \text{se } j > i \end{cases}$$

portanto

$$\det(A_{1j}) = \sum_{j=1}^{i-1} (-1)^{j-1} a_{2j} \det(A_{12|ij}) + \sum_{j=i+1}^{n+1} (-1)^j a_{2j} \det(A_{12|ij})$$

onde $A_{12|ij}$ denota a matriz $(n-1) \times (n-1)$ que se obtém de A suprimindo as primeiras duas linhas e as colunas i e j. Substituindo esta expressão em (29) vemos que há dois termos nos quais aparece $\det(A_{12|ij})$ para i, j dados com $1 \leq i < j \leq n$:

$$(-1)^{i-1} a_{1i} \cdot (-1)^{j-2} a_{2j} \det(A_{12|ij})$$

que é o $(j-1)$-ésimo termo da expansão do termo $(-1)^{i-1}a_{1i}\det(A_{1i})$ à direita do sinal de igual em (29) e

$$(-1)^{j-1}a_{1j}\cdot(-1)^{i-1}a_{2i}\det(A_{12|ij})$$

que vem da expansão do termo $(-1)^{j-1}a_{1j}\det(A_{1j})$. Uma vez que as primeiras duas linhas da matriz são iguais, temos

$$(-1)^{i-1}a_{1i}\cdot(-1)^{j-2}a_{2j}\det(A_{12|ij})+(-1)^{j-1}a_{1j}\cdot(-1)^{i-1}a_{2i}\det(A_{12|ij})=0$$

o que conclui a demonstração. □

Observação 10.6. *Uma função $f\colon M_{n\times n}(\mathbb{R})\to\mathbb{R}$ satisfazendo as propriedades (i) e (ii) na Definição 10.1 chama-se uma função multilinear alternante. O argumento usado na demonstração de unicidade do determinante aplicado a uma tal função (sem qualquer alteração) mostra que*

$$f(A)=\sum_{\sigma\in\Sigma_n}\operatorname{sgn}(\sigma)a_{1\sigma(1)}\cdots a_{n\sigma(n)}f(I_n)$$

pelo que o valor de uma tal função em qualquer matriz é completamente determinado pelo valor que assume na matriz identidade. Mas sendo $\lambda\in\mathbb{R}$ qualquer, a função $A\mapsto\lambda\det(A)$ é uma função multilinear alternante que assume o valor λ em I_n, pelo que se conclui que toda a função multilinear alternante é da forma

$$f(A)=\lambda\det(A)$$

em que $\lambda=f(I_n)$.

Vamos agora ver algumas propriedades importantes do determinante que nos ajudam a calculá-lo.

Definição 10.7. *Seja A uma matriz $n\times n$. Para $1\le i,j\le n$ designamos por A_{ij} a matriz $(n-1)\times(n-1)$ que se obtém de A omitindo a i-ésima linha e a j-ésima coluna. O* menor-ij *de A é o número* $\det A_{ij}$ *e o* cofator-ij *de A é* $(-1)^{i+j}\det A_{ij}$. *A matriz $n\times n$ cuja entrada ij é o cofator-ij diz-se* a matriz dos cofatores de A *e denota-se por* $\operatorname{cof}A$.

Proposição 10.8 (Propriedades do determinante)**.** *Sejam A e B matrizes $n\times n$.*

(i) **Expansão de Laplace** *Sendo $1\le i\le n$, temos*

$$\det(A)=\sum_{j=1}^{n}(-1)^{i+j}a_{ij}\det(A_{ij})$$

onde A_{ij} é a matriz que se obtém de A omitindo a linha i e a coluna j. A fórmula acima chama-se a expansão de Laplace para o determinante ao longo da linha i.

(ii) $\det(AB)=\det(A)\det(B)$

(iii) $\det(A^T)=\det(A)$

(iv) $A(\operatorname{cof}(A))^T=\det(A)I_n$.

Antes de vermos a demonstração destas propriedades notemos as seguintes consequências.

Corolário 10.9 (Expansão de Laplace ao longo de colunas). *Sendo* $1 \leq j \leq n$, *temos*

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij})$$

Dem. A expansão ao longo da coluna j no enunciado é exatamente a expansão ao longo da linha j de A^T. Logo calcula $\det A^T = \det A$. □

Corolário 10.10. *Uma matriz quadrada* A *é invertível sse* $\det A \neq 0$ *e nesse caso*

$$A^{-1} = \frac{1}{\det A} (\operatorname{cof} A)^T$$

Dem. Se A é invertível então $\det(A)\det(A^{-1}) = \det(AA^{-1}) = \det(I_n) = 1$ logo $\det(A) \neq 0$ e

$$\det(A^{-1}) = \frac{1}{\det A}$$

Reciprocamente se $\det A \neq 0$, a Proposição 10.8 (iv) diz-nos que

$$A\left(\frac{1}{\det A}(\operatorname{cof} A)^T\right) = I_n$$

pelo que A é invertível (cf. Proposição 7.8 (vi)) sendo a inversa descrita pela fórmula no enunciado. □

Esta fórmula para a inversa de uma matriz tem mais utilidade teórica do que prática porque não é fácil calcular determinantes de matrizes grandes. É no entanto muito útil para matrizes 2×2, caso em que afirma que

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \qquad \text{quando } ad - bc \neq 0$$

Dem. da Proposição 10.8. (i) Para $i = 1$ a expansão de Laplace é simplesmente a expressão indutiva (29) usada para demonstrar a existência do determinante. Se $i > 1$, seja $\tilde{A}$ a matriz que se obtém de A trocando a linha 1 com a linha i. Aplicando (29) obtemos

$$\det(A) = -\det(\tilde{A}) = -\sum_{j=1}^{n} (-1)^{1+j} \tilde{a}_{1j} \det(\tilde{A}_{1j}) = -\sum_{j=1}^{n} (-1)^{1+j} a_{ij} \det(\tilde{A}_{1j}) \tag{30}$$

Notamos agora que as matrizes $\tilde{A}_{1j}$ e A_{ij} diferem pela troca da $(i-1)$-ésima linha com o bloco formado pelas linhas que a precedem - o que corresponde a $(i-2)$-trocas de pares de linhas à medida que a linha $(i-1)$ "flutua até chegar à superfície". Portanto

$$\det(\tilde{A}_{1j}) = (-1)^{i-2} \det A_{ij}$$

Substitituindo em (30) obtemos a fórmula pretendida.

(ii) Fixada uma matriz B, considere-se a função $f\colon M_{n\times n}(\mathbb{R}) \to \mathbb{R}$ definida por

$$f(A) = \det(AB)$$

Trata-se de uma função multilinear e alternante das linhas de A pela definição do produto de matrizes e pelas propriedades (i) e (ii) na definição de função determinante. Uma vez que $f(I_n) = \det(B)$, a Observação 10.6 diz-nos que $f(A) = \det(A)\det(B)$.

(iii) A expressão (27) diz-nos que

$$\det(A^T) = \sum_{\sigma\in\Sigma_n} \operatorname{sgn}(\sigma) a^T_{1\sigma(1)} \cdots a^T_{n\sigma(n)} = \sum_{\sigma\in\Sigma_n} \operatorname{sgn}(\sigma) a_{\sigma(1)1} \cdots a_{\sigma(n)n}$$

Seja $\sigma^{-1}\colon \{1,\dots,n\} \to \{1,\dots,n\}$ a permutação inversa de σ (isto é, a permutação que verifica $\sigma^{-1}(\sigma(i)) = i$ para $i = 1,\dots,n$). Então

$$\sigma(i) = j \Leftrightarrow i = \sigma^{-1}(j)$$

e portanto

$$a_{\sigma(1)1} \cdots a_{\sigma(n)n} = a_{1\sigma^{-1}(1)} \cdots a_{n\sigma^{-1}(n)}$$

(do lado direito do sinal de igual aparecem as mesmas entradas da matriz que do lado esquerdo mas por outra ordem; estão agora ordenados pelo primeiro índice, enquanto que à esquerda estão ordenados pelo segundo). Temos assim

$$\det(A^T) = \sum_{\sigma\in\Sigma_n} \operatorname{sgn}(\sigma) a_{1\sigma^{-1}(1)} \cdots a_{n\sigma^{-1}(n)} \tag{31}$$

As matrizes $A(\sigma)$ associadas às permutações (ver (28)) colocam na coordenada i de um vetor coluna a coordenada que estava na posição $\sigma(i)$. Logo o efeito de $A(\sigma)A(\sigma^{-1})$ num vetor coluna é colocar na coordenada i a componente $x_{\sigma^{-1}(\sigma(i))} = x_i$. Portanto

$$A(\sigma)A(\sigma^{-1}) = I_n \Rightarrow \det(A(\sigma))\det A(\sigma^{-1}) = 1 \Rightarrow \det(A(\sigma)) = \det(A(\sigma^{-1}))$$

onde a última implicação usa que os determinantes das matrizes de permutação é ± 1. Notando que $\operatorname{sgn}(\sigma) = \det A(\sigma)$ e substituindo em (31) temos

$$\det(A^T) = \sum_{\sigma\in\Sigma_n} \operatorname{sgn}(\sigma^{-1}) a_{1\sigma^{-1}(1)} \cdots a_{n\sigma^{-1}(n)}$$

Quando σ percorre todos os elementos de Σ_n, o mesmo sucede com a sua inversa σ^{-1} pelo que a expressão à direita na igualdade acima é exatamente a fórmula (27) para o determinante de A. Isto conclui a demonstração.

(iv) A fórmula no enunciado diz-nos que o produto da linha i da matriz A pela coluna j da matriz $(\operatorname{cof} A)^T$ é $\det(A)$ se $i = j$ e 0 caso contrário. A expressão para este produto é

$$\sum_{k=1}^n a_{ik}((\operatorname{cof} A)^T)_{kj} = \sum_{k=1}^n a_{ik}(\operatorname{cof} A)_{jk} = \sum_{k=1}^n a_{ik}(1)^{j+k}\det(A_{jk})$$

Quando $i = j$, a expressão anterior é a expansão de Laplace para o determinante de A ao longo da linha i e é portanto igual a $\det A$. Para $i \neq j$, a expressão é a expansão de Laplace ao longo da linha j da matriz que se obtém de A repetindo a linha j na linha i, e é portanto igual a 0.

□

Observação 10.11. *É instrutivo pensar em escrever explicitamente a igualdade indicada na Proposição 10.8(ii) em termos das entradas das matrizes envolvidas. Mesmo para matrizes* 3×3 *a complexidade é enorme! É fácil no entanto convencer-se que, pelo menos a menos de sinal, a igualdade se deve verificar:*

Atendendo à Proposição 10.8(iii), $|\det A|$ *é o volume do paralelípipedo que tem por arestas as colunas da matriz* A*, paralelipípedo este que é a imagem do cubo com arestas unitárias em* $\mathbb{R}^n$ *pela transformação linear* $x \mapsto Ax$ *. Segue-se que a imagem de um cubo qualquer em* $\mathbb{R}^n$ *por esta transformação tem volume igual a* $|\det(A)|$ *vezes o volume do cubo original. Verão em Cálculo 2 que o volume de um subconjunto (razoável) de* $\mathbb{R}^n$ *se define aproximando esse conjunto por cubos muito pequenos e passando ao limite. Segue-se então que* $|\det A|$ *é o fator pelo qual a transformação linear* $x \mapsto Ax$ *multiplica volumes.*

Uma vez que AB *é a matriz que representa a composta das transformações lineares representadas por* A *e* B*, segue-se que o fator pela qual* AB *multiplica volumes é* $|\det(A)||\det(B)|$.

Exemplo 10.12. *Vamos calcular o determinante*

$$\begin{vmatrix} 2 & 0 & 3 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 4 & 5 & 7 \\ 1 & 8 & 9 & 3 \end{vmatrix}$$

usando a expansão de Laplace. Uma vez que a segunda linha tem 3 zeros, é mais eficiente fazer a expansão ao longo dessa linha. Obtemos

$$0\cdot(-1)^{2+1}\begin{vmatrix} 0 & 3 & 0 \\ 4 & 5 & 7 \\ 8 & 9 & 3 \end{vmatrix} + 0\cdot(-1)^{2+2}\begin{vmatrix} 2 & 3 & 0 \\ 1 & 5 & 7 \\ 1 & 9 & 3 \end{vmatrix} + 1\cdot(-1)^{2+3}\begin{vmatrix} 2 & 0 & 0 \\ 1 & 4 & 7 \\ 1 & 8 & 3 \end{vmatrix} + 0\cdot(-1)^{2+4}\begin{vmatrix} 2 & 0 & 3 \\ 1 & 4 & 5 \\ 1 & 8 & 9 \end{vmatrix}$$

e fazendo agora a expansão de Laplace do único termo não nulo ao longo da primeira linha obtém-se

$$-\begin{vmatrix} 2 & 0 & 0 \\ 1 & 4 & 7 \\ 1 & 8 & 3 \end{vmatrix} = -2\cdot(-1)^{1+1}\begin{vmatrix} 4 & 7 \\ 8 & 3 \end{vmatrix} = -2(4\cdot 3 - 7\cdot 8) = 88.$$

A fórmula para a inversa de uma matriz em termos do determinante conduz à seguinte fórmula explícita para a solução de um sistema linear quando a matriz dos coeficiente do sistema é invertível.

Proposição 10.13 (Regra de Cramer)**.** *Seja* A *uma matriz* $n \times n$ *invertível e* b *uma matriz* $n \times 1$*. Então a componente* x_i *da solução do sistema linear*

$$Ax = b$$

é dada pela fórmula

$$x_i = \frac{\det A_i}{\det A}$$

onde A_i *é a matriz que se obtém de* A *substituindo a coluna* i *de* A *por* b*.*

Dem. A componente x_i da solução do sistema é a i-ésima entrada de $A^{-1}b$ e é portanto dada por

$$x_i = \sum_{j=1}^{n} c_{ij} b_j$$

onde c_{ij} é a entrada ij da matriz A^{-1}. Pelo Corolário 10.10 esta entrada é

$$c_{ij} = (-1)^{i+j} \frac{\det(A_{ji})}{\det A}$$

pelo que

$$x_i = \frac{1}{\det A} \sum_{j=1}^{n} (-1)^{i+j} b_j \det(A_{ji})$$

O somatório na expressão anterior é exatamente o desenvolvimento de Laplace ao longo da coluna i da matriz A_i do enunciado. Isto conclui a demonstração. □

Exemplo 10.14. *Vamos achar a coordenada y da solução do sistema*

$$\begin{cases} 2x + 3y + z = 3 \\ x - y + z = 4 \\ x + 2y - z = 5 \end{cases}$$

Pela regra de Cramer temos

$$y = \frac{\begin{vmatrix} 2 & 3 & 1 \\ 1 & 4 & 1 \\ 1 & 5 & -1 \end{vmatrix}}{\begin{vmatrix} 2 & 3 & 1 \\ 1 & -1 & 1 \\ 1 & 2 & -1 \end{vmatrix}} = -\frac{11}{7}$$

Definição 10.15. *Uma matriz quadrada A diz-se* triangular superior *se $a_{ij} = 0$ para $i > j$ (isto é se todas as entradas abaixo da diagonal principal são nulas) e* triangular inferior *se $a_{ij} = 0$ para $i < j$ (isto é se todas as entradas acima da diagonal principal são nulas).*

É imediato da expansão de Laplace que o determinante de uma matriz triangular (superior ou inferior) é igual ao produto das entradas na diagonal

$$\begin{vmatrix} \lambda_1 & * & \cdots & * \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_n \end{vmatrix} = \lambda_1 \cdots \lambda_n$$

Uma generalização da última propriedade que é muito útil diz respeito ao cálculo de determinantes de matrizes escritas *por blocos*. Podemos pensar numa matriz de grande dimensão como uma "matriz de matrizes" juntando algumas entradas para formar matrizes de menor dimensão.

Por exemplo podemos escrever a matriz

$$\begin{bmatrix} 1 & 2 & 5 & 6 & 7 \\ 3 & 4 & 8 & 1 & 2 \\ 0 & 3 & 4 & 5 & 6 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \tag{32}$$

com A, B, C, D os blocos respetivamente $2 \times 2, 2 \times 3, 1 \times 2$, e 1×3 dados por

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} 5 & 6 & 7 \\ 8 & 1 & 2 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & 3 \end{bmatrix}, \quad D = \begin{bmatrix} 4 & 5 & 6 \end{bmatrix}$$

Por vezes é útil calcular com matrizes separadas por blocos. É claro que a soma e o produto por escalar se podem fazer bloco a bloco. Mais interessante é que o produto é dado pela fórmula habitual para o produto de matrizes mas com os escalares substituídos por blocos (desde que os produtos de matrizes em questão façam sentido). Isto é uma consequência imediata da definição de produto de matrizes. Por exemplo, considerando a matriz por blocos

$$\begin{bmatrix} X & Y \end{bmatrix} = \begin{bmatrix} 2 & 1 & 1 \\ 3 & 4 & 2 \end{bmatrix} \quad \text{com} \quad X = \begin{bmatrix} 2 & 1 \\ 3 & 4 \end{bmatrix} \text{ e } Y = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

O produto desta matriz pela matriz (32) é

$$\begin{bmatrix} X & Y \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} XA+YC & XB+YD \end{bmatrix}$$

Por exemplo a entrada 21 da matriz produto é igual a

$$3 \cdot 1 + 4 \cdot 3 + 2 \cdot 0 = (3 \cdot 1 + 4 \cdot 3) + 2 \cdot 0$$

onde do lado direito do sinal de igual temos dentro de parentesis a entrada 21 do bloco XA e depois a entrada 21 do bloco YC. A relevância destas decomposições para o cálculo de determinantes é a seguinte

Proposição 10.16. *O determinante de uma matriz triangular por blocos com blocos quadrados na diagonal é o produto dos determinantes dos blocos diagonais*

$$\begin{vmatrix} A_1 & * & \cdots & * \\ 0 & A_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & A_n \end{vmatrix} = |A_1| \cdots |A_n|$$

Dem. É um exercício da Ficha 11. □

Exemplo 10.17.

$$\begin{vmatrix} 1 & 0 & 5 & 11 & 6 \\ 3 & 2 & 3 & 27 & 5 \\ 0 & 0 & 4 & 2 & 2 \\ 0 & 0 & 0 & 3 & 2 \\ 0 & 0 & 0 & 1 & 4 \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 3 & 2 \end{vmatrix} \cdot 4 \cdot \begin{vmatrix} 3 & 2 \\ 1 & 4 \end{vmatrix} = 2 \cdot 4 \cdot 10 = 80$$

Finalmente, notemos que se V é um espaço vetorial de dimensão finita, pode definir-se o *determinante de uma transformação linear* $T\colon V \to V$. Sendo B uma base ordenada de V definimos

$$\det(T) = \det(A_{T,B,B})$$

Temos que verificar que este número é independente da escolha de B. Se B' é outra base e $S = S_{B\to B'}$ a matriz de mudança de coordenadas então

$$A_{T,B',B'} = SA_{T,B,B}S^{-1}$$

e portanto

$$\det(A_{T,B',B'}) = \det(S)\det(A_{T,B,B})\det(S^{-1}) = \det(S)\det(A_{T,B,B})\frac{1}{\det(S)} = \det(A_{T,B,B})$$

Não seria fácil neste momento explicar-vos como definir intrinsecamente o determinante de uma transformação linear, sem apelar à sua representação matricial.

11. Mais sobre valores e vetores próprios

Recorde-se que a nossa motivação para o estudo do determinante foi precisamente a busca de uma equação que caracterizasse os valores próprios de uma matriz (ou transformação linear). Agora que dispomos dessa equação, o seu estudo vai trazer-nos bastante informação acerca da transformação linear em questão.

