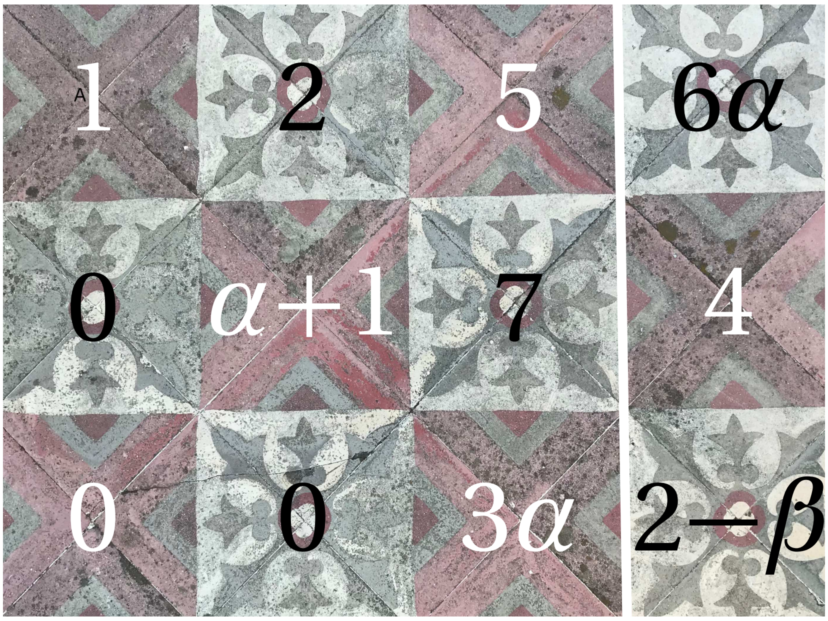
Texto de Apoio de Álgebra Linear



Pedro C. Silva • Adelino Paiva • Isabel Martins • Marta Mesquita

ISA / ULisboa
2021

Conteúdo

Prefácio

A sebenta *Texto de Apoio de Álgebra Linear* cobre a matéria teórica da UC Álgebra Linear que é lecionada no 1º Ciclo de todos os cursos de Engenharia e do curso Biologia do ISA e veio substituir os anteriores apontamentos da UC do Professor Jorge Orestes.

A influência do texto do Professor Jorge Orestes e de algumas notas e esquemas manuscritos da Professora Isabel Faria, antigos coordenadores da UC Álgebra Linear, estão patentes no presente texto.

Os autores

Capítulo 1

Cálculo Matricial e Sistemas Lineares

1.1 Vetores

Comecemos por estabelecer alguma notação e estender as operações já conhecidas para vetores do plano e do espaço para vetores com um número arbitrário de componentes.

Para qualquer inteiro $n \geq 2$, define-se o conjunto dos vetores com n componentes reais,

$$\mathbb{R}^n = \Big\{(x_1, x_2, \ldots, x_n) : x_1, x_2, \ldots, x_n \in \mathbb{R}\Big\}.$$

O caso $n = 2$ corresponde ao conjunto dos *vetores do plano* e o caso $n = 3$ ao conjunto dos *vetores do espaço* [1].

Operações algébricas sobre vetores

Dados $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$ e $\lambda \in \mathbb{R}$, definem-se as seguintes operações algébricas que estendem, para vetores com um número arbitrário de componentes, as operações já conhecidas da adição de vetores e do produto de um vetor por um escalar.

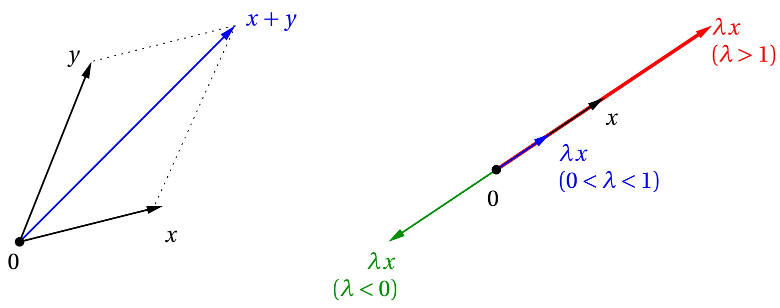
(a) *Adição de vetores*

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

(b) *Multiplicação de um vetor por um escalar*

$$\lambda x = \lambda(x_1, \ldots, x_n) = (\lambda x_1, \ldots, \lambda x_n).$$

[1] Os números reais são designados por *escalares* por oposição aos vetores.



(a) (b)

Exemplo 1. *Se* $x = (-5, 7, 0, 1)$, $y = (0, 1, 2, -1)$ *e* $\lambda = 5$, *então*

- $x + y = (-5, 7, 0, 1) + (0, 1, 2, -1) = (-5, 8, 2, 0)$
- $\lambda x = 5(-5, 7, 0, 1) = (-25, 35, 0, 5)$

As operações anteriores verificam as seguintes propriedades que decorrem das propriedades da soma e da multiplicação de números reais.

Proposição 1. *Sejam* x, y, z *vetores de* $\mathbb{R}^n$, $\vec{0} = (0, \cdots, 0) \in \mathbb{R}^n$ *e* $\lambda, \mu \in \mathbb{R}$. *Tem-se,*

1. $x + y = y + x$
2. $(x + y) + z = x + (y + z)$
3. $x + \vec{0} = x$
4. $x + (-x) = \vec{0}$
5. $\lambda(x + y) = \lambda x + \lambda y$
6. $(\lambda + \mu)x = \lambda x + \mu x$
7. $(\lambda\mu)x = \lambda(\mu x)$
8. $1x = x$.

Define-se ainda o *produto escalar* (ou *produto interno*) já conhecido para vetores do plano e do espaço.

(c) *Produto escalar de vetores*

$$x \cdot y = x|y = (x_1, \cdots, x_n) \cdot (y_1, \ldots, y_n) = x_1 y_1 + \cdots + x_n y_n.$$

Exemplo 2. Considerando novamente os vetores $x = (-5, 7, 0, 1)$ e $y = (0, 1, 2, -1)$ do exemplo anterior, obtém-se

$$x \cdot y = (-5, 7, 0, 1) \cdot (0, 1, 2, -1) = (-5) \cdot 0 + 7 \cdot 1 + 0 \cdot 2 + 1 \cdot (-1) = 0 + 7 + 0 - 1 = 6.$$

O produto escalar verifica as seguintes propriedades.

Proposição 2. *Sejam x, y, z vetores de $\mathbb{R}^n$ e $\lambda, \mu \in \mathbb{R}$. Tem-se,*

1. $x \cdot y = y \cdot x$
2. $x \cdot (y+z) = x \cdot y + x \cdot z$
3. $\lambda(x \cdot y) = (\lambda x) \cdot y = x \cdot (\lambda y)$
4. $x \cdot x = 0 \Leftrightarrow x = \vec{0}$.

As três primeiras propriedades decorrem das propriedades da adição e da multiplicação de números reais. A última propriedade decorre do facto de uma soma de quadrados só se anular se todas as parcelas forem nulas.

Observação 1. As três primeiras propriedades da proposição anterior implicam que o produto escalar se opera de modo análogo ao produto usual de números reais. Por exemplo, os casos notáveis da multiplicação de números reais

$$\begin{array}{rcl} (a+b)^2 & = & a^2+2ab+b^2 \\ (a-b)^2 & = & a^2-2ab+b^2 \\ (a-b)(a+b) & = & a^2-b^2 \end{array}$$

traduzem-se para o produto escalar como,

$$\begin{array}{rcl} (x+y)\cdot(x+y) & = & (x+y)\cdot x+(x+y)\cdot y \\ & = & x\cdot x+y\cdot x+x\cdot y+y\cdot y \\ & = & x\cdot x+2\,x\cdot y+y\cdot y \\ (x-y)\cdot(x-y) & = & x\cdot x-2\,x\cdot y+y\cdot y \\ (x+y)\cdot(x-y) & = & x\cdot x-y\cdot y. \end{array}$$

A verificação dos segundo e terceiro casos é análoga e fica como exercício.

O produto escalar permite estender conceitos bem conhecidos para vetores do plano e do espaço, como *norma*, *distância*, *ortogonalidade* e *ângulo*, para vetores com um número arbitrário de componentes.

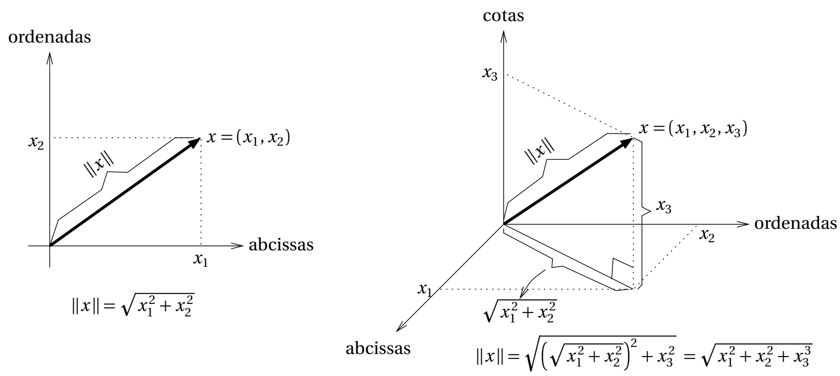
Norma e distância

Por aplicação do teorema de Pitágoras, deduz-se a norma (ou comprimento) para vetores do plano e do espaço,

(a) Se $x = (x_1, x_2) \in \mathbb{R}^2$, $\|(x_1, x_2)\| = \sqrt{x_1^2 + x_2^2}$

(b) Se $x = (x_1, x_2, x_3) \in \mathbb{R}^3$, $\|(x_1, x_2, x_3)\| = \sqrt{x_1^2 + x_2^2 + x_3^2}$,

como se pode constatar nas seguintes figuras.



(a) (b)

Note-se que, em ambos os casos, a expressão da norma pode ser definida recorrendo ao produto escalar como $\|x\|=\sqrt{x\cdot x}$. A norma de um vetor com um número arbitrário de componentes é definida de modo análogo.

Definição 1. Dado um vetor $x=(x_1,x_2,\ldots,x_n)\in\mathbb{R}^n$ define-se *norma* (ou *comprimento*) de x, denotada $\|x\|$, por

$$\|x\|=\sqrt{x\cdot x}=\sqrt{(x_1,\ldots,x_2,\ldots,x_n)\cdot(x_1,\ldots,x_2,\ldots,x_n)}=\sqrt{x_1^2+x_2^2+\cdots+x_n^2}.$$

Por exemplo,

$$\|(4,2,-1,2)\|=\sqrt{(4,2,-1,2)\cdot(4,2,-1,2)}=\sqrt{4^2+2^2+(-1)^2+2^2}=5.$$

Enunciam-se a seguir algumas propriedades da norma.

Proposição 3. *Dados* $x=(x_1,x_2,\ldots,x_n)\in\mathbb{R}^n$ *e* $\lambda\in\mathbb{R}$, *tem-se*

(ii) $\|x\|\geq 0$

(i) $\|x\|^2=x\cdot x$

(iii) $\|x\|=0$ *se e só se* $x=\vec{0}$

(iv) $\|\lambda x\|=|\lambda|\,\|x\|$.

Demonstração. As afirmações (i) e (ii) são óbvias. A propriedade (iii) resulta da afirmação 4 da Proposição 2. Relativamente a (iv), tem-se

$$\begin{aligned}\|\lambda x\| &= \sqrt{(\lambda x_1)^2+(\lambda x_2)^2+\cdots+(\lambda x_n)^2}\\ &= \sqrt{\lambda^2x_1^2+\lambda^2x_2^2+\cdots+\lambda^2x_n^2}\\ &= \sqrt{\lambda^2(x_1^2+x_2^2+\cdots+x_n^2)}\\ &= \sqrt{\lambda^2}\sqrt{(x_1^2+x_2^2+\cdots+x_n^2)}=|\lambda|\,\|x\|,\end{aligned}$$

atendendo a que $\sqrt{a^2}=|a|$ para qualquer número real a. □

Um vetor x diz-se *unitário* se $\|x\| = 1$.

Dado um vetor $x \in \mathbb{R}^n$, $x \neq \vec{0}$, define-se o *versor* de x, denotado $\text{vers}(x)$, como sendo o vetor unitário que tem a direção e sentido de x. Por outras palavras, $\text{vers}(x) = \lambda x$, com $\lambda > 0$ (para manter o sentido) tal que $\|\lambda x\| = 1$. Desta relação, concluímos, usando a propriedade (iv) da Proposição 3 e atendendo ao facto que $\lambda > 0$, que $\lambda \|x\| = 1$ e, portanto, que $\lambda = \frac{1}{\|x\|}$. Logo,

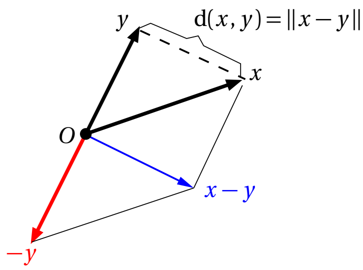
$$\text{vers}(x) = \frac{x}{\|x\|}.$$

Por exemplo, $\text{vers}(3,4) = \frac{(3,4)}{\|(3,4)\|} = \frac{(3,4)}{5} = \left(\frac{3}{5}, \frac{4}{5}\right)$.

A partir do conceito de norma define-se a noção de distância entre vetores.

Definição 2. Dados vetores de $\mathbb{R}^n$, $x = (x_1, x_2 \dots, x_n)$ e $y = (y_1, y_2, \dots, y_n)$, define-se a *distância* entre x e y, denotada $\text{d}(x, y)$, por

$$\begin{aligned} \text{d}(x,y) &= \|x - y\| \\ &= \|(x_1 - y_1, x_2 - y_2, \dots, x_n - y_n)\| \\ &= \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2} \end{aligned}$$



Enunciam-se a seguir algumas propriedades verificadas pela distância, que decorrem das propriedades da norma.

Proposição 4. *Dados $x, y, z \in \mathbb{R}^n$, tem-se*

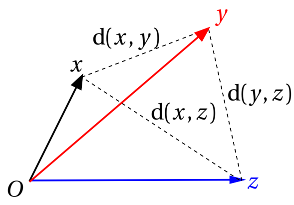
(i) $d(x, y) \geq 0$

(ii) $d(x, y) = 0$ *se e só se* $x = y$

(iii) $d(x, y) = d(y, x)$

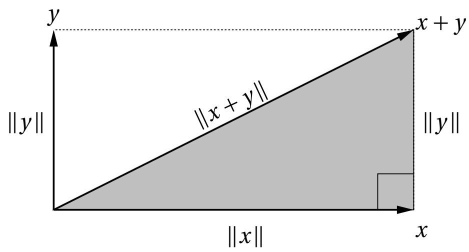
(iv) $d(x, z) \leq d(x, y) + d(y, z)$.

A propriedade (iv) designa-se por *desigualdade triangular* e traduz o facto do comprimento de um lado de um triângulo ser menor ou igual à soma dos comprimentos dos outros 2 lados (sendo que a igualdade apenas ocorre no caso de triângulos degenerados).



Ortogonalidade

Sejam $x, y \in \mathbb{R}^2$ vetores não nulos e perpendiculares entre si e considere-se o triângulo rectângulo de catetos de comprimentos $\|x\|$ e $\|y\|$ e hipotenusa de comprimento $\|x+y\|$, representado na seguinte figura.



Daqui resulta, aplicando o teorema de Pitágoras que

$$\|x+y\|^2 = \|x\|^2 + \|y\|^2. \tag{1.1}$$

Por outro lado, do primeiro caso notável da Observação 1, conclui-se que

$$\|x+y\|^2 = (x+y)\cdot(x+y) = x\cdot x + 2\,x\cdot y + y\cdot y = \|x\|^2 + 2\,x\cdot y + \|y\|^2. \tag{1.2}$$

Igualando (1.1) a (1.2) obtém-se $2\,x\cdot y = 0$. A relação de perpendicularidade entre x e y permite concluir que $x\cdot y = 0$.

A relação no outro sentido também se verifica, ou seja, se $x\cdot y = 0$ então x e y são perpendiculares entre si, uma vez que o teorema de Pitágoras só é válido para triângulos rectângulos.

Em resumo, tem-se a seguinte relação fundamental entre produto escalar e ortogonalidade,

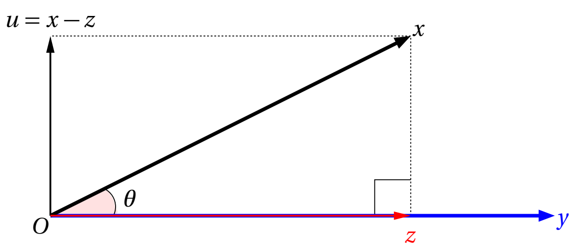
$$\boxed{x \perp y \Leftrightarrow x\cdot y = 0.}$$

Mais geralmente, dados $x, y \in \mathbb{R}^n$ diz-se que x é *ortogonal* a y e denota-se $x \perp y$, se $x\cdot y = 0$.

Ângulo de vetores

Sejam $x, y \in \mathbb{R}^2$ vetores não nulos e θ o ângulo formado por esses vetores. Vamos admitir $0 \le \theta \le \frac{\pi}{2}$ (o caso $\frac{\pi}{2} < \theta \le \pi$ é análogo).

Seja z a projeção de x sobre a reta definida por y, na direção perpendicular a y e $u = x - z$. Em particular, $u \perp y$.



Uma vez que z e y são vetores colineares com o mesmo sentido, existe $\lambda \geq 0$ tal que $z = \lambda y$. Daqui resulta que

$$\|z\| = \lambda \|y\| = \cos(\theta)\|x\|,$$

e portanto que $\lambda = \dfrac{\cos(\theta)\|x\|}{\|y\|}$. Por outro lado, como $x = u+z$ e $y \perp u$, obtém-se

$$\begin{aligned} x \cdot y &= (u+z)\cdot y = \underbrace{u \cdot y}_{0} + z \cdot y \\ &= (\lambda y)\cdot y = \lambda(y \cdot y) = \lambda\|y\|^2. \end{aligned}$$

Daqui concluímos que

$$x \cdot y = \lambda\|y\|^2 = \frac{\cos(\theta)\|x\|}{\|y\|}\|y\|^2 = \cos(\theta)\|x\|\,\|y\|.$$

isto é, obtemos a seguinte relação bem conhecida

$$x \cdot y = \cos(\theta)\|x\|\,\|y\|,$$

onde θ é o ângulo formado entre x e y.

Mais geralmente, dados vetores não nulos $x, y \in \mathbb{R}^n$, define-se de modo análogo, o *ângulo* formado por estes vetores como sendo o único $\theta \in [0,\pi]$ tal que

$$\boxed{\cos(\theta) = \tfrac{x\cdot y}{\|x\|\|y\|}.}$$

Por exemplo, se $x = (1,1,0)$ e $y = (1,0,1)$ obtém-se

$$\cos(\theta) = \frac{x \cdot y}{\|x\|\,\|y\|} = \frac{(1,1,0)\cdot(1,0,1)}{\sqrt{1^2+1^2+0^2}\sqrt{1^2+0^2+1^2}} = \frac{1}{\sqrt{2}\sqrt{2}} = \frac{1}{2}$$

e, portanto, $\theta = \dfrac{\pi}{3}$.

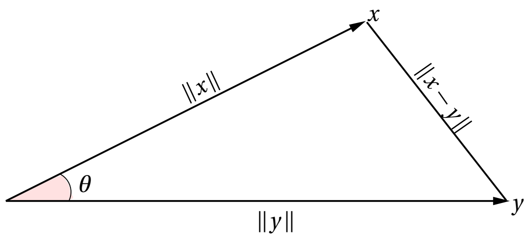
O conceito anterior de ângulo entre vetores não nulos de $\mathbb{R}^n$ está bem definida devido à seguinte relação conhecida por *desigualdade de Cauchy-Schwarz.*

Teorema 5. *Para quaisquer* $x, y \in \mathbb{R}^n$ *tem-se* $|x \cdot y| \leq \|x\|\,\|y\|$.

