Instituto Superior Técnico
Departamento de Matemática

ÁLGEBRA LINEAR

0. Introdução

Este texto tem como objetivo proporcionar um guia para as aulas de Álgebra Linear de LEBiom e LEBiol durante o primeiro semestre de 2023/2024. Não substitui os livros de texto indicados na bibliografia na página da cadeira.

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.

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.

Exemplos provavelmente menos familiares são o conjunto $\mathbb{F}_2 = \{0, 1\}$ com a soma e produto definidas tomando o resto da divisão por 2 da soma e produto usuais e $\mathbb{Q}[\sqrt{2}] = \{a + b\sqrt{2} \colon a, b \in \mathbb{Q}\} \subset \mathbb{R}$ com as operações habituais.

Date: December 21, 2023.

Mais geralmente, se p é um número primo, o conjunto $\mathbb{F}_p = \{0, 1, \ldots, p-1\}$ com a soma e produto definidos tomando o resto da divisão por p forma um corpo.

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 ou que deixe de haver variáveis, 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

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

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:

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

$$\xrightarrow{L_3 - \frac{1}{2}L_2} \left[\begin{array}{cccc|c} 0 & 1 & 0 & 3 & 1 \\ 0 & 0 & 2 & -1 & 5 \\ 0 & 0 & 0 & -\frac{9}{2} & -\frac{5}{2} \end{array}\right] \tag{4}$$

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*

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

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 quando, para todos os $1 \leq i \leq m-1$ e $0 \leq j, k \leq m$,*

$$a_{ij} = 0 \text{ para } j \leq k \quad \Rightarrow \quad a_{i+1\ j} = 0 \text{ para } j \leq k+1$$

onde estamos a deixar j, k tomar o valor 0 com o entendimento que entradas que têm um dos índices iguais a 0 são nulas.

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.

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

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 final da aplicação do 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 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, isto é, sistemas que têm exatamente as mesmas soluções.

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. Os números c_i dizem-se os *coeficientes* da combinação linear. Concretizando, a combinação linear com coeficientes 2 e -3 das equações

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

é a equação

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

[1] Em inglês "rank of a matrix".

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.

As combinações lineares envolvidas nos algoritmos 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

$$L_i' = L_j, \quad L_i' = \alpha L_i \text{ com } \alpha \neq 0, \quad \text{ ou } L_i' = L_i - \alpha L_j \text{ com } j \neq i \tag{7}$$

Note-se para utilização posterior que na terceira operação só a i-ésima linha é alterada. Em particular, $L_j' = L_j$.

A observação fulcral é que as expressões acima permitem também escrever as linhas do sistema S como combinações lineares das linhas de S':

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

(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 de uma matriz está *bem definida*).

Na realidade iremos demonstrar algo mais geral. As operações sobre as linhas de uma matriz usadas durante o método de Gauss e Gauss-Jordan designam-se por *operações elementares*. Duas matrizes A, B com as mesmas dimensões dizem-se *equivalentes mediante operações elementares* se existe uma sequência finita de operações elementares que transforma A em B. Qualquer aplicação dos métodos de Gauss e Gauss-Jordan a uma matriz A produz uma sequência de matrizes equivalentes[2] mediante operações elementares que

[2]Esta relação entre duas matrizes é um exemplo de uma *relação de equivalência* num conjunto. Dado um conjunto X, uma relação $\sim$ entre pares de elementos de X, diz-se uma relação de equivalência se para todos os $x, y, z \in X$ se verifica:

(1) Reflexividade: $x \sim x$
(2) Simetria: $x \sim y \Rightarrow y \sim x$

termina numa matriz em escada de linhas reduzida. Iremos provar que para cada matriz A existe no máximo uma matriz em escada de linhas reduzida que é equivalente a A mediante operações elementares.

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 $\neg A$ (isto é se $\neg A$ se reduzir ao absurdo). 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. Existe uma única matriz em escada de linhas reduzida equivalente a A mediante operações elementares.*

Dem. A existência é garantida pelos algoritmos de Gauss e Gauss-Jordan. Resta-nos demonstrar a unicidade. 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. As únicas matrizes em escada de linhas reduzidas com uma coluna são a matriz nula e a matriz

$$\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{8}$$

pelo que é suficiente ver que uma matriz coluna A não pode ser simultaneamente equivalente a estas duas. Isso é verdade porque a única matriz que é equivalente à matriz nula mediante operações elementares é a própria matriz nula. Isto conclui a prova da base da indução.

Para o passo da indução vamos admitir que a unicidade é válida para matrizes com n colunas. Queremos concluir que é também válida para matrizes com $n+1$ colunas. Designemos por $X_{\leq n}$ a matriz $m \times n$ que se obtém da matriz $m \times (n+1)$ X suprimindo a última coluna[3]. Observemos que:

(i) Se X está em escada de linhas reduzida, o mesmo acontece com $X_{\leq n}$.
(ii) Se X' resulta da aplicação de uma operação elementar a X então $X'_{\leq n}$ resulta da aplicação da mesma operação elementar a $X_{\leq n}$.

Seja A uma matriz $m \times (n+1)$ e suponhamos que B e C são matrizes em escada de linhas reduzida que se obtêm de A por operações elementares. Por (i), as matrizes $B_{\leq n}$ e $C_{\leq n}$ estão também em escada de linhas reduzida. Por (ii), $B_{\leq n}$ e $C_{\leq n}$ resultam da aplicação de operações elementares a $A_{\leq n}$. Por hipótese de indução conclui-se que $B_{\leq n} = C_{\leq n}$.

(3) Transitividade: $x \sim y \wedge y \sim z \Rightarrow x \sim z$

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

Vamos admitir com vista a chegar a um absurdo que $B \neq C$. Então B e C diferem numa entrada da última coluna. Seja 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

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

(pois $b_{ij} = c_{ij}$ para $j \leq n$). Como o coeficiente de x_{n+1} é não nulo, conclui-se que todas as soluções do sistema determinado por A (ou B ou C) satisfazem $x_{n+1} = 0$.

Tal só sucede num sistema de m equações a $(n+1)$ incógnitas cuja matriz dos coeficientes está em escada de linhas quando existe um pivot na última coluna (só assim x_{n+1} não é uma variável livre). Conclui-se que tanto B como C têm um pivot na coluna $n+1$.

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. Como B e C estão em escada de linhas reduzida, todas as entradas da última coluna de B e C são 0 excepto a entrada correspondente ao pivot, que é 1. Conclui-se que as últimas colunas de B e de C são iguais e portanto que $B = C$.

A afirmação anterior contradiz a nossa hipótese que $B \neq C$ e portanto conclui a demonstração do passo de indução. □

Corolário 1.13. *A característica de uma matriz está bem definida, isto é, é independente das escolhas realizadas durante a aplicação do método de Gauss.*

1.14. **Comida para pensamento.**

(1) Verifique que sabe a definição dos vários termos introduzidos (sistema equivalente, matriz, variáveis livres e dependentes, pivot, característica, matriz em escada de linhas e escada de linhas reduzida, operações elementares, combinação linear).
(2) Verifique que está à vontade com os algoritmos realizando pelo menos uma parte dos exercícios 1,7 e 10 da ficha sobre sistemas lineares.
(3) Quais são os possíveis valores para o número de soluções de um sistema linear?
(4) Se quisermos apenas saber se um dado sistema tem solução (sem necessariamente a(s) achar) como devemos proceder?
(5) Que números nos dão a informação básica sobre um sistema homogéneo?
(6) Se escolhermos aleatoriamente um sistema de 3 equações a 3 incógnitas que comportamento devemos esperar? E de um sistema de m equações a n incógnitas?
(7) Umas das operações elementares sobre as linhas é redundante. Qual?
(8) Vimos que operações elementares sobre as linhas de um sistema produzem um sistema equivalente. Será o recíproco verdade? Isto é, será que sistemas equivalentes se podem obter um do outro por operações elementares?

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 com k equações tem o aspecto seguinte

$$(9)\qquad \begin{cases} c_{11}(a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n) + \ldots + c_{1m}(a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n) = 0 \\ c_{21}(a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n) + \ldots + c_{2m}(a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n) = 0 \\ \vdots \\ c_{k1}(a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n) + \ldots + c_{km}(a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n) = 0 \end{cases}$$

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 (9). 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 (9))

$$(10)\qquad c_{i1}a_{1j} + c_{i2}a_{2j} + \ldots + c_{im}a_{mj} = \sum_{l=1}^{m} c_{il}a_{lj}$$

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* (10).

Note-se que a expressão (10) não é mais do que o *produto escalar* da linha i da matriz C com a coluna j da matriz A.

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

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 (10) 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 (10) 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

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

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 (11) consiste na multiplicação em ambos os lados da igualdade, à esquerda, 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 (11) pela matriz $m \times m$

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

2.3. **Multiplicação por blocos.** Uma outra observação sobre o produto de matrizes que é por vezes muito útil é que este pode ser realizado "por blocos". Se decompusermos duas matrizes A, B em "matrizes de matrizes", por exemplo,

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \qquad B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}$$

onde A_{11} é $m \times k$, A_{12} é $m \times l$, A_{21} é $n \times k$, A_{22} é $n \times l$, B_1 é $k \times p$ e B_2 é $l \times p$ então

$$AB = \begin{bmatrix} A_{11}B_1 + A_{12}B_2 \\ A_{21}B_1 + A_{22}B_2 \end{bmatrix}$$

Em geral, o produto de matrizes pode ser realizado desta forma desde que a decomposição em blocos dos fatores seja escolhida apropriadamente (de forma a que as multiplicações de matrizes façam sentido). Vejamos um exemplo concreto. Sejam

$$A = \begin{bmatrix} 0 & 1 & 2 \\ 3 & 4 & 5 \\ 6 & 7 & 8 \end{bmatrix} \quad B = \begin{bmatrix} 0 & 1 \\ 2 & 3 \\ 4 & 5 \end{bmatrix}$$

Decompondo A e B como

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

e

$$B = \begin{bmatrix} \begin{bmatrix} 0 \\ 2 \end{bmatrix} & \begin{bmatrix} 1 \\ 3 \end{bmatrix} \\ \\ \begin{bmatrix} 4 \end{bmatrix} & \begin{bmatrix} 5 \end{bmatrix} \end{bmatrix}$$

temos, por exemplo, que a matriz 2×1 correspondente às entradas 11 e 21 do produto AB é dada por

$$\begin{bmatrix} 0 & 1 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 0 \\ 2 \end{bmatrix} + \begin{bmatrix} 2 \\ 5 \end{bmatrix} \begin{bmatrix} 4 \end{bmatrix}$$

2.4. **Propriedades do produto de matrizes.**

Definição 2.5. *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 2.6 (Associatividade e existência de elementos neutros)**.** *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 (10) 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[4] (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 2.7 (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*:

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

Definição 2.8. *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

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

Exemplo 2.9 (A lei do corte não é válida para o produto de matrizes)**.** *Seja A a matriz $\begin{bmatrix} 2 & -1 \\ 4 & -2 \end{bmatrix}$. Então*

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

portanto, apesar de $A \neq 0$ temos

$$AA = A \cdot 0.$$

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

Definição 2.10. *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 2.11. *Seja A uma matriz $n \times n$* invertível, *C, D 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

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

$\square$

2.12. **Soma e produto por escalar.** Vamos também necessitar de outras operações com matrizes que têm uma natureza muito mais elementar do que o produto.

Definição 2.13. *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 2.14 (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 2.15 (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 2.16 (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 (10) 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. □

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 2.17. *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 2.18. *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 2.19. *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 2.20 (Unicidade da inversa)**.** *Seja A uma matriz $n \times n$. Se B e C são inversas de A então $B = C$.*

Dem. Temos

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

$\square$

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 2.21. *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 2.10 são satisfeitas:

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

e, analogamente,

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

□

2.22. **Cálculo da inversa.** 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

$$AX_i = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{12}$$

(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* (12) 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 2.23. *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:*

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

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

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

Temos

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

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

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

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* do conjunto das matrizes que troca linhas com colunas.

Definição 2.25. *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 2.26 (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 2.27. *Seja A uma matriz $n \times n$ de números reais ou complexos.*[5] *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) Para cada matriz $n \times 1$, B, o sistema $AX = B$ tem no máximo uma solução.
(iv) O sistema $AX = 0$ tem no máximo uma solução.
(v) A tem característica n

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

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

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

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

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

o que conclui a prova desta implicação.

(ii)⇒(iii) Evidente.

(iii)⇒(iv) Evidente.

(iv)⇒(v): Esta implicação é equivalente à implicação $\neg$(v) $\Rightarrow$ $\neg$(iv) 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.

(v)⇒(i): Se A tem característica n, então, aplicando os métodos de Gauss e 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.

[5] Ou mais geralmente com entradas num corpo.

- 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

$$E_k \cdots E_2 E_1 A = I_n \tag{13}$$

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 (13) 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 2.21, 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

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

Observação 2.28. *As matrizes que aparecem na demonstração da implicação ((v)⇒ (i)) no Teorema 2.27 (que correspondem às operações elementares sobre as linhas) designam-se por* matrizes elementares. *Vimos durante a demonstração que qualquer matriz invertível se pode escrever como um produto de matrizes elementares*[6].

Para terminar esta discussão sobre as matrizes observemos ainda a seguinte condição equivalente à invertibilidade.

Corolário 2.29. *Seja A uma matriz quadrada. As seguintes condições são equivalentes.*

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

Proof. Claramente (i) ⇒ (ii). Para demonstrar a implicação recíproca vamos ver que ¬(i)⇒ ¬(ii). Suponhamos então que A não é invertível. Pelo Teorema 2.27 a característica de A é menor do que n. Seja S uma matriz invertível que se obtém multiplicando sucessivamente as matrizes elementares correspondentes aos passos do método de Gauss, de forma que SA está em escada de linhas. Como a característica de A é menor do que n, a última linha de SA é nula. Escrevendo

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

vemos que o sistema

$$(SA)X = C \tag{14}$$

não tem solução. Ora (14) é equivalente ao sistema $AX = S^{-1}C$, logo o sistema $AX = B$ não tem solução quando $B = S^{-1}C$. Isto conclui a demonstração. □

Sendo A uma matriz quadrada, podemos considerar a função $X \mapsto AX$ que leva matrizes coluna em matrizes coluna. O Teorema 2.27 e o Corolário 2.29 mostram que, para uma tal função, as condições de bijetividade, injetividade e sobrejetividade são equivalentes!

2.30. **Comida para pensamento.**

(1) É importante estar à vontade com o produto de matrizes incluindo a sua interpretação em termos de operações com linhas e colunas. Façam muitos exemplos.
(2) Façam uma lista das propriedades das operações com matrizes. Que semelhanças e diferenças têm com as operações análogas para os números às quais estão habituados? Que cuidados especiais temos que ter ao resolver equações matriciais?

[6] É um bom exercício estimar o número máximo de fatores necessário para uma tal fatorização.

(3) A álgebra das matrizes permite lidar simbolicamente com sistemas. Usem este ponto de vista para verificar que a solução geral de um sistema linear se obtém a partir de uma solução particular somando-lhe a solução geral do sistema homogéneo associado (conforme o exercício 13 da ficha sobre sistemas lineares).
(4) Como se relaciona a característica de um produto com as características dos fatores? E se um dos fatores for invertível?

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 $\mathbb{R}^n = \{(x_1, \ldots, x_n) : x_i \in \mathbb{R}\}$ 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 não colineares 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 terno $(V, +, \cdot)$ consituído por um conjunto V, cujos elementos se designam por* vetores*, juntamente com duas funções*

- *Multiplicação por escalar: $\mathbb{R} \times V \xrightarrow{\cdot} V$ que a um par (α, v) associa um vetor $\alpha \cdot v$.*
- *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 $1 \cdot v = v$.
(vi) Para todos os $\alpha, \beta \in \mathbb{R}$, e $v \in V$ tem-se $\alpha \cdot (\beta \cdot v) = (\alpha\beta) \cdot v$.
(vii) Para todos os $\alpha \in \mathbb{R}$ e $v, w \in V$ tem-se $\alpha \cdot (v + w) = \alpha \cdot v + \alpha \cdot w$.
(viii) Para todos os $\alpha, \beta \in \mathbb{R}$ e $v \in V$ tem-se $(\alpha + \beta) \cdot v = \alpha \cdot v + \beta \cdot v$.

Há uma possível ambiguidade no axioma (iv) uma vez que o axioma (iii) não garante a unicidade do elemento neutro 0 para a soma. Mas na realidade é imediato verificar que este elemento (que se diz o *vetor zero*) é único: se $0'$ for um outro elemento neutro temos $0 + 0' = 0$ (aplicando o axioma (iii) ao elemento neutro $0'$ e ao elemento $v = 0$); por outro lado, $0+0' = 0'+0 = 0'$ (aplicando primeiro o axioma (ii) e depois (iii) ao elemento neutro 0 e ao elemento $v = 0'$).

Também 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.

(iii) O produto por escalar $\alpha \cdot v$ denota-se normalmente simplesmente por αv e passaremos a denotá-lo desta forma quando não haja possibilidade de confusão.

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. Mais geralmente $\mathbb{K}^n$ é um espaço vetorial sobre o corpo $\mathbb{K}$.*

(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 2.14 e 2.15. Analogamente, o conjunto das matrizes $M_{m\times n}(\mathbb{K})$ é um espaço vetorial sobre o corpo $\mathbb{K}$.

(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

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

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)*

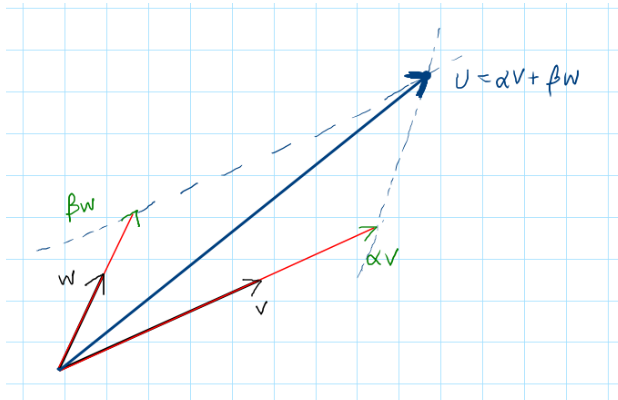


Figure 1. Uma combinação linear de dois vetores em $\mathbb{R}^3$

- *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 $F(\mathbb{R},\mathbb{R})$ 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.

3.8. Subespaços vetoriais.

Definição 3.9. *Seja V um espaço vetorial sobre $\mathbb{R}$(ou $\mathbb{C}$). Um subconjunto $W \subset V$ diz-se um* subespaço vetorial *de V se W é fechado*[7] *para as operações de V e, munido destas operações, é um espaço vetorial.*

Exemplo 3.10. *O Exemplo 3.7 (i) verifica que o conjunto dos polinómios é um subespaço vetorial de $F(\mathbb{R};\mathbb{R})$.*

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 3.11. *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 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 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 3.12. *(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 $W \neq \emptyset$ (por exemplo o polinómio 0 está em W), 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*

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

[7] Isto é, se dados $w_1, w_2 \in W$ e $\alpha \in \mathbb{K}$, temos $w_1 + w_2 \in W$ e $\alpha w_1 \in W$. Por palavras: a soma em V de vetores em W está em W, e a multiplicação em V de um vetor de W por um escalar permanece em W.

Este conjunto é um subespaço vetorial de $\mathbb{R}^n$ (o argumento é exatamente o mesmo que no exemplo anterior). Note-se que $N(A)$ é exactamente o conjunto das soluções do sistema linear homogéneo que tem A como matriz de coeficientes.

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.

3.13. **Expansão linear.**

Definição 3.14. *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 3.15. *(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

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

Os coeficientes dos polinómios em S aparecem nas primeiras três colunas. 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

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

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

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

e, como vimos acima, equação cartesiana

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

Proposição 3.16. *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 $L(S)$. Como S é não vazio, conclui-se que $L(S) \neq \emptyset$. Para ver que $L(S)$ é um subespaço vetorial precisamos agora 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 $L(S)$. 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 $L(S)$. $\square$

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.

Note-se que dado um subespaço vetorial W de V, podemos sempre encontrar um conjunto $S \subset W$ tal que $W = L(S)$: de facto podemos sempre tomar $S = W$. Esta solução não é na prática muito útil pois normalmente estaremos interessados em encontrar um conjunto de geradores tão pequeno quanto possível.

Exemplo 3.17. *(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

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

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.

3.18. **Subespaços de $\mathbb{R}^n$ associados a uma matriz.** 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

$$A = \begin{bmatrix} 2 & 0 & 1 & 4 \\ 0 & 3 & 1 & 2 \end{bmatrix} \tag{15}$$

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 3.16 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 (15) 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.

Definição 3.19. *Um espaço vetorial V diz-se* finitamente gerado *se existe um conjunto finito $S \subset V$ tal que $V = L(S)$.*

Exemplo 3.20. *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.

3.21. **Dependência linear.** Chegamos agora a um conceito fundamental da Álgebra Linear que generaliza os conceitos de colinearidade e complanaridade para vetores de $\mathbb{R}^3$.

Definição 3.22. *Seja V um espaço vetorial. Um conjunto $S \subset V$ diz-se* linearmente dependente *se existem $v_1, \ldots, v_n \in S$* todos 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 3.23. *(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 os exercícios sobre espaços vetoriais).*

(ii) Se S contém o vetor nulo então S é linearmente dependente (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 $w = -\frac{\alpha_1}{\alpha_2} v$. Reciprocamente se, por exemplo $v = \alpha w$, $v - \alpha w$ é uma combinação linear nula de v, w cujos coeficientes não são todos nulos, pelo que $\{v, w\}$ é linearmente dependente.

(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) *Sejam $e_1 = (1, 0, \ldots, 0), e_2 = (0, 1, 0, \ldots, 0), \ldots, e_n = (0, \ldots, 0, 1) \in \mathbb{R}^n$. O conjunto $B_{can} = \{e_1, e_2, \ldots, e_n\}$ 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 $L(B_{can}) = \mathbb{R}^n$ 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_{can} 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 n escalares (ou seja com $\mathbb{K}^n$).

3.24. **Bases e dimensão.** O primeiro Teorema da Álgebra Linear é que todo o espaço vetorial tem uma base e que todas as bases têm o mesmo número de elementos. Esse número chama-se a *dimensão de V*. Vamos apenas mostrar este Teorema no caso de espaços finitamente gerados, para os quais a dimensão é um número finito. O caso geral é tratado em [FIS, Secção 1.7].

Este teorema será uma consequência de certas propriedades da relação de dependência linear que passamos a explicar. 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 3.25. *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 $\alpha_1 = \cdots = \alpha_n = 0$. □

Podemos usar o resultado anterior para construir indutivamente um conjunto linearmente independente com a mesma expansão linear que um dado conjunto finito S de vetores. Mais geralmente temos o seguinte resultado.

Proposição 3.26. *Seja V um espaço vetorial e $S \subset V$ um subconjunto finito. Se $T \subset S$ é linearmente independente, podemos escolher $T' \subset S$ tal que*

- *$T \cup T'$ é linearmente independente*
- *$L(T \cup T') = L(S)$*

Dem. A demonstração é por indução no número de elementos de $S \setminus T$. Suponhamos que $S \setminus T = \{v\}$ tem apenas um elemento. Se $v \in L(T)$ então $L(S) = L(T)$ e podemos tomar $T' = \emptyset$. Se $v \notin L(T)$ podemos aplicar a Proposição 3.25 para concluir $T \cup \{v\}$ é linearmente independente e então $T' = \{v\}$ satisfaz as condições requeridas.

Suponhamos agora que a afirmação do enunciado é válida sempre que o conjunto de geradores tem mais n elementos do que o seu subconjunto linearmente independente. Suponhamos que $S \setminus T$ tem $n+1$ elementos. Se S for linearmente independente, então podemos tomar $T' = S \setminus T$. Se S é linearmente dependente podemos escolher $v_1, \ldots, v_k \in S$ distintos e escalares não todos nulos $\alpha_1, \ldots, \alpha_k$ tais que

$$\alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_k v_k = 0$$

Sem perda de generalidade podemos assumir que todos os escalares α_i são não nulos (senão podemos simplesmente suprimir os termos em que $\alpha_i = 0$). Uma vez que T é linearmente independente, nem todos os elementos v_i podem pertencer a T. Seja j tal que $v_j \in S \setminus T$ e defina-se $S' = S \setminus \{v_j\}$. Então

$$v_j = -\frac{1}{\alpha_j} \sum_{i \neq j} \alpha_i v_i \in L(S'),$$

e portanto $L(S) = L(S')$. Note-se que $T \subset S'$ e $S' \setminus T$ tem n elementos. Aplicando a hipótese de indução ao conjunto de geradores S' e ao subconjunto linearmente independente

$T \subset S'$ obtemos um conjunto linearmente independente $T' \subset S'$ tal que $T \cup T'$ é linearmente independente e $L(T \cup T') = L(S') = L(S)$. □

Por palavras, a Proposição anterior diz que dado um conjunto finito de vetores S, todo o subconjunto linearmente independente $T \subset S$ pode ser extendido a uma base para $L(S)$ contida em S. Convém apontar algumas consequências imediatas da Proposição anterior.

Corolário 3.27. *(i) Se V é um espaço vetorial e $S \subset V$ é um conjunto finito, S contém uma base de $L(S)$. Em particular todo o espaço vetorial finitamente gerado tem uma base com um número finito de elementos.*
(ii) Se V é um espaço vetorial finitamente gerado, todo o subconjunto linearmente independente $T \subset V$ está contido numa base.

Dem. (i) Aplicando a Proposição 3.26 com $T = \emptyset$ obtemos uma base T' para $L(S)$ contida em S. Se V for um espaço vetorial finitamente gerado, podemos escolher um subconjunto finito $S \subset V$ tal que $L(S) = V$, e portanto V tem uma base finita.
(ii) Sendo S' um conjunto finito de geradores para V, o conjunto $S' \cup T$ é ainda um conjunto de geradores. Aplicando a Proposição 3.26 ao conjunto $S = S' \cup T$ e ao subconjunto linearmente independente $T \subset S$, obtemos um subconjunto $T' \subset S'$ tal que $T \cup T'$ é uma base de V.

□

Vejamos agora que todas as bases de um espaço vetorial finitamente gerado têm o mesmo número de elementos. A chave da demonstração é o seguinte resultado. Recomendamos que leia o enunciado seguinte com um exemplo simples em mente: $V = \mathbb{R}^2$ ou $\mathbb{R}^3$ e $n = 1$ ou $n = 2$.

Lema 3.28. *Seja V um espaço vetorial e S um subconjunto de V com m elementos. Seja T um subconjunto linearmente independente de $L(S)$ com n elementos. Então $n \leq m$ e existe um subconjunto $T' \subset S$ com $m - n$ elementos tal que $L(T \cup T') = L(S)$.*

Dem. A demonstração é por indução no número de elementos de T. Quando $n = 0$ não há nada a provar, pois $0 \leq m$ e podemos tomar $T' = S$.

Suponhamos que o resultado é válido quando o subconjunto linearmente independente tem n elementos. Seja S um conjunto com m elementos e seja $T = \{v_1, \ldots, v_{n+1}\} \subset L(S)$ um conjunto linearmente independente com $n + 1$ elementos. Uma vez que $\{v_1, \ldots, v_n\}$ é linearmente independente, por hipótese de indução temos que $n \leq m$. Se n fosse igual a m, a hipótese de indução garantiria que $L(\{v_1, \ldots, v_n\}) = L(S)$ (pois o conjunto T' teria 0 elementos) mas então v_{n+1} pertenceria a $L(\{v_1, \ldots, v_n\})$ contradizendo a hipótese de T ser linearmente independente. Conclui-se que $n + 1 \leq m$.

Por hipótese de indução, existem $u_1, \ldots, u_{m-n}$ vetores de S tais que

$$L(\{v_1, \ldots, v_n, u_1, \ldots, u_{m-n}\}) = L(S)$$

Como $v_{n+1} \in L(S)$ existem escalares $\alpha_1, \ldots, \alpha_n$ e $\beta_1, \ldots, \beta_{m-n}$ tais que

$$v_{n+1} = \alpha_1 v_1 + \ldots + \alpha_n v_n + \beta_1 u_1 + \ldots + \beta_{m-n} u_{m-n}$$

Mas T é linearmente independente, logo algum dos coeficientes β_i tem de ser não nulo. Então

$$u_i = -\frac{1}{\beta_i}\left(\sum_{j=1}^n \alpha_j v_j + \sum_{j\neq i} \beta_j u_j\right) \in L(T \cup \{u_j \colon j \neq i\})$$

Segue-se que $L(S) = L(T \cup \{u_j \colon j \neq i\})$. Portanto podemos tomar $T' = \{u_j \colon j \neq i\}$, o que conclui a demonstração. ☐

Teorema 3.29. *Seja V um espaço vetorial finitamente gerado. Então todas as bases de V são conjuntos finitos com o mesmo número de elementos.*

Dem. Seja $S \subset V$ um conjunto finito com $L(S) = V$. Pelo Corolário 3.27 (i), S contém uma base B para V. Seja n o número de elementos de B. Uma vez que $L(B) = V$, o Lema 3.28 garante que qualquer subconjunto linearmente independente de V tem no máximo n elementos. Seja B' outra base para V. Uma vez que B' é linearmente independente, B' tem no máximo n elementos. Mas $L(B') = V$, e $B \subset V$ é linearmente independente pelo que uma nova aplicação do Lema 3.28 mostra que $n = \sharp B \leq \sharp B'$. Conclui-se que B e B' têm o mesmo número de elementos. ☐

Definição 3.30. *O número de elementos de qualquer base de um espaço finitamente gerado chama-se a* dimensão de V *e denota-se por* $\dim V$ *ou* $\dim_{\mathbb{K}}(V)$ *para enfatizar o corpo dos escalares. Se um espaço vetorial V não tem uma base finita, diz-se que tem* dimensão infinita.

Note-se que um espaço vetorial tem dimensão infinita se e só se não é finitamente gerado, ou equivalentemente é de dimensão finita se e só se é finitamente gerado. Esta consequência imediata do Corolário 3.27(i) fica como exercício.

Exemplo 3.31. *À luz do Exemplo 3.23(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 3.29 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 3.32. *A dimensão do espaço $M_{2\times 4}(\mathbb{R})$ é 8. De facto é imediato verificar que as oito matrizes*

$$E_{11} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, E_{12} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \ldots, E_{24} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

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 3.33. *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 de* V *tem pelo menos* n *elementos.*
(iii) *Qualquer conjunto linearmente independente tem no máximo* n *elementos. Equivalentemente, todo o conjunto com mais de* n *elementos é linearmente dependente.*

Dem. Pelo Corolário 3.27(i) todo o conjunto de geradores contém uma base e tem portanto pelo menos n elementos, o que prova (ii). Pelo Corolário 3.27(ii) todo o conjunto linearmente independente está contido numa base e portanto tem no máximo n elementos, o que prova (iii). Para ver (i), seja T um conjunto linearmente independente com n elementos. Pelo Corolário 3.27(ii) existe uma base B contendo T. Como todas as bases têm n elementos, $B = T$ portanto T é uma base. □

Observação 3.34. *Com excepção do Corolário 3.33(i), todos os resultados demonstrados acima que assumem que o espaço vetorial 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 no cálculo de bases e na determinação se um conjunto é ou não linearmente dependente.

Exemplo 3.35. *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 3.33(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 3.23(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 Corolário 3.33 (iii) B será linearmente dependente. Por outro lado, se $\dim L(B) = \#B$, B não pode ser linearmente dependente porque, se assim fosse, o Corolário3.27 (i) garantiria a existência de uma base para $L(B)$ com menos elementos que B o que contradiria o Teorema 3.29.*

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

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

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

Exemplo 3.36. *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 3.23(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 3.33, 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* $L(S)$*.*

3.37. **Decomposições em soma direta.** Como iremos ver mais tarde, é frequentemente útil decompor espaços vetoriais numa soma de subespaços. Trata-se de um procedimento que abstrai a familiar decomposição de um vetor no plano ou no espaço como a soma das suas componentes ao longo de eixos de um referencial dado (ver Figura 1).

Definição 3.38. *Seja* V *um espaço vetorial e* U, W *subespaços de* V*. Diz-se que* V *é a* soma direta *de* U *com* W*, e escreve-se* $V = U \oplus W$ *se*

(i) Todo o elemento $v \in V$ *se pode escrever como a soma* $v = u + w$ *com* $u \in U$ *e* $w \in W$*.*
(ii) A decomposição do ponto anterior é única, isto é se $v = u + w$ *e* $v = u' + w'$ *com* $u, u' \in U$ *e* $w, w' \in W$ *então* $u = u'$ *e* $w = w'$*.*

Na Figura 1 o plano $V = L(\{v, w\})$ é a soma direta dos seus subespaços de dimensão 1 dados por $U = L(\{v\})$ e $W = L(\{w\})$. A seguinte proposição dá critérios úteis para verificar que temos uma decomposição em soma direta.