Definição 11.1. *Seja A uma matriz $n\times n$. O* polinómio característico *de A é o polinómio definido pela expressão*

$$p(\lambda) = \det(A - \lambda I_n)$$

Tendo em conta a fórmula (27) para o determinante vemos que o polinómio característico de uma matriz $n \times n$ tem grau exatamente n (o termo de grau n resulta do produto das entradas na diagonal e é igual a $(-\lambda)^n$). A propriedade fundamental do polinómio característico é que as suas *raízes* ou *zeros* são exatamente os valores próprios de A:

$$p(\lambda) = 0 \Leftrightarrow \det(A - \lambda I_n) = 0 \Leftrightarrow (A - \lambda I_n) \text{ não é invertível} \Leftrightarrow N(A - \lambda I_n) \neq \{0\}$$

sendo que a última condição diz, por definição, que λ é um valor próprio de A.

Enunciamos agora um resultado fundamental cuja demonstração irão ver no próximo ano na cadeira de Análise Complexa, e que garante em particular que toda a matriz quadrada tem um valor próprio, pelo menos se for vista como uma matriz de números complexos.

Teorema 11.2 (Teorema Fundamental da Álgebra)**.** *Qualquer polinómio de grau k*

$$p(x) = a_0 + a_1x + \ldots a_kx^k$$

com coeficientes $a_i \in \mathbb{C}$ (e $a_k \neq 0$) pode ser escrito de forma única a menos de troca de ordem dos fatores na forma

$$a_k(x-\lambda_1)^{n_1}(x-\lambda_2)^{n_2}\cdots(x-\lambda_k)^{n_k} \tag{33}$$

com $\lambda_i \in \mathbb{C}$ distintos, e n_i números naturais.

Claro que os números λ_i na expressão (33) são as raízes do polinómio $p(x)$. O expoente n_i diz-se a *multiplicidade* da raíz λ_i.

Observação 11.3. *O Teorema Fundamental da Álgebra é análogo ao Teorema Fundamental da Aritmética que diz que qualquer número natural se pode escrever de forma única como um produto de potências de números primos a menos de troca de fatores.*

Definição 11.4. *Sendo λ um valor próprio de uma matriz quadrada A, chama-se*

- Espaço próprio *de λ ao subespaço $N(A - \lambda I_n)$ de $\mathbb{R}^n$.*
- Multiplicidade geométrica *de λ, $m_g(\lambda)$ à dimensão do espaço próprio de λ.*
- Multiplicidade algébrica *de λ, $m_a(\lambda)$ à multiplicidade de λ como raíz do polinómio característico de A.*

Os elementos do espaço próprio de λ são os vetores próprios de λ juntamente com 0, pelo que a multiplicidade geométrica é o número máximo de vetores próprios de λ linearmente independentes. Veremos em breve que $m_g(\lambda) \leq m_a(\lambda)$. Note-se que a soma das multiplicidades algébricas de todos os valores próprios é igual a n (que é o grau do polinómio característico), portanto uma matriz é diagonalizável se e só se as multiplicidades algébrica e geométrica coincidem para todos os valores próprios.

Exemplo 11.5. *Consideremos a matriz*

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

O seu polinómio característico é

$$p(\lambda) = \begin{vmatrix} 3-\lambda & 0 & 0 \\ 0 & 1-\lambda & 1 \\ 0 & -1 & 1-\lambda \end{vmatrix} = (3-\lambda)((1-\lambda)^2+1) = (3-\lambda)(\lambda-(1+i))(\lambda-(1-i))$$

Os valores próprios de A são portanto $3, 1+i, 1-i$ que têm todos multiplicidade algébrica 1. Note-se que as raízes complexas formam um par de complexos conjugados. Isto não é uma coincidência. Se

$$p(x) = a_0 + a_1 x + \ldots + a_n x^n$$

é um polinómio real, uma vez que a conjugação preserva a soma e produto de números complexos (ou seja, $\overline{z+w} = \overline{z} + \overline{w}$ e $\overline{zw} = \overline{z}\overline{w}$) temos para $\lambda \in \mathbb{C}$,

$$p(\overline{\lambda}) = a_0 + a_1\overline{\lambda} + a_n\overline{\lambda}^n = \overline{a_0 + a_1\lambda + \ldots + a_n(\lambda^n)} = \overline{p(\lambda)}$$

onde usámos que $a_i = \overline{a_i}$ porque os $a_i \in \mathbb{R}$. Assim, se λ é uma raíz complexa de p, o mesmo sucede com $\overline{\lambda}$. Adicionalmente, é um exercício simples verificar que as multiplicidades das raízes λ e $\overline{\lambda}$ são iguais.

Achemos os espaços próprios de A. O espaço próprio de 3 é $N(A-3I_3)$ que é claramente igual a $\{(x,0,0)\colon x \in \mathbb{R}\}$. Os vetores próprios de 3 são portanto os vetores não nulos que pertencem ao eixo dos xx.

O espaço próprio de $1+i$ é

$$N\left(\begin{bmatrix} 2-i & 0 & 0 \\ 0 & -i & 1 \\ 0 & -1 & -i \end{bmatrix}\right)$$

cujos elementos são as soluções do sistema

$$\begin{cases} (2-i)x = 0 \\ -iy + z = 0 \\ -y - iz = 0 \end{cases} \Leftrightarrow \begin{cases} x = 0 \\ z = iy \end{cases}$$

donde se conclui que o espaço próprio de $1+i$ é

$$\{\alpha(0,1,i) \colon \alpha \in \mathbb{C}\} \subset \mathbb{C}^3$$

Não é necessário fazer contas para achar o espaço próprio de $1-i$. De facto, se A é uma matriz real com valor próprio λ e $v \in \mathbb{C}^n$ é um vetor próprio de λ então, novamente pelas propriedades da conjugação temos

$$A\overline{v} = \overline{Av} = \overline{\lambda v} = \overline{\lambda}\overline{v}$$

onde $\overline{A}$ e $\overline{v}$ designam respetivamente a matriz e o vetor que se obtém de A e v conjugando cada componente, e na primeira igualdade usámos que $A = \overline{A}$ uma vez que A é real. Portanto v é um vetor próprio associado a λ se e só se $\overline{v}$ é um vetor próprio associado a $\overline{\lambda}$.

O espaço próprio de $(1-i)$ é portanto

$$\{\alpha(0,1,-i) \colon \alpha \in \mathbb{C}\} \subset \mathbb{C}^3$$

Proposição 11.6. *Seja A uma matriz $n \times n$ complexa e μ um valor próprio de A. Então a multiplicidade geométrica de μ é menor ou igual à multiplicidade algébrica de μ.*

Dem. Seja k a multiplicidade algébrica de μ, suponhamos por absurdo que $v_1, \ldots, v_{k+1}$ são vetores próprios de μ linearmente independentes. Sejam $v_{k+2}, \ldots, v_n$ vetores de $\mathbb{C}^n$ tais que $B = (v_1, \ldots, v_n)$ é uma base ordenada de $\mathbb{C}^n$.

Consideremos a transformação linear $T \colon \mathbb{C}^n \to \mathbb{C}^n$ determinada por $Tx = Ax$. Uma vez que $Tv_i = \mu v_i$ para $i = 1, \ldots, k+1$, a matriz que representa T na base B é triangular superior por blocos da forma

$$A_{T,B,B} = \begin{bmatrix} D & G \\ 0 & H \end{bmatrix} \qquad \text{com} \quad D = \begin{bmatrix} \mu & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \mu \end{bmatrix} \in M_{(k+1)\times(k+1)}(\mathbb{C})$$

Se $S = S_{B \to B_{can}}$ for a matriz de mudança de coordenadas da base B para a base canónica temos

$$A = SA_{T,B,B}S^{-1}$$

e portanto os polinómios caraterísticos de A e de $A_{T,B,B}$ são iguais:

$$\begin{aligned}\det(A-\lambda I_n) &= \det(SA_{T,B,B}S^{-1}-\lambda I_n)=\det(SA_{T,B,B}S^{-1}-\lambda SI_nS^{-1})\\ &= \det(S(A_{T,B,B}-\lambda I_n)S^{-1})=\det(S)\det(A_{T,B,B}-\lambda I_n)\det(S^{-1})\\ &= \det(S)\det(A_{T,B,B}-\lambda I_n)\frac{1}{\det(S)}=\det(A_{T,B,B}-\lambda I_n)\end{aligned}$$

Mas isto é uma contradição porque claramente $(\lambda-\mu)^{k+1}$ divide o polinómio característico de $A_{T,B,B}$ (que é igual a $\det(D-\lambda I_{k+1})\det(H-\lambda I_{n-k-1})=(\mu-\lambda)^{k+1}\det(H-\lambda I_{n-k-1})$) mas, por hipótese, não divide o polinómio característico de A. $\square$

Proposição 11.7. *Seja $T\colon V\to V$ uma transformação linear. Se $v_1,\ldots,v_k$ são vetores próprios de T associados a valores próprios distintos $\lambda_1,\ldots,\lambda_k$ então $\{v_1,\ldots,v_k\}$ é um conjunto linearmente independente.*

Dem. Sejam $\alpha_1,\ldots,\alpha_k$ escalares tais que

$$\alpha_1v_1+\ldots+\alpha_kv_k=0$$

Temos a ver que os α_i's são todos nulos. Uma vez que, por hipótese, $Tv_i=\lambda_iv_i$ temos

$$(T-\lambda_2\,\mathrm{Id})\circ\cdots\circ(T-\lambda_k\,\mathrm{Id})v_i=(\lambda_i-\lambda_2)\cdots(\lambda_i-\lambda_k)v_i=\begin{cases}0 & \text{se } i\geq 2\\ (\lambda_1-\lambda_2)\cdots(\lambda_1-\lambda_k)v_1 & \text{se } i=1\end{cases}$$

Portanto

$$0=(T-\lambda_2\,\mathrm{Id})\circ\cdots\circ(T-\lambda_k\,\mathrm{Id})(\alpha_1v_1+\ldots+\alpha_kv_k)=\alpha_1(\lambda_1-\lambda_2)\cdots(\lambda_1-\lambda_k)v_1$$

Donde se conclui que $\alpha_1=0$. Procedendo de forma análoga vemos que todos os coeficientes são nulos e portanto que $\{v_1,\ldots,v_k\}$ é um conjunto linearmente independente. $\square$

Exemplo 11.8. *A matriz*

$$\begin{bmatrix}1&3&-2\\0&2&1\\0&0&3\end{bmatrix}$$

é diagonalizável. De facto, tem valores próprios $1,2,3$ todos distintos. Cada valor próprio tem um vetor próprio e estes formam uma base de $\mathbb{R}^3$ pela Proposição 11.7. Mais geralmente qualquer matriz triangular superior ou inferior com entradas diagonais distintas *é diagonalizável.*

É natural perguntar como descrever uma transformação linear que não pode ser diagonalizada, isto é, tal que algum dos valores próprios tem multiplicidade geométrica inferior à multiplicidade algébrica. A resposta é que uma tal matriz pode ser "quase diagonalizada" no seguinte sentido.

Definição 11.9. *Uma matriz quadrada da forma*

$$\begin{bmatrix} \lambda & 1 & \cdots & 0 \\ 0 & \lambda & \ddots & 0 \\ 0 & 0 & \ddots & 1 \\ 0 & 0 & \cdots & \lambda \end{bmatrix}$$

com $\lambda \in \mathbb{C}$ *diz-se um* bloco de Jordan.

Diz-se que uma matriz $n \times n$ *complexa* J *está em* forma canónica de Jordan *se é diagonal por blocos e cada bloco diagonal é um bloco de Jordan. Ou seja,* J *está em forma canónica de Jordan se*

$$J = \begin{bmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & J_k \end{bmatrix} \tag{34}$$

com J_i *blocos de Jordan.*

Exemplo 11.10. *A matriz*

$$\begin{bmatrix} 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & i & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & i \end{bmatrix}$$

está em forma canónica de Jordan. Tem 4 *blocos de Jordan; um associado ao valor próprio* 2 *de tamanho* 2*, dois associados ao valor próprio* 3 *com tamanhos* 1 *e* 3 *respetivamente, e um último associado ao valor próprio* i *com tamanho* 2*.*

Teorema 11.11. *Se* A *é uma matriz* $n \times n$ *complexa, existe uma matriz invertível* S *e uma matriz* J *em forma canónica de Jordan tal que*

$$A = SJS^{-1}.$$

Observação 11.12. *Sendo* J_i *blocos de tamanho* n_i *com entrada diagonal* λ_i *temos*

$$\det(A - \lambda I) = \det(J - \lambda I) = (\lambda - \lambda_1)^{n_1} \cdots (\lambda - \lambda_k)^{n_k}$$

Logo os escalares que aparecem na diagonal da matriz J *são os valores próprios, repetidos de acordo com a sua multiplicidade algébrica. A soma dos tamanhos dos blocos com entrada diagonal* μ *é a multiplicidade algébrica do valor próprio* μ*.*

As colunas da matriz S formam uma base para $\mathbb{C}^n$. Vamos analisar o comportamento da transformação linear representada pela matriz A nesta base. Esta análise levar-nos-á a compreender como achar a forma canónica de Jordan de uma matriz A dada, assim como a matriz mudança de coordenadas S que a pôe em forma canónica de Jordan.

Sejam
$$v_1, v_2, \ldots, v_m \in \mathbb{C}^n$$
as colunas da matriz S que correspondem ao bloco de Jordan J_i e λ_i o valor que assumem as entradas diagonais de J_i. Uma vez que a coluna j de um produto de matrizes CD se obtém multiplicando a matriz C pela coluna j de D, a equação
$$AS = SJ$$
diz-nos que
$$Av_1 = \lambda_i v_1, \quad Av_2 = \lambda_i v_2 + v_1, \quad Av_3 = \lambda_i v_3 + v_2, \quad \cdots \quad Av_m = \lambda_i v_m + v_{m-1}$$
Portanto

- Os vetores que aparecem nas colunas de S correspondentes à primeira coluna de um bloco de Jordan J_i com λ_i na diagonal são vetores próprios de λ_i. Vamos dizer "primeiras colunas" para nos referir a estes vetores.
- Escrevendo um vetor $v \in \mathbb{C}^n$ na base das colunas de S vê-se imediatamente que v é um vetor próprio de A com valor próprio λ sse v é uma combinação linear de primeiras colunas correspondentes a blocos com λ na diagonal. Em particular, o número de blocos de Jordan com λ na diagonal é o número máximo de vetores próprios linearmente independentes de λ. Mais precisamente, o número de blocos de Jordan com λ na diagonal é a dimensão do espaço próprio de λ, ou seja, a multiplicidade geométrica de λ.
- Os vetores $v_1, \ldots, v_m$ satisfazem as equações

$$(A - \lambda_i I)v_2 = v_1, \quad (A - \lambda_i I)v_3 = v_2, \quad \ldots, \quad (A - \lambda_i I)v_m = v_{m-1} \tag{35}$$

  Diz-se que os vectores $v_1, \ldots, v_m$ formam uma *cadeia de Jordan.* Uma vez que $(A - \lambda_i I)v_1 = 0$ isto implica a relação
$$(A - \lambda_i I)^j v_j = 0 \quad \text{para todo o } j$$
  Diz-se que os vectores v_j são *vectores próprios generalizados* do valor próprio λ_i. O *espaço próprio generalizado* de λ é por definição o conjunto
$$\{v \in \mathbb{C}^n \colon (A - \lambda I)^k v = 0 \text{ para algum } k\}$$
  O Teorema 11.11 diz que é sempre possível obter uma base para $\mathbb{C}^n$ formada por vectores próprios generalizados de A.

Exemplo 11.13. *No Exemplo 11.10, as multiplicidades algébricas de* $2, 3$ *e* i *são* $2, 4$ *e* 2 *respetivamente, e as multiplicidades geométricas são* $1, 2$ *e* 1 *respetivamente.*

As observações anteriores podem ser usadas para determinar uma forma canónica de Jordan J e a correspondente matriz de mudança de base simultaneamente: começamos por calcular os valores próprios de A e uma base para o espaço próprio de cada valor próprio. Isso dá-nos o número de blocos de Jordan para cada um dos valores próprios de A. Se a multiplicidade algébrica de λ for superior à multiplicidade geométrica haverá pelo menos um bloco de Jordan para λ com tamanho maior do que 1. Para determinar o tamanho de cada bloco e, simultaneamente, as colunas de S que lhe correspondem tentamos resolver as equações (35) recursivamente começando com um vector próprio v_1.

Exemplo 11.14. *Vamos determinar a forma canónica de Jordan para a matriz*

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

O polinómio característico é

$$\det(A - \lambda I) = (2-\lambda)^2(1-\lambda)$$

logo os valores próprios são 2, *com multiplicidade algébrica* 2 *e* 1, *com multiplicidade algébrica* 1. *Um vector próprio para* 1 *é* $(1,-1,0)$. *Os vectores próprios de* 2 *são as soluções da equação*

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & -1 \\ 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = 0 \Leftrightarrow \begin{cases} b = 0 \\ a - c = 0 \\ a + b - c = 0 \end{cases} \Leftrightarrow \begin{cases} b = 0 \\ a = c \end{cases}$$

Uma base para os vectores próprios de 2 *é* $(1,0,1)$ *e portanto a multiplicidade geométrica de* 2 *é apenas* 1. *Isto significa que a matriz* A *não é diagonalizável. Há dois blocos de Jordan (um para cada vector próprio) e o bloco com* 2 *na diagonal tem dimensão* 2 *que é a multiplicidade algébrica de* 2. *Conclui-se que uma forma canónica de Jordan é*

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

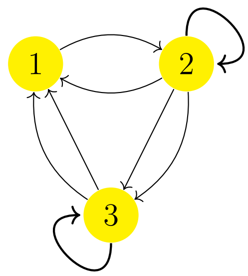
A matriz de mudança da base S *tem* $(1,-1,0)$ *na primeira coluna e* $(1,0,1)$ *na segunda. Para achar a terceira coluna resolvemos a equação* (35)*:*

$$(A - 2I)v_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \Leftrightarrow \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & -1 \\ 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \Leftrightarrow \begin{cases} b = 1 \\ a = c \end{cases}$$

Uma solução da equação anterior é, por exemplo, $v_2 = (0,1,0)$ *logo podemos tomar para matriz mudança de base*

$$S = \begin{bmatrix} 1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

11.15. **O algoritmo PageRank.** Consideremos uma internet com apenas três páginas ligadas de acordo com o diagrama



Supondo que n_1, n_2 e n_3 é o número de pessoas em cada página num dado instante e que cada pessoa clica num link ao acaso em cada página, o número de pessoas que esperaríamos estivesse numa página no instante seguinte seria

$$\begin{bmatrix} 0 & \frac{1}{4} & \frac{2}{3} \\ 1 & \frac{1}{4} & 0 \\ 0 & \frac{1}{2} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix}$$

A entrada ij da matriz é a probabilidade de uma internauta que está na página j carregar numa ligação que a leva à página i, e é portanto igual a $\frac{\ell(j,i)}{\ell(j)}$ onde $\ell(j,i)$ é o número de ligações que une a página j à página i e $\ell(j)$ é o número de total de ligações de j para outras páginas.[7]

Note-se que a soma das entradas em cada coluna da matriz é 1 (é a soma das probabilidades de ir parar a cada destino possível partindo da página correspondente à coluna). Uma tal matriz chama-se uma *matriz de Markov*. Estas matrizes são usadas para modelar sistemas nos quais há vários estados (em número igual à dimensão da matriz) e há certas probabilidades conhecidas de ocorrer uma transição entre os estados com a passagem do tempo.