Termina-se esta secção com uma generalização do teorema de Pitágoras, conhecida por *lei dos cossenos.*

Teorema 6. *Sejam x e y vetores não nulos de $\mathbb{R}^n$ e θ o ângulo por eles formado. Tem-se*

$$\|x-y\|^2 = \|x\|^2 + \|y\|^2 - 2\cos\theta\,\|x\|\,\|y\|.$$



Demonstração. Aplicando o segundo caso da Observação 1, vem

$$\begin{aligned}\|x-y\|^2 &= (x-y)\cdot(x-y) = x\cdot x - 2\,x\cdot y + y\cdot y \\ &= \|x\|^2 + \|y\|^2 - 2\cos(\theta)\|x\|\,\|y\|.\end{aligned}$$

□

1.2 Matrizes

As matrizes são uma extensão dos vetores adequada ao estudo dos sistemas lineares.

Definição 3. Sejam m, n inteiros positivos. Chama-se *matriz do tipo $m \times n$* a uma coleção $A = [a_{ij}]$ de mn números reais dispostos em m linhas e n colunas,

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}_{m\times n}.$$

O índice i percorre as linhas da matriz e designa-se por *índice de linha* e o índice j percorre as colunas da matriz e designa-se por *índice de coluna.*

Exemplos 3.

1. $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{bmatrix}_{4\times 3} = \begin{bmatrix} 1 & -1 & 2 \\ 2 & 10 & \pi \\ 0 & \sqrt{3} & -1 \\ 2 & -1 & 0 \end{bmatrix}_{4\times 3}.$

   Por exemplo, o elemento de A que se encontra na linha 3 e coluna 2 é $a_{32} = \sqrt{3}$.

2. A matriz $A = [a_{ij}]_{2\times 4}$ definida por $a_{1j} = 5$ e $a_{2j} = \sqrt{2}$, $j = 1,2,3,4$, é a matriz
$$\begin{bmatrix} 5 & 5 & 5 & 5 \\ \sqrt{2} & \sqrt{2} & \sqrt{2} & \sqrt{2} \end{bmatrix}_{2\times 4}.$$

3. A matriz $A = [a_{ij}]_{3\times 3}$ definida por $a_{ij} = i - j$, para $i, j = 1,2,3$, é a matriz
$$\begin{bmatrix} 0 & -1 & -2 \\ 1 & 0 & -1 \\ 2 & 1 & 0 \end{bmatrix}_{3\times 3}.$$

Tipos especiais de matrizes

Seja $A = [a_{ij}]_{m\times n}$ uma matriz do tipo $m \times n$.

- Se $m = 1$, $A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \end{bmatrix}$ designa-se por *matriz-linha*.

  Por exemplo, $A = \begin{bmatrix} 1 & 3 & -2 \end{bmatrix}$ é uma matriz-linha do tipo 1×3.

- Se $n = 1$, $A = \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix}$ designa-se por *matriz-coluna* ou *vetor*.

  Um vetor $x \in \mathbb{R}^m$ pode ser representado da forma usual como m-uplo de números reais ou matricialmente como uma matriz-coluna do tipo $m \times 1$:
  $$x = (x_1, \cdots, x_m) = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix}.$$

  Por exemplo, $(2,3,-1) = \begin{bmatrix} 2 \\ 3 \\ -1 \end{bmatrix} \in \mathbb{R}^3$.

  Se $v_1, v_2, \ldots, v_n$ são vetores de $\mathbb{R}^m$ representa-se por $[v_1\ v_2 \cdots v_n]$ a matriz do tipo $m \times n$ cujas colunas são os n vetores $v_1, v_2, \ldots, v_n$.

  Por exemplo, dados vetores de $\mathbb{R}^3$, $u = (2,3,-1)$ e $v = (0,0,1)$,
  $$[u\ \ v] = \begin{bmatrix} 2 & 0 \\ 3 & 0 \\ -1 & 1 \end{bmatrix}_{3\times 2}$$

- Se $m = n$, a matriz A designa-se por *matriz quadrada de ordem* n e o conjunto dos elementos a_{ii}, $i = 1, \dots n$, designa-se por *diagonal principal* de A.

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$$

A matriz quadrada A diz-se *triangular superior* se $a_{ij} = 0$ para $i > j$. Por outras palavras, uma matriz quadrada é triangular superior se todos os elementos da matriz abaixo da diagonal principal forem nulos.

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix}$$

Por exemplo, a matriz $\begin{bmatrix} 1 & 2 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ é triangular superior de ordem 3.

Analogamente, a matriz quadrada A diz-se *triangular inferior* se $a_{ij} = 0$ para $i < j$. Por outras palavras, uma matriz quadrada é triangular inferior se todos os elementos da matriz acima da diagonal principal forem nulos.

$$\begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

Por exemplo, a matriz $\begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix}$ é triangular inferior de ordem 3.

A matriz quadrada A diz-se *diagonal* se $a_{ij} = 0$ para $i \neq j$, isto é, se todos os elementos fora da diagonal principal de A forem nulos.

$$\begin{bmatrix} a_{11} & 0 & \ldots & 0 \\ 0 & a_{22} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & a_{nn} \end{bmatrix}$$

Uma matriz diagonal é simultaneamente triangular superior e triangular inferior e pode ser convenientemente representada por

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

Por exemplo, $A = \text{diag}(2, -1, 3) = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 3 \end{bmatrix}$ é diagonal de ordem 3.

A matriz quadrada A diz-se *escalar* se for diagonal e todas as entradas da diagonal principal de A forem iguais entre si, isto é, se para algum $\lambda \in \mathbb{R}$,

$$A = \text{diag}(\lambda, \lambda, \ldots, \lambda) = \begin{bmatrix} \lambda & 0 & \cdots & 0 \\ 0 & \lambda & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda \end{bmatrix}_{n \times n}.$$

Se $\lambda = 1$, A designa-se por **matriz identidade de ordem** n e denota-se por I_n (ou simplesmente por I). Por exemplo, a matriz identidade de ordem 3 é a matriz

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Igualdade entre matrizes

Diz-se que duas matrizes do mesmo tipo, $A = [a_{ij}]$ e $B = [b_{ij}]$ são *iguais* se os elementos homólogos forem iguais, isto é, se $a_{ij} = b_{ij}\ \forall i, j$.

Por exemplo,

$$\begin{bmatrix} 5 & x \\ y & 6 \end{bmatrix} = \begin{bmatrix} z & 3 \\ 2 & w \end{bmatrix} \Leftrightarrow \begin{cases} x = 3 \\ y = 2 \\ z = 5 \\ w = 6 \end{cases}.$$

Transposição de matrizes

Seja $A = \left[a_{ij}\right]$ uma matriz do tipo $m \times n$. Chama-se *transposta de* A e denota-se por A^T, à matriz do tipo $n \times m$ cujas colunas são as linhas de A pela mesma ordem. Por outras palavras, $A^T = [a_{ji}]_{n\times m}$ com $j = 1, \ldots, n$ e $i = 1, \ldots, m$. Tem-se obviamente, $(A^T)^T = A$.

Exemplos 4.

1. *Se* $A = \begin{bmatrix} 2 & 3 \\ 4 & 1 \\ 5 & 6 \end{bmatrix}$, *então* $A^T = \begin{bmatrix} 2 & 4 & 5 \\ 3 & 1 & 6 \end{bmatrix}$.

2. *Se* $x = (2,3,4,-1) = \begin{bmatrix} 2 \\ 3 \\ 4 \\ -1 \end{bmatrix} \in \mathbb{R}^4$, *então* $x^T = \begin{bmatrix} 2 & 3 & 4 & -1 \end{bmatrix}$.

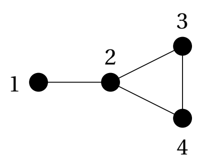
Definição 4. Uma matriz quadrada $A = [a_{ij}]$ diz-se *simétrica*, se $A^T = A$, isto é, se $a_{ij} = a_{ji}$, $\forall i, j$.

Por exemplo, $A = \begin{bmatrix} 2 & x & 4 \\ 5 & -1 & y \\ z & 8 & 3 \end{bmatrix}$ é simétrica se e só se

$$\begin{bmatrix} 2 & x & 4 \\ 5 & -1 & y \\ z & 8 & 3 \end{bmatrix}^T = \begin{bmatrix} 2 & x & 4 \\ 5 & -1 & y \\ z & 8 & 3 \end{bmatrix} \Leftrightarrow \begin{bmatrix} 2 & 5 & z \\ x & -1 & 8 \\ 4 & y & 3 \end{bmatrix} = \begin{bmatrix} 2 & x & 4 \\ 5 & -1 & y \\ z & 8 & 3 \end{bmatrix} \Leftrightarrow \begin{cases} x = 5 \\ y = 8 \\ z = 4 \end{cases}.$$

Exemplo 5. Chama-se *grafo* a um par $G = (V, E)$ onde V denota o conjunto dos *vértices* do grafo e E o conjunto das *arestas* (*edges*) que unem pares de vértices. Os grafos permitem representar objectos e relações binárias entre esses objectos. Por exemplo, os vértices podem representar píxeis numa imagem de satélite e as arestas relações de adjacência entre esses píxeis.

Um grafo (não orientado) G com n vértices $v_1, \ldots, v_n$ pode ser representado através de uma matriz simétrica $A_G = [a_{ij}]$ de ordem n, designada por *matriz de adjacência* do grafo, em que $a_{ij} = 1$ se os vértices v_i e v_j estão ligados por uma aresta e $a_{ij} = 0$ no caso contrário.



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

1.3 Operações algébricas sobre matrizes

As operações algébricas sobre matrizes estendem as operações da adição de vetores, do produto de um vetor por um escalar e do produto escalar de vetores.

Adição de matrizes

Sejam $A = \left[a_{ij}\right]_{m\times n}$ e $B = \left[b_{ij}\right]_{m\times n}$ matrizes do mesmo tipo. Define-se a *soma de A com B*, por

$$A + B = \left[a_{ij} + b_{ij}\right]_{m\times n}.$$

Por outras palavras, os elementos de $A + B$ obtêm-se somando os elementos homólogos de A e de B.

Exemplo 6. *Se* $A = \begin{bmatrix} 2 & -1 & 0 \\ 4 & 5 & 3 \end{bmatrix}$ *e* $B = \begin{bmatrix} 0 & 2 & 1 \\ -3 & 1 & 4 \end{bmatrix}$*, então*

$$A + B = \begin{bmatrix} 2+0 & -1+2 & 0+1 \\ 4-3 & 5+1 & 3+4 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 6 & 7 \end{bmatrix}.$$

Produto de uma matriz por um escalar

Sejam $\lambda \in \mathbb{R}$ e $A = \left[a_{ij}\right]_{m\times n}$. Define-se o *produto de A pelo escalar* λ, por

$$\lambda A = \left[\lambda a_{ij}\right]_{m\times n}.$$

Por outras palavras, λA é a matriz que se obtém multiplicando cada elemento de A por λ. Se $\lambda = -1$, λA denota-se mais simplesmente por $-A$.

Exemplo 7. *Se* $A = \begin{bmatrix} 2 & -7 & 3 \\ 8 & -2 & 8 \end{bmatrix}$ *e* $\lambda = 5$*, então*

$$\lambda A = 5\begin{bmatrix} 2 & -7 & 3 \\ 8 & -2 & 8 \end{bmatrix} = \begin{bmatrix} 5\cdot 2 & 5\cdot(-7) & 5\cdot 3 \\ 5\cdot 8 & 5\cdot(-2) & 5\cdot 8 \end{bmatrix} = \begin{bmatrix} 10 & -35 & 15 \\ 40 & -10 & 40 \end{bmatrix}.$$

Chama-se *matriz nula* (do tipo $m \times n$) e denota-se por $[0]_{m\times n}$, à matriz do tipo $m \times n$ cujos elementos são todos nulos.

As operações algébricas da adição de matrizes e produto de uma matriz por um escalar satisfazem as seguintes propriedades.

Proposição 7. *Sejam A, B e C matrizes do tipo $m \times n$ e $\lambda, \mu \in \mathbb{R}$. Tem-se,*

(1) $A + B = B + A$;

(2) $(A + B) + C = A + (B + C)$

(3) $A + [0]_{m\times n} = A$

(4) $A+(-A)=[0]_{m\times n}$

(5) $\lambda(A+B)=\lambda A+\lambda B$

(6) $(A+B)^T=A^T+B^T$

(7) $(\lambda+\mu)A=\lambda A+\mu A$

(8) $\lambda(\mu A)=(\lambda\mu)A$

(9) $(\lambda A)^T=\lambda A^T$.

A matriz nula é o elemento neutro da adição de matrizes. As propriedades (1)-(5), (7) e (8) decorrem imediatamente das propriedades da adição e do produto de números reais. As restantes são evidentes.

Exemplo 8. Consideremos $A=\begin{bmatrix}1 & 2\\ 3 & 4\end{bmatrix}$, $B=\begin{bmatrix}5 & 6\\ -7 & 8\end{bmatrix}$ e I_2 a matriz identidade de ordem 2. Tem-se,

$$\begin{aligned}(5(2A+B^T)-I_2)^T &= (10A+5B^T-I_2)^T=10A^T+5(B^T)^T-I_2^T\\ &= 10A^T+5B-I_2\\ &= 10\begin{bmatrix}1 & 3\\ 2 & 4\end{bmatrix}+5\begin{bmatrix}5 & 6\\ -7 & 8\end{bmatrix}-\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\\ &= \begin{bmatrix}34 & 60\\ -15 & 79\end{bmatrix}.\end{aligned}$$

Produto de matrizes

Definição 5. Diz-se que A e B são matrizes *encadeadas* se o número de colunas de A for igual ao número de linhas de B.

Exemplo 9. A matrizes $A=\begin{bmatrix}2 & 3\\ -1 & 4\\ 0 & 5\end{bmatrix}_{3\times 2}$ e $B=\begin{bmatrix}1 & -1 & 0 & 5\\ 2 & 1 & -1 & 3\end{bmatrix}_{2\times 4}$ são encadeadas pois o número de colunas de A é igual ao número de linhas de B. Observe-se no entanto que B e A não são matrizes encadeadas uma vez que o número de colunas de B é distinto do número de linhas de A.

Definição 6. Dadas <u>matrizes encadeadas</u> $A=\left[a_{ij}\right]_{m\times n}$ e $B=\left[b_{jk}\right]_{n\times p}$, define-se o *produto de A por B* e denota-se por AB, como sendo a matriz $C=\left[c_{ik}\right]_{m\times p}$, tal que o elemento que se encontra na linha i e coluna k de C é o produto escalar entre a linha i de A e a coluna k de B, ou seja,

$$c_{ik}=(\text{linha } i \text{ de } A)\cdot(\text{coluna } k \text{ de } B)$$

Note-se que o produto escalar entre a linha i de A e a coluna k de B está bem definido uma vez que ambos os vetores contêm o mesmo número de componentes (porque A e B são encadeadas). De facto,

$$\begin{aligned} c_{ik} &= (\text{linha } i \text{ de } A)\cdot(\text{coluna } k \text{ de } B) \\ &= (a_{i1}, a_{i2}, \dots, a_{in})\cdot(b_{1k}, b_{2k}, \dots, b_{nk}) \\ &= a_{i1}b_{1k} + a_{i2}b_{2k} + \cdots + a_{in}b_{nk}. \end{aligned}$$

Exemplos 10.

1. Consideremos $A = \begin{bmatrix} 2 & 4 \\ 4 & 5 \\ 1 & 8 \end{bmatrix}$ e $B = \begin{bmatrix} 1 & 1 \\ 3 & 4 \end{bmatrix}$. Então A e B são matrizes encadeadas e tem-se,

$$\begin{aligned} AB = \begin{bmatrix} 2 & 4 \\ 4 & 5 \\ 1 & 8 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 3 & 4 \end{bmatrix} &= \begin{bmatrix} (2,4)\cdot(1,3) & (2,4)\cdot(1,4) \\ (4,5)\cdot(1,3) & (4,5)\cdot(1,4) \\ (1,8)\cdot(1,3) & (1,8)\cdot(1,4) \end{bmatrix} \\ &= \begin{bmatrix} 2+12 & 2+16 \\ 4+15 & 4+20 \\ 1+24 & 1+32 \end{bmatrix} = \begin{bmatrix} 14 & 18 \\ 19 & 24 \\ 25 & 33 \end{bmatrix}. \end{aligned}$$

   Por exemplo, o elemento de AB que se encontra na linha 2 e coluna 1 vem dado pelo produto escalar da segunda linha de A (a verde) pela primeira coluna de B (a amarelo), isto é, corresponde a $(4,5)\cdot(1,3) = 19$.

2. Consideremos as matrizes $A = \begin{bmatrix} 1 & 3 & 4 \end{bmatrix}$ e $B = \begin{bmatrix} -1 \\ 10 \\ 2 \end{bmatrix}$. Ambos os produtos AB e BA estão definidos, tendo-se

   - $AB = \begin{bmatrix} 1 & 3 & 4 \end{bmatrix}_{1\times 3} \begin{bmatrix} -1 \\ 10 \\ 2 \end{bmatrix}_{3\times 1} = \begin{bmatrix} 37 \end{bmatrix}_{1\times 1} = 37$
     (as matrizes do tipo 1×1 identificam-se usualmente com o seu único elemento)
   - $BA = \begin{bmatrix} -1 \\ 10 \\ 2 \end{bmatrix}_{3\times 1} \begin{bmatrix} 1 & 3 & 4 \end{bmatrix}_{1\times 3} = \begin{bmatrix} -1 & -3 & -4 \\ 10 & 30 & 40 \\ 2 & 6 & 8 \end{bmatrix}_{3\times 3}.$

A observação seguinte mostra que a noção de produto de matrizes estende o conceito de produto escalar de vetores.

Observação 2. O produto interno de dois vetores $x, y \in \mathbb{R}^n$ pode ser escrito matricialmente como $x^T y$. De facto, se

$$x = (x_1, x_2, \ldots, x_n) = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \text{e} \quad y = (y_1, y_2, \ldots, y_n) = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix},$$

tem-se

$$\begin{aligned} x^T y &= \begin{bmatrix} x_1 & x_2 & \ldots & x_n \end{bmatrix}_{1\times n} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n\times 1} \\ &= \Big[(x_1, x_2, \ldots, x_n) \cdot (y_1, y_2, \ldots, y_n) \Big]_{1\times 1} \\ &= \Big[x_1 y_1 + x_2 y_2 + \cdots + x_n y_n \Big]_{1\times 1} \\ &= x_1 y_1 + x_2 y_2 + \cdots + x_n y_n \\ &= x \cdot y. \end{aligned}$$

Por exemplo, se $x = (1, 2, 3) = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ e $y = (-1, 1, 2) = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}$, tem-se

$$x^T y = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}_{1\times 3} \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}_{3\times 1} = \Big[(1, 2, 3) \cdot (-1, 1, 2) \Big] = 7.$$

Observação 3. Sejam $A = \big[a_{ij}\big]_{m\times n}$ e $B = \big[b_{jk}\big]_{n\times p}$ matrizes encadeadas.

(a) A i-ésima linha de AB é o produto da i-ésima linha de A por B.

(b) A k-ésima coluna de AB é o produto de A pela k-ésima coluna de B.

Exemplo 11. Vejamos a relação (b) da observação anterior num exemplo. Consideremos novamente as matrizes do Exemplo 10,

$$A = \begin{bmatrix} 2 & 4 \\ 4 & 5 \\ 1 & 8 \end{bmatrix} \quad \text{e} \quad B = \begin{bmatrix} 1 & 1 \\ 3 & 4 \end{bmatrix}.$$

Tem-se,

$$\begin{aligned} AB &= \begin{bmatrix} 2 & 4 \\ 4 & 5 \\ 1 & 8 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} (2,4)\cdot(1,3) & (2,4)\cdot(1,4) \\ (4,5)\cdot(1,3) & (4,5)\cdot(1,4) \\ (1,8)\cdot(1,3) & (1,8)\cdot(1,4) \end{bmatrix} \\ &= \left[\begin{bmatrix} 2 & 4 \\ 4 & 5 \\ 1 & 8 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} \quad \begin{bmatrix} 2 & 4 \\ 4 & 5 \\ 1 & 8 \end{bmatrix} \begin{bmatrix} 1 \\ 4 \end{bmatrix} \right] = \left[A \begin{bmatrix} 1 \\ 3 \end{bmatrix} \quad A \begin{bmatrix} 1 \\ 4 \end{bmatrix} \right]. \end{aligned}$$

Portanto a primeira [segunda] coluna de AB é o produto de A pela primeira [segunda] coluna de B, isto é,

$$AB = A \begin{bmatrix} 1 & 1 \\ 3 & 4 \end{bmatrix} = \left[A \begin{bmatrix} 1 \\ 3 \end{bmatrix} \quad A \begin{bmatrix} 1 \\ 4 \end{bmatrix} \right].$$

Deixa-se como exercício considerar a relação (a) da Observação 3 no mesmo exemplo.