Proposição 3.39. *Seja* V *um espaço vetorial e* U, W *subespaços de* V*. As seguintes afirmações são equivalentes:*

(i) $V = U \oplus W$*.*

(ii) $V = U + W$ *(isto é, todo o vetor de* V *se pode escrever como a soma de um vetor de* U *e de um vetor de* W*) e* $U \cap W = \{0\}$.

Se V *é finitamente gerado*[8] *então as afirmações anteriores são ainda equivalentes a*

(iii) Existe uma base $\{v_1, \ldots, v_n\}$ *para* V *e* $k \leq n$ *tais que* $\{v_1, \ldots, v_k\}$ *é uma base para* U *e* $\{v_{k+1}, \ldots, v_n\}$ *é uma base para* W.

Dem: $(i) \Rightarrow (ii)$: Suponhamos que $V = U \oplus W$. Então, em particular, $V = U + W$. Dado $v \in U \cap W$ temos duas decomposições de v como soma de um vetor de U e de um vetor de W, nomeadamente $v = v + 0$ e $v = 0 + v$. Uma vez que a decomposição é única conclui-se que $v = 0$ e portanto $U \cap W = \{0\}$.

$(ii) \Rightarrow (i)$: Temos a mostrar que se $V = U + W$ e $U \cap W = \{0\}$ então a decomposição de um vetor $v \in V$ como a soma de um vetor em U e outro em W é única. Suponhamos que $v \in V$ se pode decompor como

$$v = u + w \quad \text{ e } v = u' + w' \qquad \text{com } u, u' \in U,\ w, w' \in W$$

Então temos

$$u + w = u' + w' \Rightarrow u - u' = w' - w$$

O vetor do lado esquerdo do sinal de igual está em U (porque a diferença de vetores em U está em U) e o do lado direito em W. Como são iguais, trata-se de um vetor em $U \cap W = \{0\}$. Portanto $u - u' = 0 = w' - w$, ou seja, $u = u'$ e $w = w'$ pelo que a decomposição de v é única.

Suponhamos agora que V é finitamente gerado e vejamos primeiro que $(i) \Rightarrow (iii)$: Uma vez que V tem dimensão finita, o mesmo é verdade dos seus dois subespaços U, W. Sejam $\{u_1, \ldots, u_k\}$ e $\{w_1, \ldots, w_l\}$ bases de U e W. É suficiente verificar que $B = \{u_1, \ldots, u_k, w_1, \ldots, w_l\}$ é uma base para V.

Como B contém uma base de U temos $U \subset L(B)$ e analogamente $W \subset L(B)$. Uma vez que $L(B)$ é fechado para a soma conclui-se que $V = U + W \subset L(B)$ logo B gera V. Resta ver que B é linearmente independente. Sejam $\alpha_1, \ldots, \alpha_k, \beta_1, \ldots, \beta_l$ escalares tais que

$$\alpha_1 u_1 + \ldots + \alpha_k u_k + \beta_1 w_1 + \ldots + \beta_l w_l = 0$$

Então

$$\alpha_1 u_1 + \ldots + \alpha_k u_k = -\beta_1 w_1 - \ldots - \beta_l w_l$$

O vetor do lado esquerdo do sinal de igual pertence a U e o do lado direito a W. Sendo iguais pertencem a $U \cap W = \{0\}$. Uma vez que $\{u_1, \ldots, u_k\}$ e $\{w_1, \ldots, w_l\}$ são linearmente independentes segue-se que os α_i's e os β_j's são 0 e portanto B é linearmente independente.

Resta ver que $(iii) \Rightarrow (ii)$. Dado $v \in V$ podemos escrevê-lo de forma única como

$$v = \alpha_1 v_1 + \ldots + \alpha_n v_n = (\alpha_1 v_1 + \ldots + \alpha_k v_k) + (\alpha_{k+1} v_{k+1} + \ldots + \alpha_n v_n)$$

[8] Para V arbitrário pode mostrar-se que V é a soma direta de U e W se e só se existe uma base B para V e uma decomposição de B como a união de dois subconjuntos B_1 e B_2 tais que B_1 é uma base para U e B_2 é uma base para V.

logo v é a soma de um vetor em U com um vetor em W, e portanto $V = U + W$. Dado $v \in U \cap W$, têm que existir escalares $\alpha_1, \ldots, \alpha_k$ tais que $v = \alpha_1 v_1 + \ldots + \alpha_k v_k$ e escalares $\beta_1, \ldots, \beta_{n-k}$ tais que $v = \beta_1 v_{k+1} + \ldots + \beta_{n-k} v_n$. Mas então

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

e sendo B linearmente independente isto só pode acontecer se todos os α_i's e β_j's forem 0. Conclui-se que $v = 0$ e portanto $U \cap W = \{0\}$ conforme pretendido. □

Exemplo 3.40. *Seja $V = F(\mathbb{R}, \mathbb{R})$ o espaço vetorial de todas as funções reais de variável real. Se $U = \{f \in V \colon f(x) = f(-x)\}$ é o subespaço vetorial das funções pares e $W = \{f \in V \colon f(x) = -f(-x)\}$ o subespaço vetorial formado pelas funções ímpares temos*

$$V = U \oplus W$$

De facto a expressão

$$f(x) = \frac{f(x) + f(-x)}{2} + \frac{f(x) - f(-x)}{2}$$

mostra que qualquer função $f \in V$ se pode escrever como a soma de uma função par e de uma função ímpar e, dado que a única função que é simultaneamente par e ímpar é a função identicamente nula, temos $U \cap W = \{0\}$.

Podemos mais geralmente dizer que V é a soma direta dos seus subespaços $U_1, \ldots, U_k$ se todo o vetor $v \in V$ se puder escrever de forma única como uma soma de vetores $u_i \in U_i$. Fica como exercício enunciar e demonstrar a generalização natural da Proposição 3.39 a estas decomposições em mais fatores.

3.41. **Mudanças de coordenadas.**

Definição 3.42. *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,

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

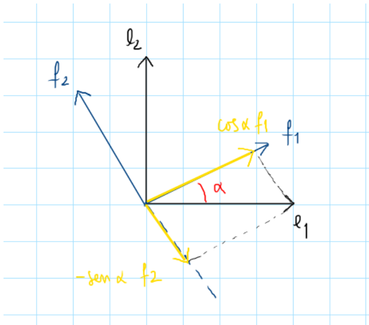
O escalar α_i diz-se *a i-ésima coordenada* de v na base ordenada B.

Exemplo 3.43. *(i) A base ordenada canónica de* $\mathbb{R}^n$ *é* $B_{can} = (e_1, \ldots, e_n)$*, 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:

$$\begin{array}{rcl} v_1 & = & a_{11}w_1 + a_{21}w_2 + \ldots + a_{n1}w_n \\ v_2 & = & a_{12}w_1 + a_{22}w_2 + \ldots + a_{n2}w_n \\ \vdots & & \\ v_n & = & a_{1n}w_1 + a_{2n}w_2 + \ldots + a_{nn}w_n \end{array}$$

Substituindo na fórmula para v obtemos

$$\begin{array}{rcl} v & = & \alpha_1(a_{11}w_1 + a_{21}w_2 + \ldots + a_{n1}w_n) + \alpha_2(a_{12}w_1 + a_{22}w_2 + \ldots + a_{n2}w_n) + \\ & & \ldots + \alpha_n(a_{1n}w_1 + a_{2n}w_2 + \ldots + a_{nn}w_n) \\ & = & (a_{11}\alpha_1 + a_{12}\alpha_2 + \ldots + a_{1n}\alpha_n)w_1 + (a_{21}\alpha_1 + a_{22}\alpha_2 + \ldots + a_{2n}\alpha_n)w_2 + \\ & & \ldots + (a_{n1}\alpha_1 + a_{n2}\alpha_2 + \ldots + a_{nn}\alpha_n)w_n \end{array}$$

Escrevendo $(\beta_1, \ldots, \beta_n)$ para as coordenadas do vetor v na base B_2 temos portanto

$$\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}$$

onde na coluna j da matriz $[a_{ij}]$ aparecem as coordenadas do vetor v_j na base B_2.

Proposição 3.44. *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 3.45. *A matriz de mudança de base da base canónica B_{can} de $\mathbb{R}^2$ para a base B do Exemplo 3.43 é dada por*

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

De facto, a primeira coluna contém as componentes do primeiro vetor da base canónica na base B como vimos no Exemplo 3.43 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 3.46. *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.

□

Exemplo 3.47. *Vamos determinar a matriz de mudança de coordenadas $S_{B_{can}\to B}$ da base canónica de $\mathbb{R}^3$ para a base $B = ((1,0,1),(1,0,-1),(1,1,1))$. É fácil escrever a matriz $S_{B\to B_{can}}$ (só temos que escrever os vetores da base B por ordem nas colunas da matriz):*

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

Pela Proposição 3.46 temos

$$S_{B_{can}\to B} = S^{-1}_{B\to B_{can}} = \begin{bmatrix} \frac{1}{2} & -1 & \frac{1}{2} \\ \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix}$$

Por exemplo, as coordenadas do vetor $(1,2,3)$ na base B são $(0,-1,2)$.

Observação 3.48. *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 verificar que qualquer matriz invertível é uma matriz de mudança de base (a partir de qualquer base dada).*

3.49. **Comida para pensamento.**

(1) Certifique-se que está à vontade com a definição e exemplos básicos dos muitos termos introduzidos nesta secção: espaço vetorial, subespaço vetorial, expansão linear de um conjunto, conjunto de geradores, conjuntos linearmente (in)dependentes, bases, dimensão, coordenadas, matriz de mudança de coordenadas. Recorde também a definição dos vários subespaços associados a uma matriz: o núcleo, espaço das linhas e espaço das colunas.
(2) Certifique-se que sabe realizar os seguintes procedimentos: verificar se um conjunto munido de duas operações apropriadas é um espaço vetorial; verificar se um subconjunto de um espaço vetorial é um subespaço; ver se um vetor pertence à expansão linear de um dado conjunto, ver se um conjunto gera um subespaço, é linearmente (in)dependente ou uma base; calcular a dimensão de um espaço vetorial; determinar as coordenadas de um vetor numa base e a matriz de mudança de coordenadas de uma base para outra.
(3) Dados dois subespaços U, V de $\mathbb{R}^n$ de dimensões m, k respetivamente, que valores pode tomar a dimensão da sua intersecção? O que acontece se os dois planos forem escolhidos aleatoriamente?
(4) Vimos que a característica de uma matriz é a dimensão do plano gerado pelas linhas da matriz. A posição dos pivots relaciona a posição desse plano no espaço com os eixos coordenados. Como? Interprete o Teorema 1.12 como uma parametrização eficiente desses conjuntos de planos.
(5) Podemos pensar no operador L de expansão linear como uma operação que leva subconjuntos de um espaço vetorial em subespaços. Que propriedades tem este operador com respeito às várias operações que podemos fazer nos conjuntos e subespaços de um espaço vetorial (intersecção, união, soma)?
(6) Consegue generalizar os teoremas sobre existência e cardinalidade das bases ao caso em que o espaço em questão tem um conjunto numerável de geradores?
(7) As propriedades dos conjuntos linearmente independentes que foram usadas acima para demonstrar os teoremas sobre existência e cardinalidade das bases podem ser abstraídos no conceito de *matróide*. Este conceito está relacionado com a teoria de grafos que as pessoas do curso de Matemática estudarão no segundo período.

4. Transformações lineares

Em cada área da Matemática estudamos um certos tipo de objetos matemáticos de natureza variável. Por exemplo, em Álgebra Linear estudamos 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 4.1. *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 4.2. *(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* $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ *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* $f\colon \mathbb{R}^n \to \mathbb{R}^m$ *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*

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

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

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

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:

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

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

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 4.3. *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 $f: V \to W$ tal que $f(v_i) = w_i$.*

Dem. Começamos por mostrar a unicidade. Suponhamos que $f: V \to W$ 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

$$f(v) = f(\alpha_1 v_1 + \ldots + \alpha_n v_n) = \alpha_1 f(v_1) + \ldots + \alpha_n f(v_n) = \alpha_1 w_1 + \ldots + \alpha_n w_n \tag{16}$$

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 (16) define efetivamente uma transformação linear com as propriedades requeridas. Seja então $f: V \to W$ a função definida pela expressão (16).

- 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

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

  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.

□

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 ((16)) 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 4.4. *A Proposição 4.3 é ainda válida quando a base de V é um conjunto infinito, sendo a demonstração essencialmente a mesma. Deixamos esta verificação como exercício.*

Exemplo 4.5. *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 4.2 (3) é na realidade exaustivo.

4.6. **Representação matricial de uma transformação linear.** 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 4.7. *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 4.8. *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 4.9. *Sejam* V, W *espaços vetoriais e* $B_1 = (v_1, \ldots, v_n)$ *e* $B_2 = (w_1, \ldots, w_m)$ *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 4.10. *(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 4.9. Vejamos primeiro 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 $m \times n$ cuja i-ésima coluna é $[f(v_i)]_{B_2}$ satisfaz a equação do enunciado. Seja $v = \alpha_1 v_1 + \ldots + \alpha_n v_n$ um vetor de V. Então

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

onde na segunda igualdade usámos o facto de f ser uma transformação linear e na terceira o Exercício 4.8. 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. □

A Proposição 4.9 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 da imagem 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 de funções diferenciáveis arbitrárias. 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.

4.11. **Operações com transformações lineares e a sua tradução em matrizes.** As transformações lineares podem ser combinadas através de várias operações que agora passamos a descrever.

Definição 4.12. *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 4.13. *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 4.14. *Sejam V, W, U espaços vetoriais e $f\colon V \to W$, e $g\colon W \to U$ transformações lineares. Então a função composta*

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

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 4.15. *Sejam V, W espaços vetoriais e $f\colon V \to W$ uma transformação linear. Se a função f é invertível (isto é, se é bijetiva) então a função inversa $f^{-1}\colon W \to V$ 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 $f^{-1}\colon W \to V$ preserva a soma de vetores. A verificação que f^{-1} preserva o produto por escalar é análoga e fica como exercício. $\square$

Observação 4.16. *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 4.17. *Sejam V, W espaços vetoriais. Uma transformação linear invertível $f\colon V \to W$ 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 da ficha sobre transformações lineares.

Exemplo 4.18. *(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*

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

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 $B = (v_1, \dots, v_n)$. 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 4.8. A função f é também bijetiva: a sobrejetividade de f traduz o facto que qualquer n-tuplo $(\alpha_1, \dots, \alpha_n)$ de escalares formar as coordenadas de um vetor de V (nomeadamente de $v = \alpha_1 v_1 + \dots \alpha_n v_n$), 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 $B_1 = (v_1, \dots, v_n), B_2 = (w_1, \dots, w_m)$ 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 4.9 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. Dado $v \in V$ temos*

$$[(f+g)(v)]_{B_2} = [f(v)+g(v)]_{B_2} = [f(v)]_{B_2} + [g(v)]_{B_2} \tag{17}$$

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

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

Das igualdades (17) *e* (18) *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 4.9 que a matriz $\Phi(f)$ tem como i-ésima coluna $[f(v_i)]_{B_2}$. Dada uma matriz A, pela Proposição 4.3 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 4.3 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. Em linguagem corrente, esta afirmação significa que os dicionários entre transformações lineares e matrizes (dados pela escolha de uma base no espaço de chegada e outra no espaço de partida) traduzem a soma e a multiplicação de um dos lados na mesma operação do outro lado.

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". O mesmo se pode dizer para um subespaço vetorial típico de $\mathbb{R}^n$. Pensar num vetor em termos das suas coordenadas numa base é análogo a pensar numa ideia através da palavra que é usada para designar a ideia numa dada língua.

Proposição 4.19. *Sejam V, W, U espaços vetoriais, B_1, B_2, B_3 bases ordenadas para V, W, U respetivamente, e $f\colon V \to W$, $g\colon W \to U$ 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 4.18(iii) da composição de funções, que é uma operação associativa.

Observação 4.20. *É possível pensar visualmente na correspondência entre transformações lineares e matrizes, e em particular na Proposição anterior da seguinte forma. Considere-se o diagrama*

(19)
$$\begin{tikzcd} V \arrow[r, "f"] \arrow[d, "{[\cdot]_{B_1}}"', "\cong"] & W \arrow[d, "{[\cdot]_{B_2}}"', "\cong"] \\ M_{n\times 1}(\mathbb{R}) \arrow[r, "A_{f,B_1,B_2}"] & M_{m\times 1}(\mathbb{R}) \end{tikzcd}$$

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 4.18(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

(20)
$$[f(v)]_{B_2} = A_{f,B_1,B_2}[v]_{B_1}$$

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 (19): do lado esquerdo *de* (20) *temos o efeito de seguir primeiro a seta de cima e depois a seta da direita; do lado direito de* (20) *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* (20) *traduz a comutatividade de* (19)*.*

Nestes termos, a Proposição 4.19 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 4.21. *Sejam V, W espaços vetoriais, $f\colon V \to W$ 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). □

Para terminar registamos a interação das operações de composição de transformações lineares com as de soma e produto por escalar.

Proposição 4.22. *Sejam V, W, U espaços vetoriais, $f, g \in L(V, W), h, k \in L(W, U)$ e $\alpha \in \mathbb{K}$. Então*

(i) $h \circ (f + g) = h \circ f + h \circ g$ *e* $(h + k) \circ f = h \circ f + k \circ f$
(ii) $h \circ (\alpha f) = (\alpha h) \circ f = \alpha(h \circ f)$

Dem. Exercício. □

Estas propriedades traduzem-se na distributividade do produto de matrizes em relação à soma e na comutatividade do produto por escalar com o produto de matrizes. Notamos ainda que estas propriedades implicam que uma transformação linear $h\colon W \to U$ determina uma transformação linear

$$h_*\colon L(V, W) \to L(V, U)$$

definida pela expressão $h_*(\phi) = h \circ \phi$ e analogamente, $f\colon V \to W$ determina uma transformação linear

$$f^*\colon L(W, U) \to L(V, U)$$

definida por $f^*(\psi) = \psi \circ f$.

4.23. **Subespaços vetoriais associados a uma transformação linear.**

Definição 4.24. *Seja $f\colon V \to W$ uma transformação linear. O* núcleo de f *é o conjunto*

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

e a imagem de f *é o conjunto*

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

Proposição 4.25. *Seja $f\colon V \to W$ 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 4.26. *Uma transformação linear $f\colon V \to W$ 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 4.26 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.

Exemplo 4.27. *Seja* $V = \{f \in F(\mathbb{R}, \mathbb{R}) \colon f$ *é diferenciável*$\}$ *o espaço vetorial das funções diferenciáveis e* $T \colon V \to F(\mathbb{R}, \mathbb{R})$ *a transformação linear definida por* $T(f) = f'$.

O núcleo de T *é o subespaço das funções constantes de* V, *e tem portanto dimensão* 1. *A transformação* T *não é portanto injetiva.* T *também não é sobrejetiva. De facto a derivada de uma função tem necessariamente a seguinte propriedade do valor médio (cuja demonstração é um bom exercício de aplicação do Teorema de Rolle): se* $f'(a) < 0$ *e* $f'(b) > 0$ *então existe um real* c *entre* a *e* b *tal que* $f'(c) = 0$.

Segue-se que, por exemplo, a função definida por

$$g(x) = \begin{cases} -1 & \text{se } x < 0 \\ 1 & \text{se } x \geq 0 \end{cases}$$

não é a derivada de nenhuma função e não está portanto na imagem de T. *É interessante notar que não foi encontrada até à data qualquer caracterização conveniente da imagem da transformação* T.

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 (19). 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 (20) 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 (20) 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.

4.28. **O Teorema da característica-nulidade.** Chegamos agora a um dos resultados básicos da Álgebra Linear, cuja importância se irá tornando clara com o desenrolar do semestre.

Teorema 4.29. *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). Pelo Corolário 3.27(ii) 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. O caso em que $n = k$ verifica-se imediatamente pelo que vamos assumir a partir de agora que $n > k$.

- $\{f(v_{k+1}), \ldots, f(v_n)\}$ 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

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

  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.

□

Exemplo 4.30. *Vamos determinar uma base para o núcleo e imagem da transformação linear $f\colon M_{2\times 2}(\mathbb{R}) \to M_{3\times 2}(\mathbb{R})$ definida por*

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

Temos

$$f\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = 0 \Leftrightarrow \begin{bmatrix} a-2c & b-2d \\ 3a-6c & 3b-6d \\ 2a-4c & 2b-4d \end{bmatrix} = 0 \Leftrightarrow \begin{cases} a = 2c \\ b = 2d \end{cases}$$

logo

$$N(f) = \left\{ \begin{bmatrix} 2c & 2d \\ c & d \end{bmatrix} : b, d \in \mathbb{R} \right\} = L\left(\left\{ \begin{bmatrix} 2 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 2 \\ 0 & 1 \end{bmatrix} \right\} \right)$$

Conclui-se que $\left\{ \begin{bmatrix} 2 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 2 \\ 0 & 1 \end{bmatrix} \right\}$ *é uma base para* $N(f)$*. Pelo Teorema 4.29*

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

Podemos obter uma base para $\mathrm{Im}(f)$ *completando uma base de* $N(f)$ *com dois elementos de forma a obter uma base de* $M_{2\times 2}(\mathbb{R})$ *e calculando a imagem por* f *desses dois novos elementos. Claramente*

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

é uma base de $M_{2\times 2}(\mathbb{R})$*, pelo que*

$$\left\{ f\left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\right), f\left(\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\right) \right\} = \left\{ \begin{bmatrix} 1 & 0 \\ 3 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 3 \\ 0 & 2 \end{bmatrix} \right\}$$

é uma base da imagem de f*.*

Definição 4.31. *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 4.29 é conhecido em inglês por "the rank-nullity Theorem". Tem o seguinte corolário extremamente útil:

Corolário 4.32. *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 4.29 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 3.33(i)) e portanto f é também sobrejetiva.

Suponhamos agora que f é sobrejetiva, ou seja que $\dim f(V) = \dim W$. Aplicando o Teorema 4.29 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 4.26, f é injetiva. □

4.33. **Aplicações ao estudo das matrizes.** Em vista da interpretação da imagem de uma transformação linear f como o espaço das colunas da matriz que a representa, o Teorema 4.29 tem a seguinte consequência importante (que está longe de ser óbvia!).[9]

Proposição 4.34. *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 4.29 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 $f(V)$.

Observação 4.35. *Também podemos deduzir a Proposição 4.34 da seguinte forma. Ela é claramente verdadeira se A é uma matriz em escada de linhas visto que as dimensões de $EC(A)$ e $EL(A)$ são nesse caso ambas iguais ao número de pivots. Ora, embora o espaço das colunas se vá alterando quando aplicamos o método de Gauss, a sua* dimensão *não se altera!*

De facto, duas matrizes A_i e A_{i+1} que estejam relacionadas por uma operação elementar sobre as linhas satisfazem $A_{i+1} = SA_i$ para alguma matriz invertível S. Sendo m o número de linhas das matrizes, a multiplicação à esquerda por S dá-nos um isomorfismo $\phi_S\colon \mathbb{R}^m \to \mathbb{R}^m$. Pela definição do produto matricial, ϕ_S leva a j-ésima coluna de A_i na j-ésima

[9] Para uma explicação conceptual desta igualdade que é independente da nossa discussão inicial dos sistemas lineares e do método de Gauss ver os exercícios da ficha sobre transformações lineares.

coluna de A_{i+1} e, em particular, transforma $EC(A_i)$ em $EC(A_{i+1})$. Uma vez que ϕ_S é um isomorfismo, a sua restrição a $EC(A_i)$ é ainda um isomorfismo e portanto temos

$$\dim EC(A_i) = \dim \phi_S(EC(A_i)) = \dim EC(A_{i+1})$$

Conclui-se que a dimensão do espaço das colunas não se altera ao longo do método de Gauss. Mais, uma vez que as colunas respetivas da matriz inicial e da matriz em escada de linhas no final do método de Gauss estão relacionadas por um isomorfismo, as colunas da matriz inicial correspondentes às colunas com pivot no final do método de Gauss formam uma base para $EC(A)$.

Exemplo 4.36. *Consideremos a matriz*

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

Aplicando o método de Gauss obtemos

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

Uma vez que os pivots ocorrem nas colunas 1 e 3, temos que uma base para o $EC(A)$ é dado por $\{(1,2,1),(-1,1,2)\}$ enquanto que uma base para $EL(A)$ é dado por $\{(1,2,-1,2),(0,0,3,-1)\}$.

Podemos agora aproveitar para atualizar os nossos critérios para a invertibilidade de uma matriz (comparem com o Teorema 2.27)

Proposição 4.37. *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 $BA = I_n$*

Dem. A equivalência das primeiras três afirmações foi já vista no Teorema 2.27 embora a equivalência de (i) com (iii) possa agora ser interpretada conceptualmente como uma consequência da Proposição 4.15 e Corolário 4.21. A equivalência de (iii), (iv) e (v) é uma consequência do Corolário 4.32 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. □

4.38. **Equações lineares.**

Definição 4.39. *Uma* equação linear *é uma equação da forma*

$$f(x) = w$$

onde $f: V \to W$ 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 4.40 (Princípio da sobreposição)**.** *Seja $f: V \to W$ uma transformação linear. Se v é uma solução da equação linear $f(x) = w$, o conjunto de todas as soluções é*

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

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

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

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

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

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 4.40 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 4.41 (O oscilador harmónico)**.** *Seja $x: \mathbb{R} \to \mathbb{R}$ 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*

$$x''(t) + kx(t) = 0 \tag{21}$$

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 (21) *(com $k = 1$) que formam portanto um subespaço vetorial de V.*

É fácil adivinhar duas soluções para a equação

$$x'' + x = 0 \tag{22}$$

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

$$x(t) = \alpha_1 \cos t + \alpha_2 \operatorname{sen} t, \qquad \textit{com } \alpha_1, \alpha_2 \in \mathbb{R} \tag{23}$$

são soluções.

Para o ano que vem irão aprender que uma solução de uma equação diferencial como (21) *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* (23) *descreve a solução geral da equação* (22).

No caso da equação (22) *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 (22). *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 (22) *com $x(0) = y(0)$ e $x'(0) = y'(0)$ então $u(t) = x(t) - y(t)$ é também uma solução de* (22) *(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 (23) *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* (22).

Suponhamos agora que queremos resolver a equação[10]

$$x'' + x = t^3 \tag{24}$$

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 4.40 diz-nos então que a solução geral da equação (24) *é*

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

4.42. **Comida para pensamento.**

(1) Certifique-se que está à vontade com a definição de transformação linear e as operações definidas (soma, produto por escalar, composição e inversão), assim como a sua representação matricial e com o conceito de isomorfismo. Idem para os subespaços associados a uma transformação linear: o núcleo e a imagem e a sua relação com a solução de equações lineares. Assegure-se que sabe o enunciado e percebe a demonstração do Teorema da característica-nulidade, e que entende a forma como simplifica a verificação que uma dada transformação linear entre espaços vetoriais finitamente gerados é um isomorfismo.
(2) Revisite a matéria coberta nas duas primeiras secções (Sistemas lineares e operações com matrizes) assim como os respetivos exercícios à luz dos novos conceitos de espaço vetorial e transformação linear.
(3) Em termos práticos, o estudo dos espaços vetoriais finitamente gerados e das transformações lineares entre eles pode ser reduzido ao estudo dos espaços $\mathbb{R}^n$ e das matrizes. Analogamente, a que se reduz o estudo dos conjuntos finitos e funções entre eles? Parece uma boa ideia pensar sempre em termos desta redução?
(4) Sendo V, W espaços finitamente gerados, quando é que duas transformações lineares $T, T' \colon V \to W$ são equivalentes (no sentido em que existem isomorfismos $\phi \colon V \to V, \psi \colon W \to W$ tais que $T\phi = \psi T'$)? Quando é que duas matrizes $m \times n$ representam a mesma transformação linear com respeito a algumas bases nos espaços de chegada e partida?
(5) Seja V um espaço vetorial de dimensão finita. Se $T \colon V \to V$ é uma transformação linear, dado um vetor $v \in V$ podemos associar ao par (T, v) um subespaço de V - o espaço gerado pelo conjunto $\{v, Tv, T^2v, \ldots\}$. Haverá alguma restrição nos subespaços que podemos obter assim? Se escolhermos v e T aleatoriamente que espaço deverá ser este?
(6) Para $V = \mathbb{R}^2$, é fácil de ver que há infinitos $T \colon V \to V$ tal que $T^2 = -I$. Consegue ver que não há soluções desta equação se V tiver dimensão 3?

[10]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).

5. O Determinante

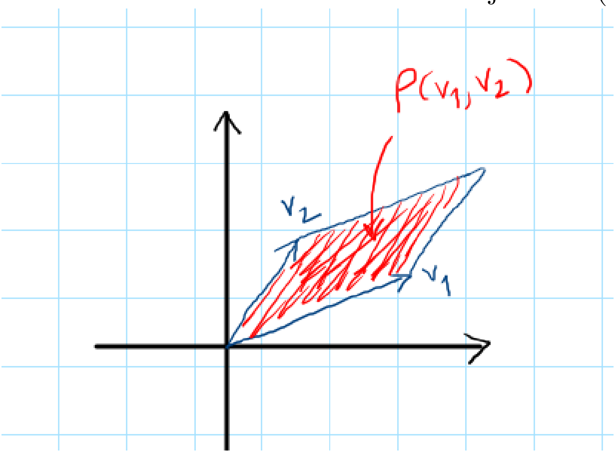
Nesta secção vamos estudar um critério para a invertibilidade de uma matriz quadrada. O critério irá dizer que uma matriz $A \in M_{n\times n}(\mathbb{K})$, onde $\mathbb{K} = \mathbb{R}$ (ou $\mathbb{C}$), é invertível se e só se uma certa expressão complicada das entradas da matriz (chamada o determinante da matriz) não se anula.

5.1. **Motivação.** Para motivar esta expressão vamos começar por discutir o caso em que o corpo $\mathbb{K}$ é o dos números reais, caso em que o determinante tem uma interpretação geométrica. Consideremos primeiro os casos $n = 2$ e $n = 3$.

A uma matriz $A \in M_{2\times 2}(\mathbb{R})$ podemos associar um paralelogramo

$$P(v_1, v_2) = \{\alpha v_1 + \beta v_2 \colon 0 \leq \alpha, \beta \leq 1\} \subset \mathbb{R}^2$$

gerado pelas linhas v_1 e v_2 da matriz. A matriz será invertível se o seu espaço das linhas for $\mathbb{R}^2$, ou, equivalentemente, se o paralelogramo $P(v_1, v_2)$ não degenerar num segmento de reta (ou até na origem). Apelando ao conceito intuitivo de área podemos dizer que a matriz será invertível se a área do conjunto $P(v_1, v_2)$ for não nula.



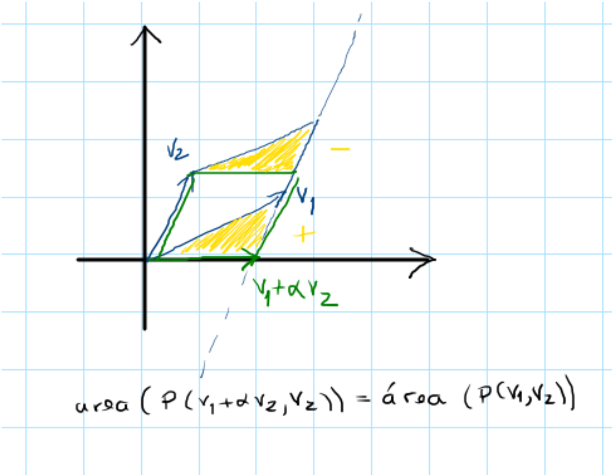
Analogamente, uma matriz 3×3 terá característica menor que 3 se e só se o paralelipípedo

$$P(v_1, v_2, v_3) = \{\alpha v_1 + \beta v_2 + \gamma v_3 \colon 0 \leq \alpha, \beta, \gamma \leq 1\}$$

gerado pelas linhas v_1, v_2, v_3 da matriz 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 para a invertibilidade de uma matriz em $M_{n\times n}(\mathbb{R})$ é que o volume n-dimensional do paralelipípedo n-dimensional $P(v_1, \ldots, v_n)$ gerado pelas linhas da matriz seja não nulo.