Quando é que o número de internautas em cada página permanece constante ao longo do tempo? Quando o vetor (n_1, n_2, n_3) é um vetor próprio da matriz

$$\begin{bmatrix} 0 & \frac{1}{4} & \frac{2}{3} \\ 1 & \frac{1}{4} & 0 \\ 0 & \frac{1}{2} & \frac{1}{3} \end{bmatrix} \tag{36}$$

com valor próprio 1. Um tal vetor próprio existe necessariamente porque a soma por linhas da matriz transposta é 1, o que significa exatamente que $(1,1,1)$ é um vetor próprio da matriz transposta com valor próprio 1. Como o polinómio característico de A^T é igual ao de A (porque $\det(A) = \det(A^T)$) a matriz (36) tem também 1 como valor próprio.

Pode mostrar-se que existe necessariamente um vetor próprio de 1 com componentes todas não negativas, e (com bastante generalidade) que se normalizarmos os vetores que indicam o estado das páginas de modo a que a soma das entradas seja 1 (isto corresponde a considerar a percentagem dos internautas em cada página em vez do número absoluto)

[7] Se uma página não tem ligações para outras assume-se que tem uma ligação para cada página.

o limite quando o tempo tende para ∞ do estado do sistema é o vetor próprio de 1 (normalizado), que é único.

Mais precisamente, se A é a matriz (36) que controla a transição entre estados e (p_1, p_2, p_3) é um estado inicial qualquer (com $p_i \geq 0$ e $p_1 + p_2 + p_3 = 1$), temos

$$\lim_{k \to \infty} A^k \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} = v$$

com v o único vetor próprio de 1 com entradas não negativas cuja soma é 1. Quando isto acontece, o significado das componentes de v é clara: v_i é a percentagem do tempo que uma internauta surfando ao acaso naquelas páginas passaria na página i. É este número que é usado como medida da relevância da página i - o seu PageRank.

No exemplo acima teríamos que os vetores próprios de 1 da matriz (36) são as soluções de

$$(A - I_3)v = 0 \Leftrightarrow \begin{bmatrix} -1 & \frac{1}{4} & \frac{2}{3} \\ 1 & -\frac{3}{4} & 0 \\ 0 & \frac{1}{2} & -\frac{2}{3} \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Leftrightarrow \begin{cases} a = \frac{3}{4}b \\ c = \frac{3}{4}b \end{cases}$$

Logo um vetor próprio de 1 é um múltiplo não nulo de $(\frac{3}{4}, 1, \frac{3}{4})$. Normalizando obtemos

$$(0.3, 0.4, 0.3)$$

Pelo que a página mais relevante é a página 2, sendo as outras duas igualmente relevantes. Uma internauta surfando aleatoriamente entre estas três páginas passaria 40% do seu tempo na página 2 e 30% em cada uma das outras duas páginas.

O algoritmo utilizado pelo Google para ordenar as páginas por relevância é seguramente muito mais complicado mas o princípio básico é o que foi explicado acima. Ao pesquisarmos um termo, o algoritmo começa por selecionar as páginas relacionadas com esse termo (utilizando as etiquetas previamente atribuídas a cada página) e analisa depois as ligações entre essas páginas conforme descrito acima, listando-as depois por ordem de relevância.

Na realidade, no algoritmo original de Larry Page e Sergey Brin é também levada em conta a possibilidade de uma internauta não seguir nenhum link na página em que se encontra (e em vez disso usar um bookmark ou escrever diretamente um URL). Esta possibilidade é considerada atribuindo uma probabilidade d de ir para qualquer outra página da internet a partir de uma dada página, sendo $(1 - d)$ a probabilidade de carregar numa das ligações da página. O parâmetro d é medido experimentalmente (e é cerca de 15%). Tente descrever analiticamente este algoritmo modificado. A solução encontra-se na página da Wikipedia do algoritmo PageRank.

12. Produtos Internos

É-vos familiar do ensino secundário o *produto interno* de vetores de $\mathbb{R}^2$ e $\mathbb{R}^3$. Trata-se de uma operação que produz um número real $\langle v, w \rangle$ a partir de dois vetores v e w. É dado pelas fórmulas

$$\langle (x_1, x_2), (y_1, y_2) \rangle = x_1 y_1 + x_2 y_2 \qquad \text{para } (x_1, x_2), (y_1, y_2) \in \mathbb{R}^2$$

e

$$\langle (x_1, x_2, x_3), (y_1, y_2, y_3)\rangle = x_1y_1 + x_2y_2 + x_3y_3 \qquad \text{para } (x_1, x_2, x_3), (y_1, y_2, y_3) \in \mathbb{R}^3$$

respetivamente. Em ambos os casos, o significado geométrico, do produto interno $\langle v, w\rangle$ é $\|v\|\|w\| \cos \alpha$ em que $\|x\|$ designa o comprimento do vetor x e α é o ângulo entre v e w.

Em qualquer espaço vetorial é possível definir uma estrutura desta natureza que é completamente caracterizada por alguns axiomas simples.

Definição 12.1. *Seja V um espaço vetorial real. Um* produto interno *em V é uma função*

$$\langle \cdot, \cdot\rangle : V \times V \to \mathbb{R}$$

satisfazendo

(1) **Bilinearidade:** *Para todos os $\alpha_1, \alpha_2 \in \mathbb{R}$ e $v_1, v_2, w \in V$.*
- $\langle \alpha_1 v_1 + \alpha_2 v_2, w\rangle = \alpha_1\langle v_1, w\rangle + \alpha_2\langle v_2, w\rangle$
- $\langle w, \alpha_1 v_1 + \alpha_2 v_2\rangle = \alpha_1\langle w, v_1\rangle + \alpha_2\langle w, v_2\rangle$

(2) **Simetria:** *$\langle v, w\rangle = \langle w, v\rangle$ para todos os $v, w \in V$.*

(3) **Positividade:** *$\langle v, v\rangle > 0$ para todo o $v \neq 0$.*

Observação 12.2. *Tendo em conta a simetria de um produto interno, para verificar a bilinearidade basta verificar a primeira (ou a segunda) das igualdades que caracterizam a bilinearidade.*

Exemplo 12.3. *O produto interno usual (ou standard) em $\mathbb{R}^n$ é definido por*

$$\langle (x_1, \dots, x_n), (y_1, \dots, y_n)\rangle = x_1y_1 + x_2y_2 + \dots + x_ny_n$$

É imediato verificar que as propriedades (1)-(3) na Definição 12.1 são verificadas. Este produto interno generaliza o produto interno já conhecido nos casos em que $n = 2$ e 3.

Exemplo 12.4. *Seja $[a, b]$ um intervalo de $\mathbb{R}$ e $V = C([a, b], \mathbb{R})$ o espaço vetorial das funções contínuas de $\mathbb{R}$ para $\mathbb{R}$ (que é um subespaço vetorial do espaço vetorial de todas as funções de $\mathbb{R}$ para $\mathbb{R}$). Define-se $\langle \cdot, \cdot\rangle : V \times V \to \mathbb{R}$ pela expressão*

$$\langle f, g\rangle = \int_a^b f(x)g(x)dx$$

A expressão anterior faz sentido porque o produto de funções contínuas é contínua e uma função contínua é integrável num intervalo compacto. Verifiquemos as propriedades (1)-(3) da Definição 12.1:

(1) $\langle \alpha_1 f_1 + \alpha_2 f_2, g\rangle = \int_a^b (\alpha_1 f_1(x) + \alpha_2 f_2(x))g(x)dx = \alpha_1 \int_a^b f_1(x)g(x)dx + \alpha_2 \int_a^b f_2(x)g(x)dx = \alpha_1\langle f_1, g\rangle + \alpha_2\langle f_2, g\rangle$

(2) É imediato uma vez que $f(x)g(x) = g(x)f(x)$.

(3) $\langle f, f\rangle = \int_a^b f^2(x)dx \geq 0$ *por monotonia do integral. Se $f(x) \neq 0$ então existe $x_0 \in [a, b]$ tal que $f(x_0) \neq 0$. Como f é contínua isso significa que existe $\epsilon > 0$ e um intervalo J contendo x_0 com interior não vazio tal que $f(x)^2 \geq \epsilon$ quando $x \in J$. Mas então* $\int_a^b f(x)^2dx \geq \int_J f(x)^2dx \geq \int_J \epsilon\, dx > 0$.

Observação 12.5. *Se pensarmos numa função f como um "vetor indexado pelos números reais" cuja componente x é o número $f(x)$, e no integral como uma "soma em x" o segundo exemplo acima é uma generalização natural do primeiro.*

Existe também uma versão do conceito de produto interno para um espaço vetorial complexo, que se chama um produto interno Hermiteano, ou simplesmente um produto interno. O modelo será $\mathbb{C}^n$, mas agora não podemos usar a fórmula que nos dá o produto interno real porque perderíamos a positividade (que é a chave para definir o comprimento de vetores). A solução é conjugar um dos argumentos coordenada a coordenada, uma vez que $\overline{z}z = |z|^2 \geq 0$. No entanto, isso afeta necessariamente os outros dois axiomas da forma seguinte.

Definição 12.6. *Seja V um espaço vetorial complexo. Um* produto interno *em V é uma função*

$$\langle \cdot, \cdot \rangle \colon V \times V \to \mathbb{C}$$

satisfazendo

(1) **Sesquilinearidade:** *Para todos os $\alpha_1, \alpha_2 \in \mathbb{C}$ e $v_1, v_2, w \in V$.*
- $\langle \alpha_1 v_1 + \alpha_2 v_2, w \rangle = \overline{\alpha_1}\langle v_1, w\rangle + \overline{\alpha_2}\langle v_2, w\rangle$
- $\langle w, \alpha_1 v_1 + \alpha_2 v_2 \rangle = \alpha_1\langle w, v_1\rangle + \alpha_2\langle w, v_2\rangle$

(2) **Simetria conjugada:** $\langle v, w\rangle = \overline{\langle w, v\rangle}$ *para todos os $v, w \in V$.*
(3) **Positividade:** $\langle v, v\rangle$ *é real e positivo para todo o $v \neq 0$.*

Observação 12.7. *Tendo em conta a simetria conjugada de um produto interno, para verificar a sesquilinearidade basta verificar a primeira (ou a segunda) das igualdades que caracterizam a sesquilinearidade.*

Exemplo 12.8. *O produto interno standard em $\mathbb{C}^n$ é a função $\langle \cdot, \cdot \rangle \colon \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$ definida pela expressão*

$$\langle (z_1, \ldots, z_n), (w_1, \ldots, w_n) \rangle = \overline{z_1}w_1 + \overline{z_2}w_2 + \ldots + \overline{z_n}w_n$$

É imediato verificar as condições (1)-(3) da Definição 12.6. Por exemplo,

$$\langle (z_1, \ldots, z_n), (z_1, \ldots, z_n) \rangle = |z_1|^2 + \ldots + |z_n|^2 \geq 0$$

e só se anula se $z_1 = \cdots = z_n = 0$.

Um produto interno num espaço vetorial real ou complexo permite-nos introduzir noções de comprimento e distância no espaço em questão.

Definição 12.9. *Seja V um espaço vetorial e $\langle \cdot, \cdot \rangle$ um produto interno em V. A* norma *ou* comprimento *de um vetor $v \in V$ é o número real não negativo $\|v\| = \sqrt{\langle v, v\rangle}$. Sendo $v, w \in V$, a* distância de v a w *é o número real não negativo $\|v - w\|$.*

Note-se que as noções de norma e comprimento para o produto interno usual em $\mathbb{R}^2$ ou $\mathbb{R}^3$ são as habituais:

$$\|(x, y, z)\| = \sqrt{x^2 + y^2 + z^2}$$

Exemplo 12.10. *Em $\mathbb{C}^2$ com o produto interno usual,*

$$\|(1+i,-1)\| = \sqrt{|1+i|^2+1} = \sqrt{2+1} = \sqrt{3}$$

Em $C([0,1],\mathbb{R}^2)$ a distância de x a 1 é

$$\|x-1\| = \sqrt{\int_0^1 (x-1)^2 dx} = \sqrt{\left.\frac{(x-1)^3}{3}\right|_0^1} = \frac{1}{\sqrt{3}}$$

Definição 12.11. *Seja V um espaço vetorial e $\langle \cdot,\cdot \rangle$ um produto interno em V. Um subconjunto $S \subset V$ diz-se* ortogonal *se $\langle v,w\rangle = 0$ para todos os $v,w \in S$ distintos. Um subconjunto $S \subset V$ diz-se* ortonormado *se S é ortogonal e $\|v\| = 1$ para todo o $v \in S$.*

Exemplo 12.12. *O conjunto $\{(1,1),(1,-1)\}$ é ortogonal em $\mathbb{R}^2$ para o produto interno usual, uma vez que $\langle (1,1),(1-1)\rangle = 1-1 = 0$. Não é ortonormado uma vez que $\|(1,1)\| = \sqrt{2} \neq 1$, mas dividindo cada um dos vetores pelo seu comprimento obtemos o conjunto ortonormado $\{\left(\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}}\right),\left(\frac{1}{\sqrt{2}},-\frac{1}{\sqrt{2}}\right)\}$.*

As funções $\operatorname{sen} x$ *e 1 são ortogonais em $C([0,2\pi],\mathbb{R})$ uma vez que*

$$\langle \operatorname{sen} x, 1\rangle = \int_0^{2\pi} \operatorname{sen} x dx = -\cos x|_0^{2\pi} = 0$$

As bases canónicas de $\mathbb{R}^n$ e $\mathbb{C}^n$ são conjuntos ortonormados para os produtos internos usuais.

12.13. **Representação matricial de um produto interno.** Seja V um espaço vetorial com produto interno $\langle \cdot,\cdot\rangle$ e suponhamos que $B = (v_1,\ldots,v_n)$ é uma base para V.

Podemos escrever dois vetores $v,w \in V$ em função da base B

$$v = \alpha_1 v_1 + \ldots + \alpha_n v_n, \qquad w = \beta_1 v_1 + \ldots + \beta_n v_n,$$

Vamos agora usar a bilinearidade/sesquilinearidade para obter uma fórmula para o produto interno em termos do produto de matrizes. Consideraremos o caso complexo mas note-se que, uma vez que para α real temos $\overline{\alpha} = \alpha$, estamos também a fazer o caso real simultaneamente. Usando linearidade conjugada na primeira variável temos

$$\langle v,w\rangle = \langle \alpha_1 v_1 + \ldots \alpha_n v_n, w\rangle = \overline{\alpha_1}\langle v_1,w\rangle + \ldots + \overline{\alpha_n}\langle v_n,w\rangle$$

Usando a linearidade na segunda coordenada temos para cada i

$$\langle v_i,w\rangle = \langle v_i, \beta_1 v_1 + \ldots \beta_n v_n\rangle = \beta_1\langle v_i,v_1\rangle + \ldots + \beta_n\langle v_i,v_n\rangle$$

e substituindo na primeira expressão obtemos a seguinte expressão para o produto interno

$$\langle v,w\rangle = \overline{\alpha_1}\beta_1\langle v_1,v_1\rangle + \ldots + \overline{\alpha_1}\beta_n\langle v_1,v_n\rangle + \overline{\alpha_2}\beta_1\langle v_2,v_1\rangle + \ldots + \overline{\alpha_n}\beta_1\langle v_n,v_1\rangle + \ldots + \overline{\alpha_n}\beta_n\langle v_n,v_n\rangle$$

Vemos assim que o produto interno é completamente determinado pelo conjunto de n^2 escalares $\langle v_i,v_j\rangle$ com $i,j = 1,\ldots,n$. Identificando escalares com matrizes 1×1 a expressão anterior pode ser escrita matricialmente na forma

$$\begin{bmatrix} \overline{\alpha_1} & \overline{\alpha_2} & \cdots & \overline{\alpha_n} \end{bmatrix} \begin{bmatrix} \langle v_1, v_1\rangle & \langle v_1, v_2\rangle & \cdots & \langle v_1, v_n\rangle \\ \langle v_2, v_1\rangle & \langle v_2, v_2\rangle & \cdots & \langle v_2, v_n\rangle \\ \vdots & \vdots & & \vdots \\ \langle v_n, v_1\rangle & \langle v_n, v_2\rangle & \cdots & \langle v_n, v_n\rangle \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} \tag{37}$$

A matriz $n \times n$ na expressão (37) chama-se a *matriz da métrica* ou *matriz de Gram* para o produto interno com respeito à base B, e iremos denotá-la por

$$G_B = [\langle v_i, v_j\rangle]$$

Note-se que para chegar à expressão (37) usámos apenas a propriedade (1) das Definições 12.1 e 12.6 pelo que a expressão matricial (37) se aplica a funções de $V \times V$ para os escalares que satisfaçam apenas o axioma (1) (ditas funções bilineares no caso real, e sesquilineares no caso complexo). As propriedades (2) e (3) das definições impõem condições adicionais sobre a matriz G_B.

Quanto à condição (2), escrevendo g_{ij} para a entrada ij da matriz G_B, temos no caso real

$$g_{ij} = \langle v_i, v_j\rangle = \langle v_j, v_i\rangle = g_{ji} \quad \Leftrightarrow \quad G_B = G_B^T$$

ou seja, a matriz da métrica é *simétrica.* No caso complexo temos

$$g_{ij} = \langle v_i, v_j\rangle = \overline{\langle v_j, v_i\rangle} = \overline{g_{ji}} \quad \Leftrightarrow \quad G_B = \overline{G_B}^T$$

Diz-se que a matriz G_B é *hermiteana.* Reciprocamente, se G é uma matriz que satisfaz estas condições é imediato verificar que a função

$$\langle v, w\rangle = \overline{[v]_B}^T G_B [w]_B$$

satisfaz as condições (1) e (2) nas definições 12.1 e 12.6.

Veremos em breve que as matrizes simétricas ou hermiteanas são sempre diagonalizáveis com valores próprios reais. Daí segue facilmente da condição (3) que os valores próprios de uma matriz da métrica têm que ser positivos: se $G_B v = \lambda v$ então

$$\langle v, v\rangle = \overline{v}^T G_B v = \lambda \overline{v}^T v > 0 \quad \Rightarrow \lambda > 0$$

Por outro lado veremos que os vetores próprios correspondentes a valores próprios distintos de uma tal matriz são ortogonais e então é fácil de ver que, reciprocamente, uma matriz simétrica ou hermiteana com valores próprios positivos determina um produto interno em $\mathbb{R}^n$ ou $\mathbb{C}^n$ respetivamente.

Exemplo 12.14. *Consideremos a restrição do produto interno usual em $\mathbb{R}^3$ ao subespaço $V = \{(x, y, z) \in \mathbb{R}^3 \colon x + y + z = 0\}$. Uma base para V é dada, por exemplo, pelos vetores $v_1 = (1, -1, 0)$ e $v_2 = (0, 1, -1)$. A matriz da métrica para o produto interno em V com respeito à base $B = (v_1, v_2)$ é portanto*

$$G_B = \begin{bmatrix} \langle v_1, v_1\rangle & \langle v_1, v_2\rangle \\ \langle v_2, v_1\rangle & \langle v_2, v_2\rangle \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

Dados vetores $v, w \in V$ *com* $[v]_B = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ *e* $[w]_B = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ *temos*

$$\langle v, w \rangle = \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} -3 \\ 3 \end{bmatrix} = 3$$

Podemos confirmar este resultado fazendo as contas em $\mathbb{R}^3$*: Temos*

$v = 1 \cdot (1, -1, 0) + 2(0, 1, -1) = (1, 1, -2), \quad w = -1 \cdot (1, -1, 0) + 1 \cdot (0, 1, -1) = (-1, 2, -1)$

logo

$$\langle v, w \rangle = 1 \cdot (-1) + 1 \cdot 2 + (-2) \cdot (-1) = -1 + 2 + 2 = 3.$$

O ponto do exemplo anterior é o seguinte. Mesmo que estejamos interessados apenas no produto interno usual em $\mathbb{R}^n$ (isto é na noção usual de comprimento e ângulo) em certas situações estaremos interessados em considerar apenas vetores que estão em certos subespaços (imaginemos por exemplo que um avião voa num dado plano) e para fazer contas nesse plano é mais prático escolher coordenadas no plano (da mesma forma que à superfície da Terra utilizamos duas coordenadas para descrever um ponto). No plano não há em geral coordenadas canónicas como em $\mathbb{R}^n$ e numas coordenadas arbitrárias que escolhamos, a expressão do produto interno não será aquela a que estamos acostumados, mesmo que o produto interno em questão provenha do produto interno usual em $\mathbb{R}^n$.