Exercício. *Qual é efeito de multiplicar uma matriz A do tipo $m \times n$ à esquerda por uma matriz diagonal $diag(\lambda_1, \ldots, \lambda_m)$? E de multiplicar A à direita por uma matriz diagonal $diag(\mu_1, \ldots, \mu_n)$?*

Definição 7. Dada uma matriz quadrada A de ordem n definem-se as *potências inteiras não negativas de A* por,

$$A^0 = I_n \qquad \text{e} \qquad A^k = \underbrace{A \cdots A}_{\text{k vezes}} \quad (k \in \mathbb{N}).$$

Têm-se as seguintes propriedades que envolvem o produto de matrizes e generalizam propriedades conhecidas da multiplicação de números reais (com excepção da última).

Proposição 8. *Sejam A, B, C matrizes, I a matriz identidade de ordem conveniente, 0 a matriz nula de tipo conveniente e $\lambda \in \mathbb{R}$. Sempre que as operações estejam definidas, tem-se:*

1. $(AB)C = A(BC)$
2. $A(B + C) = AB + AC$
3. $(A + B)C = AC + BC$
4. $AI = IA = A$
5. $A0 = 0A = 0$
6. $\lambda(AB) = (\lambda A)B = A(\lambda B)$
7. $(AB)^T = B^T A^T$.

Exemplo 12. Sejam $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}_{2\times 2}$, $B = \begin{bmatrix} 5 & 6 \\ -7 & 8 \end{bmatrix}_{2\times 2}$ $c = \begin{bmatrix} 1 & -4 \end{bmatrix}_{1\times 2}$ e $d = \begin{bmatrix} 1 \\ 2 \end{bmatrix}_{2\times 1}$.
Tem-se,

$$\begin{aligned}
(5A^T(B-dc))^T &= 5(B-dc)^T(A^T)^T = 5(B^T-(dc)^T)A \\
&= 5(B^T - c^T d^T)A \\
&= 5\left(\begin{bmatrix} 5 & -7 \\ 6 & 8 \end{bmatrix} - \begin{bmatrix} 1 \\ -4 \end{bmatrix}\begin{bmatrix} 1 & 2 \end{bmatrix}\right)\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \\
&= 5\left(\begin{bmatrix} 5 & -7 \\ 6 & 8 \end{bmatrix} - \begin{bmatrix} 1 & 2 \\ -4 & -8 \end{bmatrix}\right)\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \\
&= 5\begin{bmatrix} 4 & -9 \\ 10 & 16 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \\
&= 5\begin{bmatrix} -23 & -28 \\ 58 & 84 \end{bmatrix} = \begin{bmatrix} -115 & -140 \\ 290 & 420 \end{bmatrix}.
\end{aligned}$$

Ao contrário da adição, algumas propriedades do produto de números reais não se mantêm válidas para o produto de matrizes.

Observação 4.

- O produto de matrizes **não é comutativo**, ou seja, em geral,

  $$AB \neq BA.$$

  Por exemplo, considerando $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ e $B = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}$, tem-se

  $$AB = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \neq \begin{bmatrix} 2 & 2 \\ -2 & -2 \end{bmatrix} = BA$$

- A **lei do corte não é válida**, ou seja, dadas matrizes A, B e C, com $A \neq 0$,

  $$AB = AC \quad \not\Rightarrow \quad B = C,$$

  como se pode constatar considerando as matrizes A e B do ponto anterior e C a matriz nula do tipo 2×2.

- A **lei do anulamento do produto também não é válida** uma vez que esta propriedade implica a lei do corte. Isto significa que

  $$AB = 0 \quad \not\Rightarrow \quad (A = 0 \text{ ou } B = 0),$$

  como se pode constatar imediatamente considerando novamente as matrizes A e B definidas no primeiro ponto.

A não comutatividade do produto de matrizes tem diversas consequências importantes e inesperadas.

Observação 5. Dadas matrizes quadradas da mesma ordem A e B, tais que $AB \neq BA$, tem-se,

1. $(A+B)^2 = (A+B)(A+B) = A^2 + AB + BA + B^2 \neq A^2 + 2AB + B^2$,
2. $(A+B)(A-B) = A^2 - AB + BA - B^2 \neq A^2 - B^2$,

o que mostra, em particular, que os casos notáveis da multiplicação (cf. Observação 1) não são, em geral, válidos para o produto de matrizes. Por essa razão deve ter-se especial atenção na simplificação de expressões que envolvam o produto de matrizes.

Duas matrizes quadradas da mesma ordem, A e B, dizem-se *permutáveis* se $AB = BA$. Para matrizes permutáveis são ainda válidos os casos notáveis da Observação 1.

1.4 Sistemas lineares

Equações lineares

Definição 8. Seja n um inteiro positivo. Chama-se *equação linear a n variáveis* (ou *incógnitas*), $x_1, x_2, \dots, x_n$, a uma equação da forma

$$a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = b, \tag{1.3}$$

em que $a_1, \dots, a_n \in \mathbb{R}$, se designam por *coeficientes* das variáveis $x_1, \dots, x_n$, respetivamente e $b \in \mathbb{R}$ por *termo constante* ou *membro direito* da equação.

Definição 9. Chama-se *solução* da equação linear (1.3) a uma sequência de n números reais que substituídos nas variáveis $x_1, x_2, \dots, x_n$ transformam a equação numa proposição verdadeira.

Resolver uma equação é determinar o seu *conjunto-solução* (CS), isto é, o conjunto das soluções da equação.

Exemplo 13.

- A equação $2x_1 = 6$ é linear e o seu CS é formado por um único ponto da reta real (CS = $\{3\}$), pelo que a equação é classificada como possível e determinada (PD).
- A equação $2x_1 + 3x_2 = 4$ é linear e o seu CS define uma reta no plano ($\mathbb{R}^2$), perpendicular ao vetor $(2,3)$, pelo que a equação é classificada como possível indeterminada (PI). Resolvendo esta equação em ordem à variável x_1 obtém-se

$$\text{CS} = \left\{(x_1, x_2) \in \mathbb{R}^2 \colon x_1 = 2 - \tfrac{3}{2} x_2,\ x_2 \in \mathbb{R}\right\},$$

onde a variável x_2 pode tomar valores arbitrários sendo designada por variável *livre ou independente*: cada atribuição de valor à variável livre x_2 origina uma solução distinta da equação.

- A equação $0x_1 + x_2 + 2x_3 = 3$ é linear e define um plano no espaço ($\mathbb{R}^3$), perpendicular ao vetor $(0,1,2)$. Resolvendo esta equação em ordem à variável x_2 obtém-se

$$\text{CS} = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_2 = 3 - 2x_3,\ x_1 \in \mathbb{R}, x_3 \in \mathbb{R}\},$$

  onde x_1 e x_3 desempenham o papel de variáveis livres. Tal como no caso anterior a equação é classificada como PI.

- A equação $2x_1 + 3x_2 + x_3 - 2x_4 = 6$ é linear e define um subconjunto de $\mathbb{R}^4$, que se designa por *hiperplano* e que generaliza a noção de plano. Resolvendo esta equação em ordem à variável x_1 tem-se

$$\text{CS} = \left\{(x_1, x_2, x_3, x_4) : x_1 = 3 - \tfrac{3}{2}x_2 - \tfrac{1}{2}x_3 + x_4,\ x_2 \in \mathbb{R},\ x_3 \in \mathbb{R},\ x_4 \in \mathbb{R}\right\}.$$

  Esta equação possui 3 variáveis livres, x_2, x_3 e x_4 e é classificada como PI.

- A equação linear $0x_1 + 0x_2 = 1$ é impossível (IMP), tendo-se $\text{CS} = \emptyset$.
- A equação $0x_1 + 0x_2 + 0x_3 = 0$ é universal (todas as variáveis são livres), tendo-se $\text{CS} = \mathbb{R}^3$.
- A equação $x_1^2 + x_2^2 = 1$ não é linear pois não é da forma da expressão (1.3). Qual será o seu CS?

Em resumo, desde que um dos coeficientes a_i seja não nulo, a equação linear (1.3) pode ser resolvida em ordem à variável correspondente x_i (variável dependente), sendo as restantes varáveis livres e tem-se o seguinte:

nº var.	equação linear	CS	nº var. livres	classif.
1	$a_1x_1 = b$	ponto em $\mathbb{R}$	0	PD
2	$a_1x_1 + a_2x_2 = b$	reta em $\mathbb{R}^2$	1	PI
3	$a_1x_1 + a_2x_2 + a_3x_3 = b$	plano em $\mathbb{R}^3$	2	PI
$n \geq 4$	$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$	hiperplano em $\mathbb{R}^n$	n-1	PI

Sistemas de equações lineares

Definição 10. Sejam m, n inteiros positivos. Chama-se *sistema linear a m equações lineares e n variáveis $x_1, \ldots, x_n$* a um sistema da forma

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2 \\ \qquad\qquad \ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m \end{cases}, \tag{1.4}$$

em que $a_{ij} \in \mathbb{R}$ é o *coeficiente da variável* x_j *na* i*-ésima equação* e $b_i \in \mathbb{R}$ o respectivo *termo constante* ou *membro direito.*

Uma *solução* do sistema (1.4) é uma solução comum a todas as equações do sistema, ou seja, uma sequência de números reais $s_1, \ldots, s_n$ que verificam

$$\begin{cases} a_{11}s_1 + a_{12}s_2 + \cdots + a_{1n}s_n = b_1 \\ a_{21}s_1 + a_{22}s_2 + \cdots + a_{2n}s_n = b_2 \\ \qquad\qquad \vdots \\ a_{m1}s_1 + a_{m2}s_2 + \cdots + a_{mn}s_n = b_m \end{cases}$$

Dois sistemas lineares a m equações e n variáveis dizem-se *equivalentes* se possuem o mesmo conjunto de soluções (CS).

Geometria dos sistemas de equações lineares

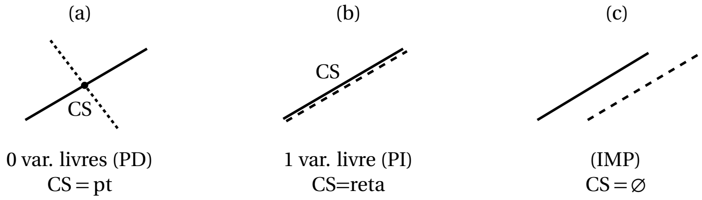
O CS de um sistema linear a m equações e n variáveis corresponde à interseção dos m conjuntos de soluções das equações que compõe esse sistema, isto é, à interseção de m retas se $n = 2$, m planos se $n = 3$ ou m hiperplanos se $n \geq 4$.

Vamos analisar a seguir a geometria dos sistemas lineares a m equações e n variáveis para alguns valores de m e n.

- Sistemas lineares a 2 equações e 2 variáveis

$$\begin{cases} a_{11}x_1 + a_{12}x_2 = b_1 \\ a_{21}x_1 + a_{22}x_2 = b_2 \end{cases}. \tag{1.5}$$

Geometricamente, estes sistemas representam a interseção de duas retas no plano ($\mathbb{R}^2$) e 3 situações distintas podem ocorrer: (a) *retas concorrentes num ponto*, (b) *retas coincidentes* e (c) *retas paralelas distintas*:



Como exemplos de sistemas para cada uma destas situações temos, respectivamente:

$$\text{(a)} \begin{cases} 2x_1 - x_2 = -1 \\ 2x_1 + 3x_2 = 6 \end{cases} \qquad \text{(b)} \begin{cases} 2x_1 - x_2 = -1 \\ 4x_1 - 2x_2 = -2 \end{cases} \qquad \text{(c)} \begin{cases} 2x_1 - x_2 = -1 \\ 4x_1 - 2x_2 = -5 \end{cases}.$$

No primeiro caso, os vetores normais às retas, $n_1 = (2,-1)$ e $n_2 = (2,3)$, são não colineares entre si, logo as retas são concorrentes num ponto, independentemente do valor dos termos constantes b_1 e b_2. Classificamos este sistema como PD.

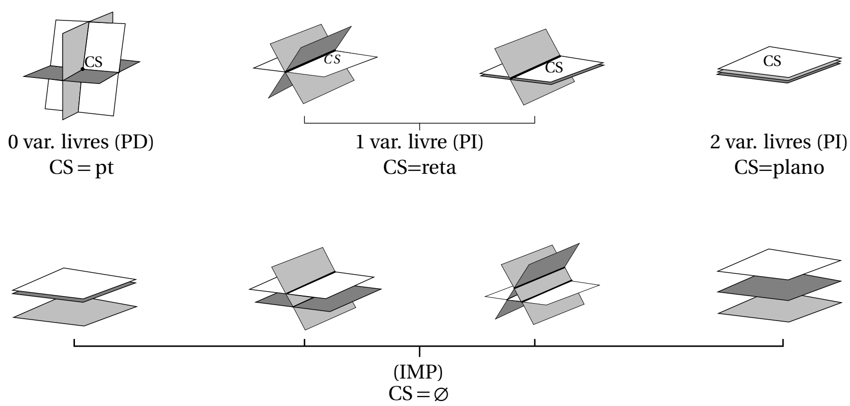
No segundo e terceiro casos os vetores normais às retas verificam a relação de colinearidade $n_2 = 2\,n_1$ e portanto as retas são coincidentes se os termos constantes verificarem $b_2 = 2\,b_1$, ou paralelas se $b_2 \neq 2\,b_1$. Estes sistemas são classificados como PI e IMP, respectivamente.

- Podemos conduzir uma análise semelhante para sistemas lineares a 2 equações e 3 variáveis. Geometricamente, estes sistemas representam a interseção de 2 planos no espaço ($\mathbb{R}^3$) e 3 casos distintos podem ocorrer: *planos concorrentes numa reta*, *planos coincidentes* e *planos paralelos distintos*. Deixa-se como exercício adaptar os exemplos de sistemas lineares do caso anterior para o caso de sistemas com 2 equações e 3 variáveis.

- Por último considere-se o caso dos sistemas a 3 equações e 3 variáveis

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 = b_3 \end{cases}. \tag{1.6}$$

Geometricamente, estes sistemas representam a interseção de 3 planos no espaço ($\mathbb{R}^3$) e 8 casos distintos podem ocorrer:



Deixa-se como exercício exibir exemplos de sistemas lineares do tipo (1.6) para cada um dos 8 casos descritos acima.

Observação 6. Se o sistema for possível o número de varáveis livres do sistema, também designado por *grau de indeterminação* do sistema, indica-nos o tipo de CS que o sistema possui. Por exemplo:

- Se o número de variáveis livres for zero, o sistema é PD e o CS é um *ponto*.
- Se o número de variáveis livres for um, o sistema é PI e o CS é uma *reta*.
- Se o número de variáveis livres for dois, o sistema é PI e o CS é um *plano*.

Estaremos particularmente interessados em interpretar geometricamente o CS de sistemas com 2 e 3 variáveis, casos esses em que o CS está contido no plano e no espaço, respectivamente.

1.5 Método de eliminação de Gauss

Uma possível estratégia para resolver um sistema de equações lineares consiste em transformá-lo num sistema equivalente mais simples cuja solução seja óbvia. O *método de eliminação de Gauss* permite reduzir um sistema linear a uma forma particularmente simples, usando sequências de operações sobre as equações do sistema. Estas operações designam-se por *elementares* e são de um dos seguintes 3 tipos:

(I) Adicionar a uma equação um múltiplo de outra.

(II) Multiplicar uma equação por um escalar não nulo.

(III) Permutar duas equações.

Proposição 9. *As operações elementares transformam sistemas lineares em sistemas lineares equivalentes.*

Demonstração. O resultado é imediato para as operações elementares II e III.

Relativamente à operação elementar I basta notar que para qualquer $\lambda \in \mathbb{R}$, se têm os seguintes factos evidentes:

1. Uma solução comum a duas equações E_i e E_j do sistema é solução da equação $E_i + \lambda E_j$.
2. Uma solução comum às equações $E_i + \lambda E_j$ e E_j, é também solução da equação $E_i = (E_i + \lambda E_j) - \lambda E_j$.

$\square$

Exemplos 14.

1. Neste exemplo aplicou-se uma operação elementar do tipo I ao primeiro sistema, que consistiu em adicionar à segunda equação a primeira equação multiplicada por -1, tendo-se obtido o sistema equivalente da direita:

$$\begin{cases} 2x_1 & - & x_2 & = & -1 \\ 2x_2 & + & 3x_2 & = & 6 \end{cases} \Leftrightarrow \begin{cases} 2x_1 & - & x_2 & = & -1 \\ & & 4x_2 & = & 7 \end{cases}.$$

2. Neste exemplo, aplicou-se uma operação elementar do tipo II ao primeiro sistema, que consistiu em multiplicar a segunda equação por $\frac{1}{4}$, tendo-se obtido o sistema equivalente da direita:

$$\begin{cases} 2x_1 & - & x_2 & = & -1 \\ & & 4x_2 & = & 7 \end{cases} . \iff \begin{cases} 2x_1 & - & x_2 & = & -1 \\ & & x_2 & = & \frac{7}{4} \end{cases}$$

Definição 11. Um sistema linear diz-se em *escada* se o primeiro coeficiente não nulo de cada equação, designado por *pivot*, se encontrar à direita do primeiro coeficiente não nulo da equação anterior. As variáveis associadas a pivots designam-se por *variáveis pivot* e as restantes variáveis por *variáveis livres*.

Um sistema linear diz-se *reduzido* se estiver em escada, todos os pivots forem iguais a 1 e cada equação não incluir mais do que uma variável pivot.

Exemplo 15.

1. Exemplo de um sistema em **escada** com os pivots assinalados a vermelho. As variáveis pivot são portanto x_1, x_2 e x_4 e a única variável livre x_3.

$$\begin{cases} {\color{red}2}x_1 & + & 2x_2 & - & x_3 & - & x_4 & = & 0 \\ & & {\color{red}-2}x_2 & + & x_3 & - & x_4 & = & -6 \\ & & & & & & {\color{red}3}x_4 & = & -1 \end{cases} .$$

2. Exemplo de um sistema **reduzido** com variáveis pivot x_1, x_2 e x_3 e variáveis livres x_4 e x_5

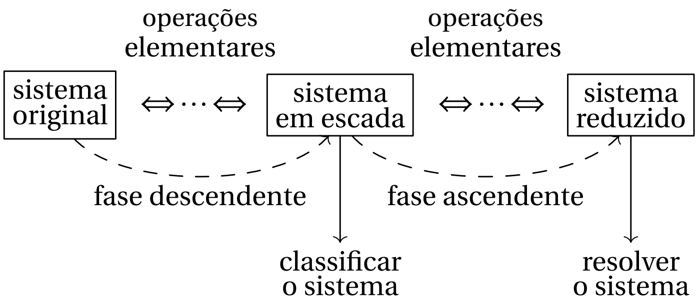
$$\begin{cases} {\color{red}x_1} & & & & & + & 4x_4 & - & 2x_5 & = & 10 \\ & & {\color{red}x_2} & & & & & + & 3x_5 & = & 0 \\ & & & & {\color{red}x_3} & - & 2x_4 & & & = & 1 \end{cases} .$$

Observe-se que é imediato obter o CS a partir de um sistema reduzido, resolvendo as equações em ordem às variáveis pivot:

$$CS = \{(x_1, x_2, x_3, x_4, x_5) : x_1 = 10 - 4x_4 + 2x_5,\ x_2 = -3x_5,\ x_3 = 1 + 2x_4,\ x_4 \in \mathbb{R},\ x_5 \in \mathbb{R}\}.$$

As variáveis pivot são as variáveis dependentes e as variáveis sem pivot correspondem às variáveis livres.

O método de eliminação de Gauss processa-se em duas fases utilizando as operações elementares sobre as equações do sistema. A primeira fase designa-se por *fase descendente* e transforma o sistema num sistema em escada equivalente. Esta fase permite classificar o sistema e determinar o tipo de CS que o sistema possui (cf. com a Observação 6). A segunda fase aplica-se aos sistemas em escada que são possíveis e designa-se por *fase ascendente*. No final desta fase o sistema está *reduzido* e o CS é conhecido. Esquematicamente,



Vejamos como se desenvolve o método de eliminação de Gauss num exemplo (a descrição detalhada deste algoritmo será feita posteriormente usando a notação matricial).