Se conseguirmos obter uma fórmula para o volume n-dimensional do paralelipípedo gerado por n vetores em $\mathbb{R}^n$ isso dar-nos-á um critério para a invertibilidade da matriz: que o volume do paralelipípedo gerado pelas linhas seja não nulo. A observação básica que nos permite obter esta fórmula é a seguinte:

> Ao deslizar em conjunto duas arestas de um paralelogramo ao longo da reta gerada por outra das outras arestas, a área do paralelogramo não se altera



ou seja

$$\text{área}(P(v_1, v_2)) = \text{área}(P(v_1 + \alpha v_2, v_2)) \tag{25}$$

(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 determinados pelas linhas das matrizes

$$\begin{bmatrix} a & 0 \\ c & d \end{bmatrix} \quad \text{e} \quad \begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix}$$

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 (25) 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

$$\text{área}\ (P((a, b), (c, d))) = |a| \cdot \left| d - \frac{bc}{a} \right| = |ad - bc|$$

É um exercício simples verificar que esta fórmula permanece válida mesmo quando $a = 0$. Obtemos assim a condição desejada nas entradas da matriz:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \text{ é invertível sse } ad - bc \neq 0$$

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, simplificando, 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.

5.2. **A função determinante.** 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, que faz sentido para um corpo arbitrário.

Definição 5.3. *Seja $\mathbb{K} = \mathbb{R}$ (ou $\mathbb{C}$). Uma função* determinante *para as matrizes $n \times n$ com entradas em $\mathbb{K}$ é uma função*

$$\det \colon M_{n \times n}(\mathbb{K}) \to \mathbb{K}$$

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,*

$$\begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ \alpha a_{i1} & & \alpha a_{in} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix} = \alpha \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{i1} & & a_{in} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix}$$

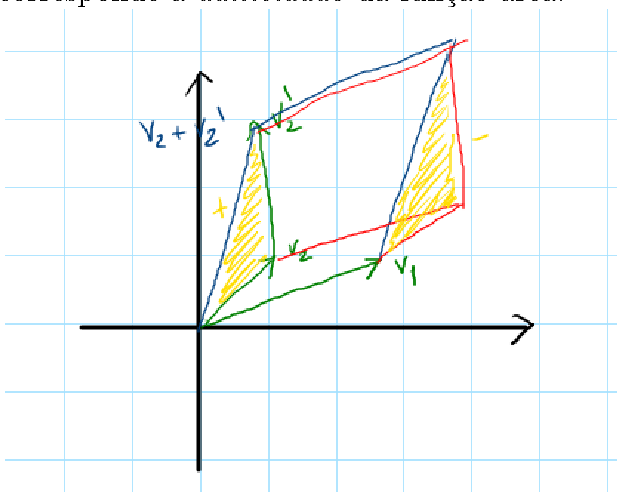
(ii) **Alternância:** $\det A = 0$ *se duas linhas da matriz* A *forem iguais.*

(iii) **Normalização:** $\det I_n = 1$.

Em concreto, no caso das matrizes 2×2, a primeira propriedade diz por exemplo que

$$\begin{vmatrix} 2 & 1 \\ 1+3 & 2+4 \end{vmatrix} = \begin{vmatrix} 2 & 1 \\ 1 & 2 \end{vmatrix} + \begin{vmatrix} 2 & 1 \\ 3 & 4 \end{vmatrix} \quad \text{e} \quad \begin{vmatrix} 2 & 1 \\ -2\cdot 3 & -2\cdot 4 \end{vmatrix} = -2\begin{vmatrix} 2 & 1 \\ 3 & 4 \end{vmatrix}$$

e corresponde à *aditividade* da função área.



Identificando as linhas de uma matriz $n \times n$ com vetores de $\mathbb{K}^n$, podemos pensar na função determinante como uma função $D\colon \mathbb{K}^n \times \cdots \times \mathbb{K}^n \to \mathbb{K}$ que associa um escalar a um n-tuplo $(v_1, \ldots, v_n)$ de vetores de $\mathbb{K}^n$ (v_i é a i-ésima linha da matriz). Deste ponto de vista, a propriedade de multilinearidade escreve-se

$$D(v_1, \ldots, \alpha v_i + \beta v_i', \ldots, v_n) = \alpha D(v_1, \ldots, v_i, \ldots, v_n) + \beta D(v_1, \ldots, v_i', \ldots, v_n) \tag{26}$$

onde $v_1, \ldots, v_n \in \mathbb{K}^n$ são vetores arbitrários e α, β escalares arbitrários. A equação (26) diz que, para cada i entre 1 e n, a função $D_i\colon \mathbb{K}^n \to \mathbb{K}$ 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 $\mathbb{K}^n$).

Em geral, uma função $D\colon V \times \cdots \times V \to \mathbb{K}$ satisfazendo (26) diz-se uma *função multilinear*[11] (é linear em cada variável independentemente).

A razão para o nome da segunda propriedade na definição de determinante é a seguinte.

Proposição 5.4. *Seja $D\colon V \times \cdots \times V \to \mathbb{K}$ 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.

$(i) \Rightarrow (ii)$ Supondo que $i < j$, e aplicando a linearidade primeiro na i-ésima variável e depois na j-ésima obtemos

$$\begin{aligned} D(v_1, \ldots, v_i + v_j, \ldots, v_i + v_j, \ldots, v_n) &= D(v_1, \ldots, v_i, \ldots, v_i + v_j, \ldots, v_n) + \\ & \qquad D(v_1, \ldots, v_j, \ldots, v_i + v_j, \ldots, v_n) = \\ &= D(v_1, \ldots, v_i, \ldots, v_i, \ldots, v_n) + D(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n) \\ &\quad + D(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_n) + D(v_1, \ldots, v_j, \ldots, v_j, \ldots, v_n) \end{aligned}$$

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 $(1+1)D(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n) = 0$.

□

5.5. **Existência e unicidade do determinante.** Vamos ver que as propriedades (i) a (iii) na definição de determinante especificam completamente uma função.

Teorema 5.6. *Existe uma única função determinante* $\det M_{n \times n}(\mathbb{K}) \to \mathbb{K}$

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

[11] Também se chama um tensor-n covariante em V.

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 $a, b, c, d \in \mathbb{K}$ 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 5.3.

Façamos agora o caso mais realista de uma matriz 3×3. Assumindo que existe a função determinante e usando linearidade na primeira linha obtemos

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a \begin{vmatrix} 1 & 0 & 0 \\ d & e & f \\ g & h & i \end{vmatrix} + b \begin{vmatrix} 0 & 1 & 0 \\ d & e & f \\ g & h & i \end{vmatrix} + c \begin{vmatrix} 0 & 0 & 1 \\ d & e & f \\ g & h & i \end{vmatrix} \tag{27}$$

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 (27) 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

$$ae \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ g & h & i \end{vmatrix} = ae \left(g \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{vmatrix} + h \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{vmatrix} + i \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} \right) = aei$$

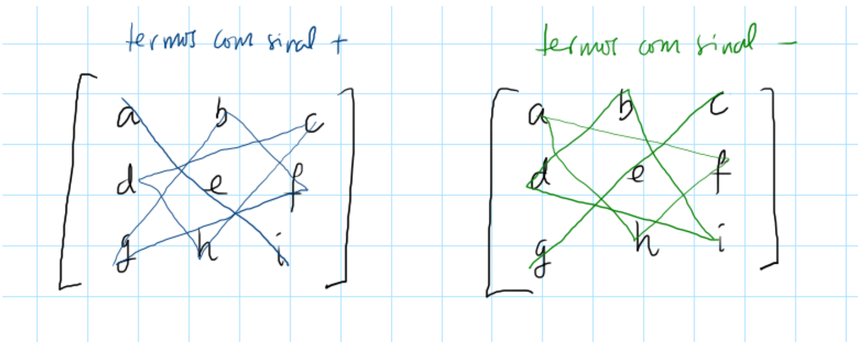
uma vez que os dois primeiros termos da soma anterior têm linhas repetidas e o determinante da matriz identidade é 1. Aplicando o mesmo raciocínio para os restantes termos não nulos na expansão até à segunda linha obtemos a seguinte expressão para o determinante:

$$aei + afh\begin{vmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{vmatrix} + bdi\begin{vmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{vmatrix} + bfg\begin{vmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{vmatrix} + cdh\begin{vmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{vmatrix} + ceg\begin{vmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{vmatrix}$$

Os determinantes das matrizes com 0s e 1s são ± 1 consoante o número de vezes que 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:

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = aei - afh - bdi + bfg + cdh - ceg$$

Os sinais da fórmula anterior podem ser memorizados usando a seguinte mnemónica (que se chama a *regra de Sarrus*):



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 em cada coluna. Para descrever estes termos por meio de uma expressão necessitamos de alguma terminologia.

Definição 5.7. *Uma* permutação *do conjunto* $\{1, \dots, n\}$ *é uma função bijetiva*

$$\sigma \colon \{1, \dots, n\} \to \{1, \dots, n\}$$

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:

$$\det(A) = \sum_{\sigma \in \Sigma_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)} \tag{28}$$

O argumento anterior torna claro que se existir uma função determinante, ela é única (tem que ser dada pela fórmula (28)!). 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

$$\operatorname{sgn}(\sigma) = \det A(\sigma) \qquad \text{com } A(\sigma) \text{ a matriz com entradas } a_{ij} = \begin{cases} 1 & \text{se } j = \sigma(i) \\ 0 & \text{caso contrário.} \end{cases} \tag{29}$$

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 $A(\sigma)$ à esquerda por um vetor linha.

Dem. do Teorema 5.6. Já vimos que se existir uma função determinante ela é única (e dada pela fórmula (28)). 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

$$\det(A) = a_{11} \det(A_{11}) - a_{12} \det(A_{12}) + \ldots + (-1)^n a_{1(n+1)} \det A_{1(n+1)} \tag{30}$$

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 5.3. A condição (i) é verificada porque a expressão (30) 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 (30) 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 (30) 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 (30) 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 (30) 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 (30) 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 (30) 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. $\square$

Observação 5.8. *Uma função $f \colon M_{n \times n}(\mathbb{K}) \to \mathbb{K}$ satisfazendo as propriedades (i) e (ii) na Definição 5.3 chama-se uma função multilinear alternante das linhas da matriz. 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 $\lambda \in \mathbb{K}$ 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 $\lambda = f(I_n)$.

5.9. **Propriedades do determinante.** Vamos agora ver algumas propriedades importantes do determinante que nos ajudam a calculá-lo.

Definição 5.10. *Seja A uma matriz $n \times n$. Para $1 \leq i, j \leq n$ 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 5.11 (Propriedades do determinante)**.** *Sejam A e B matrizes $n \times n$.*

(i) **Expansão de Laplace:** *Sendo $1 \leq i \leq n$, 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 5.12 (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 5.13. *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 5.11 (iv) diz-nos que

$$A\left(\frac{1}{\det A}(\operatorname{cof} A)^T\right) = I_n$$

pelo que A é invertível (cf. Proposição 4.37 (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

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \qquad \text{quando } ad - bc \neq 0$$

É ainda razoável aplicar a fórmula para matrizes 3×3 mas se o tamanho das matrizes é maior ou igual a 4 a fórmula começa a tornar-se impraticável e é muito mais rápido usar o método de Gauss-Jordan.

Dem. da Proposição 5.11. (i) Para $i = 1$ a expansão de Laplace é simplesmente a expressão indutiva (30) 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 (30) obtemos

$$\det(A) = -\det(\tilde{A}) = -\sum_{j=1}^{n}(-1)^{1+j}\tilde{a}_{1j}\det(\tilde{A}_{1j}) = -\sum_{j=1}^{n}(-1)^{1+j}a_{ij}\det(\tilde{A}_{1j}) \tag{31}$$

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 (31) 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 5.8 diz-nos que $f(A) = \det(A)\det(B)$.

(iii) A expressão (28) 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

$$\det(A^T) = \sum_{\sigma\in\Sigma_n} \operatorname{sgn}(\sigma) a_{1\sigma^{-1}(1)}\cdots a_{n\sigma^{-1}(n)} \tag{32}$$

As matrizes $A(\sigma)$ associadas às permutações (ver (29)) 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 o determinante de uma matriz de permutação é necessariamente ± 1. Notando que $\operatorname{sgn}(\sigma) = \det A(\sigma)$ e substituindo em (32) 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 (28) 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 5.14. *É instrutivo pensar em escrever explicitamente a igualdade indicada na Proposição 5.11(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 5.11(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 5.15. *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

$$-\begin{vmatrix} 2 & 0 & 0 \\ 1 & 4 & 7 \\ 1 & 8 & 3 \end{vmatrix} = -2\cdot(-1)^{1+1}\begin{vmatrix} 4 & 7 \\ 8 & 3 \end{vmatrix} = -2(4\cdot 3 - 7\cdot 8) = 88.$$

5.16. **Aplicação à solução de sistemas lineares.** 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 5.17 (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 5.13 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 5.18. *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

$$y = \frac{\begin{vmatrix} 2 & 3 & 1 \\ 1 & 4 & 1 \\ 1 & 5 & -1 \end{vmatrix}}{\begin{vmatrix} 2 & 3 & 1 \\ 1 & -1 & 1 \\ 1 & 2 & -1 \end{vmatrix}} = -\frac{11}{7}$$

5.19. **O determinante de uma matriz triangular por blocos.** Recorde que 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).

Usando a expansão de Laplace e indução, é imediato verificar 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*.

Proposição 5.20. *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. Exercício. □

Exemplo 5.21.

$$\begin{vmatrix} 1 & 0 & 5 & 11 & 6 \\ 3 & 2 & 3 & 27 & 5 \\ 0 & 0 & 4 & 2 & 2 \\ 0 & 0 & 0 & 3 & 2 \\ 0 & 0 & 0 & 1 & 4 \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 3 & 2 \end{vmatrix} \cdot 4 \cdot \begin{vmatrix} 3 & 2 \\ 1 & 4 \end{vmatrix} = 2 \cdot 4 \cdot 10 = 80$$

5.22. **O produto externo de vetores.**

Definição 5.23. *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 5.24.

$$(1, -3, 2) \times (5, 0, 2) = \left| \begin{matrix} \mathbf{e_1} & \mathbf{e_2} & \mathbf{e_3} \\ 1 & -3 & 2 \\ 5 & 0 & 2 \end{matrix} \right] = (-6, 8, 15)$$

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. Vamos usar a notação $\langle v, w \rangle$ para o produto interno de dois vetores de $v, w \in \mathbb{R}^3$ familiar do ensino secundário e definido por

$$\langle (v_1, v_2, v_3), (w_1, w_2, w_3) \rangle = v_1w_1 + v_2w_2 + v_3w_3$$

Mais à frente iremos estudar em detalhe este produto interno assim como as suas generalizações a espaços vetoriais reais e complexos.

Proposição 5.25 (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 também o significado geométrico do produto externo. De facto, uma vez que o determinante de uma matriz com linhas repetidas é 0, pela propriedade (iv) temos

$$\langle v, v \times w\rangle = \langle w, v \times w\rangle = 0$$

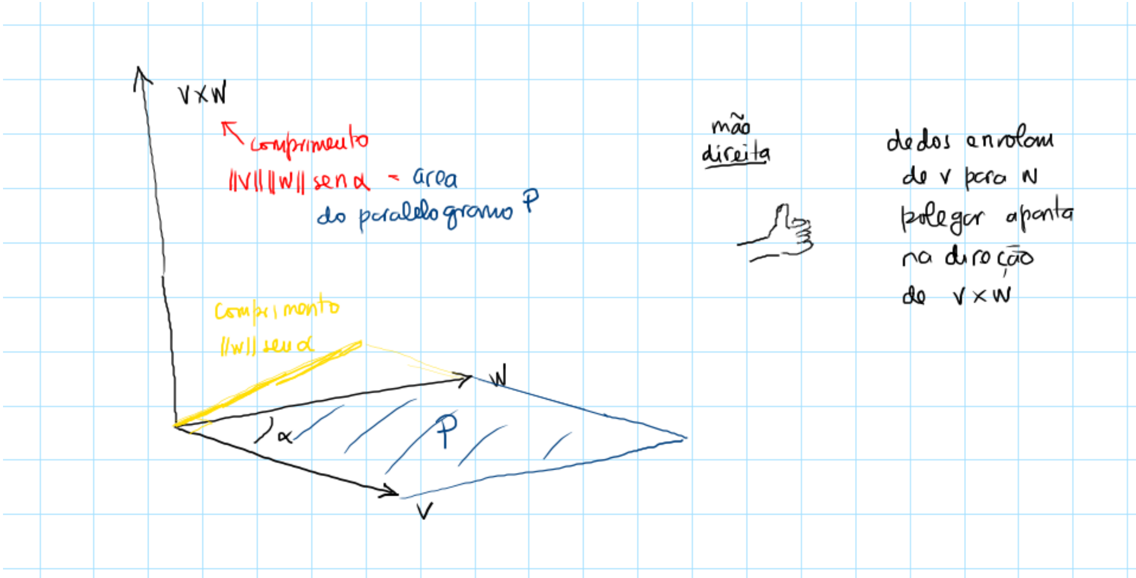
pelo que $v \times w$ *é ortogonal ao plano gerado por* v *e* w (se v e w são colineares, então as propriedade (i) e (iii) dizem-nos que o produto externo é o vetor nulo). Além disso, dada a interpretação do determinante como o volume do paralelipípedo temos que

$$\|v \times w\|^2 = \langle v \times w, v \times w\rangle = \begin{vmatrix} - & v \times w & - \\ - & v & - \\ - & w & - \end{vmatrix}$$

é o volume do paralelipípedo com base o paralelogramo formado por v e w sendo a outra aresta perpendicular ao paralelogramo e 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* 3×3 *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

$$L_i - \alpha L_j, \quad \alpha L_i, \quad L_i \leftrightarrow L_j \tag{33}$$

que, em termos da matriz dos coeficientes do sistema, corresponde à multiplicação à esquerda por uma matriz elementar. 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 (33) 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 5.26. *A fórmula da Definição 5.23 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 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 $(v_1 \times \cdots \times v_{n-1}, v_1, \ldots, v_{n-1})$ coincide com a da base canónica de $\mathbb{R}^n$.

6. Endomorfismos

Vamos agora iniciar um estudo detalhado das transformações lineares $T\colon V \to V$ em que o espaço de chegada é o mesmo que o espaço de partida. Estas transformações designam-se por endomorfismos (do grego endon - dentro) porque aplicam o espaço V para dentro de si próprio.

Estas transformações desempenham um papel especialmente importante na Matemática e nas suas aplicações, em parte porque codificam simetrias (por exemplo rotações do espaço $\mathbb{R}^3$), e talvez principalmente, porque podem ser usadas para descrever evolução temporal: se o vetor $v \in V$ codificar o estado de um sistema, podemos por vezes codificar o estado desse sistema após a passagem de uma unidade de tempo por $T(v)$, após outra unidade de tempo por $T^2(v) = T(T(v))$, etc...

Exemplo 6.1. *Suponhamos que um vetor $(x, y) \in \mathbb{R}^2$ codifica o estado de um sistema e que a evolução após uma unidade de tempo é descrita pela transformação linear*

$$T(x, y) = (2x, \tfrac{y}{2})$$

Para todo o $n \in \mathbb{Z}$ temos $T^n(x, y) = (2^n x, \frac{y}{2^n})$ pelo que os estados atingidos por um sistema com estado inicial (x_0, y_0) fora dos eixos são todos pontos da hipérbole $xy = x_0 y_0$. Quando $n \to +\infty$ o estado tende para ∞ ao longo do eixo dos xx (quando x_0 é 0, converge para $(0, 0)$).

Exemplo 6.2. *(Cadeias de Markov) Suponhamos que um sistema tem n estados possíveis e que temos uma população de tais sistemas. Seja x_i a percentagem da população que se encontra no estado i e suponhamos que podemos determinar a probabilidade p_{ij} de, ao longo de uma unidade de tempo, um sistema evoluir do estado j para o estado i.*

Por exemplo os sistemas podem ser pessoas, os estados podem ser 1-viver em Lisboa; 2-viver fora de Lisboa; x_1 é então a percentagem da população que vive em Lisboa e $x_2 = 1 - x_1$. Se cada ano, 3% da população se mudar para fora de Lisboa e 1% da população exterior se mudar para Lisboa, a evolução anual da distribuição da população é codificada pela operação

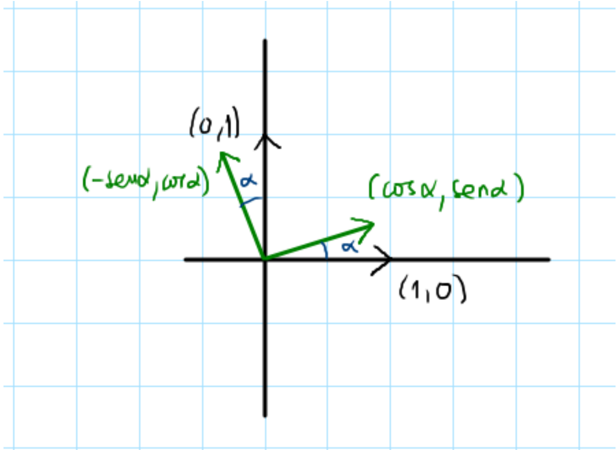
$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \mapsto \begin{bmatrix} 0.97 & 0.01 \\ 0.03 & 0.99 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Note-se que uma matriz $n \times n$ com entradas p_{ij} construída pelo procedimento descrito acima tem entradas todas não negativas e que a soma dos elementos em cada coluna é 1. Uma tal matriz chama-se uma matriz de Markov*.*

6.3. **Subespaços invariantes.** Para analisar um endomorfismo $T: V \to V$ vamos usar uma tática habitual em matemática: decompor T em objetos da mesma natureza, mas tão simples quanto possível.

Definição 6.4. *Seja $T: V \to V$ um endomorfismo. Um subespaço vetorial $W \subset V$ diz-se um* subespaço invariante de T *se $T(W) \subset W$.*

Exemplo 6.5. *(i) Se $T: \mathbb{R}^3 \to \mathbb{R}^3$ é uma rotação de um ângulo α em torno de um eixo L que passe pela origem, tanto o eixo L como o plano perpendicular que passa pela origem são subespaços invariantes.*



Concretamente, no caso em que o eixo é o dos zz, temos

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} \mapsto \begin{bmatrix} \cos\alpha & -\operatorname{sen}\alpha & 0 \\ \operatorname{sen}\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

e claramente $W_1 = \{(x, y, 0) : x, y \in \mathbb{R}\}$ e $W_2 = \{(0, 0, z) : z \in \mathbb{R}\}$ são subespaços invariantes.

(ii) *Para qualquer endomorfismo $T\colon V \to V$ temos que $\{0\}$ e V são espaços invariantes, ditos* triviais*. Exemplos mais interessantes são o núcleo $N(T)$ e a imagem* $\mathrm{Im}(T)$*. De facto, $T(N(T)) = \{0\} \subset N(T)$ e claramente $T(\mathrm{Im}(T)) \subset \mathrm{Im}(T)$.*

(iii) *Dado $v \in V$, consideremos o conjunto $S = \{v, T(v), T^2(v), \ldots\} \subset V$. Então $L(S) \subset V$ é um subespaço invariante: de facto, dado*

$$\alpha_0 v + \ldots + \alpha_n T^n(v) \in L(S)$$

temos

$$T(\alpha_0 v + \ldots + \alpha_n T^n(v)) = \alpha_0 T(v) + \ldots + \alpha_n T^{n+1}(v) \in L(S)$$

Este espaço chama-se o subespaço cíclico *determinado pelo vetor $v \in V$.*

Recorde-se dos Exercícios 11 e 21 da ficha sobre espaços vetoriais que um espaço V se diz a soma direta de dois subespaços $W_1, W_2 \subset V$ se $V = W_1 + W_2$ e $W_1 \cap W_2 = \{0\}$. Nesse caso escrevemos $V = W_1 \oplus W_2$. Cada vetor de V pode então decompor-se de forma única como a soma de um vetor de W_1 e de um vetor de W_2. No Exemplo 6.5 (i) acima, o espaço $\mathbb{R}^3$ é a soma direta do eixo de rotação e do plano ortogonal ao eixo.

Se conseguirmos decompor V como uma soma direta $W_1 \oplus W_2$ de espaços invariantes, então sendo $T_1 = T_{|W_1}\colon W_1 \to W_1$, $T_2 = T_{|W_2}\colon W_2 \to W_2$ as transformações induzidas por T nos subespaços invariantes teremos, em termos da decomposição única de um vetor $v \in V$ como a soma $v = x + y$ com $x \in W_1$ e $y \in W_2$

$$T(x + y) = T_1(x) + T_2(y)$$

e a análise do comportamento de T reduz-se à análise do comportamento de T_1 e T_2. Podemos pensar neste processo como uma "separação de variáveis": o comportamento de T como função das duas variáveis x e y é completamente determinado pelo comportamento de duas funções de apenas uma variável.

Indutivamente podemos tentar analogamente decompor T_1 e T_2 até que isso deixe de ser possível, ou seja até expressar T como uma soma direta de endomorfismos "atómicos". O nosso objetivo vai ser descrever estes últimos. Os detalhes de como fazer isso dependem muito do corpo de base $\mathbb{K}$ pois, como iremos ver, este problema está fortemente relacionado com o problema de fatorizar polinómios com coeficientes no corpo. Por questões de tempo teremos de nos concentrar nos casos em que o corpo é $\mathbb{R}$ ou $\mathbb{C}$ mas tentaremos dar alguma indicação de como o problema se pode resolver em geral.

6.6. **Valores próprios e vetores próprios.** Os subespaços invariantes (não triviais) mais simples são os que têm dimensão 1. Se v for um elemento não nulo de um tal espaço, teremos necessariamente $T(v) = \lambda v$ para algum escalar λ.

Definição 6.7. *Seja V um espaço vetorial sobre o corpo $\mathbb{K}$ e $T\colon V \to V$ um endomorfismo. Um vetor não nulo $v \in V \setminus \{0\}$ diz-se um* vetor próprio de T *se existe $\lambda \in \mathbb{K}$ tal que $T(v) = \lambda v$. Nesse caso λ diz-se um* valor próprio de T *associado ao vetor próprio v.*

Também podemos falar de valores e vetores próprios de uma matriz $A \in M_{n\times n}(\mathbb{K})$: são os vetores e valores próprios do endomorfismo de $\mathbb{K}^n$ que é representado na base canónica por A.

Exemplo 6.8. *(i) Um vetor não nulo segundo um eixo de uma rotação de $\mathbb{R}^3$ é um vetor próprio com valor próprio* 1*.*

(ii) Se $P\colon V \to V$ é uma projeção (isto é se $P^2 = P$) então todos os vetores não nulos do plano de projeção $\mathrm{Im}(P)$ são vetores próprios com valor próprio 1*.*

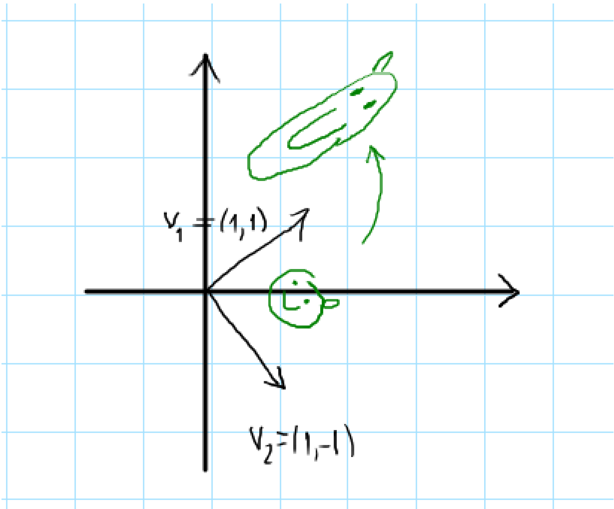
(iii) Os vetores não nulos do núcleo $N(T)$ (se existirem) são vetores próprios de T com valor próprio 0*.*

(iv) Seja $V = C^\infty(\mathbb{R})$ o espaço vetorial das funções reais de variável real indefinidamente diferenciáveis e $D\colon V \to V$ a operação de derivação definida por $D(f) = f'$. Todos os números reais λ são valores próprios de D. Os vetores próprios correspondentes a λ são as funções exponenciais $ce^{\lambda t}$ com $c \neq 0$.

Uma vez que
$$T(v) = \lambda v \Leftrightarrow T(v) = \lambda \,\mathrm{Id}(v) \Leftrightarrow (T - \lambda \,\mathrm{Id})(v) = 0$$
vemos que λ é um valor próprio se e só se $N(T - \lambda \,\mathrm{Id}) \neq \{0\}$. Desde que V tenha dimensão finita, é possível determinar os valores próprios e vetores próprios recorrendo ao determinante: escolhendo uma base B para V e considerando a matriz $A = A_{T,B,B}$ a condição acima traduz-se em
$$N(A - \lambda I) \neq \{0\} \Leftrightarrow (A - \lambda I) \text{ não é invertível} \Leftrightarrow \det(A - \lambda I) = 0$$
Quando $\det(A - \lambda I) = 0$, os elementos não nulos de $N(A - \lambda I)$ corresponderão aos vetores próprios associados a λ.

Exemplo 6.9. *Consideremos a transformação linear $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ definida por $T(x, y) = (x + 2y, 2x + y)$. Considerando a base canónica temos*
$$A - \lambda I = \begin{bmatrix} 1-\lambda & 2 \\ 2 & 1-\lambda \end{bmatrix}$$
Esta matriz não é invertível exatamente quando
$$\begin{vmatrix} 1-\lambda & 2 \\ 2 & 1-\lambda \end{vmatrix} = 0 \Leftrightarrow (1-\lambda)^2 - 4 = 0 \Leftrightarrow \lambda = -1 \text{ ou } \lambda = 3.$$
São estes os valores próprios de T. Os vetores próprios de $\lambda = -1$ são os elementos não nulos de $N(A + I)$:
$$(A + I)v = 0 \Leftrightarrow \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = 0 \Leftrightarrow b = -a$$
ou seja, os vetores da forma $a(1, -1)$ com $a \neq 0$. Analogamente, os vetores próprios de $\lambda = 3$ (os elementos não nulos do núcleo de $A - 3I$) são os vetores da forma $a(1, 1)$ com $a \neq 0$.

Os vetores $v_1 = (1, 1)$ e $v_2 = (1, -1)$ formam uma base de $\mathbb{R}^2$ em termos da qual é extremamente simples compreender o efeito que a transformação linear T tem sobre os vetores de $\mathbb{R}^2$: Ao longo da direção de v_1 (a diagonal do primeiro quadrante) T expande por um fator de 3*, enquanto que na direção ortogonal, (a diagonal do quarto quadrante), T reflete. Com base nisto é fácil descrever o efeito que T teria num desenho qualquer no plano (ver figura).*



Note-se ainda que, uma vez que $T(v_1) = 3v_1$ e $T(v_2) = -v_2$, temos que a representação matricial de T com respeito à base $B = (v_1, v_2)$ é

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

Definição 6.10. *Um endomorfismo $T\colon V \to V$ diz-se* diagonalizável *se existe uma base para V constituída por vetores próprios de T. Uma matriz $A \in M_{n\times n}(\mathbb{K})$ diz-se diagonalizável, se a transformação linear de $\mathbb{K}^n$ 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. Note-se que os valores próprios não são necessariamente distintos dois a dois.

Uma matriz quadrada A é diagonalizável se existe uma matriz invertível S (uma matriz de mudança de coordenadas de uma base formada por vetores próprios de A para a base canónica) tal que $A = SDS^{-1}$ com D uma matriz diagonal (que tem como entradas não nulas valores próprios de A). Ou seja, A é diagonalizável se é semelhante a uma matriz diagonal.

Recorde-se a nossa estratégia de decompor um endomorfismo T como uma soma direta de endomorfismos "atómicos". Um endomorfismo $T\colon V \to V$ é diagonalizável se V se decompõe numa soma direta de espaços invariantes de dimensão 1. A nossa estratégia atinge assim o seu objetivo.

Pode no entanto haver algo de arbitrário na decomposição assim obtida. Se na base formada por vetores próprios houver mais do que um vetor próprio associado a um valor próprio λ podemos substituir esses vetores por qualquer outra base para o subespaço que eles geram.