Observação 12.15. *Note-se que uma base* B *para um espaço vetorial* V *é ortogonal com respeito a um produto interno sse a matriz da métrica* G_B *é diagonal (e então as entradas diagonais são positivas e iguais às normas dos vetores da base ao quadrado) e que* B *é ortonormada (isto é um conjunto ortonormado) sse* G_B *é a matriz identidade.*

Suponhamos agora que B, B' são duas bases para o espaço vetorial V com produto interno. Como se relacionam as matrizes da métrica com respeito às duas bases?

Sendo $S = S_{B \to B'}$ a matriz de mudança de coordenadas da base B para a base B' temos para qualquer $x \in V$

$$[x]_{B'} = S[x]_B$$

substituindo na expressão para a matriz da métrica na base B' temos (novamente o caso real obtém-se omitindo os conjugados)

$$\langle v, w \rangle = \overline{[v]_{B'}}^T G_{B'} [w]_{B'} = \overline{S[v]_B}^T G_{B'} (S[w]_B) = \overline{[v]_B}^T \overline{S}^T G_{B'} S [w]_B$$

onde usámos que $\overline{AB} = \overline{A}\,\overline{B}$ e $(AB)^T = B^T A^T$. Tendo em conta a expressão

$$\langle v, w \rangle = \overline{[v]_B}^T G_B [w]_B$$

que caracteriza a matriz da métrica com respeito à base B conclui-se que

$$G_B = \overline{S}^T G_{B'} S \quad \text{ou, no caso real,} \quad G_B = S^T G_{B'} S \tag{38}$$

Estas fórmulas que traduzem como a expressão para o produto interno muda mediante uma mudança de coordenadas são inteiramente análogas à fórmula (19) que relaciona as expressões matriciais de uma transformação linear em relação a bases distintas (mas só coincidem se $S^{-1} = \overline{S}^T$).

12.16. **Projeção ortogonal de um vetor sobre um vetor não nulo.** Seja V um espaço vetorial com produto interno, $v \in V$ e $u \in V \setminus \{0\}$ um vetor não nulo. Define-se a *projeção ortogonal de v sobre u* (com respeito ao produto interno dado) por

$$\operatorname{proj}_u(v) = \langle u, v\rangle \frac{u}{\|u\|^2} = \frac{\langle u, v\rangle}{\langle u, u\rangle} u = \langle \frac{u}{\|u\|}, v\rangle \frac{u}{\|u\|} \tag{39}$$

As expressões acima são todas iguais pela definição de norma e pela linearidade na primeira variável (no caso complexo note-se que o escalar $\frac{1}{\|u\|}$ é real e portanto igual ao seu conjugado).

Quando $V = \mathbb{R}^2$ ou $\mathbb{R}^3$ com o produto interno usual, a definição anterior coincide com a noção de projeção ortogonal já estudada no ensino secundário. De facto o vetor $\frac{u}{\|u\|}$ é um versor da direção determinada por u (isto é, tem a mesma direção e sentido e comprimento 1). O escalar que multiplica este versor é

$$\langle \frac{u}{\|u\|}, v\rangle = \left\| \frac{u}{\|u\|} \right\| \|v\| \cos\alpha = 1 \cdot \|v\| \cos\alpha = \|v\| \cos\alpha$$

com α o ângulo entre u e v, pelo que a expressão 39 é, neste caso, a expressão familiar do ensino secundário.

Exemplo 12.17. *A projeção ortogonal de* $(1, -1, 2)$ *sobre o vetor* $(0, 1, 1)$ *com respeito ao produto interno usual em* $\mathbb{R}^3$ *é*

$$\frac{\langle (1, -1, 2), (0, 1, 1)\rangle}{\langle (0, 1, 1), (0, 1, 1)\rangle}(0, 1, 1) = \tfrac{1}{2}(0, 1, 1) = \left(0, \tfrac{1}{2}, \tfrac{1}{2}\right)$$

Note-se que $\operatorname{proj}_u(v)$ é colinear com u e que a projeção permite escrever um vetor qualquer v como a soma de um vetor colinear com u e outro ortogonal a u:

$$v = (v - \operatorname{proj}_u(v)) + \operatorname{proj}_u(v)$$

sendo que

$$\langle u, v - \operatorname{proj}_u(v)\rangle = \langle u, v - \frac{\langle u, v\rangle}{\langle u, u\rangle} u\rangle = \langle u, v\rangle - \frac{\langle u, v\rangle}{\langle u, u\rangle}\langle u, u\rangle = 0$$

Na realidade é fácil verificar que a função $P \colon V \to V$ definida por $P(v) = \operatorname{proj}_u(v)$ é uma projeção (com imagem a reta gerada por u), isto é, que $P^2 = P$ o mesmo acontecendo portanto com $\mathrm{Id} - P$ que calcula a componente de v ortogonal a u (ver o Exercício 9 da Ficha 11).

Da consideração da componente ortogonal a um vetor u, vêm duas desigualdades fundamentais.

Proposição 12.18. *Seja V um espaço vetorial com produto interno $\langle \cdot, \cdot\rangle$, e $u, v \in V$. Então*

(i) **Desigualdade de Cauchy-Schwarz:** $|\langle u, v\rangle| \leq \|u\|\|v\|$
(ii) **Desigualdade triangular:** $\|u + v\| \leq \|u\| + \|v\|$

A igualdade verifica-se na primeira desigualdade se e só se u e v são colineares.

Dem. (i) Podemos assumir sem perda de generalidade que $u \neq 0$ (pois nesse caso $0 = |\langle u, v\rangle| = \|u\|\|v\|$ e u, v são colineares). Nesse caso temos, pela positividade do produto interno

$$\begin{aligned} 0 \leq \|v - \operatorname{proj}_u(v)\|^2 &= \langle v - \frac{\langle u,v\rangle}{\langle u,u\rangle}u, v - \frac{\langle u,v\rangle}{\langle u,u\rangle}u\rangle \\ &= \langle v,v\rangle - \frac{\overline{\langle u,v\rangle}}{\langle u,u\rangle}\langle u,v\rangle - \frac{\langle u,v\rangle}{\langle u,u\rangle}\langle v,u\rangle + \frac{\overline{\langle u,v\rangle}}{\langle u,u\rangle}\frac{\langle u,v\rangle}{\langle u,u\rangle}\langle u,u\rangle \\ &= \|v\|^2 - \frac{|\langle u,v\rangle|^2}{\|u\|^2} \end{aligned}$$

e esta desigualdade é equivalente a

$$|\langle u,v\rangle|^2 \leq \|u\|^2\|v\|^2$$

que, tomando raízes quadradas, é a desigualdade de Cauchy-Schwarz. A igualdade verifica-se apenas quando $v - \operatorname{proj}_u(v) = 0$ e nesse caso v é um múltiplo escalar de u.

(ii) Temos

$$\|u+v\|^2 = \langle u+v, u+v\rangle = \langle u,u\rangle + \langle u,v\rangle + \langle v,u\rangle + \langle v,v\rangle \tag{40}$$

Uma vez que $z + \overline{z} = 2\operatorname{Re}(z) \leq 2|z|$ temos

$$\langle u,v\rangle + \langle v,u\rangle = 2\operatorname{Re}(\langle u,v\rangle) \leq 2|\langle u,v\rangle| \leq 2\|u\|\|v\|$$

onde na segunda desigualdade aplicámos a desigualdade de Cauchy-Schwarz. Substituindo em (40) obtemos

$$\|u+v\|^2 \leq \|u\|^2 + 2\|u\|\|v\| + \|v\|^2 = (\|u\| + \|v\|)^2$$

que é equivalente à desigualdade triangular.

□

Observação 12.19. *(i) A desigualdade triangular chama-se assim porque $v, w, v+w$ formam as arestas de um triângulo em V e a desigualdade diz precisamente que o comprimento de um dos lados de um triângulo é sempre menor ou igual à soma do comprimento dos dois outros lados.*

(ii) Quando u, v são ortogonais, a expressão (40) *é o Teorema de Pitágoras: $\|u+v\|^2 = \|u\|^2 + \|v\|^2$.*

Definição 12.20. *Seja V um espaço vetorial* real *e $v, w \in V$ vetores não nulos. Define-se o* ângulo *entre v e w como o único $\alpha \in [0, \pi]$ tal que*

$$\cos\alpha = \frac{\langle v,w\rangle}{\|v\|\|w\|}$$

(Isto faz sentido porque, pela desigualdade de Cauchy-Schwarz a expressão do lado direito do sinal de igual pertence ao intervalo $[-1, 1]$.)

Exemplo 12.21. *O ângulo entre as funções x e x^2 em $C([0,1],\mathbb{R})$ é*

$$\arccos \frac{\langle x, x^2\rangle}{\|x\|\|x^2\|} = \frac{\int_0^1 x^3 dx}{\sqrt{\int_0^1 x^2 dx \int_0^1 x^4 dx}} = \arccos \frac{\frac{1}{4}}{\sqrt{\frac{1}{3}\frac{1}{5}}} = \arccos \frac{\sqrt{15}}{4}$$

A projeção segundo um vetor dá-nos uma maneira simples de obter um conjunto ortogonal com a mesma expansão linear a partir de um conjunto linearmente independente e, em particular, de obter uma base ortogonal a partir de uma base.

Proposição 12.22 (Método de ortogonalização de Gram-Schmidt). *Seja V um espaço vetorial com produto interno e $\{v_1,\ldots,v_k\} \subset V$ um conjunto linearmente independente. Então os vetores definidos indutivamente pelas fórmulas*

$$\begin{aligned} w_1 &= v_1 \\ w_2 &= v_2 - \operatorname{proj}_{w_1}(v_2) \\ w_3 &= v_3 - \operatorname{proj}_{w_1}(v_3) - \operatorname{proj}_{w_2}(v_3) \\ \vdots \quad & \quad \vdots \\ w_k &= v_k - \operatorname{proj}_{w_1}(v_k) - \ldots - \operatorname{proj}_{w_{k-1}}(v_k) \end{aligned}$$

formam um conjunto ortogonal $\{w_1,\ldots,w_k\}$ tal que, para cada $i = 1,\ldots,k$, temos

$$L(\{v_1,\ldots,v_i\}) = L(\{w_1,\ldots,w_i\})$$

Antes de vermos a demonstração vamos fazer algumas observações.

Definição 12.23. *Seja V um espaço vetorial com um produto interno e $S \subset V$ um subconjunto. Define-se*

$$S^\perp = \{v \in V : \langle v, x\rangle = 0 \text{ para todo o } x \in S\}$$

É imediato verificar que $S^\perp$ é um subespaço vetorial de V: claramente $0 \in S^\perp$ e se $v_1, v_2 \in S^\perp$ e $\alpha_1,\alpha_2 \in \mathbb{R}$ temos $\langle \alpha_1 v_1 + \alpha_2 v_2, x\rangle = \alpha_1\langle v_1, x\rangle + \alpha_2\langle v_2, x\rangle = 0$ para todo o $x \in S$, pelo que $\alpha_1 v_1 + \alpha_2 v_2 \in S^\perp$.

Proposição 12.24. $S^\perp = L(S)^\perp$

Dem. Uma vez que $S \subset L(S)$, é evidente que $L(S)^\perp \subset S^\perp$ (se um vetor é ortogonal a todos os elementos de $L(S)$, certamente é também ortogonal a todos os vetores de S). Reciprocamente, se $v \in L(S)$, existem vetores $v_1,\ldots,v_k$ em S e escalares $\alpha_1,\ldots,\alpha_k$ tais que $v = \alpha_1 v_1 + \ldots + \alpha_k v_k$. Dado $w \in S^\perp$, temos

$$\langle w, v\rangle = \langle w, \alpha_1 v_1 + \ldots + \alpha_k v_k\rangle = \alpha_1\langle w, v_1\rangle + \ldots + \alpha_k\langle w, v_k\rangle = 0$$

Logo $w \in L(S)^\perp$. Isso mostra que $S^\perp \subset L(S)^\perp$ e conclui a demonstração. □

Exemplo 12.25. *(i) Se $A \in M_{m\times n}(\mathbb{R})$ então $N(A) = EL(A)^\perp \subset \mathbb{R}^n$ (onde o produto interno considerado é o usual). De facto, pela definição do produto de matrizes, $x \in \mathbb{R}^n$ está no núcleo de A sse é ortogonal às linhas de A para o produto interno usual em $\mathbb{R}^n$, e pela Proposição anterior isto é o mesmo que ser ortogonal ao espaço das linhas.*

(ii) Se B é uma base de V (ou mais geralmente um conjunto de geradores) então $B^\perp = \{0\}$. De facto, $B^\perp = L(B)^\perp = V^\perp$. Mas a positividade do produto interno diz-nos que o único vetor que é perpendicular a si próprio é o vetor 0. Logo $V^\perp = \{0\}$.

Dem. da Proposição 12.22. Vamos usar indução em i para ver que $\{w_1, \ldots, w_i\}$ é um conjunto ortogonal e $L(\{v_1, \ldots, v_i\}) = L(\{w_1, \ldots, w_i\})$. A base da indução é o caso $i = 1$, que é óbvio porque um conjunto com um único vetor não nulo é um conjunto ortogonal e, por definição, $w_1 = v_1$.

Seja $i > 1$ e assumamos por indução que o resultado é válido para $i - 1$. Vejamos primeiro que $L(\{v_1, \ldots, v_i\}) = L(\{w_1, \ldots, w_i\})$. Temos que verificar duas inclusões

- Por hipótese de indução $v_1, \ldots, v_{i-1} \in L(\{w_1, \ldots, w_{i-1}\}) \subset L(\{w_1, \ldots, w_i\})$. Uma vez que $\operatorname{proj}_u(v)$ é um múltiplo de u, a seguinte reformulação da definição de w_i
$$v_i = w_i + \operatorname{proj}_{w_1}(v_i) + \ldots + \operatorname{proj}_{w_{i-1}}(v_i)$$
mostra que $v_i \in L(\{w_1, \ldots, w_i\})$. Conclui-se que $L(\{v_1, \ldots, v_i\}) \subset L(\{w_1, \ldots, w_i\})$
- Novamente, por hipótese de indução, temos $L(\{w_1, \ldots, w_{i-1}\}) \subset L(\{v_1, \ldots, v_i\})$. Na expressão para w_i
$$w_i = v_i - \operatorname{proj}_{w_1}(v_i) - \ldots - \operatorname{proj}_{w_{i-1}}(v_i)$$
os termos precedidos por um sinal menos formam uma combinação linear de $w_1, \ldots, w_{i-1}$ e portanto, por hipótese de indução, de $v_1, \ldots, v_{i-1}$. Conclui-se que $w_i \in L(\{v_1, \ldots, v_i\})$ e portanto que $L(\{w_1, \ldots, w_i\}) \subset L(\{v_1, \ldots, v_i\})$.

Para ver que $\{w_1, \ldots, w_i\}$ é um conjunto ortogonal basta-nos ver que $\langle w_j, w_i \rangle = 0$ para $j < i$ pois a hipótese de indução diz-nos que $\langle w_j, w_l \rangle = 0$ para $j \neq l$ quando $j, l < i$. Ora

$$\begin{aligned}
\langle w_j, w_i \rangle &= \langle w_j, v_i - \operatorname{proj}_{w_1}(v_i) - \ldots - \operatorname{proj}_{w_{i-1}}(v_i) \rangle \\
&= \langle w_j, v_i \rangle - \langle w_j, \frac{\langle w_1, v_i \rangle}{\langle w_1, w_1 \rangle} w_1 \rangle - \ldots - \langle w_j, \frac{\langle w_{i-1}, v_i \rangle}{\langle w_{i-1}, w_{i-1} \rangle} w_{i-1} \rangle \\
&= \langle w_j, v_i \rangle - \frac{\langle w_1, v_i \rangle}{\langle w_1, w_1 \rangle} \langle w_j, w_1 \rangle - \ldots - \frac{\langle w_{i-1}, v_i \rangle}{\langle w_{i-1}, w_{i-1} \rangle} \langle w_j, w_{i-1} \rangle
\end{aligned}$$

Do lado direito do sinal de igual, novamente pela hipótese de indução que $\{w_1, \ldots, w_{i-1}\}$ é ortogonal, o único termo $\langle w_j, w_k \rangle$ que é não nulo é o termo correspondente a $k = j$ portanto

$$\langle w_j, w_i \rangle = \langle w_j, v_i \rangle - 0 - \ldots - \frac{\langle w_j, v_i \rangle}{\langle w_j, w_j \rangle} \langle w_j, w_j \rangle - \ldots - 0 = \langle w_j, v_i \rangle - \langle w_j, v_i \rangle = 0$$

o que conclui a demonstração. □

Exemplo 12.26. *Vamos achar uma base ortonormada para o subespaço*

$$V = \{(x, y, z, w) \in \mathbb{R}^4 \colon x + y + w = 0\} \subset \mathbb{R}^4$$

Uma base para este subespaço é por exemplo

$$\{(1, 0, 0, -1), (0, 1, 0, -1), (0, 0, 1, 0)\}$$

Vamos aplicar o processo de ortogonalização de Gram-Schmidt dividindo os vetores resultantes pelas suas normas para obter uma base ortonormada.

O primeiro vetor da base ortonormada será

$$w_1 = \frac{(1,0,0,-1)}{\|(1,0,0,-1)\|} = (\tfrac{1}{\sqrt{2}}, 0, 0, -\tfrac{1}{\sqrt{2}})$$

Obtemos um vetor ortogonal através da expressão

$$w_2 = (0,1,0,-1) - \langle w_1, (0,1,0,-1)\rangle w_1 = (0,1,0,-1) - \tfrac{1}{\sqrt{2}}(\tfrac{1}{\sqrt{2}}, 0, 0, -\tfrac{1}{\sqrt{2}}) = (-\tfrac{1}{2}, 1, 0, -\tfrac{1}{2})$$

Na expressão anterior não foi necessário dividir por $\langle w_1, w_1\rangle$ porque $\|w_1\| = 1$. Dividindo pela norma obtemos o segundo vetor da base ortonormada

$$\tilde{w}_2 = \frac{1}{\|w_2\|}(-\tfrac{1}{2}, 1, 0, -\tfrac{1}{2}) = \sqrt{\tfrac{2}{3}}(-\tfrac{1}{2}, 1, 0, -\tfrac{1}{2})$$

O vetor $v_3 = (0,0,1,0)$ já é ortogonal a w_1 e $\tilde{w}_2$ e tem norma 1, pelo que podemos tomar para base ortonormada de V o conjunto

$$\{(\tfrac{1}{\sqrt{2}}, 0, 0, -\tfrac{1}{\sqrt{2}}), (-\tfrac{1}{\sqrt{6}}, \sqrt{\tfrac{2}{3}}, 0, -\tfrac{1}{\sqrt{6}}), (0,0,1,0)\}$$

As bases ortogonais são extremamente úteis porque tornam os cálculos muito mais fáceis. Começamos por observar que um conjunto ortogonal sem vetores nulos é necessariamente linearmente independente

Proposição 12.27. *Seja V um espaço vetorial com produto interno e $S \subset V \setminus \{0\}$ um conjunto ortogonal de vetores não nulos. Então S é linearmente independente.*

Dem. Sejam $v_1, \ldots, v_k$ elementos de S e suponhamos que

$$\alpha_1 v_1 + \ldots + \alpha_k v_k = 0 \tag{41}$$

Queremos ver que os coeficientes α_i são todos nulos. Como S é ortogonal temos $\langle v_i, v_j\rangle = 0$ para $i \neq j$. Fazendo o produto interno da equação com v_i obtemos

$$\langle v_i, \alpha_1 v_1 + \ldots + \alpha_k v_k\rangle = \langle v_i, 0\rangle = 0$$

Do lado esquerdo temos

$$\alpha_1\langle v_i, v_1\rangle + \ldots + \alpha_i\langle v_i, v_i\rangle + \ldots + \alpha_k\langle v_i, v_k\rangle = \alpha_1 \cdot 0 + \ldots + \alpha_i\|v_i\|^2 + \ldots + \alpha_k \cdot 0$$

Portanto $\alpha_i\|v_i\|^2 = 0$. Como $v_i \neq 0$, conclui-se que $\alpha_i = 0$. □

O resultado seguinte, embora muito simples, é uma das principais razões para a utilização de bases ortogonais ou ortonormais. Juntamente com as noções de valor e vetor próprio será um dos resultados de Álgebra Linear que mais vezes será utilizado em cadeiras de engenharia. Diz essencialmente que é muito fácil calcular as coordenadas de um vetor numa base ortogonal. Não é necessário resolver um sistema linear, basta fazer uma conta muito simples.