Exemplo 16. Vamos reduzir o seguinte sistema de 3 equações e 3 variáveis usando o método de eliminação de Gauss.

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

Fase descendente

- Eliminar a variável x_1 das 2ª e 3ª equações, adicionando à 2ª equação a 1ª equação multiplicada por -2 e à 3ª equação a 1ª equação multiplicada por 2:

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

- Eliminar a variável x_2 da 3ª equação, adicionando à 3ª equação a 2ª equação multiplicada por 1:

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

  O sistema em escada *não possui equações impossíveis* e *todas as variáveis são pivot*. A primeira afirmação diz-nos que o sistema é **possível** pois a existência de inconsistências no sistema original traduz-se na existência de equações impossíveis no sistema em escada. A segunda afirmação diz-nos que o sistema é **determinado** uma vez que não possui variáveis livres. Logo o CS deste sistema reduz-se a um ponto de $\mathbb{R}^3$. Para determinar as coordenadas desse ponto aplica-se a fase ascendente.

Fase ascendente

- Eliminar a variável x_3 das 1ª e 2ª equações adicionando à 1ª equação a 3ª equação (multiplicada por 1) e à 2ª equação a 3ª equação multiplicada por -2:
$$\begin{cases} x_1 + x_2 = 2 \\ -x_2 = -1 \\ x_3 = 2 \end{cases}.$$
- Multiplicar a 2ª equação por (-1) para que o pivot da variável x_2 seja 1:
$$\begin{cases} x_1 + x_2 = 2 \\ x_2 = 1 \\ x_3 = 2 \end{cases}.$$
- Eliminar a variável x_2 da 1ª equação, adicionando à 1ª equação a 2ª equação multiplicada por -1:
$$\begin{cases} x_1 = 1 \\ x_2 = 1 \\ x_3 = 2 \end{cases}.$$

O sistema anterior encontra-se reduzido, tendo-se $CS = \{(1,1,2)\}$.

Equações matriciais

Vamos agora ver como é que certo tipo de equações envolvendo matrizes são transformadas em sistemas de equações lineares.

Vejamos um exemplo.

Consideremos a matriz $A = \begin{bmatrix} 2 & -1 \\ 2 & 3 \end{bmatrix}_{2\times 2}$, o vetor $b = (0,6) = \begin{bmatrix} 0 \\ 6 \end{bmatrix} \in \mathbb{R}^2$ e o vetor de *incógnitas*, $x = (x_1, x_2) = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2$. Tem-se,

$$\begin{aligned} Ax = b \quad &\Leftrightarrow \quad \begin{bmatrix} 2 & -1 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 6 \end{bmatrix} \\ &\Leftrightarrow \quad \begin{bmatrix} 2x_1 - x_2 \\ 2x_1 + 3x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 6 \end{bmatrix} \\ &\Leftrightarrow \quad \begin{cases} 2x_1 - x_2 = 0 \\ 2x_1 + 3x_2 = 6 \end{cases} \end{aligned}$$

Daqui conclui-se que resolver a *equação matricial* $Ax = b$ equivale a resolver o sistema de equações lineares em que os coeficientes das variáveis são os elementos de A e os termos constantes as componentes de b.

A equivalência anterior estende-se a qualquer equação matricial $Ax = b$, com $A=[a_{ij}]$, do tipo $m \times n$, $b=(b_1,\ldots,b_m)\in\mathbb{R}^m$ e $x=(x_1,\ldots,x_n)\in\mathbb{R}^n$:

$$\begin{aligned}
Ax = b \quad &\Leftrightarrow \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} \\
&\Leftrightarrow \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} \\
&\Leftrightarrow \left\{ \begin{array}{l} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \qquad\qquad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{array} \right. .
\end{aligned}$$

A equação matricial $Ax = b$ é portanto equivalente ao sistema linear

$$\left\{ \begin{array}{l} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \qquad\qquad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{array} \right. , \tag{1.7}$$

em que os coeficientes das variáveis são os elementos a_{ij} de A e os termos constantes das equações as componentes b_i de b. A matriz A designa-se por *matriz dos coeficientes* do sistema (1.7) e o vetor b por *vetor dos termos constantes* ou dos *membros direitos* do sistema. A matriz $[A|b]$, com

$$[A|b] = \left[\begin{array}{cccc|c} a_{11} & a_{12} & \ldots & a_{1n} & b_1 \\ a_{21} & a_{22} & \ldots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} & b_m \end{array}\right], \tag{1.8}$$

designa-se por *matriz ampliada* do sistema (1.7) e contém toda a informação desse sistema. Verifica-se portanto que um sistema linear do tipo (1.7) pode ser representado pela equação matricial $Ax = b$ ou pela matriz ampliada $[A|b]$. A equação matricial $Ax = b$ também se designa por sistema linear. E por abuso de linguagem designaremos ainda a matriz ampliada $[A|b]$ por sistema linear.

As definições de matriz em escada e matriz reduzida são análogas às de sistema em escada e sistema reduzido, substituindo equação por linha da matriz.

Definição 12. Uma matriz diz-se em *escada* se o primeiro elemento não nulo de cada linha, que se designa por *pivot*, estiver à direita do primeiro elemento não nulo da linha anterior. Uma matriz diz-se *reduzida* se estiver em escada, todos os pivots forem iguais a 1 e em cada coluna com pivot apenas o pivot é não nulo.

Exemplos de matrizes em escada e reduzida, respectivamente, onde os pivots se encontram assinalados a vermelho,

$$\begin{bmatrix} 0 & 1 & 2 & 1 & 4 \\ 0 & 0 & 3 & -6 & 0 \\ 0 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{e} \quad \begin{bmatrix} 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & -6 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Note-se que nas colunas sem pivot podem existir elementos não nulos.

É imediato constatar que um sistema linear está em escada [reduzido] se e somente se a respectiva matriz dos coeficientes estiver em escada [reduzida].

O método de eliminação de Gauss pode ser aplicado diretamente à matriz ampliada $[A|b]$ do sistema linear (1.7), substituindo as operações elementares sobre as equações do sistema por operações análogas sobre as linhas da matriz ampliada:

	Operação elementar	**Notação**
(I)	Adicionar à linha i a linha j multiplicada por λ	$L_i + \lambda L_j \longrightarrow L_i$
(II)	Multiplicar a linha i por $\lambda \neq 0$	$\lambda L_i \longrightarrow L_i$
(III)	Permutar a linha i com a linha j	$L_i \longleftrightarrow L_j$

Exemplo 17. Vamos considerar novamente o sistema linear do Exemplo 16

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

Aplicando sobre a matriz ampliada deste sistema a mesma sequência de operações elementares que foi aplicada no Exemplo 16, obtém-se:

Fase descendente

$$[A|b] \;=\; \left[\begin{array}{ccc|c} 1 & 1 & -1 & 0 \\ 2 & 1 & 0 & 3 \\ -2 & -1 & 1 & -1 \end{array}\right] \xrightarrow[\substack{L_2-2L_1\to L_2 \\ L_3+2L_1\to L_3}]{} \left[\begin{array}{ccc|c} 1 & 1 & -1 & 0 \\ 0 & -1 & 2 & 3 \\ 0 & 1 & -1 & -1 \end{array}\right] \xrightarrow[L_3+L_2\to L_3]{} \left[\begin{array}{ccc|c} 1 & 1 & -1 & 0 \\ 0 & -1 & 2 & 3 \\ 0 & 0 & 1 & 2 \end{array}\right] = [A'|b'].$$

A matriz dos coeficientes A' encontra-se em escada e todas as linhas de $[A'|b']$ correspondem a equações possíveis. Logo o sistema $[A|b]$ é **possível**. Por outro lado todas as colunas de A' têm pivot e portanto não existem variáveis livres. Logo o sistema é **determinado** e o seu CS define um ponto em $\mathbb{R}^3$.

Fase ascendente

$$[A'|b'] = \left[\begin{array}{ccc|c} 1 & 1 & -1 & 0 \\ 0 & -1 & 2 & 3 \\ 0 & 0 & 1 & 2 \end{array}\right] \xrightarrow[\substack{L_1+L_3\to L_1 \\ L_2-2L_3\to L_2}]{} \left[\begin{array}{ccc|c} 1 & 1 & 0 & 2 \\ 0 & -1 & 0 & -1 \\ 0 & 0 & 1 & 2 \end{array}\right] \xrightarrow[(-1)L_2\to L_2]{} \left[\begin{array}{ccc|c} 1 & 1 & 0 & 2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 2 \end{array}\right] = [A''|b''] \xrightarrow[L_1+L_2\to L_1]{} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 2 \end{array}\right] = [A''|b''].$$

A matriz dos coeficientes A'' encontra-se reduzida e $[A''|b'']$ corresponde ao sistema reduzido,

$$\begin{cases} x_1 & & & = & 1 \\ & x_2 & & = & 1 \\ & & x_3 & = & 2 \end{cases}.$$

Logo, $\text{CS} = \{(1,1,2)\}$.

Vejamos como se processa o algoritmo de Gauss na presença de sistemas indeterminados e sistemas impossíveis.

Exemplo 18. Considere-se o sistema

$$\begin{cases} & & 2x_2 & - & 5x_3 & + & x_4 & = & 0 \\ 2x_1 & + & 5x_2 & - & 8x_3 & + & 2x_4 & = & 4 \\ x_1 & + & 4x_2 & - & 4x_3 & + & x_4 & = & 5 \end{cases}.$$

Trata-se de um sistema com 3 equações e 4 variáveis que representa a interseção de 3 hiperplanos em $\mathbb{R}^4$. Aplicando o método de Gauss a este sistema, obtém-se o seguinte.

Fase descendente

$$
\left[\begin{array}{cccc|c} 0 & 2 & -5 & 1 & 0 \\ 2 & 5 & -8 & 2 & 4 \\ 1 & 4 & -4 & 1 & 5 \end{array}\right] \xrightarrow[L_1 \leftrightarrow L_3]{} \left[\begin{array}{cccc|c} 1 & 4 & -4 & 1 & 5 \\ 2 & 5 & -8 & 2 & 4 \\ 0 & 2 & -5 & 1 & 0 \end{array}\right] \xrightarrow[L_2-2L_1 \to L_2]{}
$$

$$
\left[\begin{array}{cccc|c} 1 & 4 & -4 & 1 & 5 \\ 0 & -3 & 0 & 0 & -6 \\ 0 & 2 & -5 & 1 & 0 \end{array}\right] \xrightarrow[-\frac{1}{3}L_2 \to L_2]{} \left[\begin{array}{cccc|c} 1 & 4 & -4 & 1 & 5 \\ 0 & 1 & 0 & 0 & 2 \\ 0 & 2 & -5 & 1 & 0 \end{array}\right] \xrightarrow[L_3-2L_2 \Rightarrow L_3]{}
$$

$$
\left[\begin{array}{cccc|c} 1 & 4 & -4 & 1 & 5 \\ 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & -5 & 1 & -4 \end{array}\right].
$$

No final da fase descendente, a matriz dos coeficientes está em escada, com 3 pivots correspondentes às variáveis pivot x_1, x_2 e x_3 e uma coluna sem pivot correspondente à variável livre x_4. Não há equações impossíveis. Logo o sistema é PI, sendo o respetivo CS definido como interseção de 3 hiperplanos.

Fase ascendente

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

$$
\left[\begin{array}{cccc|c} 1 & 4 & 0 & \frac{1}{5} & \frac{41}{5} \\ 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 1 & -\frac{1}{5} & \frac{4}{5} \end{array}\right] \xrightarrow[L_1-4L_2 \to L_1]{} \left[\begin{array}{cccc|c} 1 & 0 & 0 & \frac{1}{5} & \frac{1}{5} \\ 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 1 & -\frac{1}{5} & \frac{4}{5} \end{array}\right].
$$

No final da fase ascendente, o sistema reduzido vem dado por

$$
\begin{cases} x_1 \qquad + \frac{1}{5}x_4 = \frac{1}{5} \\ \quad x_2 \qquad\quad = 2 \\ \qquad x_3 - \frac{1}{5}x_4 = \frac{4}{5} \end{cases} \Leftrightarrow \begin{cases} x_1 = \frac{1}{5} - \frac{1}{5}x_4 \\ x_2 = 2 \\ x_3 = \frac{4}{5} + \frac{1}{5}x_4 \end{cases},
$$

e portanto o CS vem dado por

$$
\mathrm{CS} = \left\{ (x_1, x_2, x_3, x_4) \in \mathbb{R}^4 : x_1 = \frac{1}{5} - \frac{1}{5}x_4,\ x_2 = 2,\ x_3 = \frac{4}{5} + \frac{1}{5}x_4,\ x_4 \in \mathbb{R} \right\}.
$$

Exemplo 19. Considere o sistema de 3 equações e 5 variáveis,

$$\begin{cases} 2x_1 & - & x_2 & + & 3x_3 & + & 7x_4 & + & 5x_5 & = & 5 \\ 4x_1 & + & x_2 & - & 3x_3 & + & 2x_4 & - & x_5 & = & 4 \\ 2x_1 & + & 2x_2 & - & 6x_3 & - & 5x_4 & - & 6x_5 & = & 0 \end{cases}.$$

$$\left[\begin{array}{rrrrr|r} 2 & -1 & 3 & 7 & 5 & 5 \\ 4 & 1 & -3 & 2 & -1 & 4 \\ 2 & 2 & -6 & -5 & -6 & 0 \end{array}\right] \xrightarrow[\substack{L_2-2L_1\to L_2 \\ L_3-L_1\to L_3}]{} \left[\begin{array}{rrrrr|r} 2 & -1 & 3 & 7 & 5 & 5 \\ 0 & 3 & -9 & -12 & -11 & -6 \\ 0 & 3 & -9 & -12 & -11 & -5 \end{array}\right]$$

$$\xrightarrow[L_3-L_2\to L_3]{} \left[\begin{array}{rrrrr|r} 2 & -1 & 3 & 7 & 5 & 5 \\ 0 & 3 & -9 & -12 & -11 & -6 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right].$$

No final da fase descendente, a última linha da matriz ampliada corresponde à equação impossível

$$0x_1 + 0x_2 + 0x_3 + 0x_4 + 0x_5 = 1,$$

pelo que este sistema é impossível.

Algoritmo 1. MÉTODO DE ELIMINAÇÃO DE GAUSS

`Input:` *Matriz ampliada* $[A|b]$ *de um sistema linear* $Ax = b$

`Objectivo:` *Redução do sistema* $Ax = b$

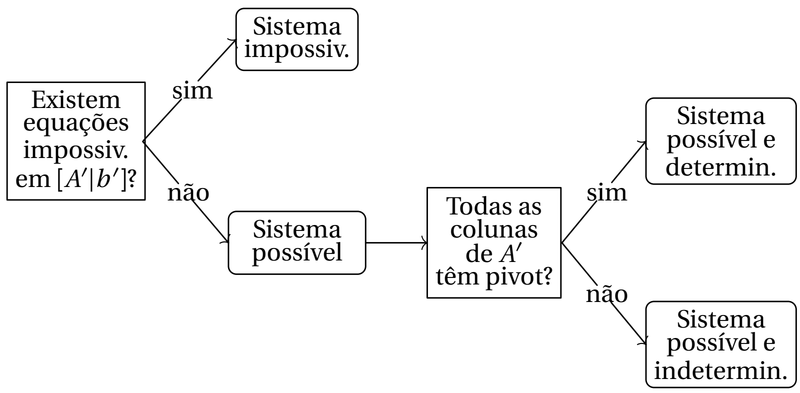
Fase descendente

1) *Se necessário, efectuar trocas de linhas em* $[A|b]$ *de modo a que o pivot da primeira linha se encontre na coluna não nula mais à esquerda da matriz dos coeficientes.*

2) *Utilizar o pivot da primeira linha para eliminar os restantes elementos da coluna desse pivot.*

3) *Repetir os procedimentos anteriores relativamente à submatriz que se obtém ignorando a primeira linha e assim sucessivamente enquanto existirem linhas não nulas na matriz dos coeficientes dessa submatriz.*

No final da fase descendente obtém-se a matriz $[A'|b']$ com A' em escada e tem-se



Note-se que a matriz $[A'|b']$ **não é única**, isto é, depende da sequência de operações elementares que foi efetuada.

Fase ascendente (apenas se aplica aos sistemas possíveis)

4) *Usando o pivot que se encontra mais à direita na matriz* A' *eliminar os restantes elementos da coluna desse pivot e fazer o pivot igual 1.*

5) *Repetir os procedimentos do passo anterior relativamente à coluna com pivot imediatamente anterior e assim sucessivamente enquanto existirem colunas com pivot (percorrendo a matriz da direita para a esquerda).*

No final da fase ascendente obtém-se a matriz $[A''|b'']$ com A'' reduzida. A matriz $[A''|b'']$ é **única**, isto é, não depende da sequência de operações elementares que foi efectuada.

Definição 13. Chama-se *característica* de uma matriz A e denota-se por $\mathrm{car}(A)$, ao número de pivots de qualquer matriz em escada que se obtenha a partir de A por aplicação do método de eliminação de Gauss.

Note-se que a característica de uma matriz A está bem definida uma vez que coincide com o número de pivots da matriz reduzida (a fase ascendente do método de Gauss não altera o número de pivots).

A característica de A pode equivalentemente ser definida como o número de linhas não nulas de qualquer matriz em escada que se obtenha a partir de A. Em particular, conclui-se que se A é do tipo $m \times n$, $\mathrm{car}(A) \leq \min\{m, n\}$.

Sistemas homogéneos

Uma família importante de sistemas lineares é dada pelos sistemas cujos termos constantes são todos nulos, isto é pelos sistemas lineares da forma $Ax = \vec{0}$. Estes sistemas designam-se por *sistemas homogéneos* e possuem sempre a solução *trivial* $x = \vec{0}$ e, portanto, **nunca são impossíveis**.

Geometricamente um sistema linear homogéneo com m equações e n variáveis corresponde à intersecção de m retas se $n = 2$, m planos se $n = 3$ ou m hiperplanos se $n \geq 4$, que passam na origem de $\mathbb{R}^n$.

Note-se que para determinar o conjunto de soluções de um sistema homogéneo, basta aplicar o método de eliminação de Gauss à matriz dos coeficientes desse sistema, uma vez que o vetor dos termos constantes se mantém sempre nulo independentemente das operações elementares que forem efetuadas.

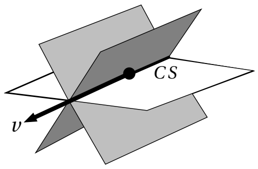
Exemplo 20. Vai-se aplicar o método de eliminação de Gauss ao seguinte sistema linear homogéneo, que representa a intersecção de 3 planos de $\mathbb{R}^3$ que passam na origem,

$$\begin{cases} x_1 - x_2 + x_3 = 0 \\ \phantom{x_1 - {}} x_2 + 2x_3 = 0 \\ x_1 \phantom{{}- x_2} + 3x_3 = 0 \end{cases} \quad \Leftrightarrow \quad \begin{bmatrix} 1 & -1 & 1 \\ 0 & 1 & 2 \\ 1 & 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Fase descendente

$$\begin{bmatrix} 1 & -1 & 1 \\ 0 & 1 & 2 \\ 1 & 0 & 3 \end{bmatrix} \xrightarrow[L_3 - L_1 \to L_3]{} \begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & 2 \\ 0 & 1 & 2 \end{bmatrix} \xrightarrow[L_3 - L_2 \to L_3]{} \begin{bmatrix} \color{red}{1} & -1 & 1 \\ 0 & \color{red}{1} & 2 \\ 0 & 0 & 0 \end{bmatrix}.$$

O sistema em escada possui uma variável livre (x_3), pelo que é indeterminado, e o seu CS define uma reta em $\mathbb{R}^3$ que passa na origem. Esta reta corresponde à intersecção de 3 planos de $\mathbb{R}^3$ concorrentes 2 a 2 entre si, uma vez que os respectivos vetores normais, $n_1 = (1, -1, 1)$, $n_2 = (0, 1, 2)$ e $n_3 = (1, 0, 3)$, são 2 a 2 não colineares:



Para determinar o vetor diretor da reta de interseção aplica-se a fase ascendente do método de Gauss.