Definição 6.11. *Seja $T\colon V \to V$ um endomorfismo. O* espaço próprio *associado ao valor próprio λ de T é*

$$E(\lambda) = N(T - \lambda\,\mathrm{Id})$$

A dimensão de $E(\lambda)$ chama-se a multiplicidade geométrica *de λ.*

O espaço próprio de λ é o conjunto de todos os vetores próprios de λ juntamente com o vetor 0. A multiplicidade geométrica de λ é o número máximo de vetores próprios de λ que são linearmente independentes. Note-se que $E(0)$ é o núcleo de T.

Uma transformação linear é diagonalizável se e só se existem valores próprios $\lambda_1, \ldots, \lambda_k$ *distintos* de T tais que

$$V = E(\lambda_1) \oplus \ldots \oplus E(\lambda_k)$$

A restrição de T a cada um dos espaços $E(\lambda)$ é simplesmente multiplicação por λ. Quando alguma das multiplicidades geométricas é maior do que um, esta decomposição é mais natural que a decomposição em espaços de dimensão 1 determinada por uma escolha de bases para cada um dos espaços próprios.

Exemplo 6.12. *Seja $T\colon \mathbb{R}^3 \to \mathbb{R}^3$ a transformação linear definida por*

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

Sendo A a matriz que representa T na base canónica temos

$$\begin{aligned} \det(A - \lambda I) &= \begin{vmatrix} 3-\lambda & -1 & 1 \\ 0 & 2-\lambda & 0 \\ 1 & -1 & 3-\lambda \end{vmatrix} \\ &= (3-\lambda)^2(2-\lambda) - (2-\lambda) = (2-\lambda)((\lambda-3)^2 - 1) \end{aligned}$$

Portanto a matriz $A - \lambda I$ tem característica < 3 se e só se $\lambda = 2$ ou $(\lambda - 3)^2 = 1 \Leftrightarrow \lambda = 2$ ou $\lambda = 4$. Os valores próprios são portanto 2 e 4. Os vetores próprios de 2 são as soluções de $(A - 2I)v = 0$:

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

Os vetores próprios de 2 são portanto os vetores não nulos da forma $(a, a + c, c)$ pelo que $E(2) = L(\{(1, 1, 0), (0, 1, 1)\})$. Os vetores próprios de 4 são as soluções de

$$\begin{bmatrix} -1 & -1 & 1 \\ 0 & -2 & 0 \\ 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = 0 \Leftrightarrow \begin{cases} b = 0 \\ c = a \end{cases}$$

Temos portanto $E(4) = L(\{(1,0,1)\})$*. Escrevendo* $v_1 = (1,1,0), v_2 = (0,1,1)$ *e* $v_3 = (1,0,1)$ *e tomando* $B = (v_1, v_2, v_3)$ *temos que*

$$A_{T,B.B} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

Temos $\mathbb{R}^3 = E(4) \oplus E(2)$*. No subespaço invariante* $E(2)$ *o efeito de* T *é multiplicar por* 2*, enquanto que em* $E(4)$ *é multiplicar por* 4*. A multiplicidade geométrica do valor próprio* 2 *é* 2*, enquanto que a multiplicidade geométrica do valor próprio* 4 *é* 1*.*

Exemplo 6.13. *Uma projeção, isto é um endomorfismo* $P\colon V \to V$ *tal que* $P^2 = P$*, é diagonalizável. De facto, vimos nos exercícios sobre transformações lineares que* V *se decompõe como a soma direta* $N(P) \oplus \mathrm{Im}(P)$*. Uma vez que* $N(P)$ *é o espaço próprio de* 0 *e que* $\mathrm{Im}(P)$ *é o espaço próprio de* 1 *temos* $V = E(0) \oplus E(1)$*. A multiplicidade geométrica de* 0 *é a dimensão do núcleo (ou nulidade) de* P*. A multiplicidade geométrica de* 1 *é a dimensão de* $\mathrm{Im}(P)$*.*

Infelizmente, nem sempre é possível diagonalizar uma transformação linear. Na realidade, nem é garantido que exista algum vetor próprio!

Exemplo 6.14. *É geometricamente óbvio (e fácil de confirmar algebricamente) que, desde que* α *não seja um múltiplo de* π*, a rotação* $R_\alpha\colon \mathbb{R}^2 \to \mathbb{R}^2$ *determinada (na base canónica) pela matriz*

$$\begin{bmatrix} \cos\alpha & -\operatorname{sen}\alpha \\ \operatorname{sen}\alpha & \cos\alpha \end{bmatrix}$$

não tem qualquer vetor próprio (não há nenhuma direção do plano que permaneça invariante pela rotação).

6.15. **Existência de valores próprios.** É no entanto um resultado básico da Álgebra Linear que um endomorfismo de um espaço vetorial complexo (de dimensão finita) tem sempre um valor próprio. Esta afirmação é uma consequência do seguinte Teorema cuja demonstração, por involver Análise (ou então Álgebra mais avançada), será omitida.

Teorema 6.16 (Teorema Fundamental da Álgebra)**.** *Todo o polinómio não constante*

$$p(x) = a_0 + a_1 x + \ldots + a_k x^k \tag{34}$$

com coeficientes $a_i \in \mathbb{C}$ *tem uma raíz complexa (isto é, existe* $z_0 \in \mathbb{C}$ *tal que* $p(z_0) = 0$*).*

Observação 6.17. *(i) O Teorema Fundamental da Álgebra 6.16 implica facilmente por indução que, assumindo* $a_k \neq 0$*, o polinómio* (34) *pode ser escrito de forma única a menos de troca de ordem dos fatores na forma*

$$a_k (x - \lambda_1)^{n_1} (x - \lambda_2)^{n_2} \cdots (x - \lambda_k)^{n_k} \tag{35}$$

com $\lambda_i \in \mathbb{C}$ *distintos, e* n_i *números naturais. Os números* λ_i *na expressão* (35) *são as raízes do polinómio* $p(x)$*. O expoente* n_i *diz-se a* multiplicidade *da raíz* λ_i*.*

Vemos assim que o Teorema Fundamental da Álgebra é análogo ao Teorema Fundamental da Aritmética que diz que qualquer número natural maior do que 1 *se pode escrever de forma única como um produto de potências de números primos a menos de troca da ordem dos fatores.*

(ii) Um corpo $\mathbb{K}$ *diz-se* algebricamente fechado *se todo o polinómio de grau* > 0 *com coeficientes em* $\mathbb{K}$ *tem uma raíz em* $\mathbb{K}$*. O Teorema 6.16 afirma que* $\mathbb{C}$ *é algebricamente fechado. Todos os resultados desta secção enunciados para espaços vetoriais complexos são verdade mais geralmente (com a mesma demonstração) para espaços vetoriais sobre corpos algebricamente fechados.*

Teorema 6.18. *Seja* $V \neq \{0\}$ *um espaço vetorial complexo de dimensão finita e* $T\colon V \to V$ *um endomorfismo. Então* T *tem um valor próprio.*

Dem. Vamos dar duas demonstrações, uma usando o determinante e outra que tem interesse conceptual e não usa o determinante.
Primeira demonstração: Seja $B = (v_1, \ldots, v_n)$ uma base para V e $A = A_{T,B,B} \in M_{n\times n}(\mathbb{C})$ a representação matricial de T com respeito a esta base. Os valores próprios de T são os valores de λ para os quais $\det(A - \lambda I) = 0$. Temos

$$\det(A-\lambda I) = \begin{vmatrix} a_{11}-\lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22}-\lambda & \cdots & a_{2n} \\ \vdots & & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn}-\lambda \end{vmatrix}$$

Se aplicarmos a fórmula (28) para o determinante, cada parcela nesta fórmula é um polinómio em λ (de grau $\leq n$). A única parcela de grau exatamente n é o produto das entradas ao longo da diagonal principal $(a_{11} - \lambda) \cdots (a_{nn} - \lambda)$. O coeficiente de λ^n é $(-1)^n$ pelo que o termo de grau n no polinómio $p(\lambda) = \det(A - \lambda I)$ é $(-1)^n\lambda^n$. Em particular $p(\lambda)$ não é um polinómio constante. O Teorema Fundamental da Álgebra 6.16 garante que $p(\lambda)$ tem pelo menos uma raíz, pelo que T tem pelo menos um valor próprio.
Segunda demonstração: Seja $v \in V \setminus \{0\}$ um vetor não nulo. Seja k o maior inteiro não negativo tal que

$$S_k = \{v, T(v), \ldots, T^k(v)\}$$

é um conjunto linearmente independente de vetores distintos. Este número existe porque S_0 é linearmente independente (logo o conjunto dos inteiros k para os quais S_k é linearmente independente é não vazio) e porque todo o subconjunto de V com $1 + \dim V$ elementos é linearmente dependente (logo $k \leq \dim(V) - 1$). Temos então

$$T^{k+1}(v) = a_0v + a_1T(v) + \ldots + a_kT^k(v)$$

para alguns $a_i \in \mathbb{C}$. Podemos escrever a expressão anterior na forma

$$(T^{k+1} - a_kT^k - \ldots - a_1T - a_0 \operatorname{Id})(v) = 0 \tag{36}$$

Seja λ uma raíz do polinómio $p(x) = x^{k+1} - a_kx^k - \ldots - a_1x - a_0$. Então

$$p(x) = x^{k+1} - a_kx^k - \ldots - a_1x - a_0 = (x - \lambda)(x^k + b_{k-1}x^{k-1} + \ldots + b_1x + b_0)$$

para alguns $b_0, \ldots, b_{k-1} \in \mathbb{C}$. Aplicando esta fatorização a (36) obtemos

$$(T - \lambda \operatorname{Id})(T^k + b_{k-1}T^{k-1} + \ldots + b_1 T + b_0 \operatorname{Id})(v) = 0$$

Seja $w = (T^k + b_{k-1}T^{k-1} + \ldots + b_1 T + b_0 \operatorname{Id})(v) = T^k(v) + b_{k-1}T^{k-1}(v) + \ldots + b_0 v$. Como S_k é linearmente independente temos que $w \neq 0$. Uma vez que $(T - \lambda \operatorname{Id})w = 0$ conclui-se que λ é um valor próprio de T. $\square$

Exemplo 6.19. *Consideremos o endomorfismo de $\mathbb{C}^2$ determinado pela matriz R_α do Exemplo 6.14. Recorde-se que estamos a assuir que $\operatorname{sen}\alpha \neq 0$. Temos*

$$\det(R_\alpha - \lambda I) = \begin{vmatrix} \cos(\alpha) - \lambda & -\operatorname{sen}\alpha \\ \operatorname{sen}\alpha & \cos(\alpha) - \lambda \end{vmatrix} = (\cos\alpha - \lambda)^2 + \operatorname{sen}^2\alpha$$

Os valores próprios de R_α são as raízes deste polinómio do segundo grau:

$$\lambda = \cos\alpha \pm i \operatorname{sen}\alpha$$

Um vetor próprio de $\cos\alpha + i\operatorname{sen}\alpha$ é um elemento não nulo do núcleo de $R_\alpha - (\cos\alpha + i\operatorname{sen}\alpha)\operatorname{Id}$:

$$(R_\alpha - (\cos\alpha + i\operatorname{sen}\alpha)\operatorname{Id})(v) = 0 \Leftrightarrow \begin{bmatrix} -i\operatorname{sen}\alpha & -\operatorname{sen}\alpha \\ \operatorname{sen}\alpha & -i\operatorname{sen}\alpha \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = 0 \Leftrightarrow b = -ia$$

O espaço próprio $E(\cos\alpha + i\operatorname{sen}\alpha)$ é portanto gerado pelo vetor $(1, -i)$. Da mesma forma vemos que $(1, i)$ gera o espaço próprio $E(\cos\alpha - i\operatorname{sen}\alpha)$.

Observação 6.20. *Não é uma coincidência que os valores próprios e vetores próprios do exemplo anterior sejam conjugados. Uma vez que dados dois números complexos z_1, z_2 temos $\overline{z_1 z_2} = \overline{z}_1\overline{z}_2$ e $\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2}$, se A é uma matriz $n \times n$ real (portanto $\overline{A} = A$) e $v \in \mathbb{C}^n$ é tal que $Av = \lambda v$ então $A\overline{v} = \overline{\lambda}\overline{v}$. Assim, os valores (e vetores) próprios complexos de um endomorfismo de $\mathbb{C}^n$ representado por uma matriz com entradas reais ocorrem aos pares. Esta observação facilita o cálculo dos valores e vetores próprios nesta situação.*

Vejamos um exemplo concreto do procedimento utilizado na segunda demonstração do Teorema 6.18.

Exemplo 6.21. *Consideremos a transformação linear $T(x, y) = (x + 2y, 2x + y)$. Considerando o vetor $v = (1, 0)$ temos $T(v) = (1, 2)$. Assim v e $T(v)$ formam uma base para $\mathbb{R}^2$. Uma vez que $T^2(v) = T(1, 2) = (5, 4) = 3v + 2T(v)$ temos*

$$(T^2 - 2T - 3\operatorname{Id})(v) = 0$$

As raízes do polinómio $x^2 - 2x - 3$ são $x = 3$ e $x = -1$. A equação acima pode portanto escrever-se na forma

$$(T - 3\operatorname{Id})(T + \operatorname{Id})(v) = 0 \quad \text{ou equivalentemente } (T + \operatorname{Id})(T - 3\operatorname{Id})(v) = 0$$

Segue-se que $(T + \operatorname{Id})(v) = (1, 2) + (1, 0) = (2, 2)$ é um vetor próprio com valor próprio 3 e que $(T - 3\operatorname{Id})(v) = (1, 2) - 3(1, 0) = (-2, 2)$ é um vetor próprio com valor próprio -1.

Experimente fazer a conta acima com outro vetor inicial v. O que conclui?

Observação 6.22. *Vale a pena sublinhar outra ideia utilizada na segunda demonstração do Teorema 6.18. Seja V um espaço vetorial sobre o corpo $\mathbb{K}$ e $T\colon V \to V$ um endomorfismo. Dado um polinómio $p(x) = a_0 + a_1 x + \ldots + a_k x^k$ com coeficientes $a_i \in \mathbb{K}$, $p(x)$ determina um endomorfismo*

$$p(T) = a_0 \operatorname{Id} + a_1 T + \ldots + a_k T^k \colon V \to V$$

Uma vez que a composição de transformações lineares é linear em cada argumento e as potências de T comutam entre si, uma fatorização $p(x) = q(x)r(x)$ determina uma fatorização $p(T) = q(T) \circ r(T)$ da transformação linear $p(T)$.

6.23. **O determinante de um endomorfismo e o polinómio caraterístico.** A primeira demonstração do Teorema 6.18 fez uso de um polinómio associado a um endomorfismo de um espaço vetorial de dimensão finita V. Começamos por notar que este polinómio é independente da escolha de base B usada na sua definição.

Definição 6.24. *Seja V um espaço vetorial de dimensão finita e $T\colon V \to V$ um endomorfismo. Sendo B uma base ordenada qualquer de V definimos o* determinante de T *por*

$$\det(T) = \det(A_{T,B,B})$$

Temos que verificar que o escalar $\det(T)$ é independente da escolha de B: se B' é outra base para V e $S = S_{B \to B'}$ a matriz de mudança de coordenadas então

$$A_{T,B',B'} = S A_{T,B,B} S^{-1}$$

e portanto

$$\det(A_{T,B',B'}) = \det(S)\det(A_{T,B,B})\det(S^{-1}) = \det(S)\det(A_{T,B,B})\frac{1}{\det(S)} = \det(A_{T,B,B})$$

Infelizmente, não seria neste momento fácil explicar-vos como definir intrinsecamente o determinante de um endomorfismo sem apelar às representações matriciais pelo que a Definição 6.24 terá de servir.

Uma vez que o determinante deteta a invertibilidade de um endomorfismo (T é invertível se e só se $A_{T,B,B}$ é invertível para qualquer base B) podemos usar o determinante para calcular os valores próprios de um endomorfismo T, como foi feito na primeira demonstração do Teorema 6.18.

Definição 6.25. *Seja V um espaço vetorial de dimensão finita sobre o corpo $\mathbb{K}$ e $T\colon V \to V$ um endomorfismo. O* polinómio caraterístico *de T é o polinómio (com coeficientes em $\mathbb{K}$) definido por*

$$p(\lambda) = \det(T - \lambda \operatorname{Id})$$

O polinómio caraterístico de uma matriz quadrada $A \in M_{n\times n}(\mathbb{K})$ é o polinómio caraterístico da transformação linear representada por A na base canónica de $\mathbb{K}^n$, ou seja,

$$p(\lambda) = \det(A - \lambda I)$$

Note-se que o termo constante do polinómio caraterístico (que é igual a $p(0)$) é o determinante de T.

Exemplo 6.26. *O polinómio caraterístico da matriz*

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

é

$$\begin{aligned} p(\lambda) &= \begin{vmatrix} \frac{5}{2}-\lambda & 0 & -\frac{1}{2} & 0 \\ \frac{1}{2} & 1-\lambda & -\frac{1}{2} & 2 \\ \frac{1}{2} & 0 & \frac{3}{2}-\lambda & 0 \\ 0 & 0 & 0 & -1-\lambda \end{vmatrix} = -(1+\lambda)(1-\lambda)\begin{vmatrix} \frac{5}{2}-\lambda & -\frac{1}{2} \\ \frac{1}{2} & \frac{3}{2}-\lambda \end{vmatrix} \\ &= (\lambda^2-1)(\lambda^2-4\lambda+4) = (\lambda^2-1)(\lambda-2)^2 \end{aligned}$$

Este polinómio anula-se quando $\lambda = 2$ ou $\lambda = \pm 1$. São portanto estes os valores próprios de A.

Registamos um par de propriedades imediatas do polinómio característico de um endomorfismo de um espaço vetorial de dimensão finita (sobre um corpo qualquer, não necessariamente o dos números complexos).

Proposição 6.27. *Seja V um espaço vetorial de dimensão finita, $T\colon V \to V$ um endomorfismo, e $p(\lambda)$ o seu polinómio característico.*

(a) Sendo $n = \dim V$, o polinómio característico é da forma

$$p(\lambda) = (-1)^n\lambda^n + a_{n-1}\lambda^{n-1} + \ldots + a_0$$

e portanto tem grau n.

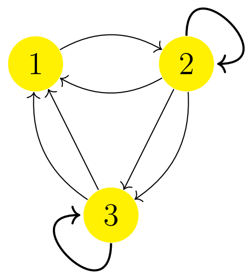
(b) As raízes do polinómio característico de T são os valores próprios de T.

Dem. (a) O argumento utilizado na primeira demonstração do Teorema 6.18 aplica-se literalmente.

(b) Conforme já explicámos, λ é um valor próprio se e só se $(T - \lambda\,\mathrm{Id})$ não é invertível, o que acontece sse $p(\lambda) = 0$.

□

6.28. **O algoritmo PageRank.** Vamos agora fazer um pequeno interlúdio para discutir uma aplicação famosa do conceito de vetor próprio. Consideremos uma internet com apenas três páginas ligadas de acordo com o diagrama



Como no Exemplo 6.2 podemos considerar a população de todos os internautas e atribuir a cada um o estado i num dado instante se estiverem a visitar a página i. Sejam n_1, n_2 e n_3 o número de pessoas em cada página num dado instante. Se cada pessoa clicar num link ao acaso em cada página, a distribuição das pessoas pelas páginas 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.[12]

Esta matriz que descreve a evolução temporal do nosso sistema é uma matriz de Markov (cf. Exemplo 6.2): as entradas são não negativas porque codificam probabilidades e a soma das entradas de cada coluna é 1 (porque é a soma das probabilidades de ir parar a cada destino possível partindo da página correspondente à coluna).

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

$$\begin{bmatrix} 0 & \frac{1}{4} & \frac{2}{3} \\ 1 & \frac{1}{4} & 0 \\ 0 & \frac{1}{2} & \frac{1}{3} \end{bmatrix} \tag{37}$$

com valor próprio 1. Um tal vetor próprio existe necessariamente porque a soma das entradas de cada coluna é um! De facto, isto significa exatamente que $(1,1,1)$ é um vetor próprio da matriz transposta, com valor próprio 1. Uma vez que as dimensões dos espaços das linhas e colunas de uma matriz quadrada coincidem, a característica de $A - \lambda\,\mathrm{Id}$ e de $(A-\lambda\,\mathrm{Id})^T = A^T - \lambda\,\mathrm{Id}$ coincidem. Isto implica que os conjuntos de valores próprios de A e A^T coincidem.

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

[12]Se uma página não tem ligações para outras assume-se que tem uma ligação para cada página.

indicam o estado das páginas de modo a que a soma das entradas[13] seja 1, o limite quando o tempo tende para $+\infty$ do estado do sistema é o vetor próprio de 1 (normalizado), que é único.

Mais precisamente, se A é a matriz (37) 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. Pode mostrar-se que (com grande generalidade) o significado das componentes de v é a seguinte: 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 (37) 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.

---

[13]Isto corresponde a considerar a percentagem dos internautas em cada página em vez do número absoluto.

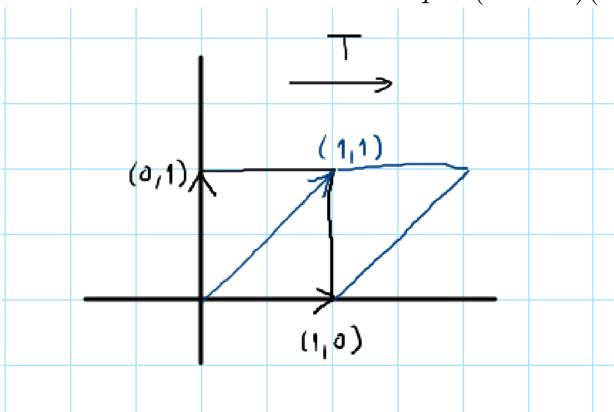
6.29. **Vetores próprios generalizados.** Apesar de todo o endomorfismo $T\colon V \to V$ de um espaço vetorial complexo ter pelo menos um valor próprio, não é em geral possível encontrar uma base para V formada por vetores próprios.

Exemplo 6.30 (A aplicação de "shear" ou deslizamento)**.** *Seja $T\colon \mathbb{C}^2 \to \mathbb{C}^2$ a transformação linear definida por $T(x,y) = (x+y, y)$. Este endomorfismo é representado na base canónica pela matriz*

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

É imediato verificar que o único valor próprio de A é 1 pelo que A não é diagonalizável (uma transformação diagonalizável com um único valor próprio é necessariamente um múltiplo escalar da identidade). O espaço próprio $E(1)$ é claramente o primeiro eixo coordenado $\{(x,0)\colon x \in \mathbb{C}\}$ pelo que o valor próprio 1 tem multiplicidade geométrica 1.

Podemos no entanto observar que $(T - \mathrm{Id})(0,1) = (1,0)$ e portanto $(T - \mathrm{Id})^2(0,1) = 0$.



Geometricamente, o efeito que T tem num vetor de $\mathbb{R}^2$ é deslizá-lo ao longo do eixo dos xx (o eixo gerado pelo vetor próprio) uma distância igual à sua ordenada.

O principal Teorema sobre endomorfismos complexos afirma que qualquer endomorfismo não diagonalizável se pode decompor nestas aplicações de "shear" ou deslizamento.

Notação 6.31. *A partir de agora, para simplificar a notação usaremos muitas vezes o escalar λ para denotar a transformação linear $\lambda\,\mathrm{Id}$. Assim, por exemplo, $(T - \lambda\,\mathrm{Id})$ aparecerá como $(T - \lambda)$. Também omitiremos alguns parêntesis quando a sua omissão não cause confusão: por exemplo um vetor $T(v)$ poderá ser denotado simplesmente por Tv.*

Definição 6.32. *Seja $T\colon V \to V$ um endomorfismo e λ um valor próprio de T. O* espaço próprio generalizado *de λ é*

$$E^g(\lambda) = \{v \in V\colon (T-\lambda)^k v = 0 \text{ para algum } k\}$$

Os elementos não nulos de $E^g(\lambda)$ designam-se por vetores próprios generalizados *associados ao valor próprio λ. A dimensão de $E^g(\lambda)$ chama-se a* multiplicidade algébrica *de λ.*

Note-se que $E(\lambda) \subset E^g(\lambda)$. É fácil verificar que $E^g(\lambda)$ é de facto um subespaço vetorial de V: claramente $0 \in E^g(\lambda)$ e $E^g(\lambda)$ é fechado para o produto por escalar. Vejamos que também é fechado para a soma: se $v_1, v_2 \in E^g(\lambda)$ então existem k_1, k_2 tais que $(T-\lambda)^{k_1}(v_1) = 0$ e $(T-\lambda)^{k_2}(v_2) = 0$. Seja k o máximo de $\{k_1, k_2\}$. Então

$$(T-\lambda)^k(v_1+v_2) = (T-\lambda)^{k-k_1}(T-\lambda)^{k_1}(v_1) + (T-\lambda)^{k-k_2}(T-\lambda)^{k_2}(v_2) = 0+0 = 0$$

Finalmente, observe-se que $E^g(\lambda)$ é um subespaço invariante: dado $v \in E^g(\lambda)$, existe k tal que $(T-\lambda)^k(v) = 0$. Logo

$$(T-\lambda)^k(Tv) = T(T-\lambda)^k(v) = T(0) = 0 \qquad \Rightarrow \qquad Tv \in E^g(\lambda) \tag{38}$$

Exemplo 6.33. *Aplicando a nossa nova terminologia ao Exemplo 6.30 vemos que nesse caso $E^g(1) = \mathbb{C}^2$ é estritamente maior do que $E(1) = L(\{(1,0)\})$. Logo a multiplicidade algébrica de $\lambda = 1$ é 2, enquanto que a multiplicidade geométrica é apenas 1.*

Exemplo 6.34. *Consideremos a transformação linear $T\colon \mathbb{R}^3 \to \mathbb{R}^3$ representada com respeito à base canónica pela matriz*

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

Calculando os valores próprios vemos que estes são $\lambda = 1$ e $\lambda = 2$. Os vectores próprios de 1 são as soluções da equação $(A-I)v = 0$, ou seja

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & -1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = 0 \Leftrightarrow \begin{cases} a+b=0 \\ a+b-c=0 \\ a+b=0 \end{cases} \Leftrightarrow \begin{cases} b=-a \\ c=0 \end{cases}$$

Assim $(1,-1,0)$ forma uma base para os valores próprios de 1. Uma vez que

$$(A-I)^2 = \begin{bmatrix} 2 & 2 & -1 \\ 1 & 1 & -1 \\ 2 & 2 & -1 \end{bmatrix}$$

tem característica 2, o seu núcleo tem dimensão 1 e portanto consiste no espaço próprio de 1. Podemos facilmente verificar que o mesmo acontece com as matrizes $(A-I)^k$ para todos os valores de k para $k \geq 2$ pelo que os únicos vetores próprios generalizados de 1 são os vetores próprios de 1. Na realidade, iremos ver em breve no Lema 6.48, o facto de $N((A-I)^2) = N(A-I)$ implica que $N((A-I)^k) = N(A-I)$ para todo o k, e não há portanto necessidade de realizar mais cálculos.

Temos assim que a multiplicidade algébrica de 1 é igual à multiplicidade geométrica, que é 1.

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 apenas dois vetores próprios independentes).

No entanto

$$(A-2I)^2 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & -1 \\ 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & -1 \\ 1 & 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -1 \\ -1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

pelo que $(A-2I)^2v = 0 \Leftrightarrow v \in L(\{(1,0,1),(0,1,0)\})$. Claramente $(A-2I)^k = \pm(A-2I)^2$ para $k \geq 2$ pelo que não há mais vetores próprios generalizados de 2. Veremos mais à frente que este último cálculo é na realidade desnecessário uma vez que a soma das dimensões dos espaços próprios generalizados é sempre menor ou igual à dimensão do espaço vetorial. Concluímos que $E^g(2) = L(\{(1,0,1),(0,1,0)\})$. Temos assim que a multiplicidade algébrica de 2 é igual a 2. Note-se que juntando as bases dos espaços próprios generalizados de 1 e 2 obtemos uma base de $\mathbb{R}^3$.

Quando V tem dimensão finita podemos escolher um expoente para $(T-\lambda)$ que anula todos os vetores de $E^g(\lambda)$ simultaneamente.

Lema 6.35. *Seja V um espaço de dimensão finita, $T\colon V \to V$ um endomorfismo e λ um valor próprio de T. Então existe $n > 0$ tal que*

$$E^g(\lambda) = \{v \in V \colon (T-\lambda)^n v = 0\} = N((T-\lambda)^n)$$

Proof. Seja $\{v_1, \ldots, v_p\}$ uma base de $E^g(\lambda)$ e $k_1, \ldots, k_p$ tais que $(T-\lambda)^{k_i}(v_i) = 0$. Tomando para n o máximo de todos os k_i e escrevendo $v \in E^g(\lambda)$ como uma combinação linear dos v_i's temos, para $v = \alpha_1 v_1 + \ldots + \alpha_p v_p$,

$$(T-\lambda)^n(\alpha_1 v_1 + \ldots + \alpha_p v_p) = \alpha_1 (T-\lambda)^n(v_1) + \ldots + \alpha_p (T-\lambda)^n(v_p) = 0$$

pelo que $E^g(\lambda) \subset N((T-\lambda)^n)$. A inclusão recíproca é imediata da definição de $E^g(\lambda)$. $\square$

6.36. **Decomposição primária de um endomorfismo.** Podemos agora dar um grande passo na direção da decomposição de um endomorfismo em "endomorfismos atómicos".

Teorema 6.37. *(Decomposição primária) Seja V um espaço vetorial complexo de dimensão finita e $T\colon V \to V$ um endomorfismo. Sejam $\lambda_1, \ldots, \lambda_k$ os valores próprios de T. Então*

$$V = E^g(\lambda_1) \oplus E^g(\lambda_2) \oplus \ldots \oplus E^g(\lambda_k)$$

A afirmação do Teorema anterior pode ser dividida em duas partes: a primeira é que a soma dos espaços próprios é direta, ou seja que cada um dos espaços próprios generalizados não interseta os vetores que se podem escrever como somas de vetores dos restantes espaços.

$$E^g(\lambda_i) \cap \left(\sum_{j \neq i} E^g(\lambda_j) \right) = 0$$

A segunda parte é a afirmação que os espaços próprios generalizados geram todo o V. Para provar a primeira afirmação vamos usar o seguinte resultado.

Lema 6.38. *Seja V um espaço vetorial de dimensão finita, $T\colon V \to V$ um endomorfismo, μ um valor próprio de T e λ um escalar diferente de μ. Então a restrição de $(T-\lambda)$ a $E^g(\mu)$ é um isomorfismo.*

Dem. Uma vez que $E^g(\mu)$ é invariante para T, também é invariante para $T-\lambda$. Como V e portanto $E^g(\mu)$ têm dimensão finita só temos que verificar que $N(T-\lambda)\cap E^g(\mu) = \{0\}$.

Suponhamos que $v \in N(T-\lambda)$. Então $Tv = \lambda v$, e portanto

$$(T-\mu)(v) = Tv - \mu v = (\lambda - \mu)v$$

Portanto, se v pertence também a $E^g(\mu)$ temos $(T-\mu)^k(v) = (\mu-\lambda)^k v = 0$ para algum $k > 0$. Uma vez que $\mu \neq \lambda$, conclui-se que $v = 0$ conforme queríamos demonstrar. □

Para demonstrar que os vetores própios generalizados geram V vamos usar a noção de espaço vetorial quociente. Recorde-se dos exercícios da ficha sobre espaços vetoriais que, dado um subespaço $W \subset V$, o conjunto V/W é o conjunto dos planos paralelos a W em V, que se escrevem na forma $v + W$ com $v \in V$. Com as operações de soma e produto por escalar definidas pelas fórmulas

$$(v_1 + W) + (v_2 + W) = (v_1 + v_2) + W, \qquad \alpha(v + W) = (\alpha v) + W$$

o conjunto V/W adquire a estrutura de um espaço vetorial - o espaço vetorial quociente de V por W. Este espaço "descarta" o subespaço W no sentido em que dois vetores v_1 e v_2 de V que difiram por um elemento de W correspondem ao mesmo elemento $v_1 + W = v_2 + W$ no espaço quociente.

Se $T\colon V \to V$ é uma transformação linear e $W \subset V$ é um subespaço invariante, então T determina um endomorfismo $\overline{T}\colon V/W \to V/W$, definido pela expressão

$$\overline{T}(v + W) = T(v) + W$$

De facto, se $v_1 + W = v_2 + W \Leftrightarrow v_1 - v_2 \in W$, então $T(v_1) - T(v_2) \in W$ pelo que $T(v_1) + W = T(v_2) + W$. Isto mostra que a fórmula acima para $\overline{T}$ define de facto uma função de V/W para V/W e é então imediato verificar que esta função preserva a soma e o produto por escalar em V/W.