Proposição 12.28. *Seja $B = (v_1, \dots, v_n)$ uma base ortogonal para o espaço com produto interno V. Então dado $v \in V$ as coordenadas de v na base B são dadas pela expressão*

$$[v]_B = \begin{bmatrix} \frac{\langle v_1, v\rangle}{\langle v_1, v_1\rangle} \\ \vdots \\ \frac{\langle v_n, v\rangle}{\langle v_n, v_n\rangle} \end{bmatrix}$$

Dem. Sendo $v \in V$, temos a mostrar que

$$v = \frac{\langle v_1, v\rangle}{\langle v_1, v_1\rangle} v_1 + \dots + \frac{\langle v_n, v\rangle}{\langle v_n, v_n\rangle} v_n \Leftrightarrow v - \frac{\langle v_1, v\rangle}{\langle v_1, v_1\rangle} v_1 - \dots - \frac{\langle v_n, v\rangle}{\langle v_n, v_n\rangle} v_n = 0$$

De acordo com o Exemplo 12.25(ii) basta ver que o vetor do lado esquerdo da segunda igualdade é ortogonal aos elementos da base B. Ora

$$\begin{aligned} \langle v_i, v - \frac{\langle v_1, v\rangle}{\langle v_1, v_1\rangle} v_1 - \dots - \frac{\langle v_n, v\rangle}{\langle v_n, v_n\rangle} v_n\rangle &= \langle v_i, v\rangle - \frac{\langle v_1, v\rangle}{\langle v_1, v_1\rangle}\langle v_i, v_1\rangle - \dots - \frac{\langle v_n, v\rangle}{\langle v_n, v_n\rangle}\langle v_i, v_n\rangle \\ &= \langle v_i, v\rangle - 0 - \dots - \frac{\langle v_i, v\rangle}{\langle v_i, v_i\rangle}\langle v_i, v_i\rangle - \dots - 0 \\ &= \langle v_i, v\rangle - \langle v_i, v\rangle = 0 \end{aligned}$$

□

Exemplo 12.29. *Numa base ortonormada as contas da Proposição anterior são ainda mais simples porque os denominadores das expressões para as coordenadas são 1. Considerando a base ortonormada*

$$B = \left((\tfrac{1}{\sqrt{2}}, 0, 0, -\tfrac{1}{\sqrt{2}}), (-\tfrac{1}{\sqrt{6}}, \sqrt{\tfrac{2}{3}}, 0, -\tfrac{1}{\sqrt{6}}), (0, 0, 1, 0) \right)$$

do Exemplo 12.26 e o vetor $(1, 1, 3, -2) \in V$ temos

$$[(1,1,3,-2)]_B = \begin{bmatrix} \langle (\frac{1}{\sqrt{2}}, 0, 0, -\frac{1}{\sqrt{2}}), (1,1,3,-2)\rangle \\ \langle (-\frac{1}{\sqrt{6}}, \sqrt{\frac{2}{3}}, 0, -\frac{1}{\sqrt{6}}), (1,1,3,-2)\rangle \\ \langle (0,0,1,0), (1,1,3,-2)\rangle \end{bmatrix} = \begin{bmatrix} \frac{3}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} + \sqrt{\frac{2}{3}} \\ 3 \end{bmatrix}$$

Uma base ortogonal para um subespaço pode ser usada para definir a projeção ortogonal nesse subespaço.

Definição 12.30. *Seja V um espaço vetorial com produto interno e $U \subset V$ um subespaço finitamente gerado. A* projeção ortogonal *de V em U é a transformação linear $P_U \colon V \to V$ definida pela fórmula*

$$P_U(v) = \operatorname{proj}_{u_1}(v) + \dots + \operatorname{proj}_{u_k}(v) \tag{42}$$

onde $\{u_1, \dots, u_k\}$ é uma base ortogonal de U.

P_U é uma transformação linear pois é uma soma de transformações lineares. Não é no entanto imediatamente óbvio que a fórmula (42) seja independente da escolha da base ortogonal para o subespaço U. Isso é uma consequência do seguinte resultado.

Proposição 12.31. *Seja V um espaço com produto interno e U um subespaço vetorial finitamente gerado. A transformação linear $P_U \colon V \to V$ definida pela expressão* (42) *verifica*

(1) $P_U^2 = P_U$ (ou seja, P_U é uma projeção).

(2) $P_U(V) = U$ e $N(P_U) = U^\perp$.

Portanto $V = U \oplus U^\perp$ (isto é $V = U + U^\perp$ e $U \cap U^\perp = \{0\}$) sendo a decomposição única de um vetor de V em vetores de U e $U^\perp$ dada pela expressão

$$v = \overbrace{P_U(v)}^{\in U} + \overbrace{(v - P_U(v))}^{\in U^\perp}$$

Dem. Exercício. $\square$

A Proposição anterior mostra que P_U é independente da escolha da base ortogonal para U que aparece na fórmula 42 uma vez que uma projeção é completamente determinada pela sua imagem e o seu núcleo. De facto, vimos na aula prática (exercício 9 da Ficha 11) que uma transformação linear é uma projeção sse é diagonalizável e tem a sua imagem como espaço próprio de 1.

Uma aplicação interessante da Proposição anterior é o cálculo da distância de um ponto x de um espaço vetorial com produto interno V a um subespaço U de V. A distância é por definição

$$d(x, U) = \inf\{\|x - u\| \colon u \in U\}$$

sendo que o ínfimo existe porque o conjunto das distâncias é limitado inferiormente (por 0). Dado um ponto qualquer $u \in U$ podemos escrever o vetor $x - u$ como

$$x - u = (x - P_U(x)) + (P_U(x) - u)$$

uma vez que $x - P_U(x) \in U^\perp$ e $P_U(x) - u \in U$, pelo Teorema de Pitágoras, temos

$$\|x - u\|^2 = \|x - P_U(x)\|^2 + \|P_U(x) - u\|^2 \geq \|x - P_U(x)\|^2 \Leftrightarrow \|x - u\| \geq \|x - P_U(x)\|$$

Uma vez que $P_U(x) \in U$, isso mostra que $d(x, U) = \|x - P_U(x)\|$ e, portanto, que $P_U(x)$ **é o ponto de U mais próximo de** x.

Este mesmo argumento pode facilmente ser adaptado para calcular distâncias de pontos a planos $v + U$ que não passam pela origem ou a distância entre planos que não se intersetem. A distância entre dois conjuntos $S, T \subset V$ é definida pela expressão

$$d(S, T) = \inf\{\|x - y\| \colon x \in S, y \in T\}.$$

Exemplo 12.32. *Vamos achar a distância (para o produto interno usual) do ponto* $(1, 2, -1)$ *ao plano $H = \{(x, y, z) \in \mathbb{R}^3 \colon x + y + 2z = 2\}$.*

A direção ortogonal ao plano é $(1, 1, 2)$*. A reta ortogonal ao plano que passa por* $(1, 2, -1)$ *tem equação paramétrica*

$$(1, 2, -1) + t(1, 1, 2) = (1 + t, 2 + t, -1 + 2t)$$

e interseta H quando

$$(1 + t) + (2 + t) + 2(-1 + 2t) = 2 \Leftrightarrow 6t = 1 \Leftrightarrow t = \tfrac{1}{6}$$

O ponto $v = (\frac{7}{6}, \frac{13}{6}, -\frac{2}{3})$ *de interseção desta reta com* H *é o ponto de* H *mais próximo de* $(1, 2, -1)$*. De facto se* $w \in H$ *for outro ponto, temos como antes, pelo Teorema de Pitágoras, que*

$$\|w - (1, 2, -1)\|^2 = \|w - v\|^2 + \|v - (1, 2, -1)\|^2 \geq \|v - (1, 2, -1)\|^2$$

uma vez que $v - (1, 2, -1)$ *(que tem a direção de* $(1, 1, 2)$*) e* $w - v$ *(que pertence ao plano paralelo a* H *que passa pela origem) são perpendiculares.*

Conclui-se que a distância de $(1, 2, -1)$ *a* H *é* $\|\frac{1}{6}(1, 1, 2)\| = \frac{1}{\sqrt{6}}$.

12.33. **O método dos mínimos quadrados.** [8] Seja A uma matriz $m \times n$. Mesmo que o sistema linear $Ax = b$ seja impossível, podemos tentar encontrar o valor de x que está mais próximo de constituir uma solução no sentido em que a distância de Ax a b é minimizada.

O conjunto $\{Ax \colon x \in \mathbb{R}^n\}$ é um subespaço de $\mathbb{R}^m$, nomeadamente o espaço das colunas de A, $EC(A)$. Como vimos acima, Ax estará o mais próximo possível de um ponto $b \in \mathbb{R}^m$ quando

$$Ax - b \in EC(A)^\perp$$

mas, uma vez que $EC(A) = EL(A^T)$, pelo Exemplo 12.25(i) temos

$$EC(A)^\perp = EL(A^T)^\perp = N(A^T)$$

Assim, Ax será o ponto mais próximo de b quando se verifica a **equação dos mínimos quadrados** para x

$$A^T(Ax - b) = 0 \Leftrightarrow A^T Ax = A^T b \tag{43}$$

Note-se que a solução pode não ser única (se $N(A) \neq 0$) mas o sistema acima tem sempre solução, que minimiza a distância a b, uma vez que o sistema traduz *exatamente* a condição de Ax ser o ponto de $EC(A)$ mais próximo de b, e este ponto existe sempre).

Este método é extremamente útil na prática. Frequentemente temos dados experimentais que queremos ajustar a uma lei conhecida, que depende de parâmetros. Os inevitáveis erros experimentais terão como consequência que nenhuma escolha dos parâmetros se adequará às medições, mas este método permite achar quais os valores dos parâmetros que melhor se adequam às medições efetuadas.

Exemplo 12.34. *Vamos determinar a reta* $y = ax + b$ *que melhor aproxima os três pontos (não colineares)* $(0, -2), (1, 3), (4, 5) \in \mathbb{R}^2$*. Se existisse uma reta que passasse pelos três pontos, os coeficientes* a, b *seriam soluções do sistema*

$$\begin{cases} a \cdot 0 + b = -2 \\ a \cdot 1 + b = 3 \\ a \cdot 4 + b = 5 \end{cases} \Leftrightarrow \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} -2 \\ 3 \\ 5 \end{bmatrix}$$

Este sistema não tem solução mas o método dos mínimos quadrados dá-nos os coeficientes a, b *tais que a soma*

$$(a \cdot 0 + b - (-2))^2 + (a \cdot 1 + b - 3)^2 + (a \cdot 4 + b - 5)^2$$

[8] Esta discussão é adaptada do tratamento deste método em [D].

é mínima (é isto que dá o nome ao método). Temos que achar a solução do sistema

$$\begin{bmatrix} 0 & 1 & 4 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0 & 1 & 4 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} -2 \\ 3 \\ 5 \end{bmatrix} \Leftrightarrow \begin{bmatrix} 17 & 5 \\ 5 & 3 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 23 \\ 6 \end{bmatrix}$$

que é

$$\begin{bmatrix} a \\ b \end{bmatrix} = \frac{1}{26} \begin{bmatrix} 3 & -5 \\ -5 & 17 \end{bmatrix} \begin{bmatrix} 23 \\ 6 \end{bmatrix} = \begin{bmatrix} \frac{39}{26} \\ -\frac{1}{2} \end{bmatrix}$$

pelo que a reta que melhor aproxima os pontos dados (no sentido dos mínimos quadrados) é

$$y = \tfrac{39}{26}x - \tfrac{1}{2}$$

Observação 12.35. *Pouco após a sua descoberta, em 1801, Ceres ficou tapado pelo Sol. Foi para prever (com sucesso) o sítio onde Ceres iria aparecer depois de passar por detrás do Sol, com base nas poucas observações que se tinham conseguido anteriormente, que Gauss inventou o método dos mínimos quadrados.*

12.36. **Transformações unitárias e (anti)-hermitianas.** Para terminar vamos falar um pouco das transformações lineares de um espaço vetorial munido de um produto interno nele próprio. Começamos por aquelas que preservam o produto interno e portanto ângulos e distâncias.

Definição 12.37. *Seja V um espaço vetorial com produto interno $\langle \cdot, \cdot \rangle$. Uma transformação linear $T\colon V \to V$ tal que*

$$\langle T(v), T(w) \rangle = \langle v, w \rangle \qquad \text{para todos os } v, w \in V$$

diz-se **ortogonal** *quando V é um espaço vetorial real e* **unitária** *quando V é um espaço vetorial complexo.*

Exemplo 12.38. *Consideremos $\mathbb{R}^n$ com o seu produto interno usual e $T\colon \mathbb{R}^n \to \mathbb{R}^n$ a transformação linear definida por $T(x) = Ax$ com A uma matriz $n \times n$ (onde, como habitualmente, estamos a identificar $\mathbb{R}^n$ com as matrizes coluna $n \times 1$). O produto interno de dois vetores x e y de $\mathbb{R}^n$ pode escrever-se matricialmente na forma $x^T y$. Portanto T é ortogonal se e só se*

$$(Ax)^T(Ay) = x^T y \Leftrightarrow x^T A^T A y = x^T y \qquad \text{para todos os } x, y \in \mathbb{R}^n \tag{44}$$

Isto acontece se e só se

$$A^T A = I_n \tag{45}$$

De facto, é claro que se A satisfaz a condição (45) então satisfaz (44). Reciprocamente se (44) é satisfeita então tomando para x e y o i-ésimo e j-ésimo vetores da base canónica de $\mathbb{R}^n$ respetivamente, a expressão $x^T A^T A y$ calcula a entrada ij da matriz $A^T A$ que é portanto 1 quando $i = j$ e 0 caso contrário, ou seja, a matriz identidade.

As matrizes de $M_{n\times n}(\mathbb{R})$ que satisfazem (45) chamam-se **matrizes ortogonais**. *Note-se que esta equação é também equivalente a dizer que A* **é invertível com inversa** *A^T.*

Uma vez que as linhas da matriz A^T são as colunas de A, a condição (45) *diz que* **uma matriz é ortogonal sse as suas colunas formam uma base ortonormada para $\mathbb{R}^n$.** *Assim, quando multiplicamos a matriz A por um vetor $x \in \mathbb{R}^n$, obtemos um vetor que tem as mesmas coordenadas que x mas numa base ortonormada diferente da canónica. Isto corresponde a uma rotação e/ou reflexão do espaço. Ver o Exemplo 12.42 abaixo.*

Consideremos agora o caso inteiramente análogo em que $V = \mathbb{C}^n$ com o produto interno usual, e $Tx = Ax$ com $x \in \mathbb{C}^n$. Temos agora que o produto interno é definido matricialmente pela expressão $\langle x, y\rangle = \overline{x}^T y$ e então T é unitária se

$$\overline{x}^T\overline{A}^T Ay = \overline{x}^T y \Leftrightarrow \overline{A}^T A = I_n$$

As matrizes que satisfazem esta condição dizem-se **unitárias**. *Novamente uma matriz é unitária sse é invertível e a sua inversa é $\overline{A}^T$, sse as suas colunas formam uma base ortonormada para $\mathbb{C}^n$.*

É conveniente simplificar a notação para a matriz transposta conjugada.

Definição 12.39. *Seja $A \in M_{m\times n}(\mathbb{C})$. A matriz transposta conjugada $\overline{A}^T$ é denotada por A^*, e é por vezes chamada a matriz transconjugada de A. Temos portanto $A^* \in M_{n\times m}(\mathbb{C})$ com entrada ij dada por $\overline{a_{ji}}$.*

Proposição 12.40. *Seja V um espaço vetorial complexo com produto interno $\langle\cdot,\cdot\rangle$ e $T\colon V \to V$ uma transformação unitária. Então*
1. *Os valores próprios de T são complexos com módulo 1.*
2. *Vetores próprios de T correspondentes a valores próprios distintos são ortogonais.*

Dem. Seja v um vetor próprio de T. Sendo $T(v) = \lambda v$ temos

$$\|T(v)\|^2 = \langle T(v), T(v)\rangle = \langle \lambda v, \lambda v\rangle = \overline{\lambda}\lambda\langle v, v\rangle = |\lambda|^2\|v\|^2$$

Por outro lado, como T é unitária temos $\langle T(v), T(v)\rangle = \langle v, v\rangle = \|v\|^2$. Portanto $\|v\|^2 = |\lambda|^2\|v\|^2$, e como $v \neq 0$, isto significa que $|\lambda| = 1$.

Suponhamos agora que $T(v) = \lambda v$ e $T(w) = \mu w$ com λ, μ distintos. Então

$$\langle v, w\rangle = \langle Tv, Tw\rangle = \langle \lambda v, \mu w\rangle = \overline{\lambda}\mu\langle v, w\rangle$$

ou seja

$$(1 - \overline{\lambda}\mu)\langle v, w\rangle = 0 \Leftrightarrow \overline{\lambda}\mu = 1 \quad \text{ou} \quad \langle v, w\rangle = 0$$

Como λ é um complexo com módulo 1, $\overline{\lambda} = \frac{1}{\lambda}$ logo a primeira condição na disjunção acima é equivalente a $\mu = \lambda$. Conclui-se que $\langle v, w\rangle = 0$, isto é, que v e w são ortogonais. $\square$

Observação 12.41. *Se encararmos uma matriz $n \times n$ real A como uma matriz complexa, dizer que A é ortogonal ou unitária é equivalente (uma vez que $\overline{A} = A$). Vemos portanto que os valores próprios de uma matriz ortogonal são complexos unitários e que os seus vetores próprios são ortogonais em $\mathbb{C}^n$.*

Exemplo 12.42. *A matriz*

$$\begin{bmatrix} \cos\alpha & -\operatorname{sen}\alpha \\ \operatorname{sen}\alpha & \cos\alpha \end{bmatrix}$$

é ortogonal, como se verifica imediatamente. Geometricamente corresponde à rotação de um ângulo α no sentido anti-horário (desenhe o efeito da matriz nos vetores da base canónica).

Note-se que os valores próprios (complexos) desta matriz são as soluções de

$$(\cos\alpha - \lambda)^2 + \operatorname{sen}^2\alpha = 0 \Leftrightarrow \lambda = \cos\alpha \pm i \operatorname{sen}\alpha$$

Os vetores próprios (também necessariamente complexos) são as soluções de

$$\begin{bmatrix} \cos\alpha - (\cos\alpha \pm i\operatorname{sen}\alpha) & -\operatorname{sen}\alpha \\ \operatorname{sen}\alpha & \cos\alpha - (\cos\alpha \pm i\operatorname{sen}\alpha) \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = 0 \Leftrightarrow \begin{bmatrix} a \\ b \end{bmatrix} = \lambda \begin{bmatrix} \pm i \\ 1 \end{bmatrix}$$

e são ortogonais para o produto interno usual em $\mathbb{C}^2$.