Fase ascendente

$$\begin{bmatrix} 1 & -1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix} \xrightarrow[L_1+L_2\to L_1]{} \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix} \to \begin{cases} x_1 & & + & 3x_3 & = & 0 \\ & x_2 & + & 2x_3 & = & 0\,. \\ & & & 0 & = & 0 \end{cases}$$

Considerando o sistema reduzido e resolvendo em ordem às variáveis pivot x_1 e x_2, obtém-se

$$\begin{aligned} \text{CS} &= \{(x_1,x_2,x_3) : x_1 = -3x_3,\ x_2 = -2x_3,\ x_3 \in \mathbb{R}\} \\ &= \{(-3x_3,-2x_3,x_3) : x_3 \in \mathbb{R}\} \\ &= \{x_3(-3,-2,1) : x_3 \in \mathbb{R}\}\,. \end{aligned}$$

Logo, a reta de interseção tem vetor director $v = (-3,-2,1)$.

Observação 7.

Se u, v forem soluções de um sistema homogéneo $Ax = \vec{0}$, então $u + v$ e λu são também soluções do mesmo sistema para todo o $\lambda \in \mathbb{R}$ (a verificação deste resultado fica como exercício). Diz-se então que o CS é *fechado* para a adição de vetores e para a multiplicação por escalares. Os subconjuntos de $\mathbb{R}^n$ que gozam desta propriedade constituem o objecto fundamental de estudo da Álgebra Linear e serão considerados em detalhe nos capítulos subsequentes.

1.6 Inversa de uma matriz

Definição 14. Uma matriz quadrada A diz-se *invertível* ou *não singular* se existir uma matriz B (da mesma ordem) tal que

$$AB = I \qquad \text{e} \qquad BA = I,$$

onde I denota a matriz identidade (da mesma ordem de A). A matriz B quando existe designa-se por *inversa* de A. Caso contrário, A diz-se *singular*.

Proposição 10. *A inversa de uma matriz A quando existe, é única.*

Demonstração. Sejam B e C inversas de A, isto é, matrizes quadradas da mesma ordem de A tais $AB = BA = I$ e $AC = CA = I$. Quer-se provar que $B = C$. De facto,

$$AC = I \Rightarrow B(AC) = BI \Rightarrow (BA)C = B \Rightarrow IC = B \Rightarrow C = B.$$

□

Se A é uma matriz invertível denota-se a respectiva inversa por A^{-1}.

Observação 8.

1. Pode-se mostrar que se $AB = I$ então $BA = I$ (e reciprocamente). Logo para averiguar se uma matriz é invertível basta verificar se existe uma matriz B tal que $AB = I$, isto é, basta averiguar se a equação matricial $AX = I$ é possível, em que X designa uma matriz de incógnitas da mesma ordem.

2. Do ponto anterior conclui-se que se A e B são matrizes da mesma ordem tais que $AB = I$, A e B são invertíveis e tem-se $A^{-1} = B$ e $B^{-1} = A$.

Exemplos 21.

1. A matriz identidade de ordem n é invertível e tem-se $I_n^{-1} = I_n$, isto é, a matriz identidade é inversa dela própria. De facto, $I_n I_n = I_n$.

2. A matriz $\begin{bmatrix} 5 & 4 & 6 \\ 4 & 3 & 4 \\ 6 & 4 & 5 \end{bmatrix}$ e invertível e tem-se $\begin{bmatrix} 5 & 4 & 6 \\ 4 & 3 & 4 \\ 6 & 4 & 5 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -4 & 2 \\ -4 & 11 & -4 \\ 2 & -4 & 1 \end{bmatrix}$.

   De facto,

$$\begin{aligned}
\begin{bmatrix} 5 & 4 & 6 \\ 4 & 3 & 4 \\ 6 & 4 & 5 \end{bmatrix} \begin{bmatrix} 1 & -4 & 2 \\ -4 & 11 & -4 \\ 2 & -4 & 1 \end{bmatrix} &= \begin{bmatrix} 5-16+12 & -20+44-24 & 10-16+6 \\ 4-12+8 & -16+33-16 & 8-12+4 \\ 6-16+10 & -24+44-20 & 12-16+5 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I_3.
\end{aligned}$$

3. A matriz nula de ordem n, $[0]_n$, é singular uma vez que $[0]_n B = [0]_n$ qualquer que seja a matriz quadrada B da mesma ordem.

Exemplo 22. Considere-se a matriz diagonal $A = \text{diag}(2,3) = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$. Pretende-se verificar se A é invertível, e em caso afirmativo, determinar a sua inversa. Para isso, vamos considerar a equação matricial $AX = I$, em que $X = [\,x \;\; y\,] =$

$\begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix}$. Tem-se

$$\begin{aligned}
AX = I \quad &\Leftrightarrow \quad \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\
&\Leftrightarrow \quad \begin{bmatrix} 2x_1 + 0x_2 & 2y_1 + 0y_2 \\ 0x_1 + 3x_2 & 0y1 + 3y_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\
&\Leftrightarrow \quad \begin{cases} 2x_1 = 1 \\ 3x_2 = 0 \\ 2y_1 = 0 \\ 3y_2 = 1 \end{cases} \quad \Leftrightarrow \quad \begin{cases} x_1 = \frac{1}{2} \\ x_2 = 0 \\ y_1 = 0 \\ y_2 = \frac{1}{3} \end{cases} \\
&\Leftrightarrow \quad X = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{3} \end{bmatrix} = \operatorname{diag}\left(\frac{1}{2}, \frac{1}{3}\right).
\end{aligned}$$

Logo A é invertível e tem-se $A^{-1} = \operatorname{diag}\left(\frac{1}{2}, \frac{1}{3}\right)$.

Exemplo 23. Em geral, uma matriz diagonal $A = \operatorname{diag}(a_1, \ldots, a_n)$ é invertível sse todos os elementos da diagonal principal são não nulos, isto é, $a_1, \ldots, a_n \neq 0$, e nessa altura tem-se

$$A^{-1} = \begin{bmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_n \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{a_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{a_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{a_n} \end{bmatrix},$$

isto é,

$$\operatorname{diag}^{-1}(a_1, \ldots, a_n) = \operatorname{diag}(a_1^{-1}, a_2^{-1}, \ldots, a_n^{-1}).$$

Vejamos algumas propriedades da matriz inversa.

Proposição 11. *Sejam A, B matrizes invertíveis da mesma ordem. Tem-se*

1. $(A^{-1})^{-1} = A$.
2. *AB é invertível e tem-se* $(AB)^{-1} = B^{-1}A^{-1}$.
3. *A^T é invertível e tem-se* $(A^T)^{-1} = (A^{-1})^T$.
4. *Se $k \in \mathbb{N}$ tem-se* $(A^k)^{-1} = (A^{-1})^k$.
5. *Se $\lambda \in \mathbb{R}$ com $\lambda \neq 0$, (λA) é invertível tendo-se* $(\lambda A)^{-1} = \lambda^{-1}A^{-1}$.

A demonstração desta proposição fica como exercício.

Note-se que se $A_1, A_2, \ldots, A_k$ são matrizes invertíveis da mesma ordem então $A_1 A_2 \cdots A_k$ é também invertível e tem-se

$$(A_1 A_2 \cdots A_k)^{-1} = A_k^{-1} \cdots A_2^{-1} A_1^{-1},$$

isto é, a inversa do produto de matrizes (invertíveis) é produto das inversas pela ordem inversa.

Se na relação anterior $A = A_1 = \cdots = A_k$ tem-se $(A^k)^{-1} = (A^{-1})^k$ o que motiva a seguinte definição.

Definição 15. Se A for invertível, definem-se as *potências inteiras negativas* de A por

$$A^{-k} = (A^{-1})^k \quad (k \in \mathbb{N}).$$

Sistemas de equações lineares e matriz inversa

A equação linear $ax = b$, com $a \neq 0$, tem solução (única) $x = \dfrac{b}{a} = a^{-1}b$ qualquer que seja $b \in \mathbb{R}$.

A tradução dos sistemas lineares em equações matriciais do tipo $Ax = b$, permite concluir um resultado análogo quando a matriz de coeficientes A é invertível. Mais precisamente, tem-se o seguinte.

Proposição 12. *Seja A uma matriz invertível de ordem n. Tem-se:*

(i) *$Ax = b$ tem solução (única) $x = A^{-1}b$, qualquer que seja $b \in \mathbb{R}^n$.*

(ii) *$Ax = \vec{0}$ é determinado, isto é, possui apenas a solução trivial, $x = \vec{0}$.*

(iii) *$\mathrm{car}(A) = n$.*

Prova. (i) Multiplicando $Ax = b$ à esquerda pela inversa de A, tem-se

$$\begin{aligned} Ax = b \quad &\Leftrightarrow \quad A^{-1}(Ax) = A^{-1}b \Leftrightarrow (AA^{-1})x = A^{-1}b \\ &\Leftrightarrow \quad I_n x = A^{-1}b \Leftrightarrow x = A^{-1}b. \end{aligned}$$

(ii) Imediata por (i), considerando $b = \vec{0}$.

(iii) Por (ii), $Ax = \vec{0}$ é determinado e portanto a matriz em escada que se obtém a partir de A tem n pivots, isto é, $\mathrm{car}(A) = n$. $\square$

Exemplo 24. Consideremos o sistema linear $Ax = b$ com matriz de coeficientes $A = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 1 & 1 \\ 2 & 1 & 1 \end{bmatrix}$, vetor de constantes $b = (b_1, b_2, b_3) = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \in \mathbb{R}^3$ e vetor de incógnitas $x = (x_1, x_2, x_3) = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \in \mathbb{R}^3$. A matriz de coeficientes tem inversa

$A^{-1} = \begin{bmatrix} 0 & -1 & 1 \\ -1 & 3 & -1 \\ 1 & -1 & 0 \end{bmatrix}$ (verifique!) e, portanto,

$$Ax = b \Leftrightarrow A^{-1}(Ax) = A^{-1}b \Leftrightarrow x = A^{-1}b = \begin{bmatrix} 0 & -1 & 1 \\ -1 & 3 & -1 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} b_3 - b_2 \\ 3b_2 - b_1 - b_3 \\ b_1 - b_2 \end{bmatrix}.$$

Daqui resulta que o sistema $Ax = b$ é PD qualquer que seja $b = (b_1, b_2, b_3) \in \mathbb{R}^3$, com solução única dada por $x = (b_3 - b_2, 3b_2 - b_1 - b_3, b_1 - b_2)$.

Algoritmo da inversa

Pretende-se deduzir um algoritmo para decidir se uma matriz A é invertível e, em caso afirmativo, calcular a sua inversa. Aplicando o método de Gauss à equação matricial $AX = I$ tem-se pela unicidade da inversa, que esta equação matricial é PD se A for invertível ou IMP se A for singular.

Exemplo 25. Consideremos as matrizes,

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}, \quad X = \begin{bmatrix} x & y \end{bmatrix} = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix} \quad \text{e} \quad I_2 = [\, e_1 \;\; e_2 \,] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Pela Observação 3, tem-se

$$\begin{aligned} AX = I_2 \quad &\Leftrightarrow \quad A\begin{bmatrix} x & y \end{bmatrix} = [\, e_1 \;\; e_2 \,] \\ &\Leftrightarrow \quad \begin{bmatrix} Ax & Ay \end{bmatrix} = [\, e_1 \;\; e_2 \,] \\ &\Leftrightarrow \quad \begin{cases} Ax = e_1 \\ Ay = e_2 \end{cases} \end{aligned}$$

Logo, a solução da equação matricial $AX = I_2$ (caso exista), vem dada pela matriz cujas colunas são as soluções dos sistemas lineares

$$Ax = e_1 \quad \text{e} \quad Ay = e_2,$$

cujas matrizes ampliadas são, respectivamente,

$$[A \,|\, e_1] \quad \text{e} \quad [A \,|\, e_2].$$

Uma vez que os dois sistemas possuem a <u>mesma matriz de coeficientes</u> A podemos aplicar o método de Gauss para reduzir <u>simultaneamente</u> os dois sistemas, ampliando A com os vetores e_1 e e_2, isto é, com a matriz identidade I_2:

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

$$\left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & -1 \end{array}\right] \xrightarrow[L_1 - 2L_2 \to L_1]{} \left[\begin{array}{cc|cc} 1 & 0 & -3 & 2 \\ 0 & 1 & 2 & -1 \end{array}\right] \longrightarrow \begin{cases} \left[\begin{array}{cc|c} 1 & 0 & -3 \\ 0 & 1 & 2 \end{array}\right] \\ \left[\begin{array}{cc|c} 1 & 0 & 2 \\ 0 & 1 & -1 \end{array}\right] \end{cases}.$$

Os sistemas lineares $Ax = e_1$ e $Ay = e_2$ são ambos PD tendo como soluções $x = (-3, 2)$ e $y = (2, -1)$, respectivamente. Daqui resulta que a equação matricial $AX = I_2$ é PD tendo como (única) solução $X = \begin{bmatrix} x & y \end{bmatrix} = \begin{bmatrix} -3 & 2 \\ 2 & -1 \end{bmatrix}$.

Portanto, A é invertível, sendo $A^{-1} = \begin{bmatrix} -3 & 2 \\ 2 & -1 \end{bmatrix}$.

O procedimento descrito no exemplo anterior pode ser estendido para uma matriz arbitrária A de ordem n, ampliando A com a matriz identidade $I_n = [\, e_1 \;\; e_2 \;\; \cdots \;\; e_n \,]$.

Se A for invertível, os n sistemas $[A\,|\,e_1], \ldots, [A\,|\,e_n]$ são todos PD e a matriz em escada A' que se obtém a partir de A tem n pivots, ou seja, $\text{car}(A) = n$. Nessa altura, as soluções destes n sistemas constituem as n colunas da matriz A^{-1}.

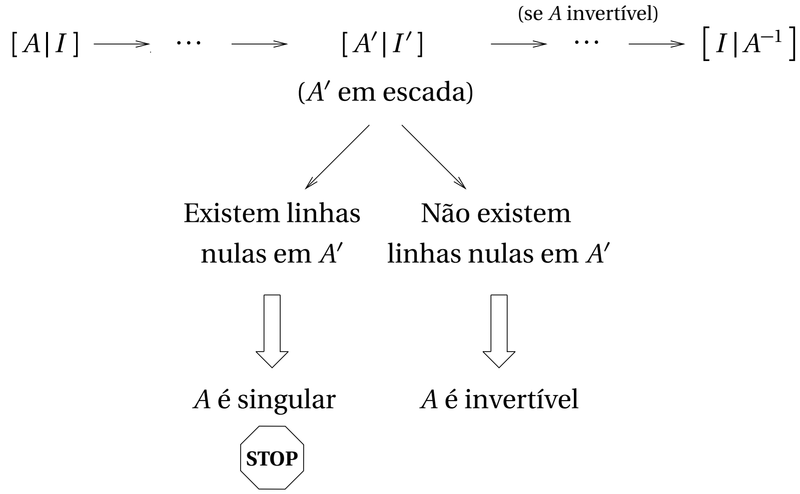
Se A for singular, existem colunas sem pivot em A', ou seja, existem linhas nulas em A' (uma vez que o número de colunas e linhas de A' são iguais) e pelo menos um dos sistemas $[A\,|\,e_i]$ é impossível (caso contrário, os sistemas $[A\,|\,e_i]$ seriam todos indeterminados, pelo que a equação matricial $AX = I_n$ também).

A partir da discussão anterior deduz-se então o seguinte algoritmo para determinar a inversa de uma matriz A (caso exista), aplicando o método de eliminação de Gauss à matriz ampliada $[\,A\,|\,I\,]$.

Algoritmo 2. ALGORITMO DA INVERSA

`Input:` *Matriz quadrada A*

`Objectivo:` *Decidir sobre a invertibilidade de A e calcular* A^{-1}



Da discussão anterior deduz-se também o seguinte importante critério para decidir sobre a invertibilidade de uma matriz.

Proposição 13. *Uma matriz quadrada A de ordem n tem inversa se e só se $car(A) = n$.*

Exemplo 26. Determinar a inversa da matriz $A = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 1 & 1 \\ 2 & 1 & 1 \end{bmatrix}$ (caso exista).

Fase descendente

$$[A\,|\,I] = \left[\begin{array}{ccc|ccc} 1 & 1 & 2 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 2 & 1 & 1 & 0 & 0 & 1 \end{array}\right] \xrightarrow[L_3-2L_1\to L_3]{L_2-L_1\to L_2} \left[\begin{array}{ccc|ccc} 1 & 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & -1 & -1 & 1 & 0 \\ 0 & -1 & -3 & -2 & 0 & 1 \end{array}\right]$$

$$\xrightarrow[L_2\leftrightarrow L_3]{} \left[\begin{array}{ccc|ccc} \color{red}{1} & 1 & 2 & 1 & 0 & 0 \\ 0 & \color{red}{-1} & -3 & -2 & 0 & 1 \\ 0 & 0 & \color{red}{-1} & -1 & 1 & 0 \end{array}\right] = \left[A'\,|\,I'\right].$$

Note-se que no final da fase descendente a matriz em escada A' não tem linhas nulas e possui 3 pivots (assinalados a vermelho), pelo que A é invertível. Para determinar a inversa de A aplica-se a fase ascendente.

Fase ascendente

$$\left[A'\,|\,I'\right] = \left[\begin{array}{ccc|ccc} \color{red}{1} & 1 & 2 & 1 & 0 & 0 \\ 0 & \color{red}{-1} & -3 & -2 & 0 & 1 \\ 0 & 0 & \color{red}{-1} & -1 & 1 & 0 \end{array}\right] \xrightarrow[-L_3\to L_3]{} \left[\begin{array}{ccc|ccc} 1 & 1 & 2 & 1 & 0 & 0 \\ 0 & -1 & -3 & -2 & 0 & 1 \\ 0 & 0 & 1 & 1 & -1 & 0 \end{array}\right]$$

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

$$\xrightarrow[L_1-L_2\to L_1]{} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & -1 & 1 \\ 0 & 1 & 0 & -1 & 3 & -1 \\ 0 & 0 & 1 & 1 & -1 & 0 \end{array}\right] = \left[I\,|\,A^{-1}\right].$$

Portanto, A é invertível, tendo-se $A^{-1} = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 1 & 1 \\ 2 & 1 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 0 & -1 & 1 \\ -1 & 3 & -1 \\ 1 & -1 & 0 \end{bmatrix}$.

Exemplo 27. Pretende-se decidir se a matriz $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$ é invertível.

Uma vez que não se pretende determinar a inversa de A não é necessário ampliar com a matriz identidade.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \xrightarrow[L_3-7L_1\to L_3]{L_2-4L_1\to L_2} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & -6 & -12 \end{bmatrix} \xrightarrow[L_3-2L_2\to L_3]{} \begin{bmatrix} \color{red}{1} & 2 & 3 \\ 0 & \color{red}{-3} & -6 \\ 0 & 0 & 0 \end{bmatrix} = A'.$$

Como A' tem 2 pivots, a $\text{car}(A) = 2$ é inferior ao número de colunas de A e, portanto, A é singular.