O nosso plano será ver que, sendo W o subespaço gerado por todos os subespaços próprios generalizados de um dado endomorfismo $T\colon V \to V$, temos $V/W = \{0\}$ e portanto $W = V$.

Dem. do Teorema 6.37. Começamos por ver que a soma dos espaços próprios generalizados é direta. Suponhamos que $v \in E^g(\lambda_1)$ e que $v = w_2 + \ldots + w_p$ com $w_j \in E^g(\lambda_j)$ vetores próprios generalizados de valores próprios distintos de λ_1. Sejam $n_2, \ldots, n_p$ tais que $(T - \lambda_j)^{n_j} w_j = 0$, e

$$q(T) = (T-\lambda_2)^{n_2} \cdots (T-\lambda_p)^{n_p}$$

Uma vez que a ordem dos fatores na fatorização de $q(T)$ é arbitrária, temos que $q(T)w_j = 0$ para todo o j. Portanto

$$q(T)v = q(T)w_2 + \ldots + q(T)w_p = 0$$

Pelo Lema 6.38 a restrição de cada um dos fatores de $q(T)$ a $E^g(\lambda_1)$ é um isomorfismo. Segue-se que a restrição de $q(T)$ a $E^g(\lambda_1)$ é também um isomorfismo e portanto

$$q(T)v = 0 \Rightarrow v = 0.$$

Conclui-se assim que $E^g(\lambda_1) \cap \sum_{j=2}^n E^g(\lambda_j) = \{0\}$ e portanto, indutivamente, que

$$V = \oplus_{i=1}^k E^g(\lambda_i)$$

Vejamos agora que os espaços próprios generalizados geram V. Seja $W = \oplus_{i=1}^k E^g(\lambda_i)$ o subespaço de V gerado por todos os espaços próprios generalizados. Claramente W é um subespaço invariante. Consideremos a transformação linear $\overline{T}\colon V/W \to V/W$ determinada por T no espaço quociente.

Se $V/W \neq 0$, pelo Teorema 6.18, o endomorfismo $\overline{T}$ tem um valor próprio α. Seja $v+W$ um vetor próprio de $\overline{T}$ associado a α. Então

$$\overline{T}(v+W) = \alpha(v+W) \Leftrightarrow T(v) = \alpha v + w \text{ com } w \in W$$

Suponhamos primeiro que α não é um dos valores próprios de T. Então, pelo lema 6.38, a restrição de $(T-\alpha)$ a W é um isomorfismo. Seja $w' \in W$ tal que $(T-\alpha)w' = w$. Então

$$(T-\alpha)(v-w') = w - (T-\alpha)w' = 0$$

Uma vez que $v - w' \neq 0$ (senão $v = w' \in W$ e $v + W = 0 + W$ contrariamente à nossa hipótese que $v+W$ é um vetor próprio de $\overline{T}$) conclui-se que α é um vetor próprio de T, o que contradiz a hipótese inicial sobre α.

Suponhamos então que α é um dos valores próprios de T. Sem perda de generalidade podemos assumir que $\alpha = \lambda_1$. Podemos escrever

$$w = w_1 + w_2 \quad \text{com } w_1 \in E^g(\lambda_1), \quad w_2 \in \bigoplus_{j=2}^k E^g(\lambda_j)$$

Pelo Lema 6.38 podemos escolher $w_2' \in \bigoplus_{j=2}^k E^g(\lambda_j)$ tal que $(T-\alpha)w_2' = w_2$. Então

$$(T-\alpha)(v-w_2') = w - w_2 = w_1$$

Sendo m tal que $(T-\lambda_1)^m(w_1) = (T-\alpha)^m(w_1) = 0$ temos então

$$(T-\alpha)^{m+1}(v-w_2') = (T-\alpha)^m(T-\alpha)(v-w_2') = (T-\alpha)^m(w_1) = 0$$

Portanto $v - w_2' \in E^g(\alpha) = E^g(\lambda_1) \subset W$. Uma vez que $w_2' \in W$ conclui-se que $v \in W$, o que novamente contradiz a hipótese de $v+W$ ser um vetor próprio. Esta contradição mostra que $V/W = \{0\}$ e conclui a demonstração. □

6.39. **Endomorfismos nilpotentes.** Tendo em conta o Teorema 6.37 resta-nos entender a restrição de uma transformação linear T a um espaço próprio generalizado $E^g(\lambda)$. Escrevendo $N = T - \lambda$, basta-nos entender a restrição de N a $E^g(\lambda)$ pois

$$T = \lambda + (T-\lambda) = \lambda + N$$

Definição 6.40. *Seja W um espaço vetorial. Um endomorfismo $N\colon W \to W$ diz-se* nilpotente *se existe $k > 0$ tal que $N^k = 0$. O menor k tal que $N^k = 0$ diz-se o* índice de nilpotência *de N.*

Note-se que $N\colon W \to W$ é nilpotente se e só $W = E^g(0)$, e portanto, de acordo com o Teorema da Decomposição Primária, N é nilpotente se e só se 0 é o único valor próprio de N.

Os endomorfismos nilpotentes podem ser encarados para certos efeitos como sendo "negligíveis". Afinal, são "raízes índice k" de 0... Num espaço próprio generalizado (que não é um espaço próprio) T expressa-se não como um múltiplo da identidade mas como um múltiplo da identidade mais um endomorfismo nilpotente.

Exemplo 6.41. *O índice de nilpotência de N é 1 se e só se N é a transformação linear identicamente nula. O endomorfismo N representado na base canónica de $\mathbb{R}^3$ pela matriz*

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

tem índice de nilpotência 3, uma vez que

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}^2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}^3 = 0$$

Mais geralmente, um endomorfismo de $\mathbb{R}^k$ representado por uma matriz $k \times k$

$$N = \begin{bmatrix} 0 & 1 & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \ddots & 1 \\ 0 & 0 & \cdots & 0 \end{bmatrix} \tag{39}$$

tem índice de nilpotência k. De facto, o efeito de multiplicar N por uma matriz com k linhas é

$$\begin{bmatrix} 0 & 1 & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \ddots & 1 \\ 0 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \cdots & L_1 & \cdots \\ \cdots & L_2 & \cdots \\ & \vdots & \\ \cdots & L_k & \cdots \end{bmatrix} = \begin{bmatrix} \cdots & L_2 & \cdots \\ \cdots & L_3 & \cdots \\ & \vdots & \\ \cdots & L_k & \cdots \\ 0 & 0 & 0 \end{bmatrix}$$

logo, à medida que o expoente i em N^i aumenta, a linha de 1s vai subindo até que finalmente desaparece quando i chega a k.

O nosso objetivo nesta secção é mostrar o seguinte resultado, que, em conjunto com o Teorema da Decomposição Primária, levará imediatamente a uma forma normal para todos os endomorfismos de espaços complexos - a forma canónica de Jordan.

Teorema 6.42. *Seja W um espaço vetorial de dimensão finita e $N\colon W \to W$ um endomorfismo nilpotente. Então existe uma base B para W tal que a matriz $A_{N,B,B}$ que representa N com respeito à base B é diagonal por blocos, sendo cada bloco da forma* (39).

Note-se que se $v_1, \dots, v_k$ são os vetores da base correspondentes às colunas de um dos blocos temos

$$Nv_1 = 0, \quad Nv_2 = v_1, \quad \dots \quad Nv_k = v_{k-1} \tag{40}$$

Definição 6.43. *Seja $N\colon W \to W$ um endomorfismo nilpotente. Uma sucessão de vetores não nulos $(v_1, \dots, v_k)$ satisfazendo* (40) *chama-se uma* cadeia de Jordan *para N.*

Proposição 6.44. *Seja $N\colon W \to W$ um endomorfismo nilpotente. Uma cadeia de Jordan $(v_1, \dots, v_k)$ para N é um conjunto linearmente independente.*

Dem. Note-se que $N^{k-i}v_k = v_i$, e que $N^j v_i = 0$ se $j \geq i$. Logo se $\alpha_1 v_1 + \ldots + \alpha_k v_k = 0$, temos

$$0 = N^{k-1}(\alpha_1 v_1 + \ldots + \alpha_k v_k) = 0 + \ldots 0 + \alpha_k v_1$$

portanto $\alpha_k = 0$. Aplicando sucessivamente N^{k-i} com $i = 2, \dots k$ obtemos $\alpha_{k-i} = 0$, o que conclui a demonstração. $\square$

Observação 6.45. *Seja $(v_1, \dots, v_k)$ uma cadeia de Jordan para N.*

(i) $W = L(\{v_1, \dots, v_k\})$ é o subespaço cíclico gerado pelo vetor v_k.
(ii) Na base $(v_1, \dots, v_k)$ a expressão matricial da restrição de N a W é (39).

Corolário 6.46. *Se V é um espaço vetorial de dimensão n, e $N\colon V \to V$ é um endomorfismo nilpotente, o índice de nilpotência de N é no máximo n. Em particular, no Lema 6.35 podemos tomar para o expoente n a dimensão de V.*

Proof. Seja k o índice de nilpotência de N e $v \in V$ tal que $N^{k-1}v \neq 0$. Então $(N^{k-1}v, \dots, Nv, v)$ forma uma cadeia de Jordan e portanto um conjunto linearmente independente. Conclui-se que $\dim V \geq k$. $\square$

A chave para entender que tamanho têm os blocos no Teorema 6.42 é a seguinte. Um endomorfismo nilpotente $N\colon W \to W$ determina uma *filtração* do espaço W, isto é uma sucessão encadeada de subespaços

$$\{0\} = W(0) \subset W(1) \subset W(2) \subset \cdots \subset W(k) = W$$

onde $W(i)$ é o núcleo de N^i, e k é o índice de nilpotência de N. De facto,

$$v \in W(i) \Leftrightarrow N^i v = 0 \Rightarrow N^{i+1}v = 0 \Leftrightarrow v \in W(i+1)$$

logo $W(i) \subset W(i+1)$. Definindo $W(-1) = \emptyset$, para cada $w \in W$, existe exatamente um $i \in \{0, \dots, k\}$ tal que $w \in W(i)$ mas $w \notin W(i-1)$. Este número chama-se a *filtração* de w (diz qual é o primeiro $W(i)$ a que w pertence). Um elemento de $W(i)$ tem portanto filtração $\leq i$.

Exemplo 6.47. *A filtração de um vetor $v \in \mathbb{R}^5$ determinada pelo endomorfismo nilpotente representado pela matriz*

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

é

- 0 *se* $v = (0,0,0,0,0)$,
- 1 *se* $v = (x,0,0,w,0)$ *com* $x \neq 0$ *ou* $w \neq 0$
- 2 *se* $v = (x,y,0,w,v)$ *com* $y \neq 0$ *ou* $v \neq 0$
- 3 *caso contrário, isto é se* $v = (x,y,z,w,v)$ *com* $z \neq 0$.

A seguinte observação é importante. Confirme-a no exemplo anterior.

Lema 6.48. *Seja $N \colon W \to W$ um endomorfismo nilpotente com índice de nilpotência k. Um elemento $w \in W$ tem filtração i, com $1 \leq i \leq k$ se e só se Nw tem filtração $i-1$.*

Dem. Para $1 \leq i \leq k$ temos $w \in W(i) \Leftrightarrow N^i w = 0 \Leftrightarrow N^{i-1}(Nw) = 0 \Leftrightarrow Nw \in W(i-1)$. Se $i = 1$ isto é a afirmação que pretendemos demonstrar. Se $i > 1$, um elemento $w \in W$ tem filtração i se e só se $w \in W(i) \wedge w \notin W(i-1) \Leftrightarrow Nw \in W(i-1) \wedge Nw \notin W(i-2) \Leftrightarrow Nw$ tem filtração $i-1$. □

Lema 6.49. *Seja $W \neq \{0\}$ e $N \colon W \to W$ um endomorfismo nilpotente com índice de nilpotência k e $d_i = \dim W(i) - \dim W(i-1)$, $n = \dim W$. Então*

$$d_1 \geq d_2 \geq \ldots \geq d_k > 0$$

Dem. Se o índice de nilpotência é 1, não temos nada a mostrar. Suponhamos que $k \geq 2$ e sejam i tal que $2 \leq i \leq k$, B uma base de $W(i-1)$ e $\{v_1, \ldots, v_{d_i}\}$ um subconjunto linearmente independente de $W(i)$ de tal forma que $B' = B \cup \{v_1, \ldots, v_{d_i}\}$ é uma base de $W(i)$ (este subconjunto existe porque podemos completar uma base B de $W(i-1)$ a uma base B' de $W(i)$). Seja

$$v = \alpha_1 v_1 + \ldots + \alpha_{d_i} v_{d_i}$$

uma combinação linear com os coeficientes não todos nulos. Trata-se de um elemento não nulo de $W(i)$ que não pode pertencer a $W(i-1) = L(B)$ devido à independência linear de B'. Então v tem filtração i e portanto pelo Lema 6.48

$$Nv = \alpha_1 N v_1 + \ldots + \alpha_{d_i} N v_{d_i}$$

tem filtração $i-1$. Em particular Nv não pode ser zero. Isto mostra que

(41) $\{Nv_1, \ldots, Nv_{d_i}\}$ é lin indep em $W(i-1)$ e $L(\{Nv_1, \ldots, Nv_{d_i}\}) \cap W(i-2) = \{0\}$

e portanto que

$$\dim W(i-2) + d_i \leq \dim W(i-1) \Leftrightarrow d_i \leq d_{i-1}$$

□

Note-se que $d_1 + d_2 + \ldots + d_k = \dim W$. Os números $d_1, \ldots, d_k$ formam aquilo a que se chama uma *partição* do inteiro $n = \dim W$. Esta pode representar-se visualmente como no exemplo seguinte, no qual $n = 17, k = 6, d_1 = 5, d_2 = d_3 = 4, d_4 = 2$ e $d_5 = d_6 = 1$.

(42)

1	1	1	1	1	1
2	2	2	2		
3	3	3			
4	4	4			
5					

O diagrama acima explica como a dimensão dos espaços $W(i)$ vai aumentando à medida que i aumenta: a dimensão de $W(i)$ é igual a $d_1 + \ldots + d_i$ que é o número total de quadrados nas primeiras i colunas do diagrama.

Dem. do Teorema 6.42. A demonstração vai consistir na construção de uma base B para W que obedeça às condições do enunciado. Para cada i entre 1 e $k-1$ seja $U(i) = W(i-1) + N(W(i+1)) \subset W(i)$ e $U(k) = W(k-1)$. Temos então

$$W(i-1) \subset U(i) \subset W(i)$$

$U(i)$ consiste nos vetores de filtração $\leq i-1$ juntamente com os que têm filtração i e estão na imagem por N de elementos de filtração $i+1$. Uma base para $U(i)$ obtém-se juntando a uma base de $W(i-i)$ a imagem por N de vetores $\{v_1, \ldots, v_{d_{i+1}}\}$ que completem uma base de $W(i)$ a uma base de $W(i+1)$ (cf. (41)). Em particular $\dim U(i) = \dim W(i-1) + d_{i+1}$

Para cada $i = 1, \ldots, k-1$ seja $b_i = d_i - d_{i+1}$ (≥ 0 pelo Lema 6.49) e seja $b_k = d_k$. Sejam $v(i)_1, \ldots v(i)_{b_i}$ vetores tais que

$$U(i) + L(\{v(i)_1, \ldots, v(i)_{b_i}) = W(i)$$

(estes vetores obtêm-se completando uma base de $U(i)$ a uma de $W(i)$). Note-se que cada vetor $v(i)_j$ tem filtração i e que o conjunto $\{v(i)_j\}$ é linearmente independente. Vamos ver que a base B pretendida é a união das cadeias de Jordan com início em $v(i)_j$ para cada i, j (em particular b_i é o número de blocos de tamanho i).

Definimos $B = \{N^m v(i)_j \colon 1 \leq i \leq k, 0 \leq m \leq i-1, 1 \leq j \leq b_i\}$. Uma vez que o número de elementos de B é

$$b_1 + 2b_2 + \ldots + kb_k = d_1 - d_2 + 2(d_2 - d_3) + \ldots + (k-1)(d_{k_1} - d_k) + kd_k = d_1 + \ldots + d_k$$

basta-nos demonstrar que B é linearmente independente. Note-se que, pelo Lema 6.48, a filtração do vetor $N^m v(i)_j$ é $i - m$. Suponhamos que

$$\sum_{i=1}^{k} \sum_{j=1}^{b_i} \sum_{m=1}^{i-1} \alpha_{i,j,m} N^m v(i)_j = 0 \tag{43}$$

Aplicando N^{k-1} à equação (43) todos os termos excepto os de filtração k são aniquilados e ficamos com

$$N^{k-1} \left(\sum_{j=1}^{b_k} \alpha_{k,j,0} v(k)_j \right) = 0$$

Como (pelo Lema 6.48) o conjunto $\{N^{k-1}v(k)_j\}$ é linearmente independente conclui-se que todos os coeficientes de vetores de B com filtração k se anulam. Aplicando N^{k-2} a (43) obtemos agora uma igualdade que envolve apenas os termos em (43) com filtração $k-1$:

$$N^{k-2}\left(\sum_{j=1}^{b_k}\alpha_{k,j,1}Nv(k)_j+\sum_{j=1}^{b_{k-1}}\alpha_{k-1,j,0}v(k-1)_j\right)=0$$

Uma vez que $U(k-1)=W(k-2)+L(\{Nv(k)_1,\dots,Nv(k)_{b_k}\})$ e $U(k-1)+L(\{v(k-1)_1,\dots,v(k-1)_{b_{k-1}}\})=W(k-1)$ vemos que o conjunto $\{Nv(k)_1,\dots,Nv(k)_{b_k},v(k-1)_1,\dots,v(k-1)_{b_{k-1}}\}$ é linearmente independente e portanto todos os coeficientes de vetores de B com filtração $k-1$ se anulam.

Aplicando sucessivamente N^{k-i} com $i=3,\dots k$ vemos como acima que todos os coeficientes se anulam, o que conclui a demonstração. □

Observação 6.50. *A demonstração anterior dá-nos um algoritmo (não muito prático) para determinar a base com respeito à qual N se expressa como uma matriz diagonal por blocos da forma* (39)*:*

(i) Determinar o índice de nilpotência k de N.

(ii) Determinar os subespaços $W(i)=N(N^i)$ e $U(i)=W(i-1)+N(W(i+1))$ e bases para estes compatíveis com as inclusões

$$0\subset U(1)\subset W(1)\subset U(2)\subset W(2)\subset W(k)=W$$

no sentido em que a base de cada subespaço contém as bases dos subespaços mais pequenos.

(iii) Tomar para base B a união das cadeias de Jordan (de comprimento i) dos elementos da base de $W(i)$ que não pertencem a $U(i)$. Em particular, o número de blocos de Jordan de tamanho i é $\dim W(i)-\dim U(i)$.

Note-se que em termos do diagrama (42) o número de blocos de tamanho i é a diferença entre os comprimentos das colunas i e $i+1$. Ou seja, os blocos de Jordan correspondem precisamente às *linhas* do diagrama, sendo o seu tamanho o comprimento da linha correspondente.

6.51. **A forma canónica de Jordan.** Podemos agora terminar a nossa classificação dos endomorfismos de um espaço vetorial complexo de dimensão finita.

Definição 6.52. *Uma matriz $n\times n$ diz-se um* bloco de Jordan*, se é da forma*

$$J=\begin{bmatrix}\lambda & 1 & \cdots & 0\\ 0 & \lambda & \ddots & 0\\ 0 & 0 & \ddots & 1\\ 0 & 0 & \cdots & \lambda\end{bmatrix} \tag{44}$$

Sendo V um espaço vetorial, $T\colon V\to V$ um endomorfismo e λ um escalar, uma cadeia de Jordan *é uma sucessão de vetores não nulos $(v_1,\dots,v_k)$ em V tal que*

$$Tv_1=\lambda v_1,\quad Tv_2=\lambda v_2+v_1,\quad \dots\quad Tv_k=\lambda v_k+v_{k-1} \tag{45}$$

As condições (45) podem ser re-escritas como

$$(T-\lambda)v_1 = 0, \quad (T-\lambda)v_2 = v_1, \quad \ldots \quad (T-\lambda)v_k = v_{k-1} \tag{46}$$

Em particular $(T-\lambda)^i v_i = 0$ pelo que os vetores de uma cadeia de Jordan são vetores próprios generalizados de λ, e formam uma cadeia de Jordan no sentido da Definição 40 para o operador nilpotente dado pela restrição de $(T-\lambda)$ a $E^g(\lambda)$.

Teorema 6.53 (Forma canónica de Jordan)**.** *Seja V um espaço vetorial complexo de dimensão finita e $T\colon V \to V$ um endomorfismo. Então existe uma base B para V tal que a matriz $A_{T,B,B}$ que representa T com respeito à base B é diagonal por blocos*

$$A_{T,B,B} = \begin{bmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & & J_m \end{bmatrix}$$

com cada J_i um bloco de Jordan.

Dem. Pelo Teorema 6.37 é suficiente ver que cada espaço próprio generalizado $E^g(\lambda)$ tem uma base com respeito à qual a restrição de T a $E^g(\lambda)$ é representada por uma matriz diagonal por blocos com cada bloco um bloco de Jordan. Pelo Teorema 6.42, a restrição de $(T-\lambda)$ a $E^g(\lambda)$ admite tal representação, tendo os blocos 0 na diagonal. Mas isso é exatamente equivalente à afirmação que a restrição de T a $E^g(\lambda)$ é representada por uma matriz diagonal por blocos, sendo os blocos de Jordan com λ na diagonal. □

Observação 6.54. *Se aplicarmos o Teorema 6.53 à transformação linear $T\colon \mathbb{C}^n \to \mathbb{C}^n$ representada na base canónica pela matriz $A \in M_{n\times n}(\mathbb{C})$ ele diz-nos que existe uma matriz invertível S (a matriz de mudança de base da base B para a base canónica) e uma matriz diagonal por blocos J cujos blocos são de Jordan, tais que*

$$A = SJS^{-1}$$

O Teorema é muitas vezes enunciado nesta forma.

Observação 6.55. *É fácil ver que a matriz diagonal por blocos no enunciado do Teorema 6.53 é única a menos de troca da ordem dos blocos. Isso está longe de ser verdade para a base B, como aliás já era verdade no caso em que a transformação é diagonalizável.*

O Teorema 6.53 classifica *todos os endomorfismos de espaços vetoriais complexos de dimensão finita a menos de isomorfismo. Se $T_1, T_2\colon V \to V$ têm a mesma forma canónica de Jordan, então sendo S o isomorfismo que envia os vetores da base B_1 dada pelo Teorema para T_1, por ordem, para os vetores da base B_2 correspondente a T_2 teremos $T_2 = S \circ T_1 \circ S^{-1}$, pelo que o isomorfismo S "traduz" T_1 em T_2.*

Reciprocamente, se existe um isomorfismo S tal que $T_2 = ST_1S^{-1}$, os endomorfismos T_1 e T_2 têm a mesma forma canónica de Jordan (exercício).

Para determinar a forma canónica de Jordan de um endomorfismo T (o que significa calcular a matriz diagonal por blocos e a base) podemos começar por calcular os valores próprios de T. Estes dizem-nos que escalares aparecem na diagonal da matriz $A_{T,B,B}$. Embora seja possível determinar os espaços próprios generalizados de cada valor próprio λ e depois aplicar o algoritmo da Observação 6.50 à restrição de $T - \lambda$ a cada $E^g(\lambda)$, é mais prático proceder da seguinte forma.

Uma vez que os elementos da base correspondentes às primeiras colunas de cada bloco J_i são vetores próprios, a multiplicidade geométrica de cada valor próprio λ dá-nos o número de blocos com λ na diagonal.

Para determinar o resto da base B e o comprimento dos blocos tentamos resolver recursivamente as equações (46) começando com um vetor próprio v_1. Isto pode requerer algum cuidado na escolha do vetor v_1 como iremos ver nos exemplos que se seguem.

Finalmente, o polinómio característico de T dá informação sobre a forma canónica de Jordan que pode facilitar a determinação da base:

Proposição 6.56. *Seja V um espaço vetorial de dimensão finita, $T\colon V \to V$ um endomorfismo, e $p(\lambda)$ o seu polinómio característico.*

(a) A multiplicidade do valor próprio λ_i enquanto raíz de p (isto é, o maior expoente m tal que $(\lambda - \lambda_i)^m$ divide $p(\lambda)$) é igual à multiplicidade algébrica de λ_i, $\dim E^g(\lambda_i)$.

(b) Se $\mathbb{K} = \mathbb{C}$ então

$$p(\lambda) = (\lambda_1 - \lambda)^{n_1}(\lambda_2 - \lambda)^{n_2} \cdots (\lambda_k - \lambda)^{n_k}$$

onde $\lambda_1, \ldots, \lambda_k$ são os valores próprios (distintos) de T e n_i as suas multiplicidades algébricas.

Dem. (a) Quando o corpo de base é $\mathbb{C}$ esta afirmação é uma consequência imediata da forma canónica de Jordan. De facto a multiplicidade de λ_i como raíz é igual aos número de vezes que λ_i aparece na diagonal na forma canónica de Jordan, que é precisamente o número de elementos numa base para o espaço próprio generalizado de λ_i.

Para demonstrar a afirmação em geral (para um espaço vetorial sobre um corpo qualquer) note-se que pelo Teorema 6.42 existe uma base B_1 para $E^g(\lambda_i)$ formada por cadeias de Jordan para o valor próprio λ_i. Completando esta base com um conjunto B_2 de vetores de V tal que $B = B_1 \cup B_2$ é uma base de B obtemos uma representação matricial diagonal por blocos

$$A_{T,B,B} = \begin{bmatrix} J & X \\ 0 & T' \end{bmatrix}$$

onde as colunas de J correspondem aos vetores do conjunto B_1 (J é diagonal por blocos, com blocos de Jordan correspondentes a λ_i na diagonal) e as restantes colunas correspondem aos vetores de B_2. Pela Proposição 5.20 o polinómio caraterístico de T é igual a $(\lambda_i - \lambda)^m q(\lambda)$, onde $m = \#B_1 = \dim E^g(\lambda_i)$, e $q(\lambda)$ é o polinómio caraterístico de T'.

Argumentando como na demonstração do Teorema 6.37 vemos que λ_i não pode ser um valor próprio de T', e portanto $(\lambda_i - \lambda)$ não divide $q(\lambda)$. Isto conclui a demonstração.

(b) É uma consequência imediata do facto de podermos calcular o polinómio caraterístico usando a forma canónica de Jordan de T.

□

Observação 6.57. *A Proposição 6.56(a) justifica a terminologia* multiplicidade algébrica.

A Proposição anterior ajuda a calcular a forma canónica de Jordan de uma matriz complexa. De facto, uma fatorização do polinómio caraterístico em fatores do primeiro grau dá-nos a multiplicidade algébrica de cada valor próprio λ_i e portanto o número de vezes que λ_i aparece na diagonal da forma canónica de Jordan (que é a soma dos comprimentos dos blocos correspondentes a λ_i).

Exemplo 6.58. *No Exemplo 6.26 a fatorização $p(\lambda) = (\lambda-1)(\lambda+1)(\lambda-2)^2$ do polinómio caraterístico implica que a multiplicidade algébrica de 2 é igual a 2. Assim, ao verificar que a multiplicidade geométrica de $\lambda = 2$ é apenas 1, podemos concluir, sem quaisquer outros cálculos adicionais, que a forma canónica de Jordan desta matriz é*

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

Exemplo 6.59. *Retomando o Exemplo 6.34 vemos que, uma vez que a multiplicidade geométrica de cada valor próprio é 1, teremos um bloco para cada valor próprio. À partida haveria duas possibilidades para a matriz diagonal por blocos (a menos de troca de ordem dos blocos):*

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \qquad ou \qquad \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

No entanto vimos já que o espaço próprio generalizado de $\lambda = 2$ tem dimensão 2 (e isto é também uma consequência da Proposição 6.56(a)) pelo que teremos necessariamente a segunda opção. A base $B = (v_1, v_2, v_3)$ terá que ser formada por um vetor próprio v_1 de $\lambda = 2$, um vetor próprio v_3 de $\lambda = 1$ e um vetor próprio generalizado v_2 de 2 satisfazendo

$$(A - 2I)v_2 = v_1$$

Neste exemplo as escolhas possíveis para v_1 e v_3 são únicas a menos de um escalar não nulo. Se tomarmos $v_1 = (1, 0, 1)$ e $v_3 = (1, -1, 0)$ temos que resolver a equação

$$(A - 2I)v_2 = v_1 \Leftrightarrow \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & -1 \\ 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

Uma solução é por exemplo $v_2 = (0, 1, 0)$ (mas poderíamos somar a este vetor qualquer elemento do núcleo de $(A - 2I)$, isto é, qualquer vetor próprio de 2). Conclui-se assim que neste exemplo podemos tomar para a base B no Teorema 6.53

$$B = ((1, 0, 1), (0, 1, 0), (1, -1, 0))$$

Exemplo 6.60. *Seja A uma matriz com forma canónica de Jordan*

$$J = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{47}$$

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 pôe na primeira coluna da matriz S. De facto, só será possível resolver a equação

$$(A - I)v_2 = v$$

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

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

Verifica-se facilmente que 1 é o único valor próprio. 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 $\lambda = 1$ tem multiplicidade geométrica 2. Há portanto dois blocos de Jordan e a forma canónica de Jordan de A é necessariamente (47).

Não é no entanto possível resolver a equação

$$(A - I)v_2 = v_1$$

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

$$\begin{bmatrix} -1 & -1 & 2 \\ -1 & -1 & 2 \\ -1 & -1 & 2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \Leftrightarrow 2c = a + b + 1$$

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* v_3 *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 base*

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

e temos

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

Exemplo 6.61. *Suponhamos que* V *é um espaço vetorial de dimensão* 5 *e* $T\colon V \to V$ *tem um único valor próprio* λ *com multiplicidade geométrica* 2. *As possíveis formas canónicas de Jordan (a menos de troca de blocos) são*

$$\begin{bmatrix} \lambda & 1 & 0 & 0 & 0 \\ 0 & \lambda & 1 & 0 & 0 \\ 0 & 0 & \lambda & 1 & 0 \\ 0 & 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & 0 & \lambda \end{bmatrix} \qquad \text{ou} \qquad \begin{bmatrix} \lambda & 1 & 0 & 0 & 0 \\ 0 & \lambda & 1 & 0 & 0 \\ 0 & 0 & \lambda & 0 & 0 \\ 0 & 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & 0 & \lambda \end{bmatrix}$$

Sendo $B = (v_1, v_2, v_3, v_4, v_5)$ *uma base na qual* T *fica em forma de Jordan, podemos distinguir os dois casos da seguinte forma. Na matriz da direita, uma vez que os blocos têm dimensão* ≤ 3 *o endomorfismo* $(T-\lambda)^3$ *é identicamente nulo. Isso não acontece na matriz da esquerda, onde* $(T-\lambda)^3 v_4 = v_1$.

Para acharmos a base B *resolvendo as equações* (46) *indutivamente teremos de ter o cuidado de começar com um vetor próprio* v_1 *de* λ *que esteja na imagem de* $(T-\lambda)^3$ *no caso da matriz da esquerda, e na imagem de* $(T-\lambda)^2$ *no caso da matriz da direita. O segundo vetor próprio (*v_5 *no caso da esquerda e* v_4 *no caso da direita) tem unicamente que ser escolhido de forma a gerar o espaço próprio de* λ *juntamente com* v_1.

6.62. **O Teorema de Cayley-Hamilton.** Terminamos a matéria desta secção com um resultado fundamental que é muitas vezes útil para fazer cálculos com matrizes quadradas. Seja V um espaço vetorial de dimensão finita n e $T\colon V \to V$ um endomorfismo. Dado um vetor $v \in V$, sabemos que existem necessariamente escalares não todos nulos $c_0, c_1, \ldots, c_n$ tais que

$$c_0 v + c_1 T v + \ldots + c_n T^n v = 0$$

(ou porque há alguma repetição na sequência $v, Tv, \ldots, T^n v$, ou porque um conjunto com mais de n vetores de V é necessariamente linearmente dependente). O próximo Teorema garante, em particular, que existe uma escolha para os coeficientes c_i que funciona para todos os vetores $v \in V$ simultaneamente.