Definição 12.43. *Sejam V, W espaços vetoriais reais ou complexos de dimensão finita com produto interno e $T\colon V \to W$ uma transformação linear. A* transformação adjunta de T *é a única transformação linear $T^*\colon W \to V$ tal que*

$$\langle Tv, w\rangle = \langle v, T^*w\rangle \quad \text{para todos os } v \in V, w \in W. \tag{46}$$

Temos que verificar que esta definição faz sentido. Começamos por notar que, dado $w \in W$, há um único vetor T^*w que satisfaz a relação $\langle Tv, w\rangle = \langle v, T^*w\rangle$ para todo o $v \in V$. De facto sendo $B = (v_1, \ldots, v_n)$ uma base ortonormada para V, o escalar $\langle v_i, T^*w\rangle$, que é a componente i de $[T^*w]_B$, tem de ser necessariamente igual a $\langle Tv_i, w\rangle$. Portanto

$$T^*w = \langle Tv_1, w\rangle v_1 + \ldots + \langle Tv_n, w\rangle v_n$$

Por outro lado, é um exercício simples verificar que se definirmos T^*w pela fórmula anterior, T^* é uma transformação linear que satisfaz a igualdade na definição de transformação adjunta.

No caso em que $V = \mathbb{C}^n$, $W = \mathbb{C}^m$, e $Tv = Av$ é determinada por uma matriz $A \in M_{m\times n}(\mathbb{C})$, a equação (46) traduz-se em

$$(\overline{Av})^T w = \overline{v}^T(T^*w) \text{ para todo o } v, w \in \mathbb{C}^n \Leftrightarrow \overline{v}^T A^* w = \overline{v}^T(T^*w)$$

pelo que a transformação adjunta T^* é a transformação linear de $\mathbb{C}^m$ para $\mathbb{C}^n$ determinada pela matriz transposta conjugada A^*. No caso real, a transformação adjunta é a transformação determinada pela matriz transposta.

Observação 12.44. *Em termos da identificação do dual de um espaço vetorial com um produto interno com o próprio espaço vetorial, explicada nos exercícios da Ficha 12, a transformação adjunta de T corresponde à transformação induzida por T entre os espaços duais.*

Definição 12.45. *Seja V um espaço vetorial de dimensão finita com produto interno. Uma transformação linear $T\colon V \to V$ diz-se* auto-adjunta *se $T = T^*$ e* anti-adjunta *se $T = -T^*$.*

Segue imediatamente da discussão anterior que uma matriz $A \in M_{n\times n}(\mathbb{R})$ determina uma transformação auto-adjunta de $\mathbb{R}^n$ sse é simétrica e anti-adjunta sse é *anti-simétrica*, isto é se $A^T + A = 0$. Analogamente, uma matriz $A \in M_{n\times n}(\mathbb{C})$ determina uma transformação auto-adjunta sse A é hermitiana e uma transformação anti-adjunta sse é *anti-hermitiana*, isto é se $A^* + A = 0$.

Proposição 12.46. *Os valores próprios de uma transformação linear auto-adjunta são reais, e os de uma transformação linear anti-adjunta são imaginários puros. Em qualquer dos casos, vetores próprios de valores próprios distintos são ortogonais.*

Dem. Suponhamos que T é auto-adjunta e v é um vetor próprio de T então

$$\overline{\lambda}\langle v, v\rangle = \langle \lambda v, v\rangle = \langle Tv, v\rangle = \langle v, Tv\rangle = \langle v, \lambda v\rangle = \lambda\langle v, v\rangle$$

Como $v \neq 0$ temos que $\overline{\lambda} = \lambda$ pelo que λ é real. No caso anti-adjunto obteríamos a igualdade $\overline{\lambda} + \lambda = 0$ que diz que λ é imaginário puro.

Sejam λ e μ valores próprios distintos de T auto-adjunta com vetores próprios v, w. Então

$$\lambda\langle v, w\rangle = \langle Tv, w\rangle = \langle v, Tw\rangle = \mu\langle v, w\rangle$$

onde na primeira igualdade usámos o facto de λ ser real e portanto igual ao seu conjugado. A igualdade anterior traduz-se em $(\lambda - \mu)\langle v, w\rangle = 0$. Uma vez que $\lambda \neq \mu$, conclui-se que v e w são ortogonais.

No caso anti-adjunto obtemos análogamente $(\overline{\lambda} + \mu)\langle v, w\rangle = 0$. Como λ e μ são imaginários puros $\overline{\lambda} + \mu = -\lambda + \mu$ pelo que novamente vemos que v, w são ortogonais. □

Teorema 12.47 (Teorema espetral). *(i) Seja V um espaço vetorial complexo de dimensão finita com produto interno e $T\colon V \to V$ uma transformação linear unitária, auto-adjunta ou anti-adjunta. Então T é diagonalizável por uma base ortogonal de V.*

(ii) Seja V um espaço vetorial real de dimensão finita com produto interno e $T\colon V \to V$ uma transformação linear auto-adjunta. Então T é diagonalizável por uma base ortogonal de V.

Dem. As demonstrações são todas análogas pelo que vamos apenas ilustrar o caso de uma transformação auto-adjunta deixando os outros como exercício.

A demonstração é por indução na dimensão do espaço V, sendo que o caso de dimensão 1 é trivial. Supondo que a dimensão de V é maior do que 1, seja v um vetor próprio de T e consideremos o subespaço $W = v^\perp \subset V$. Então $T_{|W}$ é também auto-adjunta para a restrição do produto interno de T a W. De facto, a igualdade

$$\overline{\lambda}\langle v, w\rangle = \langle Tv, w\rangle = \langle v, T^*w\rangle$$

mostra que, se $w \in v^\perp$ então $T^*w \in v^\perp$. É então imediato que $(T_{|W})^* = T^*_{|W}$ e portanto $T_{|W}$ é auto-adjunta. Por hipótese de indução, existe uma base ortogonal de W formada por valores próprios de $T_{|W}$ que juntamente com v forma a base ortogonal desejada para V.

Se V é um espaço vetorial real, com respeito a uma base ortonormada B para V, o produto interno em V é calculado da mesma forma que o produto interno usual em $\mathbb{R}^n$. Uma transformação auto-adjunta T é representada na base B por uma matriz simétrica

$A = A_{T,B,B}$. A transformação linear $\tilde{T}\colon \mathbb{C}^n \to \mathbb{C}^n$ representada por A é portanto auto-adjunta (com respeito ao produto interno usual em $\mathbb{C}^n$). Como tal é diagonalizável (sobre $\mathbb{C}$) por uma base ortogonal. No entanto, como os valores próprios de $\tilde{T}$ são reais, e a matriz A é real, os vetores próprios de $\tilde{T}$ são também reais. Os vetores de V correspondentes aos vetores próprios de $\tilde{T}$ são a base ortogonal desejada para V. $\square$

Sumarizamos agora a informaçao sobre matrizes quadradas que resulta do Teorema anterior, aplicando-o à transformação linear definida por $Tx = Ax$ com A uma matriz $n \times n$ real ou complexa. Em cada caso a matriz A pode ser escrita na forma

$$A = SDS^{-1}$$

com S uma matriz, unitária quando A é diagonalizável sobre $\mathbb{C}$, e ortogonal quando A é diagonalizável sobre $\mathbb{R}$, (cujas colunas formam uma base ortonormada para $\mathbb{C}^n$ ou $\mathbb{R}^n$ consoante o caso, constituída por vetores próprios de A), e D uma matriz diagonal cujas entradas são os valores próprios de A.

Tipo de matriz	Definição	Diagonalizável	Valores próprios
ortogonal	$AA^T = I_n$	sobre $\mathbb{C}$	$\lambda \in \mathbb{C}, \lvert\lambda\rvert = 1$
simétrica	$A = A^T$	sobre $\mathbb{R}$	reais
anti-simétrica	$A + A^T = 0$	sobre $\mathbb{C}$	imaginários puros

Matrizes $n \times n$ reais especiais.

Tipo de matriz	Definição	Valores próprios
unitária	$AA^* = I_n$	$\lambda \in \mathbb{C}, \lvert\lambda\rvert = 1$
hermitiana	$A = A^*$	reais
anti-hermitiana	$A + A^* = 0$	imaginários puros

Matrizes $n \times n$ complexas especiais.

Observação 12.48. *Embora não haja nenhum critério útil para ver se uma matriz é diagonalizável, há um critério muito simples para ver se uma matriz complexa A é diagonalizável por uma base ortonormada. Isto acontece sse $AA^* = A^*A$. Ver os exercícios da Ficha 13 para uma demonstração.*

12.49. **Formas quadráticas.** Como aplicação do resultado anterior vamos aproveitar para classificar a menos de mudança de variável linear os polinómios homogéneos de grau 2 de várias variáveis. Podemos pensar nestes como as funções de várias variáveis mais simples a seguir às funções lineares.

Definição 12.50. *Uma* forma quadrática *em $\mathbb{R}^n$ é uma função $f\colon \mathbb{R}^n \to \mathbb{R}$ da forma*

$$f(x) = x^T A x \tag{47}$$

com $A \in M_{n\times n}(\mathbb{R})$ (onde identificamos como habitualmente uma matriz 1×1 com um escalar).

Por exemplo

$$f(x,y) = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 2 & 5 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 2x^2 + 6xy + 4y^2 \tag{48}$$

é uma forma quadrática em $\mathbb{R}^2$. Note-se que a forma quadrática depende apenas da *parte simétrica* $\frac{A+A^T}{2}$ da matriz A. De facto uma vez que a transposição de matrizes 1×1 não tem qualquer efeito temos $x^TAx = (x^TAx)^T = x^TA^Tx$. Substituindo a matriz A em (47) por $\frac{A+A^T}{2}$ obtemos portanto a mesma expressão. Por outro lado, uma vez que a soma das entradas ij e ji da matriz A é o coeficiente de x_ix_j na expressão (47) matrizes simétricas distintas dão azo a formas quadráticas distintas. Há assim uma *correspondência biunívoca* entre formas quadráticas e matrizes quadradas reais simétricas.

Tendo em conta o Teorema espetral, dada uma matriz simétrica A, existe uma matriz ortogonal S e uma matriz diagonal (real) D tal que

$$A = SDS^{-1}$$

E dado que S é ortogonal, $S^{-1} = S^T$. Usando coordenadas y na base ortonormada formada pelas colunas de S a expressão para a forma quadrática simplifica-se muito. Temos $x = Sy$ e então

$$f(x) = x^TAx = (y^TS^T)A(Sy) = (y^TS^T)SDS^T(Sy) = y^TDy = \lambda_1 y_1^2 + \ldots + \lambda_n y_n^2 \tag{49}$$

onde $\lambda_1, \ldots, \lambda_n$ são as entradas diagonais de D, ou seja, os valores próprios de A. Nas aplicações (por exemplo para a determinação de extremos de funções de várias variáveis como verão em Cálculo 2) é importante determinar o "sinal" de uma forma quadrática no seguinte sentido.

Definição 12.51. *Uma forma quadrática $f\colon \mathbb{R}^n \to \mathbb{R}$ diz-se*

(i) definida positiva *se $f(x) > 0$ para $x \neq 0$.*
(ii) semi-definida positiva *se $f(x) \geq 0$ para todo o $x \in \mathbb{R}^n$.*
(iii) definida negativa *se $f(x) < 0$ para $x \neq 0$.*
(iv) semi-definida negativa *se $f(x) \leq 0$ para todo o $x \in \mathbb{R}^n$.*
(v) indefinida *se $f(x)$ assume valores positivos e negativos.*

Da discussão anterior obtemos imediatamente o seguinte resultado.

Proposição 12.52. *Uma forma quadrática $f(x) = x^TAx$ com $A \in M_{n\times n}(\mathbb{R})$ simétrica é*

(i) definida positiva sse todos os valores próprios de A são positivos.
(ii) semidefinida positiva sse todos os valores próprios de A são maiores ou iguais a zero.
(iii) definida negativa sse todos os valores próprios de A são negativos.
(iv) semidefinida negativa sse todos os valores próprios de A são menores ou iguais a zero.
(v) indefinida sse A tem valores próprios de sinal contrário.

Exemplo 12.53. *A forma quadrática* (48) *é indefinida uma vez que a matriz simétrica que a representa*

$$\begin{bmatrix} 2 & 3 \\ 3 & 4 \end{bmatrix}$$

tem determinante negativo e portanto valores próprios de sinais contrários.

Observação 12.54. *A expressão* (49) *mostra também que toda a matriz simétrica com valores próprios positivos é a matriz da métrica de um produto interno, pois a positividade do produto interno corresponde precisamente ao facto da forma quadrática determinada pela matriz ser definida positiva.*

Appendix A. Mais sobre a forma canónica de Jordan

O procedimento descrito acima para achar a forma canónica de Jordan funciona bastante bem para matrizes pequenas mas, em geral, pode ser difícil encontrar os vectores próprios v_1 para os quais se consegue resolver as equações (35) recursivamente. O seguinte exemplo ilustra as dificuldades no caso mais simples.

Exemplo A.1. *Seja A uma matriz com forma canónica de Jordan*

$$J = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{50}$$

O espaço próprio de 1 *tem dimensão* 2. *Seja $\{v_1, v_1'\}$ uma base para o espaço próprio de* 1. *Tem que se ter cuidado na escolha do vector próprio v de* 1 *que se põe na primeira coluna da matriz S. De facto, só será possível resolver a equação* (35)

$$(A - I)v_2 = v$$

para achar a segunda coluna se v estiver no espaço das colunas da matriz $(A - I)$, que tem dimensão 1. *É portanto necessário achar uma combinação linear $v = \alpha v_1 + \beta v_1'$ que pertença ao espaço das colunas de $A - I$. A terceira coluna poderá ser qualquer vector próprio de* 1 *que juntamente com v forme uma base para o espaço próprio.*

Vejamos um exemplo concreto. Considere-se a matriz

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

O polinómio característico é

$$\det(A - \lambda I) = -\lambda^3 + 3\lambda^2 - 3\lambda + 1 = -(\lambda - 1)^3$$

logo o único valor próprio é 1, *com multiplicidade algébrica* 3. *Os vectores próprios de* 1 *são as soluções de*

$$\begin{bmatrix} -1 & -1 & 2 \\ -1 & -1 & 2 \\ -1 & -1 & 2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = 0 \Leftrightarrow 2c = a + b$$

O espaço próprio de 1 *é portanto o conjunto dos vectores*

$$\begin{bmatrix} a \\ b \\ \frac{1}{2}(a+b) \end{bmatrix} = a \begin{bmatrix} 1 \\ 0 \\ \frac{1}{2} \end{bmatrix} + b \begin{bmatrix} 0 \\ 1 \\ \frac{1}{2} \end{bmatrix}$$

e 1 *tem multiplicidade geométrica* 2*. Há portanto dois blocos de Jordan e a forma canónica de Jordan de* A *é necessariamente* (50)*.*

Não é no entanto possível resolver a equação (35)

$$(A-I)v_2 = v_1$$

quando v_1 *é um dos vectores* $\left(1,0,\frac{1}{2}\right)$ *ou* $\left(0,1,\frac{1}{2}\right)$ *da base "natural" do espaço próprio de* 1*. Como observámos acima, para que a equação tenha solução é necessário que* v_1 *pertença ao espaço das colunas de* $A-I$*, que é o espaço gerado por* $(1,1,1)$*. A soma dos dois vectores da "base natural" é exactamente* $(1,1,1)$*. Resolvendo a equação*

$$\begin{bmatrix} -1 & -1 & 2 \\ -1 & -1 & 2 \\ -1 & -1 & 2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = 0 \Leftrightarrow 2c = a+b+1$$

obtemos as soluções

$$\begin{bmatrix} a \\ b \\ \frac{1}{2}+\frac{a}{2}+\frac{b}{2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \frac{1}{2} \end{bmatrix} + a \begin{bmatrix} 1 \\ 0 \\ \frac{1}{2} \end{bmatrix} + b \begin{bmatrix} 0 \\ 1 \\ \frac{1}{2} \end{bmatrix}$$

Podemos tomar por exemplo $v_2 = \left(0,0,\frac{1}{2}\right)$*. Para terceira coluna de* S *podemos tomar qualquer vector próprio de* 1 *que juntamente com* $(1,1,1)$ *forme uma base do espaço próprio, por exemplo,* $\left(1,0,\frac{1}{2}\right)$*. Obtemos assim a matriz de mudança de base*

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

O exemplo anterior é bastante simples e foi resolvido facilmente mas a situação complica-se à medida que o número e tamanho dos blocos relativos ao mesmo valor próprio aumenta. Por exemplo, se a multiplicidade algébrica de λ é 4 e a multiplicidade geométrica é 2, não sabemos à partida qual é a dimensão dos dois blocos de Jordan associados a λ. As possibilidades são 1 e 3 ou 2 e 2. No primeiro caso teremos novamente que ter cuidado na seleção do vector próprio com o qual iniciar a resolução recursiva das equações (35) (terá de estar na imagem de $(A-\lambda I)^2$). No final deste apêndice descreve-se um algoritmo geral (nada prático em termos de contas) para achar a forma canónica J juntamente com a matriz S (ver a secção A.5).

A.2. **Demonstração do Teorema 11.11.** Seja A uma matriz $n \times n$ complexa. Se $\lambda \in \mathbb{C}$ é um valor próprio de A o conjunto

$$V(\lambda) = \{v \in \mathbb{C}^n \colon (A-\lambda I)^k v = 0 \text{ para algum } k \geq 1\}$$

chama-se o *espaço próprio generalizado* de A associado ao vector próprio λ. É fácil ver que $V(\lambda)$ é um subespaço vectorial de $\mathbb{C}^n$.

Para cada $j \geq 0$ seja

$$V(\lambda, j) = \{v \in C^n : (A - \lambda I)^j v = 0\}$$

Temos então uma sucessão de subespaços

$$0 = V(\lambda, 0) \subset V(\lambda, 1) \subset V(\lambda, 2) \subset \cdots \subset V(\lambda, k) \subset \cdots \subset V(\lambda)$$

Por definição, $V(\lambda, 1)$ é o espaço próprio de λ e $V(\lambda)$ é a união de todos os $V(\lambda, j)$. Uma vez que $V(\lambda)$ tem dimensão finita, existe $n(\lambda) \geq 1$ tal que

$$V(\lambda, n(\lambda) - 1) \subsetneq V(\lambda, n(\lambda)) = V(\lambda)$$

Note-se também que, uma vez que a matriz A comuta com a matriz $(A - \lambda I)$, os espaços $V(\lambda, j)$ são invariantes, isto é, $AV(\lambda, j) \subset V(\lambda, j)$.

Definição A.3. *Seja A uma matriz $n \times n$ complexa e $\lambda \in \mathbb{C}$ um valor próprio de A. O* índice *de um vector próprio generalizado $v \in V(\lambda)$ é o menor $i \geq 0$ tal que $v \in V(\lambda, i)$.*

Por exemplo o vector 0 tem indíce 0 e um vector próprio tem índice 1.

Lema A.4. *Se $v \in V(\lambda)$ tem índice i então dado $0 \leq j \leq i$, o vector $(A - \lambda I)^j$ tem índice $i - j$. Além disso o conjunto*

$$(A - \lambda I)^{i-1} v, \ldots, (A - \lambda I)v, v$$

é linearmente independente (isto é, é uma cadeia de Jordan).

Dem. Uma vez que $(A-\lambda I)^{i-j}\left((A - \lambda I)^j v\right) = 0$, o vector $(A-\lambda I)^j v$ pertence a $V(\lambda, i-j)$. Se estivesse contido em $V(\lambda, m)$ com $m < i - j$ então $(A - \lambda I)^{j+m} v = 0$ e portanto $v \in V(\lambda, m + j)$ contradizendo o facto de v ter índice i.