Anexo - Propriedades das operações com matrizes

SOMA DE MATRIZES

$A.\ B,\ C$ do mesmo tipo

1. Comutativa : $A+B=B+A$
2. Associativa : $(A+B)+C=A+(B+C)$
3. Existência de el. neutro : $A+[0]=A$
4. Existência de el. simétrico : $A+(-A)=[0]$

MULTIPLICAÇÃO POR ESCALAR

$A,\ B$ do mesmo tipo

$\lambda, \mu \in \mathbb{R}$

1. Distributivas : $\lambda(A+B)=\lambda A+\lambda B$
   $(\lambda+\mu)A=\lambda A+\mu A$
2. Associativa : $\lambda(\mu A)=(\lambda\mu)A$
3. $1\cdot A=A$

PRODUTO DE MATRIZES

$A,\ B,\ C$ de tipo conveniente

$\lambda, \mu \in \mathbb{R}$

1. Associativas : $(AB)C=A(BC)$
   $A(\lambda B)=(\lambda A)B=\lambda(AB)$
2. Distributivas : $(A+B)C=AC+BC$
   $C(A+B)=CA+CB$
3. Existência de el neutro : $AI=A$ e $IA=A$

4. Não comutativo : em geral $AB\neq BA$
5. Não é válida a lei do anulamento do produto : $AB=0 \not\Rightarrow (A=0$ ou $B=0)$
6. Não é válida a lei do corte : $(AB=AC$ com $A\neq 0) \not\Rightarrow B=C$

TRANSPOSTA

$A,\ B$ de tipo conveniente

$\lambda \in \mathbb{R}$

1. $(A^T)^T=A$
2. $(A+B)^T=A^T+B^T$
3. $(AB)^T=B^TA^T$
4. $(\lambda A)^T=\lambda A^T$

INVERSA

$A,\ B$ invertíveis da mesma ordem

$k\in\mathbb{N}$

$\lambda\in\mathbb{R},\ \lambda\neq 0$

1. A inversa, quando existe, é única
2. $(A^{-1})^{-1}=A$
3. $(AB)^{-1}=B^{-1}A^{-1}$
4. $(A^T)^{-1}=(A^{-1})^T$
5. $(A^k)^{-1}=(A^{-1})^k$
6. $(\lambda A)^{-1}=\lambda^{-1}A^{-1}$

Capítulo 2

Espaços vetoriais

2.1 Subespaços vetoriais de $\mathbb{R}^n$

Um *espaço vetorial*[1] (real) é um conjunto $E \neq \varnothing$ munido de uma soma $x, y \in E \mapsto x + y \in E$, com as propriedades,

1. $\forall x, y \in E \;\; x + y = y + x$
2. $\forall x, y, z \in E \;\; (x + y) + z = x + (y + z)$
3. $\exists \vec{0} \in E\colon \; \forall x \in E \;\; x + \vec{0} = x$
4. $\forall x \in E \; \exists (-x) \in E\colon \; x + (-x) = \vec{0}$

e de um produto por escalar, $\alpha \in \mathbb{R}$, $x \in E \mapsto \alpha x \in E$, com as propriedades

5. $\forall x, y \in E,\; \forall \alpha \in \mathbb{R} \;\; \alpha(x + y) = \alpha y + \alpha x$
6. $\forall x \in E,\; \forall \alpha, \beta \in \mathbb{R} \;\; (\alpha + \beta)x = \alpha x + \beta x$
7. $\forall x \in E,\; \forall \alpha, \beta \in \mathbb{R} \;\; (\alpha\beta)x = \alpha(\beta x)$
8. $\forall x \in E \;\; 1x = x$

O conjunto $\mathbb{R}^n = \{(x_1, \ldots, x_n)\colon x_1, \ldots, x_n \in \mathbb{R}\}$ é um espaço vetorial com as operações usuais da adição de vetores e da multiplicação por escalares, isto é, verifica as propriedades (1), ..., (8), como vimos na Proposição 1.

Neste capítulo vamos estudar os subconjuntos $V \subset \mathbb{R}^n$ para os quais se podem adicionar vetores de V e multiplicar vetores de V por escalares, de modo a obter ainda vetores de V. Nessa altura, pode-se mostrar que são ainda verificadas as propriedades (1), ..., (8) relativamente aos vetores de V e portanto que V define um espaço vetorial com as operações da adição de vetores e da multiplicação de vetores por escalares.

[1] Também designado por *espaço linear*.

Definição 16. Diz-se que $V \subset \mathbb{R}^n$ é um *subespaço vetorial* de $\mathbb{R}^n$ se verificar as seguintes condições:

(i) $V \neq \emptyset$

(ii) V é *fechado para a adição*, isto é, para todo o $u, v \in V$ tem-se $u+v \in V$

(iii) V é *fechado para o produto por escalar*, isto é, para todo o $u \in V$ e $\alpha \in \mathbb{R}$ tem-se $\alpha u \in V$.

Vamos analisar estas condições nalguns exemplos de subconjuntos $V \subset \mathbb{R}^n$ e verificar se estes conjuntos definem subespaços vetoriais.

Exemplos 28.

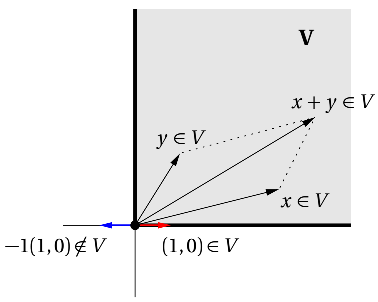
1. $V = \{(x_1, x_2) \in \mathbb{R}^2 \colon x_1, x_2 \geq 0\}$ (1^{o} quadrante incluindo eixos coordenados):

   (i) $V \neq \emptyset$ (por exemplo, $(0,0) \in V$).

   (ii) V **é fechado para a adição**, isto é, a soma de vetores do 1^{o} quadrante ainda é um vetor do 1^{o} quadrante. De facto, dados vetores de V, $x = (x_1, x_2)$ e $y = (y_1, y_2)$, tem-se que $x_1, x_2 \geq 0$ e $y_1, y_2 \geq 0$. Logo, $x_1 + x_2 \geq 0$ e $y_1 + y_2 \geq 0$ e, portanto, $x + y = (x_1 + y_1, x_2 + y_2) \in V$.

   (iii) V **não é fechado para o produto por escalar**. De facto, considerando $x = (1,0) \in V$ e $\alpha = -1 \in \mathbb{R}$, obtém-se $\alpha x = (-1,0) \notin V$.

   Portanto, V não é subespaço vetorial de $\mathbb{R}^2$.



2. $V = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \colon x_1 = x_2 = 0\} = \{(0,0,x_3) \colon x_3 \in \mathbb{R}\}$ (eixo da terceira coordenada):

   (i) $V \neq \emptyset$ ($\vec{0} \in V$).

   (ii) V **é fechado para a adição**. De facto, $\forall x, y \in V$, tem-se $x = (0,0,x_3)$ e $y = (0,0,y_3)$ para algum $x_3, y_3 \in \mathbb{R}$ e, portanto, $x + y = (0,0,x_3 + y_3)$ que ainda pertence a V uma vez que as duas primeiras componentes são nulas.

   (iii) V **é fechado para o produto por escalar**. Tem-se de modo análogo, que se $\alpha \in \mathbb{R}$ e $x = (0,0,x_3) \in V$, então $\alpha x = \alpha(0,0,x_3) = (0,0,\alpha x_3) \in V$.

Logo V define um subespaço vetorial de $\mathbb{R}^3$.

3. $V = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 = 1\}$ (reta vertical de abcissa 1):

   (i) $V \neq \emptyset$ ($(1,0) \in V$).

   (ii) Considerando os vetores de V, $(1,0)$ e $(1,1)$, tem-se $(1,0)+(1,1) = (2,1) \notin V$ e, portanto, V não é fechado para a adição.

   Logo V não é subespaço vetorial de $\mathbb{R}^2$.

Observação 9. As condições (i), (ii) e (iii) da definição de subespaço vetorial são trivialmente verificadas para os seguintes subconjuntos de $\mathbb{R}^n$:

- $V = \{\vec{0}\}$ subespaço vetorial *minimal* ou *trivial*.
- $V = \mathbb{R}^n$ subespaço vetorial *maximal*.

O seguinte resultado dá-nos uma condição necessária para que um subconjunto de $\mathbb{R}^n$ possa definir um subespaço vetorial.

Proposição 14. *Se V é subespaço vetorial de $\mathbb{R}^n$ então $\vec{0} \in V$.*

Demonstração. Como $V \neq \emptyset$, existe um elemento $v \in V$. Como V é fechado para o produto por escalar, $\alpha v \in V$ para qualquer $\alpha \in \mathbb{R}$. Em particular, considerando $\alpha = 0$ tem-se $0v = \vec{0} \in V$. □

Considere-se novamente o Exemplo 28.3. Verifica-se imediatamente que a a reta vertical de abcissa 1 não contém a origem e portanto este conjunto não é um subespaço vetorial de $\mathbb{R}^2$.

Espaço nulo de uma matriz

Sejam A matriz do tipo $m \times n$, $b \in \mathbb{R}^m$ e $V = \{x \in \mathbb{R}^n : Ax = b\}$ o conjunto de soluções do sistema linear $Ax = b$.

Se $b \neq \vec{0}$, $Ax = b$ é um sistema não homogéneo e o vetor $x = \vec{0}$ não é solução deste sistema (justifique!) e portanto $\vec{0} \notin V$. Logo, pela proposição anterior, V **não define um subespaço vetorial de** $\mathbb{R}^n$.

Se $b = \vec{0}$, tem-se o seguinte resultado.

Proposição 15. *Se A é uma matriz do tipo $m \times n$, o conjunto de soluções do sistema linear homogéneo $Ax = \vec{0}$, $V = \{x \in \mathbb{R}^n : Ax = \vec{0}\}$, define um subespaço vetorial de $\mathbb{R}^n$.*

Demonstração. Temos que verificar as 3 condições da definição de subespaço vetorial (Definição 16):

(i) $V \neq \emptyset$ pois $\vec{0} \in V$.

(ii) Se u, v são soluções de $Ax = \vec{0}$, então $Au = \vec{0}$ e $Av = \vec{0}$. Logo $A(u+v) = Au + Av = \vec{0} + \vec{0} = \vec{0}$ e portanto, $u + v$ é também solução de $Ax = \vec{0}$. Por conseguinte, V é fechado para a adição.

(iii) Seja $\alpha \in \mathbb{R}$. Se u é solução de $Ax = \vec{0}$ então $Au = \vec{0}$ e por conseguinte $A(\alpha u) = \alpha(Au) = \alpha\vec{0} = \vec{0}$. Portanto, αu é também solução de $Ax = \vec{0}$. Logo V é fechado para o produto por escalar.

De (i),(ii) e (iii) concluímos que V é um subespaço vetorial de $\mathbb{R}^n$. □

Definição 17. Seja A uma matriz do tipo $m \times n$. Chama-se *espaço nulo de A*, e denota-se por $\mathcal{N}(A)$, ao subespaço vetorial de $\mathbb{R}^n$ definido pelas soluções do sistema linear homogéneo $Ax = \vec{0}$, ou seja,

$$\mathcal{N}(A) = \left\{ x \in \mathbb{R}^n : Ax = \vec{0} \right\}.$$

Exemplo 29. Sejam $A = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 3 & 0 \end{bmatrix}$ e $v = (-3, 1, 1)$. Vejamos que $v \in \mathcal{N}(A)$.

Ora, mostrar que $v \in \mathcal{N}(A)$ equivale a mostrar que $Av = \vec{0}$. De facto, tem-se

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

Alternativamente, pode-se provar que $v = (-3, 1, 1) \in \mathcal{N}(A)$, verificando que $x_1 = -3$, $x_2 = 1$ e $x_3 = 1$ é solução do sistema linear homogéneo com matriz de coeficientes A,

$$\begin{cases} x_1 + 2x_2 + x_3 &= 0 \\ x_1 + 3x_2 &= 0 \end{cases}.$$

Exemplos 30.

Vamos calcular o espaço nulo para alguns exemplos de matrizes.

1. Se $A = \begin{bmatrix} 1 & -1 \end{bmatrix}$ então

$$\begin{aligned} \mathcal{N}(A) &= \left\{ x \in \mathbb{R}^2 : Ax = \vec{0} \right\} = \left\{ (x_1, x_2) \in \mathbb{R}^2 : \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0 \right\} \\ &= \left\{ (x_1, x_2) \in \mathbb{R}^2 : x_1 - x_2 = 0 \right\} = \left\{ (x_1, x_2) \in \mathbb{R}^2 : x_1 = x_2,\ x_2 \in \mathbb{R} \right\}, \end{aligned}$$

isto é, $\mathcal{N}(A)$ define a bissectriz dos quadrantes ímpares.

2. Se $A = \begin{bmatrix} 1 & 2 & -1 \\ 2 & 5 & -1 \end{bmatrix}$ então

$$\mathcal{N}(A) = \left\{ x \in \mathbb{R}^3 : Ax = \vec{0} \right\} = \left\{ (x_1, x_2, x_3) \in \mathbb{R}^3 : \begin{bmatrix} 1 & 2 & -1 \\ 2 & 5 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\}.$$

Resolvendo o sistema homogéneo $Ax = \vec{0}$ obtém-se,

$$\begin{bmatrix} 1 & 2 & -1 \\ 2 & 5 & -1 \end{bmatrix} \xrightarrow[L_2-2L_1\to L_2]{} \begin{bmatrix} 1 & 2 & -1 \\ 0 & 1 & 1 \end{bmatrix} \xrightarrow[L_1-2L_2\to L_1]{} \begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & 1 \end{bmatrix}.$$

A última matriz corresponde ao sistema reduzido,

$$\begin{cases} x_1 - 3x_3 = 0 \\ x_2 + x_3 = 0 \end{cases} \Leftrightarrow \begin{cases} x_1 = 3x_3 \\ x_2 = -x_3 \end{cases}$$

e portanto

$$\begin{aligned} \mathcal{N}(A) &= \{(x_1, x_2, x_3) \colon x_1 = 3x_3, x_2 = -x_3, x_3 \in \mathbb{R}\} \\ &= \{(3x_3, -x_3, x_3) \colon x_3 \in \mathbb{R}\} \\ &= \{x_3(3, -1, 1) \colon x_3 \in \mathbb{R}\}. \end{aligned}$$

Geometricamente, $\mathcal{N}(A)$ é formado pelos múltiplos do vetor $(3, -1, 1)$, logo define a reta de $\mathbb{R}^3$ que passa na origem e tem direção $(3, -1, 1)$. Esta reta foi obtida como interseção dos 2 planos que passam na origem, $x_1 + 2x_2 - x_3 = 0$ e $2x_1 + 5x_2 - x_3 = 0$.

3. Se $A = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix}$ então

$$\mathcal{N}(A) = \left\{x \in \mathbb{R}^2 \colon Ax = \vec{0}\right\} = \left\{(x_1, x_2) \in \mathbb{R}^2 \colon \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right\}.$$

Resolvendo o sistema homogéneo $Ax = \vec{0}$ tem-se:

$$\begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix} \xrightarrow[L_2-2L_1\to L_2]{} \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \xrightarrow[L_1-2L_2\to L_1]{} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

A última matriz corresponde ao sistema reduzido

$$\begin{cases} x_1 = 0 \\ x_2 = 0 \end{cases}.$$

Logo $Ax = \vec{0}$ tem como única solução o vetor $x = \vec{0} = (0, 0)$ e portanto que $\mathcal{N}(A) = \{\vec{0}\}$ (subespaço minimal). Este ponto foi obtido como intersecção das retas que passam na origem $x_1 + 2x_2 = 0$ e $2x_1 + 5x_2 = 0$. Note-se que no final da fase descendente podemos logo concluir que o sistema $Ax = \vec{0}$ é determinado (porque não possui variáveis livres) e portanto que o conjunto de soluções se reduz a $\{\vec{0}\}$.

4. Se $A = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = [0]_{2\times 2}$ então

$$\mathcal{N}(A) = \{x \in \mathbb{R}^2 : Ax = \vec{0}\} = \left\{(x_1, x_2) \in \mathbb{R}^2 : \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right\} = \mathbb{R}^2.$$

Logo $\mathcal{N}(A) = \mathbb{R}^2$ (subespaço maximal).

Observação 10. Mais geralmente, se A é uma matriz de tipo $m \times n$ então

- $\mathcal{N}(A) = \{\vec{0}\}$ (subespaço minimal) se e só se $Ax = \vec{0}$ é um sistema determinado.
- $\mathcal{N}(A) = \mathbb{R}^n$ (subespaço maximal) se e só se $A = [0]_{m\times n}$.

Os subconjuntos de $\mathbb{R}^n$ definidos como espaços nulos de matrizes (não nulas), representam intersecções de retas ($n = 2$), planos ($n = 3$) e hiperplanos ($n \geq 4$), que passam na origem e pode-se mostrar que qualquer subespaço vetorial (não maximal) pode ser definido desta forma. Daqui conclui-se que os subespaços vetoriais de $\mathbb{R}^2$ e de $\mathbb{R}^3$ são de um dos seguintes tipos:

n	subespaços vetoriais de $\mathbb{R}^n$
2	$\{\vec{0}\}$, retas que passam na origem, $\mathbb{R}^2$
3	$\{\vec{0}\}$, retas e planos que passam na origem, $\mathbb{R}^3$

2.2 Combinação linear e espaço gerado

A partir de um conjunto finito de vetores, $v_1, \ldots, v_n$, de um subespaço vetorial V podem ser obtidos outros vetores de V, *combinando* os vetores $v_1, \ldots, v_n$ através das operações algébricas da adição e da multiplicação por escalares. Se além disso, todos os vetores de V puderem ser obtidos dessa forma diz-se que $v_1, \ldots, v_n$ *geram* V. Estas e outras ideias fundamentais de Álgebra Linear são desenvolvidas nesta secção.

Combinação linear

Definição 18. Sejam $v_1, \ldots, v_n, b \in \mathbb{R}^m$. Diz-se que b é *combinação linear* de $v_1, \ldots, v_n$, se existirem escalares $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ tais que

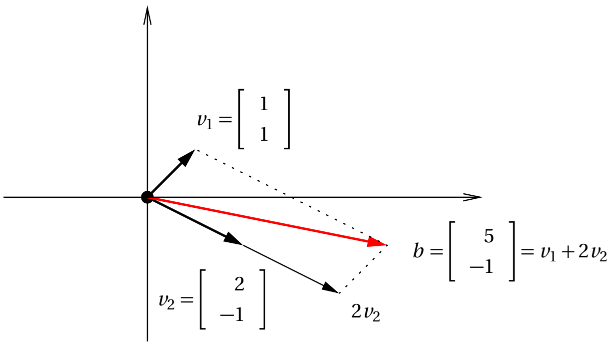
$$b = \alpha_1 v_1 + \cdots + \alpha_n v_n.$$

Os escalares $\alpha_1, \ldots, \alpha_n$ designam-se por *coeficientes* da combinação linear.

Note-se que se $v_1, \ldots, v_n$ pertencerem a um mesmo subespaço vetorial V então b ainda pertence a V pois é obtido como soma de múltiplos dos vetores $v_1, \ldots, v_n$ e V é fechado para a adição e para o produto por escalares.

Exemplos 31.

1. Sejam $b = (2,2)$ e $v = (1,1)$. Então $b = 2v$ e portanto b é combinação linear de v uma vez que é múltiplo de v. De igual forma, v é combinação linear de b, pois $v = \frac{1}{2}b$.
2. Sejam $b = (5,-1)$, $v_1 = (1,1)$ e $v_2 = (2,-1)$. Geometricamente, é possível ver que $b = (5,-1) = 1(1,1) + 2(2,-1) = v_1 + 2v_2$.



   Logo b é combinação linear de v_1 e v_2, pois é *soma de múltiplos* de v_1 e v_2.

   Deixa-se ao leitor o cuidado de justificar que v_1 é combinação linear de b e v_2 e que v_2 é combinação linear de b e v_1.
3. Considerem-se os vetores $b = (3,-3,0)$, $v_1 = (2,-1,-1)$ e $v_2 = (1,1,-2)$.

   Pretende-se averiguar se b é combinação linear de v_1 e v_2, ou seja, se existem escalares $\alpha_1, \alpha_2 \in \mathbb{R}$ tais que $b = \alpha_1 v_1 + \alpha_2 v_2$. Ora,

$$b = \alpha_1 v_1 + \alpha_2 v_2 \Leftrightarrow \begin{bmatrix} 3 \\ -3 \\ 0 \end{bmatrix} = \alpha_1 \begin{bmatrix} 2 \\ -1 \\ -1 \end{bmatrix} + \alpha_2 \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} \Leftrightarrow \begin{bmatrix} 3 \\ -3 \\ 0 \end{bmatrix} = \begin{bmatrix} 2\alpha_1 + \alpha_2 \\ -\alpha_1 + \alpha_2 \\ -\alpha_1 - 2\alpha_2 \end{bmatrix}$$

$$\Leftrightarrow \begin{cases} 2\alpha_1 + \alpha_2 = 3 \\ -\alpha_1 + \alpha_2 = -3 \\ -\alpha_1 - 2\alpha_2 = 0 \end{cases}.$$

   Logo (α_1, α_2) tem que ser solução do sistema linear com matriz ampliada,

$$\left[\begin{array}{cc|c} 2 & 1 & 3 \\ -1 & 1 & -3 \\ -1 & -2 & 0 \end{array}\right] = \left[\begin{array}{cc|c} v_1 & v_2 & b \end{array}\right].$$

Daqui resulta que b pode ser escrito como combinação linear de v_1 e v_2, da forma $b = \alpha_1 v_1 + \alpha_2 v_2$, se e somente se o sistema anterior, nas incógnitas α_1, α_2, for possível:

$$\left[\begin{array}{rr|r} 2 & 1 & 3 \\ -1 & 1 & -3 \\ -1 & -2 & 0 \end{array}\right] \xrightarrow[L_1 \leftrightarrow L_3]{} \left[\begin{array}{rr|r} -1 & -2 & 0 \\ -1 & 1 & -3 \\ 2 & 1 & 3 \end{array}\right] \xrightarrow[\substack{L_2 - L_1 \to L_2 \\ L_3 + 2L_1 \to L_3}]{} \left[\begin{array}{rr|r} -1 & -2 & 0 \\ 0 & 3 & -3 \\ 0 & -3 & 3 \end{array}\right]$$
$$\xrightarrow[L_3 + L_2 \to L_3]{} \left[\begin{array}{rr|r} -1 & -2 & 0 \\ 0 & 3 & -3 \\ 0 & 0 & 0 \end{array}\right]$$

Logo o sistema é possível e portanto b é combinação linear de v_1 e v_2. Aplicando a fase ascendente do método de Gauss obtém-se,

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

A última matriz corresponde ao sistema reduzido,

$$\begin{cases} \alpha_1 = 2 \\ \alpha_2 = -1, \end{cases}$$

e portanto b é combinação linear de v_1 e v_2 da forma $b = 2v_1 - v_2$, isto é, $(3,-3,0) = 2(2,-1,-1) - (1,1,-2)$.