Teorema 6.63 (Teorema de Cayley-Hamilton)**.** *Seja V um espaço vetorial de dimensão finita, $T\colon V \to V$ um endomorfismo e $p(\lambda)$ o seu polinómio caraterístico. Então $p(T) = 0$ (isto é, $p(T)\colon V \to V$ é o endomorfismo identicamente nulo).*

Dem. Seja $A \in M_{n\times n}(\mathbb{K})$ a matriz que representa T com respeito a uma qualquer base B de V. É suficiente ver que $p(A) = 0$. Pela Proposição 5.11(iv) temos[14]

$$(A - \lambda I)\operatorname{cof}(A - \lambda I)^T = p(\lambda)I \tag{48}$$

Seja

$$p(\lambda) = (-1)^n\lambda^n + c_{n-1}\lambda^{n-1} + \ldots + c_1\lambda + c_0$$

Cada entrada da matriz $\operatorname{cof}(A-\lambda I)^T$ é o determinante de uma submatriz $(n-1)\times(n-1)$ de $(A-\lambda I)^T$ e portanto é uma expressão polinomial de grau $\leq (n-1)$ em λ. Pondo as potências de λ em evidência vemos que existem matrizes $B_i \in M_{n\times n}(\mathbb{K})$ tais que

$$\operatorname{cof}(A - \lambda I)^T = B_0 + B_1\lambda + \ldots + B_{n-1}\lambda^{n-1}$$

Substituindo em (48) obtemos

$$(A - \lambda I)(B_0 + B_1\lambda + \ldots + \lambda^{n-1}B_{n-1}) = ((-1)^n\lambda^n + \ldots + c_1\lambda + c_0)I$$

Distribuindo o produto do lado direito do sinal de igual e igualando os coeficientes de λ^k obtemos as seguintes igualdades

$$\begin{cases} AB_0 = c_0 I \\ AB_1 - B_0 = c_1 I \\ AB_2 - B_1 = c_2 I \\ \vdots \\ AB_{n-1} - B_{n-2} = c_{n-1} I \\ -B_{n-1} = (-1)^n I \end{cases}$$

Multiplicando a i-ésima equação por A^{i-1} e somando obtemos

$$\begin{aligned} AB_0 + A(AB_1 - B_0) + A^2(AB_2 - B_1) + \ldots + A^{n-1}(AB_{n-1} - B_{n-2}) - A^n B_{n-1} = \\ = c_0 I + c_1 A + \ldots + c_{n-1}A^{n-1} + (-1)^n A^n \end{aligned}$$

A soma telescópica do lado esquerdo do sinal de igual anula-se, o que conclui a demonstração. □

O seguinte exemplo ilustra os cálculos efetuados na demonstração anterior.

[14] O significado de (48) é algo subtil. No caso em que o corpo $\mathbb{K}$ tem infinitos elementos, os polinómios são determinados pelas funções $\mathbb{K} \to \mathbb{K}$ que definem e portanto a validade de (48) segue do facto de a igualdade se verificar para todo o $\lambda \in \mathbb{K}$. Quando $\mathbb{K}$ é um corpo finito, é necessário observar que, na dedução da igualdade descrita na Proposição 5.11(iv), nunca usámos a invertibilidade dos elementos não nulos de $\mathbb{K}$ mas apenas os restantes axiomas de corpo. O resultado é portanto válido para matrizes cujas entradas são polinómios com coeficientes em $\mathbb{K}$.

Exemplo 6.64. *Seja A a matriz $\begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix}$ que tem polinómio caraterístico $p(\lambda) = \lambda^2 - 3\lambda - 4$. A identidade (48) afirma neste caso que*

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

$$\left(\begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix} - \lambda\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)\left(\begin{bmatrix} 2 & -2 \\ -3 & 1 \end{bmatrix} - \lambda\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) = \lambda^2\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \lambda\begin{bmatrix} -3 & 0 \\ 0 & -3 \end{bmatrix} + \begin{bmatrix} -4 & 0 \\ 0 & -4 \end{bmatrix}$$

Distribuindo a soma à esquerda e igualando os coeficientes de λ^k para $k = 0, 1, 2$ obtemos três equações matriciais que implicam a relação $A^2 - 3A - 4I = 0$.

Observação 6.65. *Para endomorfismos de espaços complexos de dimensão finita, o Teorema de Cayley-Hamilton é uma consequência imediata da forma canónica de Jordan.*

Exemplo 6.66. *Consideremos a matriz*

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

O seu polinómio caraterístico é

$$p(\lambda) = \begin{vmatrix} 1-\lambda & 0 & -1 \\ 0 & 2-\lambda & 1 \\ -1 & 0 & 2-\lambda \end{vmatrix} = -\lambda^3 + 5\lambda^2 - 7\lambda + 2$$

Portanto

$$-A^3 + 5A^2 - 7A + 2I = 0 \Leftrightarrow A(-A^2 + 5A - 7I) = -2I \Rightarrow A^{-1} = \tfrac{1}{2}(A^2 - 5A + 7I)$$

É fácil calcular o valor de $t(A)$ para qualquer polinómio $t(x)$. Dividindo $t(x)$ pelo polinómio caraterístico obtemos

$$t(x) = q(x)p(x) + r(x)$$

sendo o grau de r menor ou igual a 2. Uma vez que $p(A) = 0$, temos $t(A) = q(A)p(A) + r(A) = r(A)$.

6.67. **Leitura opcional: Endomorfismos de espaços vetoriais reais.** Vamos agora usar o nosso entendimento dos endomorfismos de espaços complexos para descrever os endomorfismos de espaços vetoriais reais de dimensão finita.

Uma vez que os números reais são um subconjunto dos números complexos, temos uma inclusão natural $M_{n\times n}(\mathbb{R}) \subset M_{n\times n}(\mathbb{C})$ e portanto uma matriz real pode ser encarada como uma matriz complexa cujas entradas têm parte imaginária nula. Uma matriz $A \in M_{n\times n}(\mathbb{R})$ determina portanto um endomorfismo $x \mapsto Ax$ de $\mathbb{C}^n$ ao qual podemos aplicar o Teorema 6.53.

O processo descrito no parágrafo anterior tem uma versão para transformações lineares: todo o espaço vetorial real pode ser extendido a um espaço vetorial complexo - a sua complexificação - de uma maneira que generaliza a inclusão $\mathbb{R}^n \subset \mathbb{C}^n$ e um endomorfismo de um espaço vetorial real determina um endomorfismo da sua complexificação.

Definição 6.68. *Seja V um espaço vetorial real. A* complexificação *de V é o espaço vetorial complexo $V^{\mathbb{C}}$ que tem por conjunto subjacente $V \times V$ e operações de soma e multiplicação por escalar definidas por*

$$(v_1, v_2) + (w_1, w_2) = (v_1 + w_1, v_2 + w_2) \qquad (a + bi) \cdot (v_1, v_2) = (av_1 - bv_2, av_2 + bv_1)$$

Definimos as partes real e imaginária de $(v_1, v_2) \in V^{\mathbb{C}}$ por $\mathrm{Re}((v_1, v_2)) = v_1$ e $\mathrm{Im}((v_1, v_2)) = v_2$ e a operação de conjugação *em $V^{\mathbb{C}}$ por*

$$\overline{(v_1, v_2)} = (v_1, -v_2)$$

Note-se que $(0+1i)\cdot(v, 0) = (0, v)$. Podemos identificar V com o subconjunto $\{(v, 0) \colon v \in V\} \subset V^{\mathbb{C}}$ e mediante esta identificação é habitual escrever o elemento $(v_1, v_2) \in V^{\mathbb{C}}$ como $v_1 + iv_2$. Nestes termos a multiplicação por escalar é definida pela fórmula familiar

$$(a + bi)(v_1 + iv_2) = (av_1 - bv_2) + i(av_2 + bv_1)$$

e a conjugação escreve-se

$$\overline{v_1 + iv_2} = v_1 - iv_2$$

Deixamos como exercício a verificação que com as operações de soma e multiplicação definidas acima $V^{\mathbb{C}}$ é de facto um espaço vetorial complexo.

Note-se que a conjugação é um isomorfismo entre o espaço vetorial real subjacente a $V^{\mathbb{C}}$ e ele próprio.

Definição 6.69. *Seja V um espaço vetorial real e $T \colon V \to V$ um endomorfismo. A* complexificação *de T é a transformação linear complexa $T^{\mathbb{C}} \colon V^{\mathbb{C}} \to V^{\mathbb{C}}$ definida por*

$$T^{\mathbb{C}}(v_1 + iv_2) = T(v_1) + iT(v_2)$$

Deixamos como exercício a verificação que $T^{\mathbb{C}}$ é de facto uma transformação linear de espaços vetoriais complexos.

Exemplo 6.70. *(i) Se $V = \mathbb{R}^n$ podemos identificar $V^{\mathbb{C}}$ naturalmente com $\mathbb{C}^n$ usando o isomorfismo de espaços vetoriais complexos dado por*

$$((x_1, x_2, \ldots, x_n), (y_1, y_2 \ldots, y_n)) \mapsto (x_1 + iy_1, x_2 + iy_2, \ldots, x_n + iy_n)$$

Se $T \colon \mathbb{R}^2 \to \mathbb{R}^2$ for a transformação linear definida pela expressão

$$T(x, y) = (2x + y, x - 3y)$$

mediante a identificação acima de $(\mathbb{R}^2)^{\mathbb{C}}$ com $\mathbb{C}^2$, a transformação linear $T^{\mathbb{C}} \colon \mathbb{C}^2 \to \mathbb{C}^2$ é dada pela mesma expressão. De facto, temos

$$T^{\mathbb{C}}(1 + 0i, 0 + 0i) = T(1, 0) + iT(0, 0) = (2 + 0i, 1 + 0i = (2, 1))$$

$$T^{\mathbb{C}}(0 + 0i, 1 + 0i) = T(0, 1) + iT(0, 0) = (1 + 0i, -3 + 0i) = (1, -3)$$

logo

$$T^{\mathbb{C}}(x, y) = (2x + y, x - 3y)$$

(ii) Se V for o espaço dos polinómios reais, $V^{\mathbb{C}}$ identifica-se naturalmente com o espaço dos polinómios com coeficientes complexos mediante o isomorfismo

$$(a_0 + a_1x + \ldots + a_nx^n, b_0 + b_1x + \ldots + b_nx^n) \mapsto (a_0 + ib_0) + (a_1 + ib_1)x + \ldots + (a_n + ib_n)x^n$$

Se $B = (v_1, \ldots, v_n)$ é uma base para V é imediato verificar que B (ou mais precisamente o conjunto correspondente $((v_1, 0), \ldots, (v_n, 0)) \subset V^{\mathbb{C}}$) é também uma base para $V^{\mathbb{C}}$ (agora como espaço vetorial complexo) e que a representação matricial de $T^{\mathbb{C}}$ na base B coincide com a representação matricial de T na base B:

$$A_{T^{\mathbb{C}},B,B} = A_{T,B,B}$$

sendo a matriz real $A_{T,B,B}$ encarada como uma matriz complexa cujas entradas têm parte imaginária nula. No Exemplo (6.70)(i) acima temos

$$A_{T^{\mathbb{C}},B_{can},B_{can}} = A_{T,B_{can},B_{can}} = \begin{bmatrix} 2 & 1 \\ 1 & -3 \end{bmatrix}$$

Proposição 6.71. *Seja V um espaço vetorial real e $T\colon V \to V$ um endomorfismo.*

(i) *$\lambda \in \mathbb{C}$ é um valor próprio de $T^{\mathbb{C}}$ se e só se $\overline{\lambda}$ é um valor próprio de $T^{\mathbb{C}}$*

(ii) *$v \in V^{\mathbb{C}}$ é um vetor próprio (generalizado) de $T^{\mathbb{C}}$ se e só se $\overline{v}$ é um vetor próprio (generalizado) de $T^{\mathbb{C}}$.*

(iii) *$(v_1, v_2, \ldots, v_k)$ formam uma cadeia de Jordan em $V^{\mathbb{C}}$ associada ao valor próprio λ se e só se $(\overline{v_1}, \overline{v_2}, \ldots, \overline{v_k})$ formam uma cadeia de Jordan em $V^{\mathbb{C}}$ associada ao valor próprio $\overline{\lambda}$.*

Dem. Escrevendo $\lambda = a + bi$ e $v = x + iy \in V^{\mathbb{C}}$ com $x, y \in V$ temos

$$(49)\quad T(x+iy) = (a+ib)(x+iy) \Leftrightarrow Tx+iTy = ax-by+i(ay+bx) \Leftrightarrow \begin{cases} Tx = ax - by \\ Ty = ay + bx \end{cases}$$

o que claramente é equivalente a

$$T(x - iy) = (a - ib)(x - iy) \Leftrightarrow Tx - iTy = ax - by - i(ay + bx)$$

logo λ é um valor próprio se e só se $\overline{\lambda}$ é um valor próprio, e $v = x + iy$ é um vetor próprio associado a λ se e só se $\overline{v} = x - iy$ é um vetor próprio associado a $\overline{\lambda}$. Isto mostra (i) e uma parte de (ii). A demonstração de (iii) é inteiramente análoga e é deixada como exercício. Uma vez que um vetor $v \neq 0$ é um vetor próprio generalizado de λ se e só se $(v, (T-\lambda)v, \ldots, (T-\lambda)^k v)$ é uma cadeia de Jordan para algum k, (iii) implica a afirmação que falta em (ii). $\square$

O resultado anterior diz que os espaços próprios generalizados de $T^{\mathbb{C}}$ correspondentes a valores próprios complexos (isto é, com parte imaginária não nula) ocorrem em pares conjugados e que podemos tomar para base de $E^g(\overline{\lambda})$ os vetores conjugados de uma base de $E^g(\lambda)$. Assim, na forma canónica de Jordan para $T^{\mathbb{C}}$ há uma correspondência bijetiva entre os blocos com λ na diagonal e os blocos com $\overline{\lambda}$ na diagonal.

A observação anterior juntamente com o resultado seguinte permite-nos obter uma forma canónica para os endomorfismos reais.

Proposição 6.72. *Seja V um espaço vetorial real e $\{v_1, \ldots, v_n\}$ um subconjunto de vetores (distintos) de $V^{\mathbb{C}}$. Se $\{v_1, \ldots, v_n, \overline{v_1}, \ldots, \overline{v_n}\}$ é um subconjunto linearmente independente de $V^{\mathbb{C}}$ então $\{\mathrm{Re}(v_1), \mathrm{Im}(v_1), \ldots, \mathrm{Re}(v_n), \mathrm{Im}(v_n)\}$ é um subconjunto linearmente independente de V.*

Dem. Usando a identificação de V com o conjunto $\{(x,0)\colon x\in V\}$ de $V^{\mathbb{C}}$, temos

$$\operatorname{Re}(v)=\frac{v+\overline{v}}{2}\qquad \operatorname{Im}(v)=\frac{v-\overline{v}}{2i}$$

Dados escalares reais $\alpha_1,\dots,\alpha_n,\beta_1,\dots,\beta_n$, a expressão

$$\alpha_1\operatorname{Re}(v_1)+\ldots+\alpha_n\operatorname{Re}(v_n)+\beta_1\operatorname{Im}(v_1)+\ldots+\beta_n\operatorname{Im}(v_n)=0$$

é equivalente a

$$\begin{aligned} &\alpha_1\frac{v_1+\overline{v_1}}{2}+\ldots+\alpha_n\frac{v_n+\overline{v_n}}{2}+\beta_1\frac{v_1-\overline{v_1}}{2i}+\ldots+\beta_n\frac{v_n-\overline{v_n}}{2i})=0\\ \Leftrightarrow\quad &\frac{\alpha_1-i\beta_1}{2}v_1+\ldots+\frac{\alpha_n-i\beta_n}{2}v_n+\frac{\alpha_1+i\beta_1}{2}\overline{v_1}+\ldots+\frac{\alpha_n+i\beta_n}{2}\overline{v_n}=0\end{aligned}$$

logo se $\{\operatorname{Re}(v_1),\operatorname{Im}(v_1),\dots,\operatorname{Re}(v_n),\operatorname{Im}(v_n)\}$ é linearmente dependente em V, então o conjunto $\{v_1,\dots,v_n,\overline{v_1},\dots,\overline{v_n}\}$ é linearmente dependente em $V^{\mathbb{C}}$, o que conclui a demonstração. $\square$

Definição 6.73. *Um* bloco de Jordan real *é uma matriz quadrada da forma* (44) *com* λ *real ou da forma*

$$(50)\qquad \begin{bmatrix} a & -b & 1 & 0 & \cdots & & 0 & 0\\ b & a & 0 & 1 & 0 & \cdots & & \vdots\\ 0 & 0 & a & -b & 1 & 0 & \ddots & \\ 0 & 0 & b & a & 0 & 1 & \ddots & \\ \vdots & & \ddots & & \ddots & & 1 & 0\\ & & & & & & 0 & 1\\ 0 & 0 & \cdots & & 0 & 0 & a & -b\\ 0 & 0 & \cdots & & 0 & 0 & b & a \end{bmatrix}$$

com $a,b\in\mathbb{R}$ *e* $b\neq 0$.

Teorema 6.74. *(Forma canónica de Jordan real) Seja* V *um espaço vetorial real de dimensão finita e* $T\colon V\to V$ *um endomorfismo. Então existe uma base* B *para* V *tal que a matriz* $A_{T,B,B}$ *que representa* T *com respeito à base* B *é diagonal por blocos*

$$A_{T,B,B}=\begin{bmatrix} J_1 & 0 & \cdots & 0\\ 0 & J_2 & \ddots & \vdots\\ \vdots & & \ddots & 0\\ 0 & \cdots & & J_m\end{bmatrix}$$

com cada J_i *um bloco de Jordan real.*

Dem. Seja $T^{\mathbb{C}}\colon V^{\mathbb{C}}\to V^{\mathbb{C}}$ a complexificação de T.

Se $\lambda\in\mathbb{R}$ então $(T^{\mathbb{C}}-\lambda)^k v=0\Leftrightarrow (T-\lambda)^k\operatorname{Re}(v)=(T-\lambda)^k\operatorname{Im}(v)=0$ pelo que o espaço próprio generalizado de λ para $T^{\mathbb{C}}$ é a complexificação do espaço próprio generalizado de λ para T. Em particular, podemos assumir que as cadeias de Jordan para $T^{\mathbb{C}}$ que correspondem a blocos com $\lambda \in \mathbb{R}$ na diagonal são formadas por vetores de $V \subset V^{\mathbb{C}}$.

Pela Prop 6.71, os blocos de Jordan de $T^{\mathbb{C}}$ correspondentes a valores próprios complexos $\lambda = a + bi$ com $b \neq 0$ ocorrem em pares conjugados e para cada par $J, \overline{J}$ podemos escolher cadeias de Jordan conjugadas $(v_1, \ldots, v_k)$ e $(\overline{v_1}, \ldots, \overline{v_k})$ com

$$T^{\mathbb{C}} v_1 = \lambda v_1, \ldots, T^{\mathbb{C}} v_k = \lambda v_k + v_{k-1}$$

Seja B' a base de $V^{\mathbb{C}}$ formada por todas as cadeias de Jordan para $V^{\mathbb{C}}$ escolhidas da forma acima (isto é, escolhendo vetores de V para as cadeias correspondentes a valores próprios reais e escolhendo cadeias conjugadas para blocos de valores próprios conjugados correspondentes). Tomamos para B o subconjunto de V formado por

(i) cadeias de Jordan em B' correspondentes a valores próprios reais,
(ii) $\{\text{Re}(v_1), \text{Im}(v_1), \ldots, \text{Re}(v_k), \text{Im}(v_k)\}$ para cada par conjugado de cadeias $(v_1, \ldots, v_k)$ e $(\overline{v_1}, \ldots, \overline{v_k})$

Pelo Lema 6.72, os conjuntos $\{\text{Re}(v_1), \text{Im}(v_1), \ldots, \text{Re}(v_k), \text{Im}(v_k)\}$ são linearmente independentes. Contas inteiramente análogas a (49) mostram que a restrição de T a

$$L(\{\text{Re}(v_1), \text{Im}(v_1), \ldots, \text{Re}(v_k), \text{Im}(v_k)\})$$

na base $(\text{Re}(v_1), \text{Im}(v_1), \ldots, \text{Re}(v_k), \text{Im}(v_k))$ é dado pelo bloco de Jordan real $(2k) \times (2k)$ (50) com a e b tais que $v_1 \in V^{\mathbb{C}}$ é um vetor próprio de $a - bi$.

B é um conjunto linearmente independente porque a soma dos subespaços gerados pelos vários conjuntos de vetores de tipo (i) e (ii) é direta (isso é claramente verdade para as suas complexificações pelo Teorema da Decomposição Primária). Uma vez que o número de elementos de B é igual à dimensão de V conclui-se que B é uma base para V com as propriedades requeridas. $\square$

Observação 6.75. *Ao contrário da forma canónica de Jordan complexa, a forma real não é única a menos da ordem dos blocos. Nos blocos correspondentes a valores próprios complexos, o sinal da parte imaginária b pode ser escolhido arbitrariamente (mas a escolha tem de ser constante em cada bloco). A troca do sinal de b corresponde à simetria dada pela conjugação que troca a ordem do par $(\lambda, \overline{\lambda})$. Esta indeterminação pode ser resolvida convencionando que nos blocos com entrada complexa temos sempre $b > 0$.*

Exemplo 6.76. *Vamos determinar a forma canónica de Jordan real da transformação linear $T\colon \mathbb{R}^4 \to \mathbb{R}^4$ representada na base canónica pela matriz*

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

Calcula-se que os valores próprios são 0, e $1 \pm i$. A multiplicidade geométrica de 0 é 2, sendo uma base para o espaço próprio de 0 dada pelos vetores $(0,1,1,0)$ e $(1,-1,0,1)$. A matriz A é portanto diagonalizável enquanto matriz complexa. A forma canónica de

Jordan real de A é

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

Um vetor próprio de $1-i$ é $(1-i, 0, -i, 1)$ logo uma matriz S tal que $A = SJS^{-1}$ é, por exemplo,

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

6.77. **Cultura geral: classificação geral dos endomorfismos de espaços vetoriais de dimensão finita.** Nesta secção de leitura opcional vamos, por uma questão de cultura geral, indicar brevemente como se generaliza a classificação dos endomorfismos que vimos para $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$ a corpos arbitrários. Seja V um espaço vetorial de dimensão finita sobre o corpo $\mathbb{K}$ e $T\colon V \to V$ um endomorfismo qualquer.

Todo o polinómio com coeficientes em $\mathbb{K}$ pode ser fatorizado de forma única como o produto de polinómios ditos *irredutíveis* que desempenham um papel análogo ao dos números primos nos inteiros. Quando $\mathbb{K} = \mathbb{C}$ o Teorema Fundamental da Álgebra garante que os únicos polinómios (mónicos) irredutíveis são da forma $(x - \lambda)$ com $\lambda \in \mathbb{C}$ e daqui segue fácilmente que os únicos polinómios (mónicos) reais irredutíveis são $(x - \lambda)$ com $\lambda \in \mathbb{R}$ e $(x - a)^2 + b^2$ para $a + bi \in \mathbb{C} \setminus \mathbb{R}$.

A versão geral do Teorema da Decomposição Primária 6.37 afirma que todo o endomorfismo T de um espaço vetorial de dimensão finita pode ser decomposto numa soma direta de endomorfismos cujos polinómios caraterísticos são potências de polinómios irredutíveis, havendo uma parcela para cada polinómio irredutível que divide o polinómio caraterístico de T. Quando $\mathbb{K} = \mathbb{C}$ este resultado é precisamente o Teorema da Decomposição Primária 6.37. Quando $\mathbb{K} = \mathbb{R}$, este resultado corresponde à decomposição do endomorfismo em blocos de Jordan reais e complexos (cada valor próprio real e cada par de valores próprios complexos conjugados corresponde a uma parcela).

Vejamos que forma tomam os blocos de uma decomposição primária. Dado $v \in V \setminus \{0\}$, associamos a v o subespaço cíclico $W = L(\{T^i v\colon i \in \mathbb{N}_0\})$. Sendo $p(x) = c_0 + c_1 x + \ldots + x^k$ um polinómio com coeficientes em $\mathbb{K}$ de grau mínimo (dito o polinómio mínimo de v) tal que

$$p(T)v = 0$$

temos que $B = (v, Tv, T^2 v, \ldots, T^{k-1} v)$ é uma base para W e, nessa base para W, a matriz que representa $T_{|W}$ é

$$A_{T_{|W},B,B} = \begin{bmatrix} 0 & 0 & \ldots & -c_0 \\ 1 & 0 & & -c_1 \\ 0 & \ddots & \ddots & \vdots \\ 0 & \ddots & 1 & -c_{k-1} \end{bmatrix}$$

que se chama a *matriz companheira* do polinómio p. Estas matrizes têm uma forma bastante simples: a grande maioria das entradas é igual a zero. O seu polinómio característico é exatamente o polinómio $p(\lambda)$ como verão nos exercícios sobre o determinante. Usando o Teorema de Cayley-Hamilton, não é díficil verificar que o polinómio mínimo de um dado $v \in V \setminus \{0\}$ é sempre um fator do polinómio característico de T.

Um Teorema fundamental da Álgebra Linear afirma que para todo o endomorfismo T de um espaço vetorial de dimensão finita, existe uma decomposição de V como a soma direta de subespaços cíclicos, ou alternativamente, que existe uma base B de V tal que $A_{T,B,B}$ é diagonal por blocos sendo cada bloco diagonal uma matriz companheira de um fator do polinómio característico[15].

Como viram nos exercícios sobre endomorfismos, uma fatorização de um polinómio leva a uma decomposição em blocos da sua matriz companheira. Aplicando o resultado mencionado no parágrafo anterior aos blocos de uma decomposição primária (cujos polinómios caraterísticos são potências de polinómios irredutíveis) obtemos uma decomposição diagonal por blocos na qual, ao longo da diagonal, temos matrizes companheiras de polinómios irredutíveis. Quando $\mathbb{K} = \mathbb{C}$ isto é exatamente a forma canónica de Jordan, enquanto que no caso em que $\mathbb{K} = \mathbb{R}$ obtemos uma matriz semelhante à forma canónica de Jordan real.

Para um excelente tratamento elementar dos temas descritos nos parágrafos anteriores recomendamos [HK]. Quem continuar a estudar Álgebra irá provavelmente ver todos estes resultados como um caso particular do *Teorema de classificação dos módulos finitamente gerados sobre domínios de ideais principais*. Um endomorfismo $T\colon V \to V$ dá a V a estrutura de um *módulo*[16] sobre o domínio de ideais principais $\mathbb{K}[t]$ dos polinómios com coeficientes em $\mathbb{K}$: define-se a multiplicação de $p(t) \in \mathbb{K}[t]$ por um elemento $v \in V$ como $p(t) \cdot v = p(T)v$. As formas normais para os endomorfismos brevemente descritas acima são uma consequência imediata da classificação dos módulos finitamente gerados sobre $\mathbb{K}[t]$.

7. Espaços vetoriais com produto interno

Neste capítulo final vamos introduzir mais estrutura geométrica nos objetos que temos vindo a estudar. Esta estrutura incluirá por exemplo noções de distância e ângulo entre vetores. Isto é feito com recurso à noção de *produto interno*, um conceito que generaliza a noção de produto escalar em $\mathbb{R}^2$ e $\mathbb{R}^3$ familiar do ensino secundário. Nesta secção o corpo de base será sempre $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$.

7.1. **Definição de produto interno num espaço vetorial.** Recordemos o *produto interno* (ou escalar) de vetores de $\mathbb{R}^2$ e $\mathbb{R}^3$. Trata-se de uma operação que produz um número

[15] Além disso é possível escolher estes fatores de uma forma natural, o que leva à chamada *Forma canónica racional* de um endomorfismo (assim conhecida porque é válida, em particular, sobre o corpo dos racionais)

[16] Se suprimirmos na definição de corpo o requisito da invertibilidade dos elementos não nulos obtemos a noção de anel. Os números inteiros e os polinómios com coeficientes num corpo são exemplos de anéis. Um módulo sobre um anel é o análogo neste contexto de um espaço vetorial sobre um corpo.

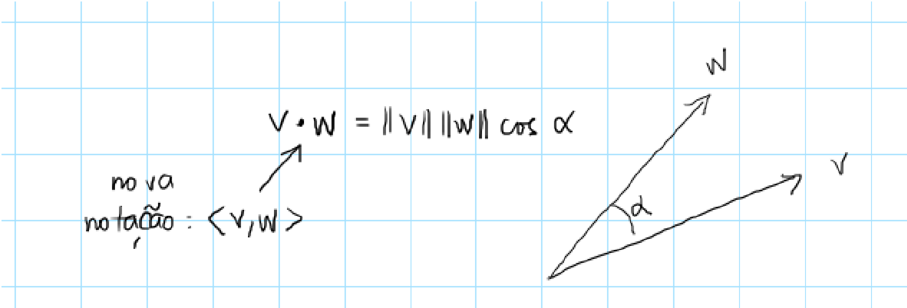
real $\langle v, w\rangle$ a partir de dois vetores v e w. É dado pelas fórmulas

$$\langle (x_1, x_2), (y_1, y_2)\rangle = x_1y_1 + x_2y_2 \qquad \text{para } (x_1, x_2), (y_1, y_2) \in \mathbb{R}^2$$

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.



As propriedades destes produtos podem ser abstraídas nos seguintes axiomas simples.

Definição 7.2. *Seja V um espaço vetorial* real. *Um* produto interno *em V é uma função*

$$\langle \cdot, \cdot\rangle \colon V \times V \to \mathbb{R}$$

satisfazendo

(1) **Bilinearidade:** *Para todos os $\alpha_1, \alpha_2 \in \mathbb{R}$ e $v_1, v_2, w \in V$.*
- $\langle \alpha_1v_1 + \alpha_2v_2, w\rangle = \alpha_1\langle v_1, w\rangle + \alpha_2\langle v_2, w\rangle$
- $\langle w, \alpha_1v_1 + \alpha_2v_2\rangle = \alpha_1\langle w, v_1\rangle + \alpha_2\langle w, v_2\rangle$

(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 7.3. *Tendo em conta a simetria do produto interno, para verificar a bilinearidade basta verificar a primeira (ou a segunda) das igualdades que caracterizam a bilinearidade.*

Exemplo 7.4. *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 7.2 são verificadas. Este produto interno generaliza o produto interno já conhecido nos casos em que $n = 2$ e 3.

Exemplo 7.5. *Seja $[a, b]$ um intervalo limitado e fechado de $\mathbb{R}$ e $V = C([a, b], \mathbb{R})$ o espaço vetorial das funções contínuas de $[a, b]$ para $\mathbb{R}$ (que é um subespaço vetorial do espaço vetorial de todas as funções de $\mathbb{R}$ para $\mathbb{R}$). Define-se $\langle \cdot, \cdot\rangle \colon V \times V \to \mathbb{R}$ 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 7.2:

(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 7.6. *Se pensarmos numa função* f *como um "vetor com componentes indexadas 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 Hermitiano*, ou simplesmente um produto interno. O modelo será $\mathbb{C}^n$, mas agora não podemos usar exatamente a mesma fórmula que nos dá o produto interno standard em $\mathbb{R}^n$ porque perderíamos a propriedade da positividade (que é a chave para definir o comprimento de vetores). A solução é conjugar um dos argumentos coordenada a coordenada, uma vez que $\overline{z}z = |z|^2 \geq 0$. Esta solução requer a modificação dos restantes dois axiomas da forma seguinte.

Definição 7.7. *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 ou linearidade conjugada:** *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 7.8. *Tendo em conta a simetria conjugada de um produto interno complexo, para verificar a sesquilinearidade basta verificar a primeira (ou a segunda) das igualdades que caracterizam a sesquilinearidade.*

Exemplo 7.9. *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 7.7. 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 7.10. *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. Por exemplo:

$$\|(x, y, z)\| = \sqrt{x^2 + y^2 + z^2}$$

Exemplo 7.11. *Em* $\mathbb{C}^2$ *com o produto interno usual,*

$$\|(1 + i, -1)\| = \sqrt{|1 + i|^2 + 1} = \sqrt{2 + 1} = \sqrt{3}$$

Em $C([0, 1], \mathbb{R})$ *a distância entre as funções* x *e* 1 *é*

$$\|x - 1\| = \sqrt{\int_0^1 (x - 1)^2 dx} = \sqrt{\left. \frac{(x - 1)^3}{3} \right|_0^1} = \frac{1}{\sqrt{3}}$$

Além da distância, um outro conceito fundamental de que passamos a dispôr num espaço com produto interno é a noção de ortogonalidade ou perpendicularidade.