Suponhamos que $\alpha_l \in \mathbb{C}$ são tais que

$$\alpha_{i-1}(A - \lambda I)^{i-1} v + \ldots + \alpha_1(A - \lambda I)v + \alpha_0 v = 0$$

Aplicando $(A - \lambda I)^{i-1}$ à combinação linear anterior obtemos

$$0 + \ldots + 0 + \alpha_0 (A - \lambda I)^{i-1} v = 0$$

logo $\alpha_0 = 0$. Aplicando $(A - \lambda I)^{i-2}$ à combinação linear obtemos agora $\alpha_1 = 0$, e continuando da mesma maneira vemos que todos os α_l são 0. $\square$

O plano da demonstração do Teorema 11.11 é o seguinte. Iremos demonstrar que

(1) A afirmação do Teorema é verdadeira quando $\mathbb{C}^n = V(\lambda)$ para algum λ.
(2) Dados valores próprios distintos $\lambda_1, \ldots, \lambda_k$ tem-se $V(\lambda_1) \cap \left(\oplus_{j=2}^k V(\lambda_k)\right) = \{0\}$ logo a afirmação do Teorema é válida quando $\mathbb{C}^n = \oplus_{\lambda \in \sigma(A)} V(\lambda)$ onde $\sigma(A)$ denota o conjunto dos valores próprios de A.
(3) $\mathbb{C}^n = \oplus_{\lambda \in \sigma(A)} V(\lambda)$

Dem. do Teorema 11.11. (1) O nosso objectivo é escrever uma base para $\mathbb{C}^n = V(\lambda)$ formada por cadeias de Jordan (35). Seja $l = n(\lambda)$. Começamos por escolher uma base $v_{l,1}, \dots, v_{l,k_l}$ para um espaço $U(l)$ complementar a $V(\lambda, l-1)$ em $V(\lambda, l) = V(\lambda)$.

Qualquer combinação linear não nula dos vectores $v_{l,m}$ tem indíce l e portanto o argumento usado na demonstração do Lema A.4 mostra que o conjunto

$$v_{l,1}, \dots, v_{l,k_l}, (A-\lambda I)v_{l,1}, \dots, (A-\lambda I)v_{l,k_l}, \dots, (A-\lambda I)^{l-1}v_{l,1}, \dots, (A-\lambda I)^{l-1}v_{l,k_l} \tag{51}$$

é linearmente independente e portanto forma uma base para o subespaço

$$W(l) = U(l) + (A-\lambda I)U(l) + \dots + (A-\lambda I)^{l-1}U(l) \subset V(\lambda)$$

Este espaço é invariante para $(A - \lambda I)$ e portanto para A. Na base (51) a transformação linear A é representada por uma matriz diagonal por blocos sendo todos os blocos, blocos de Jordan de dimensão l. O número de blocos é k_l. Os vectores $v_{l,m}$ são as colunas de S correspondentes às colunas mais à direita destes blocos e os restantes vectores da base (51) são as restantes colunas de S correspondentes a estes blocos.

Seja $U(l-1)$ um complementar para o subespaço $V(\lambda, l-2) + (A-\lambda I)U(l) \subset V(\lambda, l-1)$. Escolhemos uma base $v_{l-1,1}, \dots, v_{l-1,k_{l-1}}$ para $U(l-1)$ (que pode ser vazia se $U_{l-1} = 0$). Note-se que qualquer combinação linear não nula destes vectores tem índice $l-1$. O argumento do Lema A.4 mostra novamente que o conjunto

$$v_{l-1,1}, \dots, v_{l-1,k_{l-1}}, \dots, (A-\lambda I)^{l-2}v_{l-1,1}, \dots, (A-\lambda I)^{l-2}v_{l-1,k_{l-1}} \tag{52}$$

é linearmente independente. Mais geralmente, notando que qualquer combinação linear de $(A-\lambda I)v_{l,1}, \dots, (A-\lambda I)v_{l,k_l}, v_{l-1,1}, \dots, v_{l-1,k_{l-1}}$ tem índice $l-1$, vemos que a união dos conjuntos (51) e (52) é linearmente independente.

Seja

$$W(l-1) = U(l-1) + (A-\lambda I)U(l-1) + \dots + (A-\lambda I)^{l-2}U(l-1)$$

Este espaço é invariante para A e na base (52) a transformação linear A é representada por uma matriz diagonal por blocos. Todos os blocos são blocos de Jordan de dimensão $l-1$ e há k_{l-1} blocos.

Escolhemos agora um complementar $U(l-2)$ para o subespaço $V(l-3) + (A-\lambda I)^2U(l) + (A-\lambda I)U(l-1)$. No espaço $W(l-2) = U(l-2) + \dots + (A-\lambda I)^{l-3}U(l-2)$ a transformação linear é diagonal por blocos sendo todos estes blocos de Jordan de dimensão $l-2$.

Prosseguindo desta forma obtemos uma decomposição

$$V(\lambda) = W(l) \oplus W(l-1) \oplus \dots \oplus W(1)$$

e uma base de $V(\lambda)$ na qual a tranformação linear A é representada por uma matriz em forma canónica de Jordan.

(2) Sejam $\lambda_1, \dots, \lambda_k$ valores próprios distintos. Para verificar que

$$V(\lambda_1) \cap (V(\lambda_2) \oplus \dots \oplus V(\lambda_k)) = 0$$

basta ver que a transformação linear $(A-\mu I)$ restrita a $V(\lambda_1)$ é invertível se $\mu \neq \lambda_1$. Admitindo essa afirmação, a transformação linear $(A-\lambda_2 I)^{n(\lambda_2)}\cdots(A-\lambda_k I)^{n(\lambda_k)}$ é invertível em $V(\lambda_1)$ e 0 em $V(\lambda_2)\oplus\ldots\oplus V(\lambda_k)$ logo a interseção dos dois espaços é nula.

Mas $(A-\mu I) = (A-\lambda_1 I)+(\lambda_1-\mu)I = (\lambda_1-\mu)\left(\frac{1}{\lambda_1-\mu}(A-\lambda_1)+I\right)$ tem inverso

$$\frac{1}{\lambda_1-\mu}\left(I-\frac{1}{\lambda_1-\mu}(A-\lambda_1 I)+\ldots+(-1)^{n(\lambda_1)-1}\frac{1}{(\lambda_1-\mu)^{n(\lambda_1)-1}}(A-\lambda_1)^{n(\lambda_1)-1}\right)$$

como se verifica facilmente.

(3) Seja $\sigma(A)$ o conjunto dos valores próprios de A e suponhamos por absurdo que

$$Z = \oplus_{\lambda\in\sigma(A)}V(\lambda) \neq \mathbb{C}^n$$

Seja W um complemento para o espaço $Z \subset \mathbb{C}^n$ e sejam $n_1 = \dim Z$ e $n_2 = \dim W$. Podemos escolher uma base para $\mathbb{C}^n$ tal que os primeiros n_1 elementos da base pertencem a Z e os restantes a W. Nessa base a transformação linear definida por A tem a forma

$$A = \begin{bmatrix} J & B \\ 0 & C \end{bmatrix}$$

(onde J é uma matriz $n_1\times n_1$ e C é uma matriz $n_2\times n_2$). Seja $w\in W$ um vector próprio da matriz C e λ o valor próprio correspondente. Então $\lambda\in\sigma(A)$ e temos

$$Aw = \lambda w + z \Leftrightarrow (A-\lambda I)w = z \qquad \text{para algum } z\in Z$$

Sejam $\lambda_2,\ldots,\lambda_k$ os elementos de $\sigma(A)\setminus\{\lambda\}$. Então

$$(A-\lambda_2)^{n(\lambda_2)}\cdots(A-\lambda_k)^{n(\lambda_k)}z \in V(\lambda)$$

logo

$$v = (A-\lambda_2)^{n(\lambda_2)}\cdots(A-\lambda_k)^{n(\lambda_k)}w$$

é tal que

$$(A-\lambda I)v \in V(\lambda)$$

e portanto

$$v\in V(\lambda)\subset Z.$$

Mas por outro lado

$$v = (\lambda-\lambda_2)^{n(\lambda_2)}\cdots(\lambda-\lambda_k)^{n(\lambda_k)}w + z'$$

com $z'\in Z$, o que é uma contradição.

□

A.5. **Algoritmo para a determinação da forma canónica de Jordan.** A demonstração do passo (1) do Teorema 11.11 contém implicitamente o seguinte algoritmo (nada prático) para o cálculo da matriz J e S:

(1) Achar o conjunto $\sigma(A)$ dos valores próprios de A. Para cada $\lambda \in \sigma(A)$:

(2) Determinar os espaços
$$V(\lambda, i) = N(A - \lambda I)^i \text{ para } i = 1, 2, \ldots$$
e em particular determinar
$$n(\lambda) = \min\{k \colon N(A-\lambda I)^k = N(A - \lambda I)^{k+1}\}$$

(3) Seja $l = n(\lambda)$. Determinar uma base $v_{l,1}, \ldots, v_{l,k_l}$ para um espaço $U(\lambda, l)$ complementar a $V(\lambda, l-1)$ em $V(\lambda, l)$. O número k_l é o número de blocos de Jordan com λ na diagonal de tamanho l. Cada $v_{l,i}$ gera uma cadeia de Jordan de tamanho l
$$v_{l,i}, (A - \lambda I)v_{l,i}, \ldots, (A - \lambda I)^{l-1} v_{l,i}$$
que dão as colunas de A correspondentes a um bloco de Jordan de tamanho l.

(4) Determinar uma base $v_{l-1,1}, \ldots, v_{l-1,k_{l-1}}$ para um espaço $U(\lambda, l-1)$ complementar a $V(\lambda, l-2) + (A - \lambda I)U(\lambda, l)$ em $V(\lambda, l-1)$. Cada $v_{l-1,i}$ gera uma cadeia de Jordan de tamanho $l-1$ que dá as colunas de A correspondentes a um bloco de Jordan de tamanho $l-1$.

(5) Determinar uma base $v_{l-2,1}, \ldots, v_{l-1,k_{l-2}}$ para um espaço $U(\lambda, l-2)$ complementar a $V(\lambda, l-3) + (A-\lambda I)U(\lambda, l-1) + (A-\lambda I)^2 U(\lambda, l)$ em $V(\lambda, l-2)$. Cada $v_{l-2,i}$ gera uma cadeia de Jordan de tamanho $l-2$ que dá as colunas de S correspondentes a um bloco de Jordan de tamanho $l-2$.

(6) Continuando desta forma obtêm-se as colunas de S correspondentes aos blocos de Jordan com λ na diagonal (assim como o número destes blocos de cada dimensão).

Appendix B. O produto externo de vetores

Definição B.1. *Sejam $v, w \in \mathbb{R}^3$. O* produto externo *de v e w é o vetor $v \times w \in \mathbb{R}^3$ definido por*
$$\begin{aligned} v \times w &= \begin{vmatrix} \mathbf{e_1} & \mathbf{e_2} & \mathbf{e_3} \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} = (v_2w_3 - v_3w_2)\mathbf{e_1} + (v_3w_1 - v_1w_3)\mathbf{e_2} + (v_1w_2 - v_2w_1)\mathbf{e_3} \\ &= (v_2w_3 - v_3w_2, v_3w_1 - v_1w_3, v_1w_2 - v_2w_1) \end{aligned}$$
onde $\mathbf{e_i}$ designa o i-ésimo vetor da base canónica de $\mathbb{R}^3$ e a expressão à direita se obtém expandindo o determinante ao longo da primeira linha.

Exemplo B.2.
$$(1, -3, 2) \times (5, 0, 2) = \left| \begin{matrix} \mathbf{e_1} & \mathbf{e_2} & \mathbf{e_3} \\ 1 & -3 & 2 \\ 5 & 0 & 2 \end{matrix} \right] = (-6, 8, 15)$$

O produto externo tem inúmeras aplicações em Matemática e Física. Será usado em Cálculo 2 para calcular fluxos de campos vetoriais através de superfícies. Em Mecânica aparece por exemplo na expressão para o momento angular de uma partícula em torno de um ponto, que é dado pela expressão $\vec{L} = \vec{r} \times \vec{p}$ com $\vec{r}$ o vetor de posição e $\vec{p}$ o momento linear. A Força de Lorentz a que uma carga elétrica em movimento é sujeita ao interagir com um campo magnético $\vec{B}$ é $\vec{F} = q\vec{v} \times \vec{B}$, com $\vec{v}$ a velocidade e q a carga da partícula em questão.

As propriedades do determinante implicam imediatamente certas propriedades do produto externo.

Proposição B.3 (Propriedades do produto externo)**.** *(i) O produto externo é linear em cada um dos seus argumentos.*

(ii) $v \times w = -w \times v$

(iii) $v \times v = 0$

(iv) $\langle u, (v \times w) \rangle = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}$

Proof. A primeira afirmação é verdadeira porque o determinante é multilinear, a segunda porque o determinante troca de sinal quando se trocam linhas, e a terceira porque o determinante é zero se houver uma linha repetida. A quarta é uma consequência da definição do produto interno e da expansão de Laplace ao longo da primeira linha. □

A Proposição anterior dá-nos o significado geométrico do produto externo. De facto, por (iv) temos

$$\langle v, v \times w \rangle = \langle w, v \times w \rangle = 0$$

pelo que $v \times w$ *é ortogonal ao plano gerado por* v *e* w (se v e w são colineares, então as propriedade (i) e (iii) dizem-nos que o produto externo é o vetor nulo). Além disso, dada a interpretação do determinante como o volume do paralelipípedo temos que

$$\|v \times w\|^2 = \langle v \times w, v \times w \rangle = \begin{vmatrix} - & v \times w & - \\ - & v & - \\ - & w & - \end{vmatrix}$$

é o volume do paralelipípedo com base o paralelogramo formado por v e w sendo a outra aresta perpendicular ao paralelogramo com comprimento $\|v \times w\|$. Este volume é a área da base vezes o comprimento da aresta perpendicular à base pelo que $\|v \times w\|$ *é a área do paralelogramo com arestas* v *e* w. Note-se que no caso degenerado em que v e w são colineares a afirmação anterior continua a ser válida.

Em suma, quando v, w não são colineares, o produto externo $v \times w$ é um vetor perpendicular ao plano determinado por v e w, cujo comprimento é a área do paralelogramo com arestas v e w. Se α for a o ângulo entre v e w, a área do paralelogramo é a mesma que a área do retângulo com arestas de comprimento $\|v\|$ e $\|w\| \operatorname{sen} \alpha$ (isto vê-se deslizando a aresta w ao longo de uma reta paralela a v até que fique perpendicular a v - movimento que não afeta a área do paralelogramo). Portanto

$$\|v \times w\| = \|v\|\|w\| \operatorname{sen} \alpha \qquad \text{com } \alpha \text{ o ângulo entre } v \text{ e } w$$

Há dois vetores com a propriedade que acabámos de descrever, que diferem apenas no seu sentido. O sentido do produto externo é dado pela *regra da mão direita:* se colocarmos a mão *direita* aberta, com os dedos que não o polegar juntos apontando na direção de v e a rodarmos de modo a que esses dedos apontem para w, o polegar aponta na direção de $v \times w$.

A razão pela qual isto é assim prende-se com o *significado geométrico do sinal do determinante de uma matriz invertível*, que é precisamente

$$\begin{vmatrix} - & v_1 & - \\ - & v_2 & - \\ - & v_3 & - \end{vmatrix} > 0 \quad \Leftrightarrow \quad v_1, v_2 \text{ e } v_3 \text{ satisfazem a regra da mão direita.}$$

Nesse caso diz-se que a *orientação* do referencial (v_1, v_2, v_3) é positiva. Note-se que o referencial canónico formado pela base canónica de $\mathbb{R}^3$ tem esta propriedade. Assim podemos pensar nos referenciais positivamente orientados como sendo "semelhantes" ao referencial habitual.

Para perceber a afirmação anterior recorde-se que podemos transformar a matriz com linhas v_1, v_2 e v_3 na matriz identidade aplicando o método do Gauss-Jordan. Cada passo do método consiste numa operação

$$L_i - \alpha L_j, \quad \alpha L_i, \quad L_i \leftrightarrow L_j \tag{53}$$

que, em termos da matriz dos coeficientes do sistema, corresponde à multiplicação à esquerda por uma matriz simples. No primeiro caso trata-se de uma matriz triangular com uma única entrada não nula fora da diagonal, no segundo caso por uma matriz diagonal com α na posição i e 1 nas restantes, e no último por uma matriz de permutação que troca as linhas i e j. O sinal do determinante da matriz dos coeficientes não é alterado pelas operações do primeiro tipo, permanece igual ou é alterado pelas do segundo tipo consoante α é positivo ou negativo, e é sempre alterado por operações do terceiro tipo (com $i \neq j$).

Resta agora observar que o efeito que as operações (53) têm relativamente à verificação da regra da mão direita por um referencial é exatamente o mesmo: operações do primeiro tipo não têm efeito no que diz respeito à verificação da regra da mão direita pelas linhas da matriz; operações do segundo tipo não têm efeito se $\alpha > 0$ e têm efeito se $\alpha < 0$; as operações do terceiro tipo têm sempre efeito. Conclui-se que o determinante é positivo sse as linhas satisfazem a regra da mão direita.

Observação B.4. *A fórmula da Definição B.1 pode ser usada para definir o produto externo de $(n-1)$ vetores em $\mathbb{R}^n$, para $n \geq 1$. Sendo $\mathbf{e_1}, \ldots, \mathbf{e_n}$ a base canónica de $\mathbb{R}^n$ e $v_1, \ldots v_{n-1}$ vetores de $\mathbb{R}^n$, define-se*

$$v_1 \times \cdots \times v_{n-1} = \begin{vmatrix} \mathbf{e_1} & \cdots & \mathbf{e_n} \\ - & v_1 & - \\ - & \cdots & - \\ - & v_{n-1} & - \end{vmatrix}$$

Por exemplo, se $n = 2$, o produto externo de um único vetor $v_1 \in \mathbb{R}^2$ dá o vetor que se obtém de v_1 rodando 90 graus no sentido anti-horário. Em geral, os argumentos acima mostram que o produto externo é nulo sse os vetores $v_1, \ldots, v_{n-1}$ forem linearmente dependentes e senão é perpendicular ao plano $(n-1)$-dimensional gerado por $v_1, \ldots, v_{n-1}$. Além disso, o comprimento do produto externo é o volume $(n-1)$-dimensional do paralelipípedo com arestas $v_1, \ldots, v_{n-1}$ e o seu sentido é tal que a orientação do referencial $v_1, \ldots, v_{n-1}, v_1 \times \cdots \times v_{n-1}$ coincide com a da base canónica de $\mathbb{R}^n$.

Para terminar mencionamos ainda outra fórmula para o volume k-dimensional de um paralelipípedo de dimensão k em $\mathbb{R}^n$ que será útil em Cálculo 2 quando se estudar a integração em superfícies (k-dimensionais) curvas.

Proposição B.5. *Sejam $v_1, \ldots, v_k \in \mathbb{R}^n$ vetores linearmente independentes. Então o volume k-dimensional do paralelipípedo P com arestas $v_1, \ldots, v_k$ é*

$$\mathrm{Vol}_k(P) = \sqrt{\det A^T A}$$

onde $A \in M_{n\times k}(\mathbb{R})$ é a matriz que tem $v_1, \ldots, v_k$ por colunas.