4. Seja $b = (1,1,0)$ e consideremos novamente os vetores $v_1 = (2,-1,-1)$ e $v_2 = (1,1,-2)$ do exemplo anterior. Aplicando o método de Gauss à matriz ampliada $\left[\begin{array}{cc|c} v_1 & v_2 & b \end{array}\right]$, obtém-se

$$\left[\begin{array}{rr|r} 2 & 1 & 1 \\ -1 & 1 & 1 \\ -1 & -2 & 0 \end{array}\right] \xrightarrow[L_1 \leftrightarrow L_2]{} \left[\begin{array}{rr|r} -1 & 1 & 1 \\ 2 & 1 & 1 \\ -1 & -2 & 0 \end{array}\right] \xrightarrow[\substack{L_2 + 2L_1 \to L_2 \\ L_3 - L_1 \to L_3}]{}$$
$$\left[\begin{array}{rr|r} -1 & 1 & 1 \\ 0 & 3 & 3 \\ 0 & -3 & -1 \end{array}\right] \xrightarrow[L_3 + L_2 \to L_3]{} \left[\begin{array}{rr|r} -1 & 1 & 1 \\ 0 & 3 & 3 \\ 0 & 0 & 2 \end{array}\right].$$

Como o sistema neste caso não é possível, b não é combinação linear de v_1 e v_2. A explicação geométrica para a diferença entre este exemplo e o exemplo anterior será feita mais à frente no exemplo 32.

5. Sejam $b=(4,4)$, $v_1=(1,4)$, $v_2=(-2,6)$ e $v_3=(2,2)$. Para determinar os coeficientes $\alpha_1,\alpha_2,\alpha_3\in\mathbb{R}$ (caso existam) da combinação linear $b=\alpha_1 v_1+\alpha_2 v_2+\alpha_3 v_3$ temos que aplicar o método de Gauss à matriz ampliada $\left[\begin{array}{ccc|c} v_1 & v_2 & v_3 & b\end{array}\right]$:

$$\left[\begin{array}{ccc|c} 1 & -2 & 2 & 4\\ 4 & -6 & 2 & 4\end{array}\right] \xrightarrow[L_2-4L_1\to L_2]{} \left[\begin{array}{ccc|c} 1 & -2 & 2 & 4\\ 0 & 2 & -6 & -12\end{array}\right] \xrightarrow[\frac{1}{2}L_2\to L_2]{}$$
$$\left[\begin{array}{ccc|c} 1 & -2 & 2 & 4\\ 0 & 1 & -3 & -6\end{array}\right] \xrightarrow[L_1+2L_2\to L_1]{} \left[\begin{array}{ccc|c} 1 & 0 & -4 & -8\\ 0 & 1 & -3 & -6\end{array}\right].$$

A última matriz corresponde ao sistema reduzido

$$\begin{cases}\alpha_1 \qquad - 4\alpha_3 = -8\\ \qquad \alpha_2 - 3\alpha_3 = -6\end{cases} \Leftrightarrow \begin{cases}\alpha_1 = 4\alpha_3-8\\ \alpha_2 = 3\alpha_3-6\end{cases},$$

com $\alpha_3\in\mathbb{R}$. Por conseguinte,

$$b=(4\alpha_3-8)v_1+(3\alpha_3-6)v_2+\alpha_3 v_3, \qquad \alpha_3\in\mathbb{R},$$

o que significa que b se escreve como combinação linear de v_1, v_2 e v_3, de infinitas maneiras distintas. Por exemplo, tomando $\alpha_3=0$ obtém-se a combinação linear $b=-8v_1-6v_2+0v_3$, isto é

$$\begin{bmatrix}4\\4\end{bmatrix}=-8\begin{bmatrix}1\\4\end{bmatrix}-6\begin{bmatrix}-2\\-6\end{bmatrix}+0\begin{bmatrix}2\\2\end{bmatrix}.$$

Considerando agora $\alpha_3=2$, por exemplo, obtém-se a combinação linear $b=0v_1+0v_2+2v_3$, isto é,

$$\begin{bmatrix}4\\4\end{bmatrix}=0\begin{bmatrix}1\\4\end{bmatrix}+0\begin{bmatrix}-2\\-6\end{bmatrix}+2\begin{bmatrix}2\\2\end{bmatrix}.$$

Observação 11. Mais geralmente, $b\in\mathbb{R}^n$ pode ser escrito como combinação linear de $v_1,\ldots,v_n\in\mathbb{R}^n$ com coeficientes $\alpha_1,\ldots,\alpha_n\in\mathbb{R}$, isto é,

$$b=\alpha_1 v_1+\cdots+\alpha_n v_n,$$

se e só se $(\alpha_1,\ldots,\alpha_n)$ for solução do sistema linear $Ax=b$ com $A=[v_1\ v_2\ \cdots\ v_n]$. Assim:

- Se o sistema $Ax=b$ for possível, b é combinação linear de $v_1,\ldots,v_n$ e tem-se que
  - se o sistema $Ax=b$ for determinado, b é combinação linear de $v_1,\ldots,v_n$ de uma **única** forma.

- se o sistema $Ax = b$ for indeterminado, b é combinação linear de $v_1, \dots, v_n$ **de infinitas** formas distintas.

- Se o sistema $Ax = b$ for impossível, b não é combinação linear de $v_1, \dots, v_n$.

Espaço gerado

Definição 19. Sejam $v_1, \dots, v_n \in \mathbb{R}^m$. Chama-se *espaço gerado* por $v_1, \dots, v_n$, e denota-se por $\langle\{v_1, \dots, v_n\}\rangle$, ou simplesmente por $\langle v_1, \dots, v_n\rangle$, ao conjunto de todos os vetores de $\mathbb{R}^m$ que se podem escrever como combinação linear de $v_1, \dots, v_n$, isto é,

$$\begin{aligned}\langle v_1, \dots, v_n\rangle &= \{b \in \mathbb{R}^m : \exists \alpha_1, \dots, \alpha_n \in \mathbb{R} \text{ tal que } b = \alpha_1 v_1 + \cdots + \alpha_n v_n\}\\ &= \{\alpha_1 v_1 + \cdots + \alpha_n v_n : \alpha_1, \dots, \alpha_n \in \mathbb{R}\}.\end{aligned}$$

Observação 12.

1. $\vec{0} \in \langle v_1, \dots, v_n\rangle$ uma vez que $\vec{0} = 0v_1 + 0v_2 + \cdots + 0v_n$.

2. $v_i \in \langle v_1, \dots, v_n\rangle$ para $i = 1, \dots, n$, uma vez que

$$v_i = 0v_1 + \cdots + 0v_{i-1} + 1v_i + 0v_{i+1} + \cdots + 0v_n.$$

Note-se que as combinações lineares anteriores podem não ser únicas.

Antes de mostrarmos que o espaço gerado por um conjunto de vetores define um subespaço vetorial vamos considerar alguns exemplos.

Exemplo 32. Sejam $v_1 = (2, -1, -1)$ e $v_2 = (1, 1, -2)$ os vetores dos exemplos 31.3 e 31.4 e seja $A = [v_1\ v_2]$. Por definição,

$$\begin{aligned}\langle v_1, v_2\rangle &= \left\{b = (b_1, b_2, b_3) \in \mathbb{R}^3 : b \text{ é combinação linear de } v_1 \text{ e } v_2\right\}\\ &= \left\{b = (b_1, b_2, b_3) \in \mathbb{R}^3 : Ax = b \text{ é possível}\right\}.\end{aligned}$$

Ora,

$$\left[\begin{array}{rr|l} 2 & 1 & b_1 \\ -1 & 1 & b_2 \\ -1 & -2 & b_3 \end{array}\right] \xrightarrow[L_1 \leftrightarrow L_2]{} \left[\begin{array}{rr|l} -1 & 1 & b_2 \\ 2 & 1 & b_1 \\ -1 & -2 & b_3 \end{array}\right] \xrightarrow[\substack{L_2 + 2L_1 \to L_2 \\ L_3 - L_1 \to L_3}]{}$$

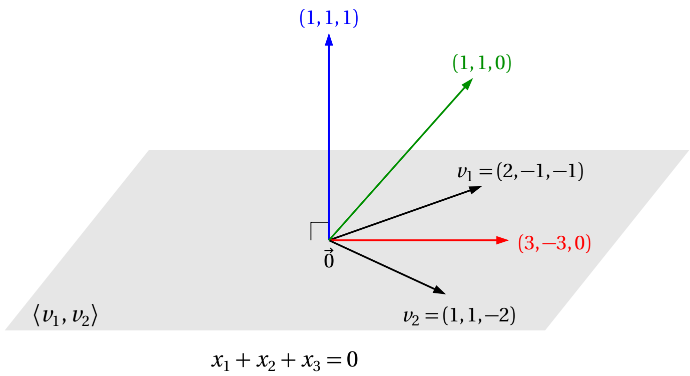
$$\left[\begin{array}{rr|l} -1 & 1 & b_2 \\ 0 & 3 & b_1 + 2b_2 \\ 0 & -3 & -b_2 + b_3 \end{array}\right] \xrightarrow[L_3 + L_2 \to L_3]{} \left[\begin{array}{rr|l} -1 & 1 & b_2 \\ 0 & 3 & b_1 + 2b_2 \\ 0 & 0 & b_1 + b_2 + b_3 \end{array}\right].$$

Logo o sistema $Ax = b$ é possível se e só se $b_1 + b_2 + b_3 = 0$ e, portanto,

$$\langle v_1, v_2\rangle = \{b = (b_1, b_2, b_3) \in \mathbb{R}^3 : b_1 + b_2 + b_3 = 0\}. \tag{2.1}$$

Geometricamente $\langle v_1, v_2\rangle$ define o plano de $\mathbb{R}^3$ que passa na origem e é ortogonal ao vetor $(1, 1, 1)$. Este plano é constituído pelos vetores que se podem escrever como combinação linear de v_1 e v_2.

O vetor $b = (3, -3, 0)$ do exemplo 31.3 satisfaz a equação $b_1 + b_2 + b_3 = 0$ e portanto é um vetor do plano $\langle v_1, v_2\rangle$. Logo, é combinação linear desses 2 vetores. Por outro lado, o vetor $b = (1, 1, 0)$ do exemplo 31.4, não satisfaz a equação anterior e portanto não pertence ao plano $\langle v_1, v_2\rangle$. Logo, não é combinação linear de v_1 e v_2.



O exemplo anterior mostra ainda que um mesmo subespaço vetorial pode estar definido por um conjunto de geradores ou como conjunto de soluções de um sistema linear homogéneo.

Exemplo 33. Consideremos o vetor $v = (2, 2, 3)$. Por definição

$$\begin{aligned} \langle v\rangle &= \left\{b = (b_1, b_2, b_3) \in \mathbb{R}^3 : b \text{ é combinação linear de } v\right\} \\ &= \left\{b = (b_1, b_2, b_3) \in \mathbb{R}^3 : (b_1, b_2, b_3) = \lambda(2, 2, 3),\ \lambda \in \mathbb{R}\right\}. \end{aligned}$$

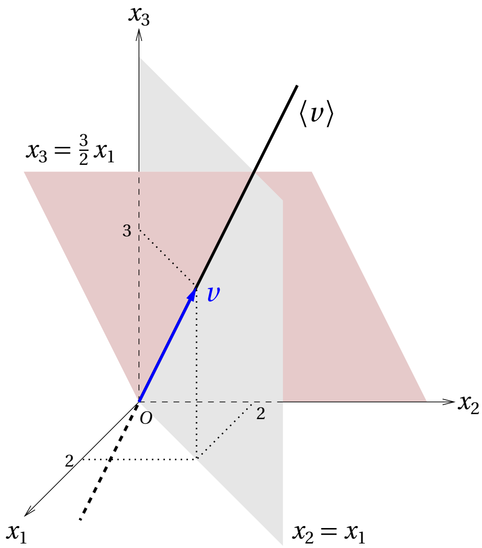
Logo $\langle v\rangle$ consiste no conjunto dos múltiplos do vetor $v = (2, 2, 3)$, isto é, na reta de $\mathbb{R}^3$ que passa na origem e contém a direção $(2, 2, 3)$. Pretende-se redefinir esta reta através de um sistema de equações lineares homogéneas. Ora,

$$[v \mid b] = \left[\begin{array}{c|c} 2 & b_1 \\ 2 & b_2 \\ 3 & b_3 \end{array}\right] \xrightarrow[L_3 - \frac{3}{2}L_1 \to L_3]{L_2 - L_1 \to L_2} \left[\begin{array}{c|c} 2 & b_1 \\ 0 & b_2 - b_1 \\ 0 & b_3 - \frac{3}{2}b_1 \end{array}\right].$$

Logo, $\langle v\rangle = \left\{(b_1, b_2, b_3) : b_2 - b_1 = 0,\ b_3 - \dfrac{3}{2}b_1 = 0\right\}$, isto é, a reta $\langle v\rangle$ corresponde à intersecção dos planos

$$\begin{cases} -x_1 + x_2 = 0 \\ -\dfrac{3}{2}x_1 + x_3 = 0, \end{cases}$$

representados na seguinte figura.



Observação 13. Não é demais realçar a diferença entre as notações $\{v_1, \ldots, v_n\}$ e $\langle v_1, \ldots, v_n \rangle$. No primeiro caso, a notação refere-se ao conjunto formado pelos n vetores $v_1, \ldots, v_n$ e, no segundo caso, a notação refere-se ao conjunto dos vetores que se podem obter como somas de mútliplos dos n vetores, $v_1, \ldots, v_n$, contendo por isso uma infinidade de elementos.

Se considerarmos novamente o exemplo anterior, $\{(2,2,3)\}$ representa o conjunto formado pelo único vetor $(2,2,3)$, enquanto que $\langle(2,2,3)\rangle$ representa a reta que passa na origem de $\mathbb{R}^3$ definida pela direção $(2,2,3)$, contendo portanto todos os múltiplos do vetor $(2,2,3)$!

Teorema 16. *Sejam $v_1, \ldots, v_n \in \mathbb{R}^m$. Tem-se que $\langle v_1, \ldots, v_n \rangle$ define um subespaço vetorial de $\mathbb{R}^m$.*

Demonstração. Vejamos que as três condições da definição de subespaço vetorial (Definição 16) são satisfeitas.

Tem-se que $\langle v_1, \ldots, v_n \rangle \neq \emptyset$, uma vez que $\vec{0} \in \langle v_1, \ldots, v_n \rangle$ como vimos na Observação 12.

Consideremos agora $u, v \in \langle v_1, \ldots, v_n \rangle$ e $\lambda \in \mathbb{R}$. Vamos provar que $u+v$ e λu ainda pertencem a $\langle v_1, \ldots, v_n \rangle$, isto é, que $\langle v_1, \ldots, v_n \rangle$ é fechado para a adição e para o produto por escalar.

Por definição, existem escalares $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ tais que $u = \alpha_1 v_1 + \cdots + \alpha_n v_n$ e escalares $\beta_1, \ldots, \beta_n \in \mathbb{R}$ tais que $v = \beta_1 v_1 + \cdots + \beta_n v_n$. Tem-se então,

$$\begin{aligned} u + v &= (\alpha_1 v_1 + \cdots + \alpha_n v_n) + (\beta_1 v_1 + \cdots + \beta_n v_n) \\ &= \underbrace{(\alpha_1 + \beta_1)}_{\in \mathbb{R}} v_1 + \cdots + \underbrace{(\alpha_1 + \beta_1)}_{\in \mathbb{R}} v_n, \end{aligned}$$

e portanto $u + v$ é combinação linear de $v_1, \dots, v_n$, isto é, $u + v \in \langle v_1, \dots, v_n \rangle$. Analogamente tem-se,

$$\begin{aligned} \lambda u &= \lambda(\alpha_1 v_1 + \cdots + \alpha_n v_n) \\ &= \underbrace{(\lambda\alpha_1)}_{\in\mathbb{R}} v_1 + \cdots + \underbrace{(\lambda\alpha_n)}_{\in\mathbb{R}} v_n, \end{aligned}$$

e portanto λu é combinação linear de $v_1, \dots, v_n$, isto é, $\lambda u \in \langle v_1, \dots, v_n \rangle$. □

Corolário 17. *Sejam $u_1, \dots, u_k, v_1, \dots, v_n \in \mathbb{R}^m$ e consideremos os subespaços vetoriais $U = \langle u_1, \dots, u_k \rangle$ e $V = \langle v_1, \dots, v_n \rangle$. Tem-se que $U \subset V$ se e só se u_i for combinação linear de $v_1, \dots, v_n$, $\forall i = 1, \dots, k$.*

Demonstração. Suponhamos que $U \subset V$. Como $u_i \in U \subset V = \langle v_1, \dots, v_n \rangle$ para $i = 1, \dots, k$, u_i é combinação linear de $v_1, \dots, v_n$ para todo o i.

Reciprocamente, suponhamos que u_i é combinação linear de $v_1, \dots, v_n$, para $i = 1, \dots, k$, isto é, $u_i \in V = \langle v_1, \dots, v_n \rangle$ para todo o i. Se $u \in U$ tem-se que u é soma de múltiplos de $u_1, \dots, u_k$. Como $u_1, \dots, u_k \in V$ e V é um subespaço vetorial, logo fechado para a adição e para o produto por escalares, conclui-se que $u \in V$. Logo $U \subset V$. □

Exemplo 34. Consideremos os subespaços vetoriais de $\mathbb{R}^3$,

$$U = \langle (1,-1,-2), (2,3,1) \rangle \quad \text{e} \quad V = \langle (1,1,0), (-1,0,1) \rangle.$$

Vejamos que $U = V$. Para isso temos que verificar as duas inclusões,

$$U \subset V \quad \text{e} \quad V \subset U.$$

Vejamos a primeira inclusão. Pelo corolário anterior basta verificar que

$$(1,-1,-2), (2,3,1) \in V = \langle (1,1,0), (-1,0,1) \rangle,$$

isto é, que os sistemas lineares com matrizes ampliadas,

$$\left[\begin{array}{cc|c} 1 & -1 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & -2 \end{array}\right] \quad \text{e} \quad \left[\begin{array}{cc|c} 1 & -1 & 2 \\ 1 & 0 & 3 \\ 0 & 1 & 1 \end{array}\right]$$

são ambos possíveis. Uma vez que estes sistemas possuem a **mesma matriz de coeficientes** podemos verificar simultaneamente ambas as condições ampliando a matriz dos coeficientes com os vetores $(1,-1,-2)$ e $(2,3,1)$:

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

Logo ambos os sistemas são possíveis o que mostra que $U \subset V$.

A inclusão $V \subset U$ pode ser verificada de modo análogo e é deixada como exercício.

Observação 14. Alternativamente, podemos mostrar que dois subespaços vetoriais são iguais, mostrando que ambos os subespaços podem ser definidos por sistemas de equações homogéneas equivalentes.

Consideremos novamente os subespaços vetoriais do exemplo anterior, $U = \langle u_1, u_2\rangle$, com $u_1 = (1,-1,-2)$ e $u_2 = (2,3,1)$ e $V = \langle v_1, v_2\rangle$, com $v_1 = (1,1,0)$ e $v_2 = (-1,0,1)$. Sejam $A = [u_1 \;\; u_2]$ e $B = [v_1 \; v_2]$.