Definição 7.12. *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 7.13. *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.*

7.14. **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 tratar o caso real simultaneamente (simplesmente omitimos a conjugação no caso real). Usando linearidade conjugada na primeira variável temos

$$\langle v, w\rangle = \langle \alpha_1 v_1 + \dots \alpha_n v_n, w\rangle = \overline{\alpha_1}\langle v_1, w\rangle + \dots + \overline{\alpha_n}\langle v_n, w\rangle$$

Usando a linearidade na segunda coordenada temos para cada i

$$\langle v_i, w\rangle = \langle v_i, \beta_1 v_1 + \dots \beta_n v_n\rangle = \beta_1\langle v_i, v_1\rangle + \dots + \beta_n\langle v_i, v_n\rangle$$

e substituindo na primeira expressão obtemos a seguinte expressão para o produto interno

$$\langle v, w\rangle = \overline{\alpha_1}\beta_1\langle v_1, v_1\rangle + \dots + \overline{\alpha_1}\beta_n\langle v_1, v_n\rangle + \overline{\alpha_2}\beta_1\langle v_2, v_1\rangle + \dots + \overline{\alpha_n}\beta_1\langle v_n, v_1\rangle + \dots + \overline{\alpha_n}\beta_n\langle v_n, v_n\rangle$$

Vemos assim que o produto interno é completamente determinado pelo conjunto de n^2 escalares $\langle v_i, v_j\rangle$ com $i, j = 1, \dots, n$. Identificando escalares com matrizes 1×1 a expressão anterior pode ser escrita matricialmente na forma

$$\begin{bmatrix} \overline{\alpha_1} & \overline{\alpha_2} & \cdots & \overline{\alpha_n} \end{bmatrix} \begin{bmatrix} \langle v_1, v_1\rangle & \langle v_1, v_2\rangle & \cdots & \langle v_1, v_n\rangle \\ \langle v_2, v_1\rangle & \langle v_2, v_2\rangle & \cdots & \langle v_2, v_n\rangle \\ \vdots & \vdots & & \vdots \\ \langle v_n, v_1\rangle & \langle v_n, v_2\rangle & \cdots & \langle v_n, v_n\rangle \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} \tag{51}$$

Proposição 7.15. *Seja V um espaço vetorial de dimensão finita sobre $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$ e seja $\langle \cdot, \cdot\rangle$ um produto interno em V. Dada uma base ordenada $B = (v_1, \dots, v_n)$ para V, existe uma única matriz $G_B \in M_{n\times n}(\mathbb{K})$, chamada a* matriz da métrica *ou* matriz de Gram *tal que*

$$\langle v, w\rangle = \begin{cases} [v]_B^T G_B [w]_B & \text{se } \mathbb{K} = \mathbb{R} \\ \overline{[v]_B}^T G_B [w]_B & \text{se } \mathbb{K} = \mathbb{C} \end{cases}$$

Esta matriz é dada pela fórmula

$$G_B = [\langle v_i, v_j\rangle] \tag{52}$$

Dem. Resta-nos apenas provar a unicidade. Consideremos apenas o caso real uma vez que o caso complexo é inteiramente análogo: se $A \in M_{n\times n}(\mathbb{R})$ é tal que $\langle v, w\rangle = [v]_B^T A [w]_B$ então tomando $v = v_i$, $w = w_j$ obtemos

$$\langle v_i, w_j\rangle = \begin{bmatrix} 0 & \cdots & 1 & \cdots & 0 \end{bmatrix} A \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = a_{ij}$$

pelo que a matriz A é necessariamente dada pela fórmula (52). □

Observação 7.16. *Para chegar à expressão* (51) *usámos apenas a propriedade (1) das Definições 7.2 e 7.7 pelo que a expressão matricial* (51) *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) na definição de produto interno impõem algumas condições 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 *hermitiana*. Reciprocamente, se G é uma matriz que satisfaz estas condições é imediato verificar que a função

$$\langle v, w \rangle = \overline{[v]_B}^T G[w]_B$$

satisfaz as condições (1) e (2) nas definições 7.2 e 7.7 (uma tal função diz-se uma forma bilinear (resp. sesquilinear) simétrica).

Quanto à condição (3), ela claramente implica que os valores próprios de uma matriz da métrica têm que ser *reais* positivos (mesmo no caso em que $\mathbb{K} = \mathbb{C}$): se $G_B[v]_B = \lambda[v]_B$ então

$$\|v\|^2 = \langle v, v \rangle = \overline{[v]_B}^T G_B[v]_B = \lambda\overline{[v]_B}^T[v]_B = \lambda\|[v]_B\|^2 > 0 \quad \Rightarrow \lambda > 0$$

Observação 7.17. *Veremos em breve que uma matriz simétrica ou hermitiana é sempre diagonalizável. Admitindo este facto, é imediato verificar que dada uma base B para V e uma matriz G simétrica (resp. hermitiana) com valores próprios todos positivos, a expressão $\langle v, w \rangle = \overline{[v]_B}^T G[w]_B$ define um produto interno em V.*

Assim, uma base B para o espaço vetorial V estabelece uma correspondência biunívoca entre os produtos internos em V e as matrizes simétricas (resp. hermitianas) com valores próprios todos positivos.

Exemplo 7.18. *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 7.19. *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 $\overline{A}\ \overline{B} = \overline{AB}$ 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

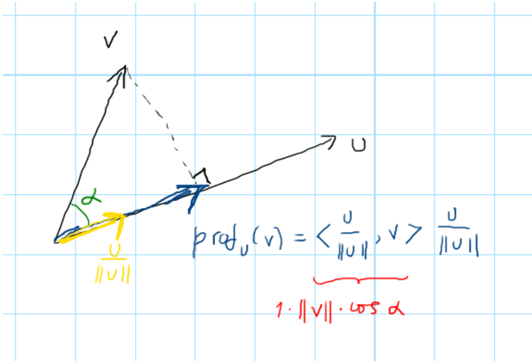
$$G_B = \overline{S}^T G_{B'} S \quad \text{ou, no caso real,} \quad G_B = S^T G_{B'} S \tag{53}$$

Estas fórmulas explicam como a expressão para o produto interno é alterada por uma mudança de coordenadas e são inteiramente análogas à fórmula que descreve a maneira como as expressões matriciais de um endomorfismo em duas bases distintas estão relacionadas.

7.20. As desigualdades triangular e de Cauchy-Schwarz.

Definição 7.21. *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*

$$\operatorname{proj}_u(v) = \langle u, v\rangle \frac{u}{\|u\|^2} = \frac{\langle u, v\rangle}{\langle u, u\rangle} u = \langle \frac{u}{\|u\|}, v\rangle \frac{u}{\|u\|} \tag{54}$$



As expressões do lado direito do primeiro sinal de igual em (54) 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 54 é, neste caso, a expressão familiar do ensino secundário.

Observação 7.22. *No caso complexo é importante que o produto interno na expressão para* $\operatorname{proj}_u(v)$ *seja realizado na ordem indicada, caso contrário a expressão não será linear em* v*.*

Exemplo 7.23. *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 (por definição, $\operatorname{proj}_u(v)$ é um múltiplo escalar de u). Claramente proj_u é uma transformação linear (pois o produto interno é linear na segunda coordenada) e, como

$$\operatorname{proj}_u(\lambda u) = \langle u, \lambda u\rangle \frac{u}{\|u\|^2} = \lambda \|u\|^2 \frac{u}{\|u\|^2} = \lambda u$$

temos que

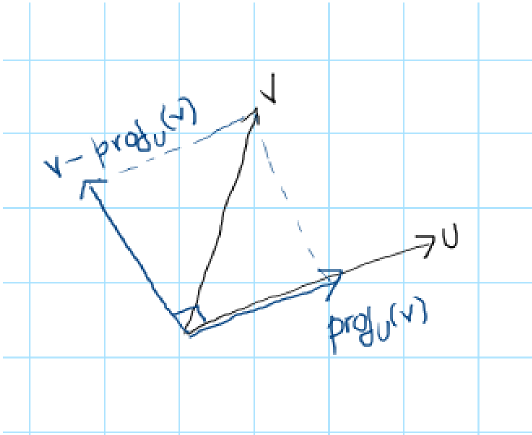
- A imagem de proj_u é exatamente $L(\{u\})$
- proj_u é a identidade na sua imagem, isto é $\operatorname{proj}_u^2 = \operatorname{proj}_u$ (isto é, proj_u é uma projeção).

A transformação linear proj_u é portanto uma projeção na reta gerada por u. As direções de projeção são as que estão contidas no núcleo de proj_u que é precisamente o plano perpendicular a u:

$$\mathrm{proj}_u(v) = 0 \Leftrightarrow \langle u, v\rangle \frac{u}{\|u\|^2} = 0 \Leftrightarrow \langle u, v\rangle = 0$$

A transformação proj_u permite assim escrever qualquer vetor v como a soma de um vetor colinear com u e outro ortogonal a u:

$$v = (v - \mathrm{proj}_u(v)) + \mathrm{proj}_u(v)$$



Da consideração da componente ortogonal a um vetor u, vêm duas desigualdades fundamentais.

Proposição 7.24. *Seja V um espaço vetorial (real ou complexo) 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

$$\begin{aligned}
0 \leq \|v - \mathrm{proj}_u(v)\|^2 &= \langle v - \frac{\langle u, v\rangle}{\langle u, u\rangle} u, v - \frac{\langle u, v\rangle}{\langle u, u\rangle} u\rangle \\
&= \langle v, v\rangle - \frac{\overline{\langle u, v\rangle}}{\langle u, u\rangle}\langle u, v\rangle - \frac{\langle u, v\rangle}{\langle u, u\rangle}\langle v, u\rangle + \frac{\overline{\langle u, v\rangle}}{\langle u, u\rangle}\frac{\langle u, v\rangle}{\langle u, u\rangle}\langle u, u\rangle \\
&= \|v\|^2 - \frac{|\langle u, v\rangle|^2}{\|u\|^2}
\end{aligned}$$

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 $v - \mathrm{proj}_u(v) = 0$, o que acontece se e só se v é um múltiplo escalar de u.

(ii) Temos

$$\|u+v\|^2 = \langle u+v, u+v\rangle = \langle u,u\rangle + \langle u,v\rangle + \langle v,u\rangle + \langle v,v\rangle \tag{55}$$

Uma vez que $z + \overline{z} = 2\,\mathrm{Re}(z) \leq 2|z|$ temos

$$\langle u,v\rangle + \langle v,u\rangle = 2\,\mathrm{Re}(\langle u,v\rangle) \leq 2|\langle u,v\rangle| \leq 2\|u\|\|v\|$$

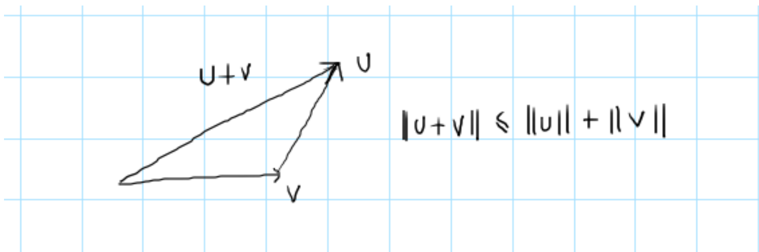
onde na segunda desigualdade aplicámos a desigualdade de Cauchy-Schwarz. Substituindo em (55) obtemos

$$\|u+v\|^2 \leq \|u\|^2 + 2\|u\|\|v\| + \|v\|^2 = (\|u\| + \|v\|)^2$$

que é equivalente à desigualdade triangular.

□

Observação 7.25. *(i) A desigualdade triangular chama-se assim porque $u, v, u+v$ 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 (55) é o **Teorema de Pitágoras:** $\|u+v\|^2 = \|u\|^2 + \|v\|^2$.

7.26. **A definição de ângulo. Complementos ortogonais.**

Definição 7.27. *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 $[-1,1]$.)

Note-se que $v, w \in V$ são ortogonais se e só se o ângulo entre v e w é $\frac{\pi}{2}$.

Exemplo 7.28. *O ângulo entre as funções x e x^2 em $C([0,1], \mathbb{R})$ é*

$$\arccos\frac{\langle x, x^2\rangle}{\|x\|\|x^2\|} = \frac{\int_0^1 x^3 dx}{\sqrt{\int_0^1 x^2 dx \int_0^1 x^4 dx}} = \arccos\frac{\frac{1}{4}}{\sqrt{\frac{1}{3}\frac{1}{5}}} = \arccos\frac{\sqrt{15}}{4}$$

Definição 7.29. *Seja V um espaço vetorial com um produto interno e $S \subset V$ um subconjunto qualquer. Define-se o subespaço ortogonal a S por*

$$S^\perp = \{v \in V : \langle v, x\rangle = 0 \text{ para todo o } x \in S\}$$

Se $W \subset V$ é um subespaço, chama-se a $W^\perp$ o complemento ortogonal *de W em V.*

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 $\alpha_1 v_1 + \alpha_2 v_2 \in S^\perp$.

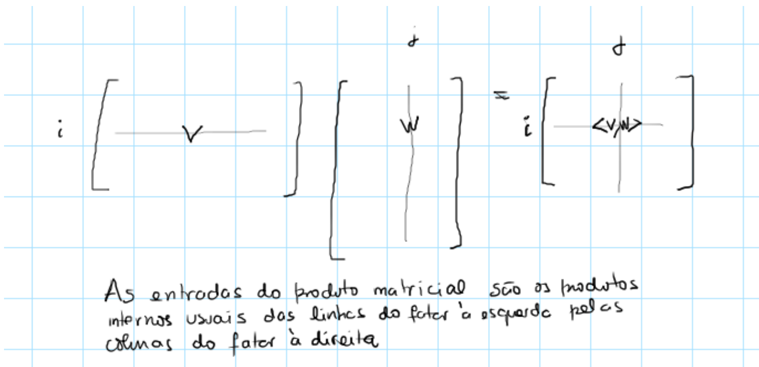
Proposição 7.30. $S^\perp = L(S)^\perp$

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 7.31. *(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 $V^\perp = \{0\}$.

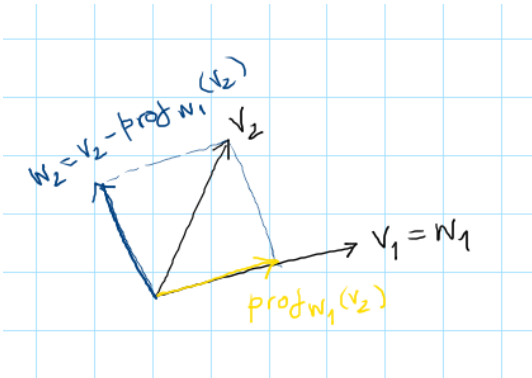
7.32. **O método de ortogonalização de Gram-Schmidt.** A operação de projeção ortogonal segundo um vetor permite-nos facilmente achar uma base ortogonal a partir de uma base qualquer.

Proposição 7.33 (**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*

$$\begin{aligned} w_1 &= v_1 \\ w_2 &= v_2 - \mathrm{proj}_{w_1}(v_2) \\ w_3 &= v_3 - \mathrm{proj}_{w_1}(v_3) - \mathrm{proj}_{w_2}(v_3) \\ \vdots & \quad \vdots \\ w_k &= v_k - \mathrm{proj}_{w_1}(v_k) - \ldots - \mathrm{proj}_{w_{k-1}}(v_k) \end{aligned}$$

formam um conjunto ortogonal $\{w_1, \ldots, w_k\}$ tal que, para cada $i = 1, \ldots, k$, temos

$$L(\{v_1, \ldots, v_i\}) = L(\{w_1, \ldots, w_i\})$$



Dem. 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 $\mathrm{proj}_u(v)$ é um múltiplo de u, a seguinte reformulação da definição de w_i

$$v_i = w_i + \mathrm{proj}_{w_1}(v_i) + \ldots + \mathrm{proj}_{w_{i-1}}(v_i)$$

  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

$$w_i = v_i - \mathrm{proj}_{w_1}(v_i) - \ldots - \mathrm{proj}_{w_{i-1}}(v_i)$$

  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 7.34. *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

$$\{(\tfrac{1}{\sqrt{2}}, 0, 0, -\tfrac{1}{\sqrt{2}}), (-\tfrac{1}{\sqrt{6}}, \sqrt{\tfrac{2}{3}}, 0, -\tfrac{1}{\sqrt{6}}), (0, 0, 1, 0)\}$$

7.35. **Cálculos em bases ortogonais e projeção ortogonal.** 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 7.36. *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 $v_1, \dots, v_k$ elementos de S e suponhamos que

$$\alpha_1 v_1 + \ldots + \alpha_k v_k = 0 \tag{56}$$

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 ortonormadas. Juntamente com as noções de valor e vetor próprio será um dos resultados de Álgebra Linear que mais vezes será utilizado em aplicações. 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 7.37. *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

$$v = \frac{\langle v_1, v \rangle}{\langle v_1, v_1 \rangle} v_1 + \ldots + \frac{\langle v_n, v \rangle}{\langle v_n, v_n \rangle} v_n \Leftrightarrow v - \frac{\langle v_1, v \rangle}{\langle v_1, v_1 \rangle} v_1 - \ldots - \frac{\langle v_n, v \rangle}{\langle v_n, v_n \rangle} v_n = 0$$

De acordo com o Exemplo 7.31(ii) basta ver que o vetor do lado esquerdo da segunda igualdade é ortogonal aos elementos da base B. Ora

$$\begin{aligned} \langle v_i, v - \frac{\langle v_1, v \rangle}{\langle v_1, v_1 \rangle} v_1 - \ldots - \frac{\langle v_n, v \rangle}{\langle v_n, v_n \rangle} v_n \rangle &= \langle v_i, v \rangle - \frac{\langle v_1, v \rangle}{\langle v_1, v_1 \rangle} \langle v_i, v_1 \rangle - \ldots - \frac{\langle v_n, v \rangle}{\langle v_n, v_n \rangle} \langle v_i, v_n \rangle \\ &= \langle v_i, v \rangle - 0 - \ldots - \frac{\langle v_i, v \rangle}{\langle v_i, v_i \rangle} \langle v_i, v_i \rangle - \ldots - 0 \\ &= \langle v_i, v \rangle - \langle v_i, v \rangle = 0 \end{aligned}$$

□

Exemplo 7.38. *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*

$$B = \left(\left(\tfrac{1}{\sqrt{2}}, 0, 0, -\tfrac{1}{\sqrt{2}}\right), \left(-\tfrac{1}{\sqrt{6}}, \sqrt{\tfrac{2}{3}}, 0, -\tfrac{1}{\sqrt{6}}\right), (0,0,1,0) \right)$$

do Exemplo 7.34 e o vetor $(1,1,3,-2) \in V$ *temos*

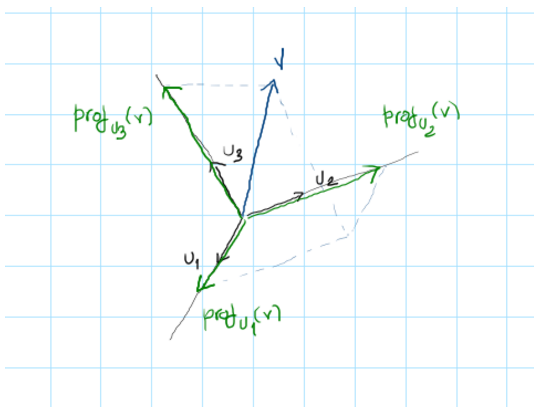
$$[(1,1,3,-2)]_B = \begin{bmatrix} \langle (\tfrac{1}{\sqrt{2}}, 0, 0, -\tfrac{1}{\sqrt{2}}), (1,1,3,-2) \rangle \\ \langle (-\tfrac{1}{\sqrt{6}}, \sqrt{\tfrac{2}{3}}, 0, -\tfrac{1}{\sqrt{6}}), (1,1,3,-2) \rangle \\ \langle (0,0,1,0), (1,1,3,-2) \rangle \end{bmatrix} = \begin{bmatrix} \frac{3}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} + \sqrt{\frac{2}{3}} \\ 3 \end{bmatrix}$$

Uma base ortogonal para um subespaço pode ser usada para definir a projeção ortogonal nesse subespaço.

Definição 7.39. *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* $P_U : V \to V$ *definida pela fórmula*

$$P_U(v) = \operatorname{proj}_{u_1}(v) + \ldots + \operatorname{proj}_{u_k}(v) \tag{57}$$

onde $\{u_1, \ldots, u_k\}$ *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 (57) seja independente da escolha da base ortogonal para o subespaço U. Isso é uma consequência do seguinte resultado.

Proposição 7.40. *Seja* V *um espaço com produto interno e* U *um subespaço vetorial finitamente gerado. A transformação linear* $P_U : V \to V$ *definida pela expressão* (57) *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. □

A Proposição anterior mostra que P_U é independente da escolha da base ortogonal para U que aparece na fórmula 57 pois uma projeção é completamente determinada pela sua imagem e o seu núcleo (como vimos nos exercícios sobre transformações lineares).

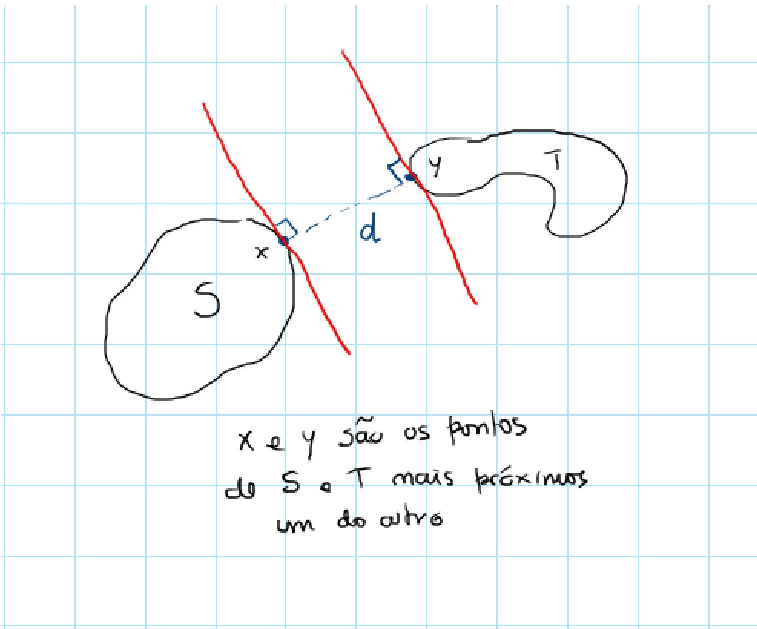
Uma aplicação interessante das projeções ortogonais é o cálculo da distância de um ponto x de um espaço vetorial com produto interno V a um subespaço U de V.

Definição 7.41. *Seja V um espaço vetorial com produto interno $x \in V$ e $S \subset V$ um subconjunto não vazio. A* distância de x a S *é*

$$d(x, S) = \inf\{\|x - u\| : u \in S\}$$

Mais geralmente, a distância entre dois subconjuntos não vazios $S, T \subset V$ define-se por

$$d(S, T) = \inf\{\|x - y\| : x \in S, y \in T\}$$



Note-se que o ínfimo existe porque o conjunto $\{\|x-y\| : x \in S, y \in T\}$ é não vazio (porque S e T são não vazios) e limitado inferiormente (por 0).

Slogan: As distâncias medem-se na perpendicular

Veremos agora algumas instâncias do aforismo anterior deixando outras para os exercícios. Verão outras instâncias quando estudarem curvas e superfícies em Cálculo 2.

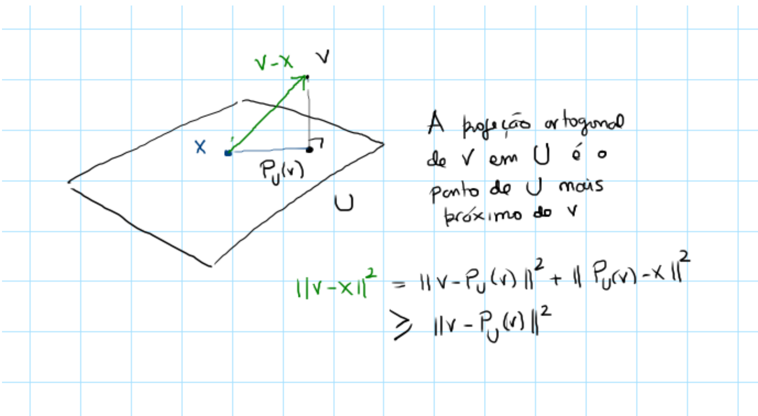
No caso em que $S = U$ é um subespaço vetorial de V e $v \notin U$, dado $x \in U$ podemos escrever o vetor $v - x$ como

$$v - x = (v - P_U(v)) + (P_U(v) - x)$$

uma vez que $v - P_U(v) \in U^\perp$ e $P_U(v) - x \in U$, pelo Teorema de Pitágoras, temos

$$\|v - x\|^2 = \|v - P_U(v)\|^2 + \|P_U(v) - x\|^2 \geq \|v - P_U(v)\|^2 \Leftrightarrow \|v - x\| \geq \|v - P_U(v)\|$$

Uma vez que $P_U(v) \in U$, isso mostra que $d(v, U) = \|v - P_U(v)\|$ e, portanto, que $P_U(v)$ **é o ponto de U mais próximo de v.**



Este 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.

Exemplo 7.42. *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*

$$(1, 2, -1) + t(1, 1, 2) = (1 + t, 2 + t, -1 + 2t)$$

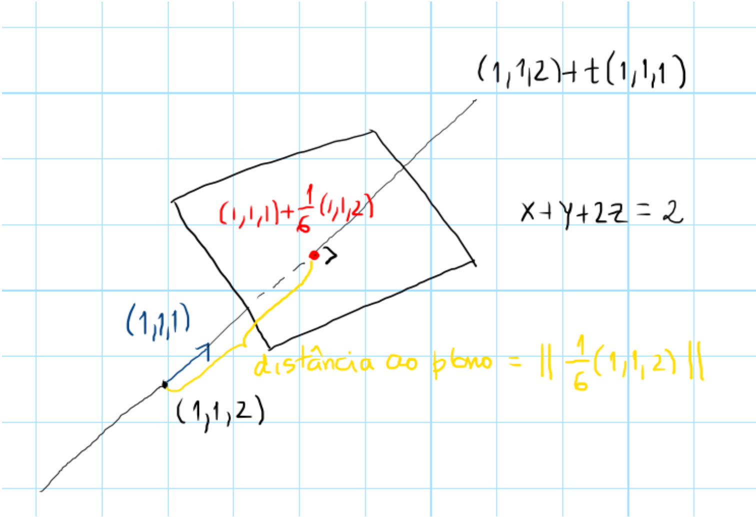
e interseta H quando

$$(1 + t) + (2 + t) + 2(-1 + 2t) = 2 \Leftrightarrow 6t = 1 \Leftrightarrow t = \tfrac{1}{6}$$

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*

$$\|w - (1, 2, -1)\|^2 = \|w - v\|^2 + \|v - (1, 2, -1)\|^2 \geq \|v - (1, 2, -1)\|^2$$

pois $v - (1, 2, -1)$ *(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* $\|\frac{1}{6}(1, 1, 2)\| = \frac{1}{\sqrt{6}}$.

7.43. **O método dos quadrados mínimos.** [17] 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

$$Ax - b \in EC(A)^\perp$$

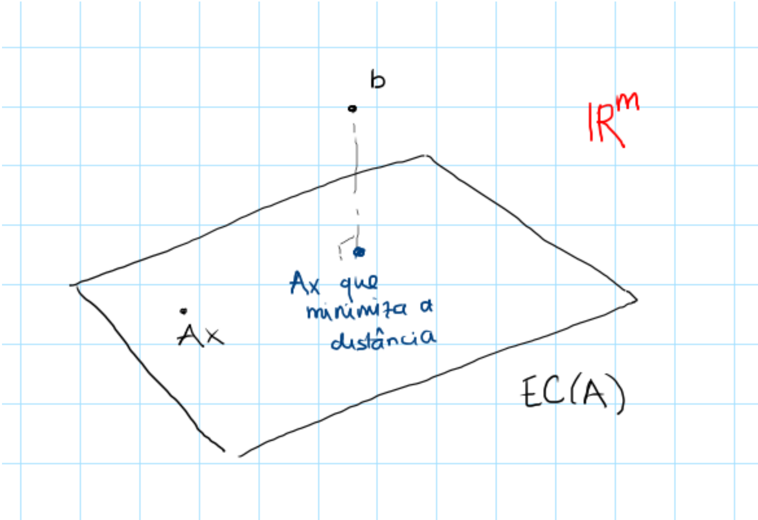
mas, uma vez que $EC(A) = EL(A^T)$, pelo Exemplo 7.31(i) temos

$$EC(A)^\perp = EL(A^T)^\perp = N(A^T)$$

Assim, Ax será o ponto mais próximo de b quando se verifica a **equação dos quadrados mínimos** para x

$$A^T(Ax - b) = 0 \Leftrightarrow A^T Ax = A^T b \tag{58}$$

[17] Esta discussão é adaptada do tratamento deste método em [D].



Note-se que a solução pode não ser única (se $N(A) \neq 0$) mas o sistema (58) tem sempre solução pois 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 exatamente às medições, mas este método permite achar quais os valores dos parâmetros que melhor se adequam às medições efetuadas.

Exemplo 7.44. *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 quadrados mínimos dá-nos os coeficientes a, b tais que a soma

$$(a \cdot 0 + b - (-2))^2 + (a \cdot 1 + b - 3)^2 + (a \cdot 4 + b - 5)^2$$

é mínima (é isto que dá o nome ao método). Temos que achar a solução do sistema

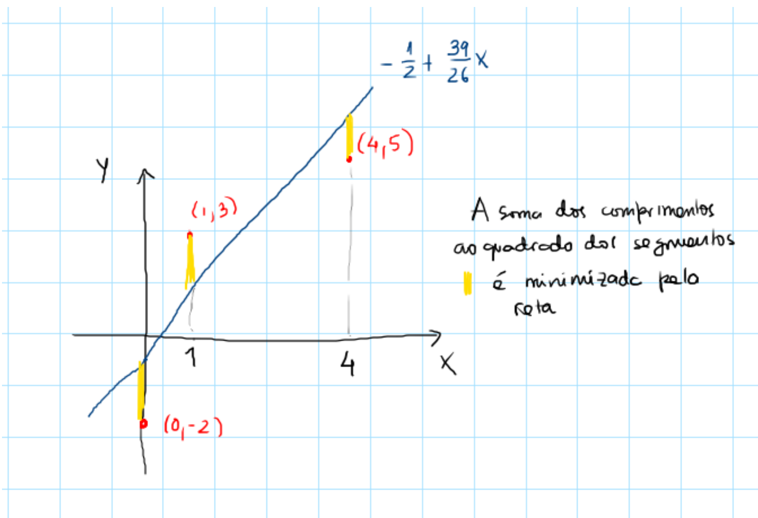
$$\begin{bmatrix} 0 & 1 & 4 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0 & 1 & 4 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} -2 \\ 3 \\ 5 \end{bmatrix} \Leftrightarrow \begin{bmatrix} 17 & 5 \\ 5 & 3 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 23 \\ 6 \end{bmatrix}$$

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 7.45. *Pouco após a sua descoberta, em 1801, o planeta anão Ceres (na cintura dos asteróides) 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 quadrados mínimos.*

7.46. **Uma fórmula para o volume de um paralelipípedo k-dimensional.** Vamos descrever uma 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 7.47. *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*

$$\operatorname{Vol}_k(P) = \sqrt{\det A^T A}$$

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 (para qualquer noção razoável de volume k-dimensional) 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 $C^TC = I_{n-k}$ porque os vetores w_i constituem uma base ortonormada para o plano que geram). Portanto

$$(\det B)^2 = \det(B^TB) = \det(A^TA) \quad \Leftrightarrow \quad \sqrt{\det A^TA} = |\det B|$$

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^TA 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 7.48. *A área do paralelogramo em* $\mathbb{R}^3$ *com arestas* $(1, -2, 1)$ *e* $(2, 3, 0)$ *é*

$$\sqrt{\det \begin{bmatrix} 1 & -2 & 1 \\ 2 & 3 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ -2 & 3 \\ 1 & 0 \end{bmatrix}} = \sqrt{\begin{vmatrix} 6 & -4 \\ -4 & 13 \end{vmatrix}} = \sqrt{62}$$

7.49. **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 $f\colon [0, \pi] \to \mathbb{R}$ (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 $C([0, \pi], \mathbb{R})$ das funções contínuas em $[0, \pi]$ definido por

$$\langle f, g \rangle = \int_0^\pi f(x)g(x)dx$$

o conjunto

$$\{\operatorname{sen} x, \operatorname{sen}(2x), \ldots, \operatorname{sen}(nx), \ldots\}$$

é ortogonal. Fourier descobriu que é possível expressar qualquer função contínua como "combinação linear" destas funções[18] - 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 $[0, \pi]$.