Dem. Sejam $w_{k+1}, \ldots, w_n$ uma base ortonormada para o complemento ortogonal do plano gerado por $v_1, \ldots, v_k$. Então o volume do paralelipípedo n-dimensional com arestas $v_1, \ldots, v_k, w_{k+1}, \ldots, w_n$ é igual ao volume k-dimensional que queremos calcular. Sendo $B \in M_{n\times n}(\mathbb{R})$ a matriz que tem por colunas os vetores $v_1, \ldots, v_k, w_{k+1}, \ldots, w_n$ (por ordem) e escrevendo B por blocos na forma $[A \mid C]$ com A a matriz formada pelas primeiras k colunas, temos

$$B^T B = \begin{bmatrix} A^T A & 0 \\ 0 & I_{n-k} \end{bmatrix}$$

(onde $C^T C = I_{n-k}$ porque os vetores w_i constituem uma base ortonormada para o plano que geram). Portanto

$$(\det B)^2 = \det(A^T A) \quad \Leftrightarrow \quad \sqrt{\det A^T A} = |\det B|$$

e, uma vez que $|\det B|$ é o volume do paralelipípedo n-dimensional com arestas $v_1, \ldots, v_k, w_{k+1}, \ldots, w_n$, isto conclui a demonstração. $\square$

Notamos que a matriz $A^T A$ no enunciado anterior é exatamente a *matriz da métrica* com respeito à base $(v_1, \ldots, v_k)$ para a restrição do produto interno usual ao plano gerado por $\{v_1, \ldots, v_k\}$.

Exemplo B.6. *A área do paralelogramo em $\mathbb{R}^3$ com arestas $(1, -2, 1)$ e $(2, 3, 0)$ é*

$$\sqrt{\det \begin{bmatrix} 1 & -2 & 1 \\ 2 & 3 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ -2 & 3 \\ 1 & 0 \end{bmatrix}} = \sqrt{\begin{vmatrix} 6 & -4 \\ -4 & 13 \end{vmatrix}} = \sqrt{62}$$

Appendix C. Projeção ortogonal e compressão de dados

A ideia fundamental utilizada na compressão de dados (por exemplo som, ou imagem) é a projeção ortogonal e baseia-se na descoberta por Joseph Fourier, um engenheiro, matemático e físico do século XIX, durante as suas investigações sobre a propagação do calor, que é possível descrever funções por meio de somas de funções trigonométricas.

Na sua expressão mais simples, suponhamos que pretendemos descrever uma função real contínua $f\colon [0, 2\pi] \to \mathbb{R}$ (que pode representar por exemplo, uma linha numa imagem, ou a intensidade do som). É fácil verificar, que com respeito ao produto interno $\langle \cdot, \cdot \rangle$ no espaço vetorial $C([0, 2\pi], \mathbb{R})$ das funções contínuas em $[0, 2\pi]$ definido por

$$\langle f, g \rangle = \int_0^{2\pi} f(x)g(x)dx$$

o conjunto

$$\{1, \operatorname{sen} x, \operatorname{sen}(2x), \ldots, \operatorname{sen}(nx), \ldots\}$$

é ortogonal. Fourier descobriu que é possível expressar qualquer função contínua como "combinação linear" destas funções[9] - aquilo a que se chama hoje uma *série de Fourier*. Intuitivamente isto significa que o conjunto acima forma uma "base ortogonal" para o espaço das funções contínuas em $[0, 2\pi]$.

A ortogonalidade permite determinar facimente os coeficientes da combinação linear correspondente a uma função f: o coeficiente segundo $\operatorname{sen}(nx)$ da função f é dado pela expressão

$$P_{\operatorname{sen}(nx)}(f) = \frac{\langle \operatorname{sen}(nx), f(x) \rangle}{\| \operatorname{sen}(nx) \|^2}$$

A ideia básica da compressão de dados é que, para armazenar a informação contida no gráfico de f basta armazenar um número suficientemente grande destes coeficientes. Quanto maior o número de coeficientes, maior a fidelidade com que conseguimos reproduzir a função f. Dados os coeficientes, reproduzir a função f corresponde em somar a expressão com os coeficientes armazenados. Desde que o número de coeficientes utilizado seja suficientemente grande será impossível ao ouvido ou olho humano distinguir entre a função original e a soma de funções trigonométricas usada para a aproximar.

[9]Trata-se de uma combinação linear infinita, ou desenvolvimento em série. A análise da convergência destas séries é delicada e constitui ainda hoje uma área da Matemática que se designa por Análise Harmónica.

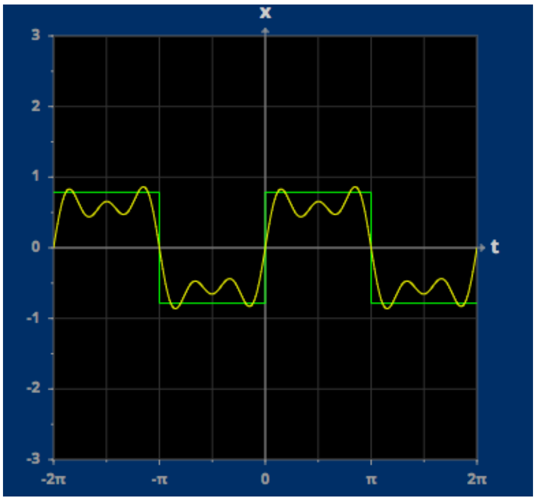


Figure 1. Aproximação de um sinal retangular por uma soma de Fourier com 5 termos.

Recomendamos a utilização do applet disponível em `http://mathlets.org/mathlets/fourier-coefficients/` (parte dos MIT Mathlets) para explorar esta ideia, que será descrita com mais detalhe e utilizada no próximo ano, no curso de Análise Complexa e Equações Diferenciais.

Appendix D. O critério de Sylvester

Seja A uma matriz simétrica $n \times n$. Dado $1 \leq i \leq n$ escrevemos A_i para a matriz que se obtém de A tomando apenas as primeiras i linhas e colunas de A. Os determinantes destas submatrizes de A chamam-se os *menores principais* de A.

Proposição D.1 (Critério de Sylvester)**.** *Seja $f\colon \mathbb{R}^n \to \mathbb{R}$ a forma quadrática determinada pela matriz simétrica $A \in M_{n\times n}(\mathbb{R})$. Então*

- *f é definida positiva sse $\det A_i > 0$ para $i = 1, \ldots, n$.*
- *f é definida negativa sse $\det A_i$ é positivo para i par e negativo para i ímpar.*

Dem. Note-se que $f(x) = x^T Ax$ é definida positiva sse $-f(x) = x^T(-A)x$ é definida negativa. Uma vez que $\det(-A_i) = (-1)^i \det A_i$ (em geral, a multilinearidade do determinante implica que $\det(\lambda A)_i = \lambda^i \det A_i$), vemos que as duas afirmações do enunciado são equivalentes. Basta portanto demonstrar a primeira.

Se f é definida positiva, a sua restrição a $\mathbb{R}^i = \{(x_1, \ldots, x_i, 0, \ldots, 0)\colon x_1, \ldots, x_i \in \mathbb{R}\}$ será também definida positiva. Mas claramente esta restrição é dada pela fórmula (com $x \in \mathbb{R}^i$)

$$f_{|\mathbb{R}^i}(x) = x^T A_i x$$

logo, para que f seja definida positiva, é necessário que $\det A_i > 0$.

Reciprocamente, suponhamos que $\det A_i > 0$ para cada $i = 1, \dots n$. Seja $i > 1$ e suponhamos indutivamente que já verificámos que $f_{|\mathbb{R}^k}$ é definida positiva para todo $k < i$ (para $k = 1$ é claro que se $\det A_1 = a_{11} > 0$ então $f_{|\mathbb{R}^1}(x_1) = a_{11}x_1^2$ é definida positiva).

Suponhamos por absurdo que $f_{|\mathbb{R}^i}$ não era definida positiva. Uma vez que, por hipótese, $\det A_i > 0$, a matriz A_i teria que ter pelo menos dois valores próprios negativos (contados com multiplicidade). Sendo $W \subset \mathbb{R}^i$ um plano gerado por dois vetores próprios independentes de A_i com valores próprios negativos, teríamos $f_{|W}(y) < 0$ para $y \in W \setminus \{0\}$.

Mas a interseção de W com $\mathbb{R}^{i-1} \subset \mathbb{R}^i$ tem dimensão pelo menos 1 pelo que existiria um vetor $y \in \mathbb{R}^{i-1} \setminus \{0\}$ com $f(y) < 0$, contradizendo a hipótese de indução que $f_{|\mathbb{R}^{i-1}}$ é definida positiva. $\square$

Exemplo D.2. *Consideremos a forma quadrática $f\colon \mathbb{R}^3 \to \mathbb{R}$ definida por*

$$f(x, y, z) = 10x^2 + 10y^2 + 10z^2 + 2xy + 2yz$$

A matriz simétrica que lhe está associada é

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

Os menores principais

$$|10| = 10, \quad \begin{vmatrix} 10 & 1 \\ 1 & 10 \end{vmatrix} = 99, \quad \begin{vmatrix} 10 & 1 & 0 \\ 1 & 10 & 1 \\ 0 & 1 & 10 \end{vmatrix} = 1000 - 20 = 980$$

são todos positivos, pelo que a forma quadrática é definida positiva.

Appendix E. A classificação das quádricas

Uma *quádrica* é uma curva em $\mathbb{R}^2$ ou uma superfície em $\mathbb{R}^3$ definida por uma equação quadrática. Podemos usar a diagonalização de matrizes simétricas para entender geometricamente estas curvas e superfícies (que irão ser exemplos básicos em Cálculo 2).

E.1. **Quádricas em $\mathbb{R}^2$.** A expressão geral de uma quádrica é

$$ax^2 + bxy + cy^2 + dx + ey + f = 0 \tag{54}$$

em que $a, b, c, d, e, f \in \mathbb{R}$. Devemos excluir alguns casos degenerados: se $a = b = c = 0$ então o conjunto descrito pela expressão anterior é uma reta se $(d, e) \neq (0, 0)$, vazio se $d = e = 0$ e $f \neq 0$, e todo o plano se $d = e = f = 0$. Consideremos portanto o caso em que os termos de grau 2 não se anulam todos. Temos

$$ax^2 + bxy + cy^2 = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & \frac{b}{2} \\ \frac{b}{2} & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

Sejam $\lambda_1, \lambda_2 \in \mathbb{R}$ os valores próprios da matriz associada à forma quadrática anterior e $(u_1, u_2), (v_1, v_2)$ os vetores próprios correspondentes, que podemos assumir formarem uma base ortonormada para $\mathbb{R}^2$. Sendo (u,v) as coordenadas no referencial determinado pelos vetores próprios temos

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} u_1 & v_1 \\ u_2 & v_2 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \tag{55}$$

Nestas coordenadas, temos

$$\begin{aligned} ax^2+bxy+cy^2 &= \begin{bmatrix} u & v \end{bmatrix} \begin{bmatrix} u_1 & u_2 \\ v_1 & v_2 \end{bmatrix} \begin{bmatrix} a & \frac{b}{2} \\ \frac{b}{2} & c \end{bmatrix} \begin{bmatrix} u_1 & v_1 \\ u_2 & v_2 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \\ &= \begin{bmatrix} u & v \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \lambda_1 u^2 + \lambda_2 v^2 \end{aligned}$$

O termo linear em (54) transforma-se mediante a mudança de coordenadas (55) num termo linear em u e v, pelo que esta mudança de coordenadas transforma (54) na seguinte equação:

$$\lambda_1 u^2 + \lambda_2 v^2 + d'u + e'v + f = 0 \tag{56}$$

Temos agora a considerar três casos:

- λ_1, λ_2 ambos diferentes de 0, com o mesmo sinal: Multiplicando (56) por -1 se necessário podemos assumir que λ_1 e λ_2 são positivos. Completando os quadrados podemos escrever a expressão na forma
$$\lambda_1(u+\tfrac{d'}{2\lambda_1})^2 + \lambda_2(v+\tfrac{e'}{2\lambda_2})^2 = f'$$
onde $f' = -f + \frac{d'^2}{4\lambda_1^2} + \frac{e'^2}{4\lambda_2^2}$. Se $f' < 0$ este conjunto é vazio, se $f' = 0$ este conjunto consiste no ponto $(-\frac{d'}{2\lambda_1}, -\frac{e'}{2\lambda_2})$, e se $f' > 0$, o conjunto é uma elipse com centro em $(-\frac{d'}{2\lambda_1}, -\frac{e'}{2\lambda_2})$ (ou uma circunferência quando $\lambda_1 = \lambda_2$).
- λ_1, λ_2 ambos diferentes de 0, com sinais opostos: Multiplicando (56) por -1 se necessário podemos assumir que λ_1 é positivo. Com uma manipulação semelhante à do caso anterior obtemos uma expressão da forma
$$\lambda_1(u-u_0)^2 + \lambda_2(v-v_0)^2 = f'$$
que, para $f' \neq 0$ é a equação de uma hipérbole[10] com "centro" em (u_0, v_0) e assíntotas dadas pelas retas
$$v - v_0 = \pm\frac{\lambda_1}{\lambda_2}(u-u_0)$$
Quando $f' = 0$ a equação reduz-se à equação das retas definidas pela expressão anterior.
- λ_1 ou λ_2 são 0: Sem perda de generalidade podemos assumir que $\lambda_2 = 0$ e que $\lambda_1 > 0$. Manipulando a expressão (56) como antes obtemos
$$\lambda_1(u-u_0) + e'v + f' = 0$$

[10]Note-se que a equação $x^2 - y^2 = 1$ se pode escrever na forma $(x-y)(x+y) = 1$ e portanto, mediante a mudança de variável linear $u = x-y, v = x+y$ é equivalente à equação mais familiar para uma hipérbole $uv = 1$.

Se $e' \neq 0$, trata-se da equação de uma parábola, cujo sentido é determinado pelo sinal de e'. Se $e' = 0$ obtemos o conjunto vazio, a reta $u = u_0$, ou duas retas paralelas a esta última, consoante $f' > 0$, $f' = 0$ ou $f' < 0$ respetivamente.

Exemplo E.2. *Consideremos o exemplo concreto da equação*

$$x^2 + 2xy + y^2 + x + 2y + 3 = 0$$

A matriz simétrica associada à forma quadrática determinada pelos termos quadráticas é

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

que tem valores próprios 0 *e* 2 *com vetores próprios* $(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})$ *e* $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$. *Fazendo a mudança de coordenadas*

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \Leftrightarrow \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

obtemos a equação

$$0 \cdot u^2 + 2v^2 + \tfrac{1}{\sqrt{2}}u + \tfrac{1}{\sqrt{2}}v + 2(-\tfrac{1}{\sqrt{2}}u + \tfrac{1}{\sqrt{2}}v) + 3 = 0 \Leftrightarrow 2v^2 - \tfrac{1}{\sqrt{2}}u + \tfrac{3}{\sqrt{2}}v + 3 = 0$$

que se pode escrever na forma

$$u = 2\sqrt{2}(v + \tfrac{3}{4\sqrt{2}})^2 + 3\sqrt{2} - \tfrac{9}{8\sqrt{2}}$$

E.3. **Quádricas em $\mathbb{R}^3$.** A expressão geral de uma quádrica é

$$ax^2 + by^2 + cz^2 + dxy + exz + fyz + gx + hy + iz + j = 0 \tag{57}$$

Novamente a análise desta superfície baseia-se na análise dos termos de grau 2 (se estes se anulam identicamente a equação define um plano, o vazio ou todo o $\mathbb{R}^3$) que constituem a forma quadrática

$$\begin{bmatrix} x & y & z \end{bmatrix} \begin{bmatrix} a & \frac{d}{2} & \frac{e}{2} \\ \frac{d}{2} & b & \frac{f}{2} \\ \frac{e}{2} & \frac{f}{2} & c \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

Num referencial ortonormado formado por vetores próprios da matriz simétrica que ocorre na expressão acima, a expressão (57) transforma-se em

$$\lambda_1 u^2 + \lambda_2 v^2 + \lambda_3 w^2 + g'u + h'v + i'z + j = 0$$

Módulo translações nos eixos dos u, v, w podemos assumir que as constantes $g' = h' = i'$ se anulam, desde que o λ_i correspondente não se anule. Temos então os seguintes casos:

- $\lambda_1, \lambda_2, \lambda_3$ todos diferentes de 0 e com sinais iguais (que podemos assumir positivos): A equação define o conjunto vazio se $j < 0$, um ponto se $j = 0$ e um *elipsóide* se $j > 0$ (trata-se da superfície que se obtém de uma superfície esférica reescalando os eixos).

- $\lambda_1, \lambda_2, \lambda_3$ todos diferentes de 0 e com sinais não todos iguais (podemos assumir que $\lambda_1, \lambda_2 > 0$ e $\lambda_3 < 0$: Os protótipos destas superfícies são as definidas pelas equações
$$x^2 + y^2 - z^2 = 1, \qquad x^2 + y^2 - z^2 = 0, \qquad x^2 + y^2 - z^2 = -1$$
Para entender a sua forma convém observar que o significado geométrico de $\sqrt{x^2 + y^2}$ é (pelo Teorema de Pitágoras) a distância do ponto (x, y, z) ao eixo dos zz. Num qualquer semiplano limitado pelo eixo dos zz podemos usar $r = \sqrt{x^2 + y^2} \geq 0$ como coordenada ao longo do semi-eixo perpendicular a Oz e a equação da interseção da nossa superfície com esse semiplano é determinada pela equação
$$r^2 - z^2 = 1, \qquad r^2 - z^2 = 0, \qquad r^2 - z^2 = -1$$
ou seja, trata-se de uma hipérbole nos casos em que o termo direito é ± 1 e de um par de semi-retas no caso restante. As superfícies que pretendemos descrever obtêm-se *rodando estas curvas em torno do eixo Oz*. Denominam-se respetivamente um *hiperbolóide*, um *cone* e um *hiperbolóide de duas folhas*.
- $\lambda_1 = 0$ e $\lambda_2, \lambda_3 \neq 0$ com o mesmo sinal que podemos assumir positivo: Os protótipos são agora da forma
$$x^2 + y^2 = j', \quad x^2 + y^2 - z = j'$$
que são respetivamente o vazio, o eixo dos zz ou um *cilindro* em torno do eixo dos zz no primeiro caso, ou um *parabolóide* (uma parábola $z = r^2 - j'$ rodada em torno do eixo dos zz).
- $\lambda_1 = 0$ e $\lambda_2, \lambda_3 \neq 0$ com sinais diferentes (podemos assumir $\lambda_2 > 0, \lambda_3 < 0$): Os protótipos são
$$x^2 - y^2 = j', \quad x^2 - y^2 - z = j'$$
No primeiro caso trata-se de um *cilindro hiperbólico*, isto é, de uma hipérbole transladada ao longo do eixo dos zz (ou no caso degenerado em que $j' = 0$, da união de dois planos concorrentes no eixo dos zz), enquanto que no segundo a superfície designa-se por uma *sela* uma vez que tem o aspeto de uma sela de um cavalo (há uma parábola virada para cima ao longo do eixo dos xx e uma decrescente ao longo do eixo dos yy).
- $\lambda_1 = \lambda_2 = 0$ e $\lambda_3 > 0$. Os protótipos são agoras as equações da forma
$$z^2 + g'x + h'y = j'$$
Se $g' = h' = 0$ esta equação define o vazio, um plano ou dois planos paralelos consoante o sinal de j'. No caso em que $(g', h') \neq 0$ define *um cilindro parabólico*, isto é a translação de uma parábola, ao longo de um eixo no plano xy perpendicular ao vetor (g', h').

References

[D] E. Dias, *Álgebra Linear*, `https://www.math.tecnico.ulisboa.pt/~edias/TextosNet/ALbookfin_Net.pdf`

[H] J. Hefferon, *Linear Algebra*, `http://joshua.smcvt.edu/linearalgebra/book.pdf`

[HK] K. Hoffman and R. Kunze, *Linear Algebra*, Prentice-Hall (1961)