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

$$P_{\operatorname{sen}(nx)}(f) = \frac{\langle \operatorname{sen}(nx), f(x) \rangle}{\|\operatorname{sen}(nx)\|^2}$$

[18] 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 vibrante da Matemática que se designa por Análise Harmónica.

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.

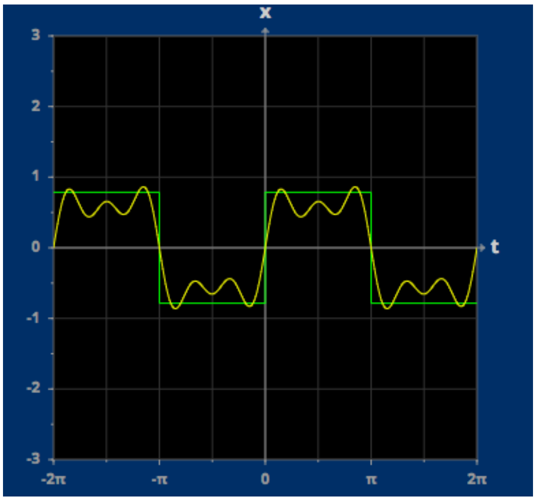


FIGURE 2. Aproximação de um sinal retangular por uma soma de Fourier com 5 termos.

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 Cálculo 3.

7.50. **Transformações unitárias e (anti)-hermitianas.** Para terminar vamos falar um pouco das transformações lineares entre espaços vetoriais munidos de produto interno. Começamos por estudar as transformações que preservam o produto interno e portanto ângulos e distâncias.

Definição 7.51. *Seja V um espaço vetorial com produto interno $\langle\cdot,\cdot\rangle$. Uma transformação linear $T\colon V \to V$ tal que*

$$\langle T(v), T(w)\rangle = \langle v, w\rangle \qquad \textit{para todos os } v, w \in V$$

diz-se **ortogonal** *quando V é um espaço vetorial real e* **unitária** *quando V é um espaço vetorial complexo.*

É um exercício da ficha sobre produtos internos ver que as transformações unitárias são exatamente as que preservam a distância determinada pelo produto interno - isto é as *isometrias*. O caso mais importante é o dos produtos internos usuais em $\mathbb{R}^n$ e $\mathbb{C}^n$.

Exemplo 7.52. *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*

$$(Ax)^T(Ay) = x^T y \Leftrightarrow x^T A^T A y = x^T y \qquad \textit{para todos os } x, y \in \mathbb{R}^n \tag{59}$$

Isto acontece se e só se

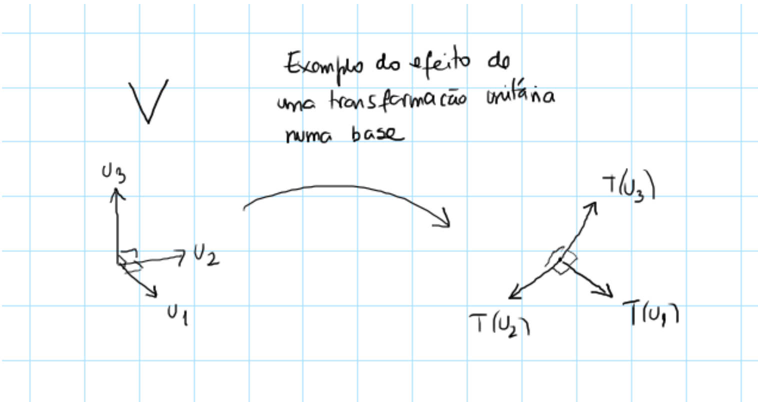
$$A^T A = I_n \tag{60}$$

De facto, é claro que se A *satisfaz a condição* (60) *então satisfaz* (59)*. Reciprocamente se* (59) *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.* $A^T A$ *é portanto a matriz identidade.*

As matrizes de $M_{n\times n}(\mathbb{R})$ *que satisfazem* (60) *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* (60) *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. Podemos pensar que o referencial ortonormado inicial foi "rodado" para um novo referencial ortornormado. Isto generaliza o conceito de rotação e reflexão no espaço tridimensional como veremos nos exercícios sobre o produto interno.*



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*

$$\overline{x}^T \overline{A}^T A y = \overline{x}^T y \Leftrightarrow \overline{A}^T A = I_n$$

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 7.53. *Seja $A \in M_{m\times n}(\mathbb{C})$. A matriz transposta conjugada $\overline{A}^T$ é denotada por A^* e designada pela matriz* adjunta *de A* [19]. *Assim, $A^* \in M_{n\times m}(\mathbb{C})$ é a matriz com entrada ij dada por $\overline{a_{ji}}$.*

Proposição 7.54. *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. ☐

Observação 7.55. *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 (encarada como uma matriz complexa) são complexos unitários e que os seus vetores próprios são ortogonais em $\mathbb{C}^n$.*

Exemplo 7.56. *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.

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

$$\begin{bmatrix} \cos\alpha - (\cos\alpha \pm i\operatorname{sen}\alpha) & -\operatorname{sen}\alpha \\ \operatorname{sen}\alpha & \cos\alpha - (\cos\alpha \pm i\operatorname{sen}\alpha) \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = 0 \Leftrightarrow \begin{bmatrix} a \\ b \end{bmatrix} = \gamma \begin{bmatrix} \pm i \\ 1 \end{bmatrix}$$

(com $\gamma \in \mathbb{C}$) e são ortogonais para o produto interno usual em $\mathbb{C}^2$.

[19] Por vezes também a matriz transconjugada de A

Vamos agora falar de outras duas classes importantes de endomorfismos de espaços vetoriais com produto interno que são "bem comportados" com respeito ao produto interno. A verdadeira justificação para considerarmos estas classes é que endomorfismos deste tipo são ubíquos em Matemática e nas suas aplicações, pelo que é importante conhecer as suas propriedades.

Definição 7.57. *Seja V um espaço vetorial de dimensão finita com produto interno. Uma transformação linear $T\colon V \to V$ diz-se* auto adjunta[20] *se para todos os $v, w \in V$ temos*

$$\langle T(v), w\rangle = \langle v, T(w)\rangle \tag{61}$$

e anti-adjunta[21] *se para todos os $v, w \in V$ temos*

$$\langle T(v), w\rangle = -\langle v, T(w)\rangle \tag{62}$$

No caso em que $V = \mathbb{C}^n$, $W = \mathbb{C}^m$ com os produtos internos usuais, e $Tv = Av$ é determinada por uma matriz $A \in M_{m\times n}(\mathbb{C})$, as equações (61) e (62) traduzem-se em

$$(\overline{Av})^T w = \pm \overline{v}^T (Aw) \Leftrightarrow \overline{v}^T A^* w = \pm \overline{v}^T A w \qquad \text{para todos os } v, w \in \mathbb{C}^n$$

Vemos assim que uma matriz $A \in M_{n\times n}(\mathbb{C})$ determina uma transformação auto-adjunta sse $A = A^*$, isto é, se A é uma matriz hermitiana. Uma transformação é anti-adjunta sse $A = -A^*$ e diz-se então que A é uma matriz *anti-hermitiana*. Analogamente uma matriz $A \in M_{n\times n}(\mathbb{R})$ determina uma transformação auto-adjunta de $\mathbb{R}^n$ sse é uma matriz simétrica e anti-adjunta sse é *anti-simétrica*, isto é se $A^T + A = 0$.

Proposição 7.58. *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. ☐

[20] Ver os exercícios sobre endomorfismos para a justificação desta terminologia.

[21] Em inglês *skew-adjoint*.

Teorema 7.59 (Teorema espetral). [22]

(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 vetorial complexo 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 $T^{\mathbb{C}}\colon \mathbb{C}^n \to \mathbb{C}^n$ 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 ortonormada. Como os valores próprios de $T^{\mathbb{C}}$ são reais, e a matriz A é real, os vetores próprios de $T^{\mathbb{C}}$ podem ser escolhidos como tendo coordenadas reais e formam então 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.

Observação 7.60. *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*

[22] O *espetro* de um endomorfismo é o conjunto dos seus valores próprios. O nome vem de espetro de cores - as cores são frequências de ondas eletromagnéticas que por sua vez são os valores próprios de um certo endomorfismo - o operador das ondas.

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.

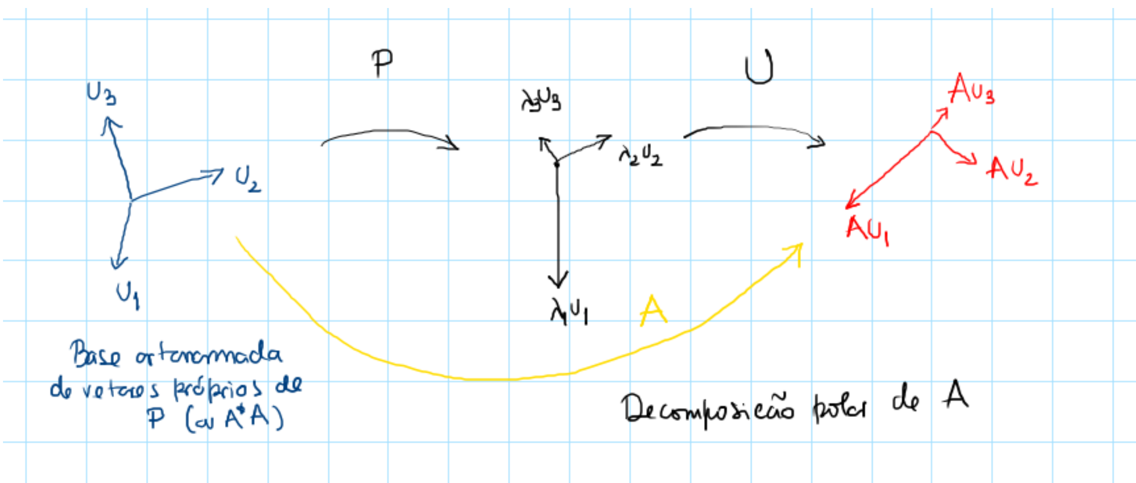
Tipo de matriz	Definição	Valores próprios
unitária	$AA^* = I_n$	$\lambda \in \mathbb{C}, \lvert\lambda\rvert = 1$
hermitiana	$A = A^*$	reais
anti-hermitiana	$A + A^* = 0$	imaginários puros

Matrizes $n \times n$ complexas especiais.

por uma base ortonormada. *Isto acontece sse $AA^* = A^*A$, caso em que se diz que a matriz A é* normal. *Vejam os exercícios sobre o produto interno para uma demonstração.*

7.61. **A decomposição em valores singulares.** Vamos agora usar o Teorema espetral para obter uma decomposição muito útil para uma transformação linear de $\mathbb{K}^n$ em $\mathbb{K}^m$ para $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$. Deixamos como exercício a adaptação desta discussão ao caso em que os espaços $\mathbb{K}^n$ são substituído por espaços de dimensão finita com produto interno.

Começamos por considerar o caso de uma matriz quadrada. O seguinte resultado generaliza a decomposição polar de um número complexo não nulo. Geometricamente, exprime um endomorfismo invertível de $\mathbb{K}^n$ como a composição de um conjunto de expansões e contrações (reais) nas direções de uma base ortonormada (pelo Teorema Espetral é esse o efeito de uma matriz hermitiana com todos os valores próprios positivos) seguido de uma "rotação" (uma transformação unitária).



Proposição 7.62 (Decomposição polar)**.** *Seja $A \in M_{n\times n}(\mathbb{C})$ uma matriz invertível. Então existem uma matriz hermitiana P com valores próprios todos positivos e uma matriz unitária U únicas tais que $A = UP$.*

Dem. A matriz A^*A é hermitiana e todos os seus valores próprios são positivos uma vez que $A^*Av = \lambda v \Rightarrow \|Av\|^2 = v^*A^*Av = \lambda\|v\|^2$. Pelo Teorema espetral existe uma matriz unitária S e uma matriz diagonal D com entradas diagonais positivas tais que $A^*A = SDS^*$. Seja $P = S\sqrt{D}S^*$, onde $\sqrt{D}$ denota a matriz que se obtém de D tomando a raíz quadrada não negativa de cada uma das entradas de D.

É imediato verificar que P é hermitiana com todos os valores próprios positivos (os seus valores próprios são as raízes quadradas dos valores próprios de D). Além disso

$$P^*P = (S\sqrt{D}S^*)^*(S\sqrt{D}S^*) = S\sqrt{D}^*S^*S\sqrt{D}S^* = S\sqrt{D}\sqrt{D}S^* = SDS^* = A^*A$$

Para concluir a existência da decomposição polar, verifiquemos que $U = AP^{-1}$ é unitária:

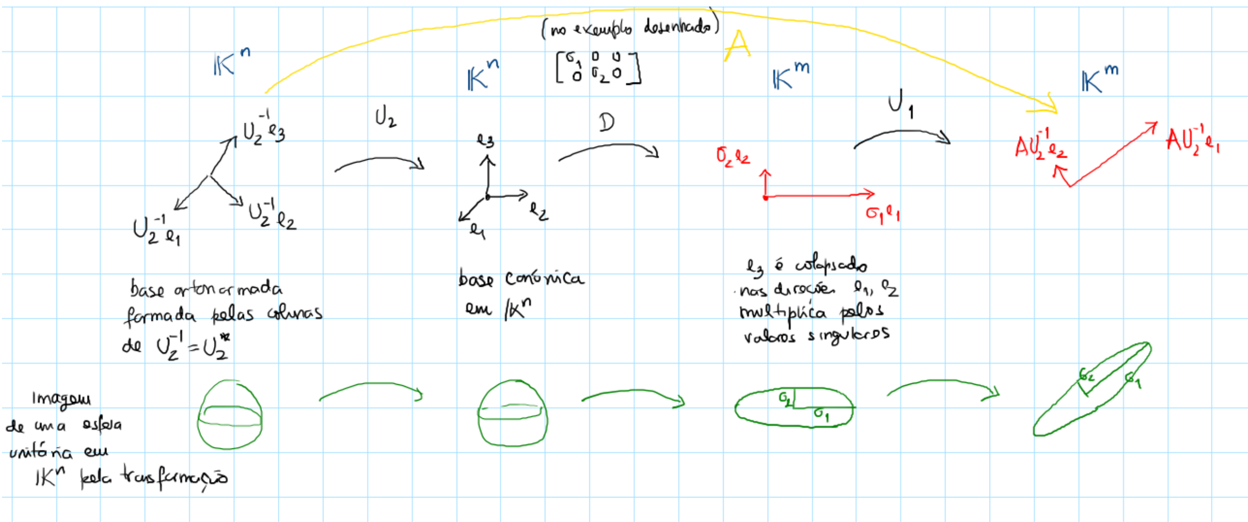
$$U^*U = (P^{-1})^*A^*AP^{-1} = (P^{-1})^*P^*PP^{-1} = I$$

Resta-nos demonstrar a unicidade das matrizes U e P. Para tal basta notar que se $A = UP$ é uma decomposição polar então $A^*A = P^*P = P^2$. Se B for uma base ortonormada que diagonaliza P então a mesma base ortonormada diagonaliza A^*A. Conclui-se que o espaço próprio de λ para P coincide com o espaço próprio de λ^2 para A^*A. Logo P é completamente determinada por A, e portanto o mesmo sucede com U. $\square$

Observação 7.63. *No Teorema anterior, se $A \in M_{n\times n}(\mathbb{R})$ então a matriz P na demonstração será real (pelo Teorema espetral para matrizes simétricas) e portanto U será também real. Conclui-se assim que toda a matriz real invertível A se fatoriza de forma única como $A = OP$ com O uma matriz ortogonal e P uma matriz simétrica com todos os valores próprios positivos.*

Exemplo 7.64. *Se $A = [a + bi] \in M_{1\times 1}(\mathbb{C})$ então $A^*A = [a^2 + b^2]$, $P = [\sqrt{a^2 + b^2}]$ e escrevendo $a + bi = r(\cos\theta + i \operatorname{sen}\theta)$ temos que $U = P^{-1}A = [\cos\theta + i\operatorname{sen}\theta]$. Vemos assim que a decomposição polar generaliza a forma trigonométrica dos números complexos.*

O seguinte Teorema pode ser visto como uma generalização da decomposição polar a transformações lineares arbitrárias. Afirma que qualquer transformação linear de $\mathbb{K}^n$ para $\mathbb{K}^m$ pode ser decomposta numa rotação inicial em $\mathbb{K}^n$, seguida de uma transformação que expande ou contrai ao longo de alguns eixos coordenados, colapsando os restantes eixos coordenados, seguida de outra rotação em $\mathbb{K}^m$.



Teorema 7.65 (Decomposição em valores singulares)**.** *Seja $A \in M_{m\times n}(\mathbb{C})$. Existe uma fatorização de A na forma*

$$A = U_1 D U_2$$

com $U_1 \in M_{m\times m}(\mathbb{C})$ e $U_2 \in M_{n\times n}(\mathbb{C})$ unitárias e $D \in M_{m\times n}(\mathbb{R})$ uma matriz cujas únicas entradas não nulas são reais positivos $\sigma_1 \geq \ldots \geq \sigma_k > 0$ ao longo da diagonal, isto é

$$d_{ij} = \begin{cases} \sigma_i & \text{se } j = i \text{ e } i \leq k, \\ 0 & \text{caso contrário} \end{cases}$$

Os números σ_i chamam-se os valores singulares *de A.*

Proof. A matriz $A^*A \in M_{n\times n}(\mathbb{C})$ é hermitiana e tem todos os valores próprios ≥ 0 (uma vez que $\langle A^*Av, v\rangle = \|Av\|^2 \geq 0$). Sejam U_2 uma matriz unitária e Λ uma matriz diagonal tais que

$$A^*A = U_2^{-1}\Lambda U_2 \tag{63}$$

sendo as entradas diagonais não nulas de Λ por ordem $\sigma_1^2 \geq \ldots \geq \sigma_k^2 > 0$. Seja $D \in M_{m\times n}(\mathbb{R})$ uma matriz cujas únicas entradas não nulas são $\sigma_1 \geq \ldots \geq \sigma_k > 0$ ao longo da diagonal (isto é $d_{ii} = \sigma_i$ para $1 \leq i \leq k$ e todas as outras entradas são nulas). Temos $D^T D = \Lambda$.

Recorde que $U_i^{-1} = U_i^*$ uma vez que as matrizes U_i são unitárias. A equação (63) diz-nos que $(AU_2^{-1})^*(AU_2^{-1}) = \Lambda$ pelo que

- As últimas $n-k$ colunas de AU_2^{-1} são nulas.
- As primeiras k colunas $v_1, \ldots, v_k \in \mathbb{C}^m$ são um conjunto ortogonal em $\mathbb{C}^m$ sendo o comprimento da coluna i igual a σ_i

As colunas não nulas de AU_2^{-1} são uma base ortogonal para o espaço das colunas de A (uma vez que geram $EC(A)$ e são ortogonais).

Seja U_1 uma matriz unitária cujas primeiras k colunas são $\frac{v_i}{\|v_i\|}$ (tal matriz obtém-se completando a base ortonormada $\{\frac{v_1}{\|v_1\|}, \ldots, \frac{v_k}{\|v_k\|}\}$ para $EC(A)$ a uma base ortonormada de $\mathbb{C}^n$.

Então $A = U_1 D U_2$ conforme desejado: esta equação é equivalente a $AU_2^{-1} = U_1 D$ que se verifica por definição da matriz U_1! □

Observação 7.66. *(i) Se A for uma matriz real, podemos tomar para U_1 e U_2 matrizes ortogonais obtendo assim a decomposição em valores singulares real.*

(ii) Embora a decomposição em valores singulares não seja única, os valores singulares *de A são completamente determinados por A: são as raízes quadradas positivas dos valores próprios de A^*A.*

(iii) Em termos muito concretos:
- *As colunas de U_2^{-1} são uma base ortonormada $B_1 = \{w_1, \ldots, w_n\}$ para $\mathbb{C}^n$ (os elementos chamam-se vetores singulares à direita)*
- *As colunas de U_1 são uma base ortonormada $B_2 = \{x_1, \ldots, x_m\}$ para $\mathbb{C}^m$ (os seus elementos chamam-se vetores singulares à esquerda)*
- *O vetor w_i é enviado por A em $\sigma_i x_i$ para $1 \leq i \leq k$ e para 0 se $i > k$.*

Exemplo 7.67. *Consideremos a matriz $A = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & -1 \end{bmatrix}$. Temos*

$$\begin{aligned} A^T A &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & -1 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 1 & -1 \\ 0 & -1 & 2 \end{bmatrix} = U_2^{-1} \Lambda U_2 \\ &= \begin{bmatrix} -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \end{bmatrix}^T \end{aligned}$$

Portanto

$$U_2 = \begin{bmatrix} -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \end{bmatrix}^T, \qquad D = \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & \sqrt{2} & 0 \end{bmatrix}$$

Os valores singulares são $\sigma_1 = \sqrt{3}$, $\sigma_2 = \sqrt{2}$. Neste caso, uma vez que o espaço das colunas é todo o $\mathbb{R}^2$, a matriz U_1 é dada por

$$AU_2^{-1} = \begin{bmatrix} 0 & \sqrt{2} & 0 \\ -\sqrt{3} & 0 & 0 \end{bmatrix} \qquad U_1 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

A decomposição singular de A é portanto

$$A = U_1 D U_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & \sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix}$$

O efeito de A é primeiro colapsar $\mathbb{R}^3$ ortogonalmente no plano gerado pelas duas primeiras colunas de U_2^{-1}. Um círculo unitário neste plano é enviado numa elipse com eixos $\sqrt{3}$ e $\sqrt{2}$ segundo as direções dadas pelas colunas de U_1 (neste caso o eixo dos yy, que é a imagem da reta gerada pela primeira coluna de U_2^{-1} em $\mathbb{R}^3$ e e o eixo dos xx que é a imagem da segunda coluna de U_2^{-1})

A decomposição em valores singulares desempenha um papel importante na análise de dados. Os dados podem normalmente ser organizados numa matriz (gigante) e a decomposição em valores singulares ajuda a organizá-los. A decomposição em valores singulares também pode ser usada para comprimir dados, descartando os valores singulares de tamanho inferior a um dado limite fixado.

7.68. **Formas quadráticas.** Como outra aplicação do Teorema espetral 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 7.69. *Uma* forma quadrática *em $\mathbb{R}^n$ é uma função $f\colon \mathbb{R}^n \to \mathbb{R}$ da forma*

$$f(x) = x^T A x \tag{64}$$

com $A \in M_{n\times n}(\mathbb{R})$ (estamos como habitualmente a identificar uma matriz 1×1 com um escalar).

Por exemplo

$$f(x,y) = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 2 & 5 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 2x^2 + 6xy + 4y^2 \tag{65}$$

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 $x^T A x = (x^T A x)^T = x^T A^T x$. Substituindo a matriz A em (64) 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 $x_i x_j$ na expressão (64) 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

$$f(x) = x^T A x = (y^T S^T) A (Sy) = (y^T S^T) SDS^T (Sy) = y^T D y = \lambda_1 y_1^2 + \ldots + \lambda_n y_n^2 \tag{66}$$

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 7.70. *Uma forma quadrática* $f : \mathbb{R}^n \to \mathbb{R}$ *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.

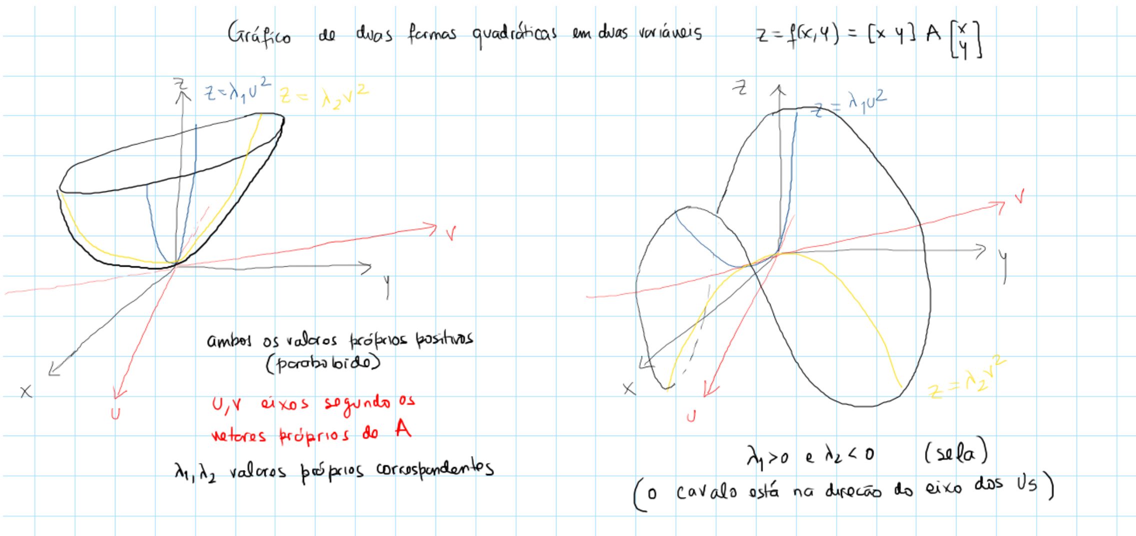
Proposição 7.71. *Uma forma quadrática* $f(x) = x^T Ax$ *com* $A \in M_{n\times n}(\mathbb{R})$ *simétrica é*

(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 7.72. *A forma quadrática* (65) *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 7.73. *A expressão* (66) *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.*

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 7.74 (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 $i = 1, \ldots n$. 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 intersecção 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 7.75. *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.

7.76. **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).

Quádricas em $\mathbb{R}^2$. A expressão geral de uma quádrica é

$$ax^2 + bxy + cy^2 + dx + ey + f = 0 \tag{67}$$

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

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} u_1 & v_1 \\ u_2 & v_2 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \tag{68}$$

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 (67) transforma-se mediante a mudança de coordenadas (68) num termo linear em u e v, pelo que esta mudança de coordenadas transforma (67) na seguinte equação:

$$\lambda_1 u^2 + \lambda_2 v^2 + d'u + e'v + f = 0 \tag{69}$$

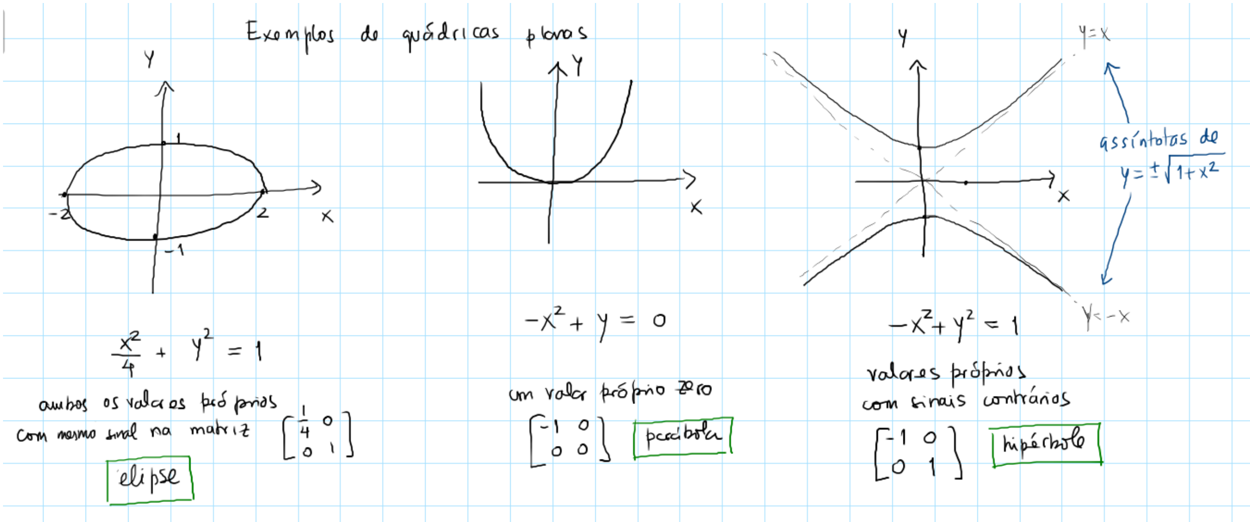
Temos agora a considerar três casos:

- λ_1, λ_2 ambos diferentes de 0, com o mesmo sinal: Multiplicando (69) 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 (69) 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[23] 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 (69) como antes obtemos
$$\lambda_1(u-u_0)^2 + e'v + f' = 0$$
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 7.77. *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
$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

[23]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 $uv = 1$.

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

$$u = 2\sqrt{2}(v + \tfrac{3}{4\sqrt{2}})^2 + 3\sqrt{2} - \tfrac{9}{8\sqrt{2}}$$

Quádricas em $\mathbb{R}^3$. A expressão geral de uma quádrica é

$$ax^2 + by^2 + cz^2 + dxy + exz + fyz + gx + hy + iz + j = 0 \tag{70}$$

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 (70) 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

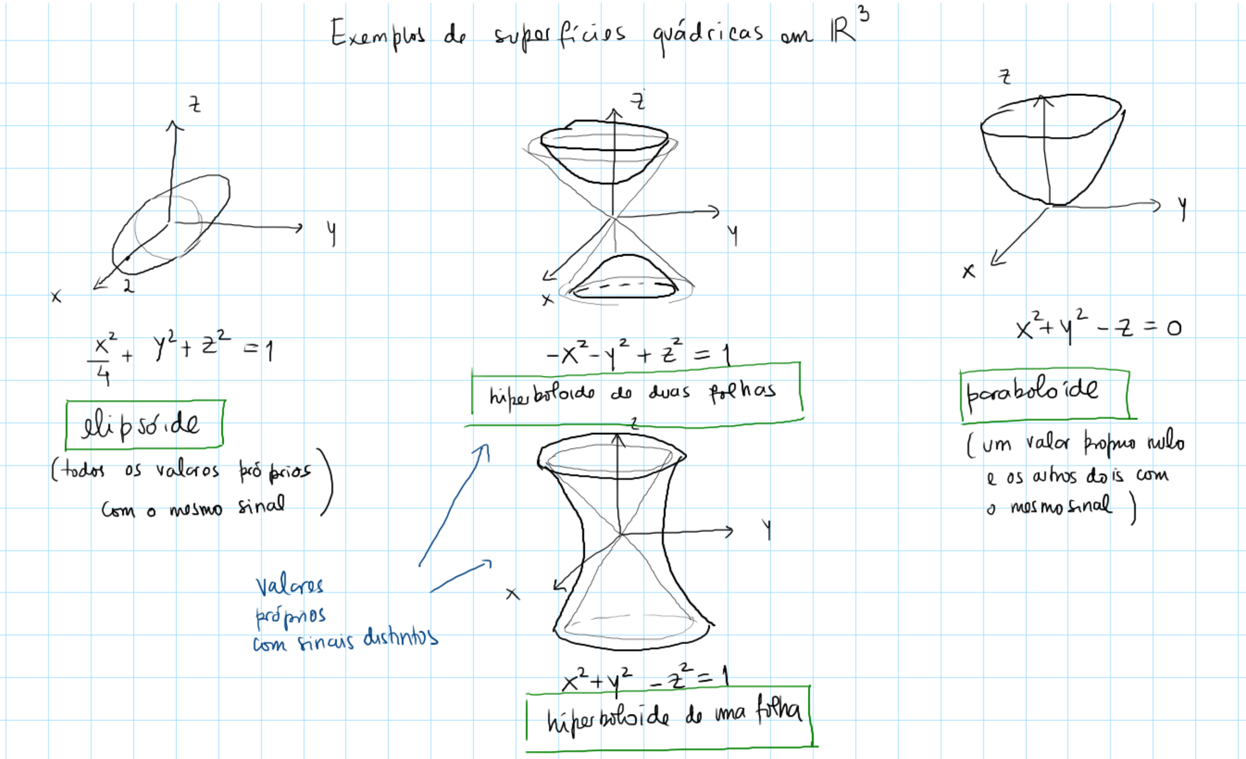
$$x^2 - y^2 = j', \quad x^2 - y^2 - z = j'$$

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

[Ax] S. Axler, *Linear Algebra Done Right*, Springer UTM (1997).

[D] E. Dias, *Álgebra Linear*, `https://www.math.tecnico.ulisboa.pt/~edias/TextosNet/ALbookfin_Net.pdf`

[FIS] S. Friedberg, A. Insel and L. Spence, *Linear Algebra* (4th edition), Pearson Education (2003).

[HK] K. Hoffman and R. Kunze, *Linear Algebra*, Prentice-Hall (1961)