Tem-se $U = \langle u_1, u_2\rangle = \{b = (b_1, b_2, b_3) \in \mathbb{R}^3 \colon Ax = b$ é possível $\}$. Ora,

$$\left[\begin{array}{rr|l} 1 & 2 & b_1 \\ -1 & 3 & b_2 \\ -2 & 1 & b_3 \end{array}\right] \xrightarrow[L_3+2L_1\to L_3]{L_2+L_1\to L_2} \left[\begin{array}{rr|l} 1 & 2 & b_1 \\ 0 & 5 & b_1+b_2 \\ 0 & 5 & 2b_1+b_3 \end{array}\right] \xrightarrow[L_3-L_2\to L_3]{} \left[\begin{array}{rr|l} 1 & 2 & b_1 \\ 0 & 5 & b_1+b_2 \\ 0 & 0 & b_1-b_2+b_3 \end{array}\right]$$

e, portanto,

$$U = \langle u_1, u_2\rangle = \{(b_1, b_2, b_3) \in \mathbb{R}^3 \colon b_1 - b_2 + b_3 = 0\}.$$

Analogamente, $V = \langle v_1, v_2\rangle = \{b = (b_1, b_2, b_3) \in \mathbb{R}^3 \colon Bx = b$ é possível $\}$. Ora,

$$\left[\begin{array}{rr|l} 1 & -1 & b_1 \\ 1 & 0 & b_2 \\ 0 & 1 & b_3 \end{array}\right] \xrightarrow[L_2-L_1\to L_2]{} \left[\begin{array}{rr|l} 1 & -1 & b_1 \\ 0 & 1 & b_2-b_1 \\ 0 & 1 & b_3 \end{array}\right] \xrightarrow[L_3-L_2\to L_3]{} \left[\begin{array}{rr|l} 1 & -1 & b_1 \\ 0 & 1 & b_2-b_1 \\ 0 & 0 & b_1-b_2+b_3 \end{array}\right]$$

e, portanto

$$V = \langle\langle v_1, v_2\rangle = \{(b_1, b_2, b_3) \in \mathbb{R}^3 \colon b_1 - b_2 + b_3 = 0\}.$$

Dos cálculos anteriores concluímos que apesar de U e V serem gerados por vetores distintos, ambos definem o plano de $\mathbb{R}^3$ que passa na origem de equação $x_1 - x_2 + x_3 = 0$. Logo $U = V$.[2]

Espaço das colunas de uma matriz

Vamos apresentar o segundo subespaço vetorial fundamental associado a uma matriz.

Definição 20. Seja $A = \begin{bmatrix} v_1 & \dots & v_n \end{bmatrix}$ uma matriz de tipo $m \times n$. Chama-se *espaço das colunas* de A, e denota-se por $\mathscr{C}(A)$, ao subespaço vetorial de $\mathbb{R}^m$ gerado pelos vetores que constituem as n colunas de A, isto é,

$$\mathscr{C}(A) = \langle v_1, \dots, v_n\rangle = \Big\{ b \in \mathbb{R}^m \colon Ax = b \text{ é possível} \Big\}.$$

[2]Note-se que para provar que dois sistemas lineares homogéneos constituídos por mais que uma equação definem o mesmo subespaço vetorial é geralmente necessário reduzir previamente ambos os sistemas.

Uma vez que o espaço das colunas de uma matriz corresponde ao espaço gerado pelos vetores que constituem as colunas dessa matriz aplicam-se todos resultados e observações referidos anteriormente para o espaço gerado.

Exemplo 35. Consideremos a matriz

$$A = \begin{bmatrix} 1 & 2 & 1 & 0 & -1 \\ 0 & -1 & -2 & 2 & 3 \\ 2 & 3 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix} = [v_1 \ v_2 \ v_3 \ v_4 \ v_5].$$

Por definição,

$$\mathscr{C}(A) = \langle v_1, v_2, v_3, v_4, v_5 \rangle = \left\{ b = (b_1, b_2, b_3, b_4) \in \mathbb{R}^4 \, : \, Ax = b \text{ é possível} \right\}.$$

Aplicando o método de eliminação de Gauss obtém-se (verifique!),

$$[A|b] = \left[\begin{array}{ccccc|c} 1 & 2 & 1 & 0 & -1 & b_1 \\ 0 & -1 & -2 & 2 & 3 & b_2 \\ 2 & 3 & 0 & 1 & 1 & b_3 \\ 1 & 1 & 1 & 1 & 0 & b_4 \end{array}\right] \to \cdots \to \left[\begin{array}{ccccc|c} 1 & 2 & 1 & 0 & -1 & b_1 \\ 0 & -1 & -2 & 2 & 3 & b_2 \\ 0 & 0 & 2 & -1 & -2 & b_4 - b_1 - b_2 \\ 0 & 0 & 0 & -1 & 0 & b_3 - 2b_1 - b_2 \end{array}\right] = [A'|b'].$$

Uma vez que a matriz em escada A' não possui linhas nulas, car$(A) = 4$ e o sistema $Ax = b$ é possível qualquer que seja $b \in \mathbb{R}^4$. Logo,

$$\mathscr{C}(A) = \langle v_1, v_2, v_3, v_4, v_5 \rangle = \mathbb{R}^4.$$

Mais geralmente tem-se o seguinte resultado.

Proposição 18. *Sejam $v_1, \ldots, v_n \in \mathbb{R}^m$, $A = [v_1 \ \cdots \ v_n]$ do tipo $m \times n$ e A' matriz em escada obtida a partir de A por aplicação do método de eliminação de Gauss. As seguintes afirmações são equivalentes:*

(i) $\langle v_1, \ldots, v_n \rangle = \mathbb{R}^m$*, isto é,* $\mathscr{C}(A) = \mathbb{R}^m$.

(ii) A' *não tem linhas nulas.*

(iii) *car*$(A) = m$.

Uma vez que car$(A) \leq \min\{m, n\}$, deduz-se da Proposição 18 que $m \leq n$, e portanto tem-se o seguinte resultado.

Corolário 19. *Um conjunto de vetores de $\mathbb{R}^m$ que gere $\mathbb{R}^m$ possui pelo menos m vetores.*

2.3 Independência linear

Definição 21. Sejam $v_1,\dots,v_n \in \mathbb{R}^m$. O conjunto $\{v_1,\dots,v_n\}$ diz-se *linearmente independente* (l.i.) se

$$\forall \alpha_1,\dots,\alpha_n \in \mathbb{R}: \alpha_1 v_1 + \cdots + \alpha_n v_n = \vec{0} \Rightarrow \alpha_1 = \cdots = \alpha_n = 0.$$

Caso contrário, $\{v_1,\dots,v_n\}$ diz-se *linearmente dependente* (l.d.)

Por outras palavras, um conjunto de vetores $\{v_1,\dots,v_n\}$ é linearmente independente se a combinação linear nula *trivial*, isto é, com todos os coeficientes nulos, for a única combinação linear nula de $v_1,\dots,v_n$.

Observação 15. A caracterização da (in)dependência linear para conjuntos com um ou dois vetores é imediata:

(i) *$\{v\}$ é linearmente dependente se e só se $v = \vec{0}$.*

De facto, se $v = \vec{0}$ então $\alpha v = \vec{0}$ constitui uma combinação linear nula *não trivial* para todo $\alpha \neq 0$ e portanto $\{\vec{0}\}$ é linearmente dependente.

Reciprocamente, se $v \neq \vec{0}$, então $\alpha v = \vec{0} \Rightarrow \alpha = 0$ e, portanto, $\{v\}$ é linearmente independente.

(ii) *$\{v_1, v_2\}$ é linearmente dependente se e só se um dos vetores é múltiplo do outro.*

Com efeito, se v_1 for múltiplo de v_2, digamos, $v_1 = \lambda v_2$ tem-se a combinação linear nula não trivial $v_1 - \lambda v_2 = \vec{0}$ e portanto $\{v_1, v_2\}$ é linearmente dependente. O caso v_2 múltiplo de v_1 é análogo.

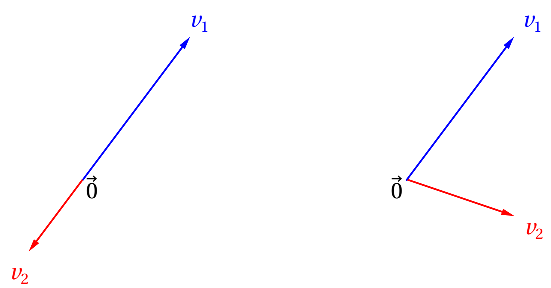
Reciprocamente, se $\{v_1, v_2\}$ for linearmente dependente então existe uma combinação linear nula

$$\alpha_1 v_1 + \alpha_2 v_2 = \vec{0},$$

com pelo um dos $\alpha_i \neq 0$. Se $\alpha_1 \neq 0$, então

$$v_1 = -\frac{\alpha_2}{\alpha_1} v_2,$$

e portanto v_1 é múltiplo de v_2. De modo análogo se conclui que se $\alpha_2 \neq 0$ então v_2 é múltiplo de v_1.

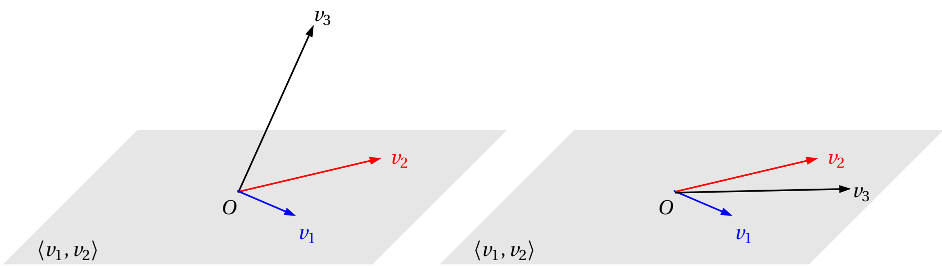


$\{v_1, v_2\}$ l.d. $\qquad$ $\{v_1, v_2\}$ l.i.

O tipo de argumentos usado na observação anterior para conjuntos com dois vetores pode ser generalizado para conjuntos com mais que dois vetores. De facto, tem-se o seguinte resultado que estabelece uma caracterização mais intuitiva da noção de dependência linear.

Proposição 20. *Um conjunto (finito) com dois ou mais vetores é linearmente dependente se e só se pelo menos um dos vetores for combinação linear dos restantes vetores do conjunto.*

Observação 16. Sejam v_1, v_2, v_3 vetores de $\mathbb{R}^m$. Pela Proposição 20, $\{v_1, v_2, v_3\}$ é linearmente independente se e só se nenhum dos vetores é combinação linear dos restantes 2 vetores, isto é, se e só se nenhum dos vetores pertence ao espaço gerado pelos outros 2 vetores, o que é o mesmo que dizer que os três vetores são *não complanares.*



$\{v_1, v_2, v_3\}$ l.i. $\{v_1, v_2, v_3\}$ l.d.

Ao contrário do que sucede com os conjuntos com um ou dois vetores, a verificação da independência linear para conjuntos com três ou mais vetores não é, em geral, imediata. Vejamos como aplicar o método de eliminação de Gauss para decidir sobre a independência linear de conjuntos com um número arbitrário de vetores.

Sejam $v_1, \ldots, v_n \in \mathbb{R}^m$ e $A = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$. Duas situações podem ocorrer:

(i) O sistema $Ax = \vec{0}$ é determinado e nessa altura $\alpha_1 = \cdots = \alpha_n = 0$ é a única solução do sistema, isto é, $\vec{0} = 0v_1 + \cdots + 0v_n$ é a única combinação linear nula de $v_1, \ldots, v_n$. Nesta situação $\{v_1, \ldots, v_n\}$ é um conjunto linearmente independente de vetores.

(ii) O sistema $Ax = \vec{0}$ é indeterminado e nessa altura $\alpha_1 = \cdots = \alpha_n = 0$ não é a única solução do sistema, ou seja existem combinações lineares nulas $\vec{0} = \alpha_1 v_1 + \cdots + \alpha_n v_n$, em que pelo menos um dos α_i é não nulo. Neste caso $\{v_1, \ldots, v_n\}$ é um conjunto linearmente dependente de vetores.

Da discussão anterior resultam os seguintes critérios que permitem averiguar a independência linear de um conjunto de vetores aplicando o método de Gauss à matriz definida por esses vetores.

Proposição 21. *Consideremos $v_1, \ldots, v_n \in \mathbb{R}^m$ e sejam $A = [v_1 \ldots v_n]$ e A' matriz em escada obtida a partir de A por aplicação do método de eliminação de Gauss. As seguintes afirmações são equivalentes:*

(i) $\{v_1, \ldots, v_n\}$ *é linearmente independente.*

(ii) $Ax = 0$ *é determinado, isto é,* $\mathcal{N}(A) = \{\vec{0}\}$.

(iii) Todas as colunas de A' *têm pivot.*

(iv) $\mathrm{car}(A) = n$.

Uma vez que a $\mathrm{car}(A) \leq \min\{m, n\}$, deduz-se da proposição anterior que $n \leq m$ e portanto tem-se o seguinte resultado.

Corolário 22. *Um conjunto linearmente independente de vetores de $\mathbb{R}^m$ contém no máximo m vetores.*

Exemplos 36.

1. Consideremos $\{v_1, v_2 v_3\}$ com $v_1 = (1,2,1)$, $v_2 = (0,1,-1)$ e $v_3 = (2,3,4)$. Aplicando o método de Gauss à matriz $A = [v_1 \;\; v_2 \;\; v_3]$, obtém-se

$$A = \begin{bmatrix} 1 & 0 & 2 \\ 2 & 1 & 3 \\ 1 & -1 & 4 \end{bmatrix} \xrightarrow[\substack{L_2-2L_1\to L_2 \\ L_3-L_1\to L_3}]{} \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & -1 \\ 0 & -1 & 2 \end{bmatrix} \xrightarrow[L_3+L_2\to L_3]{} \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} = A',$$

e, portanto, $\{v_1, v_2, v_3\}$ é linearmente independente, uma vez que todas as colunas de A' têm pivot. Em particular, os 3 vetores que constituem as colunas de A são não complanares.

2. Pelo Corolário 22 o conjunto $\{(1,2,1),(0,1,-1),(2,3,4),(1,5,5)\}$ é linearmente dependente uma vez que é constituído por 4 vetores de $\mathbb{R}^3$.

Pela Proposição 20, um conjunto com dois ou mais vetores é linearmente dependente se e só se pelo menos um dos vetores desse conjunto for *redundante*, isto é, puder ser obtido como combinação linear dos restantes vetores do conjunto. Daqui resulta em particular, que qualquer conjunto de vetores que contenha o vetor nulo é linearmente dependente.

Em geral, temos o seguinte resultado que enuncia características importantes da (in)dependência linear.

Proposição 23.

(i) Qualquer conjunto (finito) de vetores que contenha um conjunto linearmente dependente de vetores é ainda linearmente dependente.

(ii) Reciprocamente, um subconjunto não vazio de um conjunto linearmente independente de vetores é ainda linearmente independente.

Demonstração. Se $\{v_1,\dots,v_n\}$ é um conjunto linearmente dependente de vetores de $\mathbb{R}^m$, existe uma combinação linear nula,

$$\alpha_1 v_1 + \cdots + \alpha_n v_n = \vec{0}$$

em que $\alpha_1,\dots,\alpha_n$ não são todos nulos. Por conseguinte, qualquer conjunto $\{v_1,\dots,v_n,v_{n+1},\dots,v_k\}$ que contenha o conjunto $\{v_1,\dots,v_n\}$ $(k>n)$, é ainda linearmente dependente, uma vez que

$$\alpha_1 v_1 + \cdots + \alpha_n v_n + 0v_{n+1} + \cdots + 0v_k = \vec{0}$$

é uma combinação linear nula não trivial de $v_1,\dots,v_n,v_{n+1},\dots,v_k$. □

2.4 Base e dimensão

Base de um subespaço vetorial

Intuitivamente uma base de um subespaço vetorial V é um conjunto de vetores de V que não contém vetores *redundantes* e tal que todo o vetor de V se pode obter como combinação linear dos vetores desse conjunto.

Mais precisamente temos a seguinte definição.

Definição 22. Sejam V subespaço vetorial de $\mathbb{R}^m$ e $v_1,\dots,v_n \in V$. Diz-se que $\{v_1,\dots,v_n\}$ é *base* de V, se as seguintes condições forem verificadas:

(i) $V = \langle v_1,\dots,v_n\rangle$, isto é, $\{v_1,\dots,v_n\}$ é um conjunto de geradores de V.

(ii) $\{v_1,\dots,v_n\}$ é linearmente independente.

Convenciona-se que $\{\}$ é a base de $\{\vec{0}\}$.

Exemplo 37. Consideremos os vetores de $\mathbb{R}^3$,

$$e_1 = (1,0,0), \quad e_2 = (0,1,0), \quad e_3 = (0,0,1).$$

Vejamos que $\{e_1,e_2,e_3\}$ é uma base de $\mathbb{R}^3$:

(i) Por um lado, tem-se para qualquer $b = (b_1,b_2,b_3) \in \mathbb{R}^3$,

$$(b_1,b_2,b_3) = b_1(1,0,0) + b_2(0,1,0) + b_3(0,0,1) = b_1 e_1 + b_2 e_2 + b_3 e_3,$$

e, portanto, $b \in \langle e_1,e_2,e_3\rangle$. Logo $\langle e_1,e_2,e_3\rangle = \mathbb{R}^3$.

(ii) Por outro lado, $[e_1 \;\; e_2 \;\; e_3] = I_3$, matriz que já está em escada e tem todas as colunas com pivot. Logo, $\{e_1,e_2,e_3\}$ é linearmente independente.

De (i) e (ii) conclui-se que $\{e_1,e_2,e_3\}$ é uma base de $\mathbb{R}^3$. Esta base é usualmente designada por *base canónica* de $\mathbb{R}^3$.

Observação 17. Mais geralmente, chama-se *base canónica* de $\mathbb{R}^n$, à base constituída pelos n vetores,

$$e_1 = (1,0,\ldots,0), \quad e_2 = (0,1,\ldots,0), \quad \ldots, \quad e_n = (0,0,\ldots,1),$$

isto é, formada pelos vetores que constituem as colunas da matriz identidade de ordem n, I_n.

Exemplo 38. Considerermos o conjunto $V = \{(x,y,z)\colon x+y+z=0\}$. Então, V é um subespaço vetorial de $\mathbb{R}^3$, pois corresponde ao conjunto de soluções de uma equação linear homogénea[3]. Vamos determinar uma base de V.

(i) Por um lado, tem-se

$$\begin{aligned} V &= \{(x,y,z)\colon x=-y-z,\ y,z\in\mathbb{R}\} \\ &= \{(-y-z,y,z)\colon y,z\in\mathbb{R}\} \\ &= \{(-y,y,0)+(-z,0,z)\colon y,z\in\mathbb{R}\} \\ &= \{y(-1,1,0)+z(-1,0,1)\colon y,z\in\mathbb{R}\} \\ &= \langle(-1,1,0),(-1,0,1)\rangle, \end{aligned}$$

e, portanto $\{(-1,1,0),(-1,0,1)\}$ é um conjunto de geradores de V.

(ii) Por outro lado, $(-1,1,0)$, $(-1,0,1)$ não são múltiplos entre si e, portanto, $\{(-1,1,0),(-1,0,1)\}$ é linearmente independente.

De (i) e (ii) conclui-se que $\{(-1,1,0),(-1,0,1)\}$ é uma base de V.

Vejamos uma propriedade importante das bases de um subespaço vetorial que permite compreender melhor as duas condições da definição de base.

Teorema 24. *Sejam V um subespaço vetorial de $\mathbb{R}^m$ e $\{v_1,\ldots,v_n\}$ uma base de V. Então qualquer vetor $b \in V$ pode ser escrito como combinação linear de $v_1,\ldots,v_n$ de uma única forma.*

Demonstração. Seja $b \in V$ um vetor arbitrário. Como $\{v_1,\ldots,v_n\}$ é uma base de V, tem-se que $\langle v_1,\ldots,v_n\rangle = V$. Logo $b \in \langle v_1,\ldots,v_n\rangle$ o que significa que b é combinação linear de $v_1,\ldots,v_n$. Suponhamos agora que temos combinações lineares,

$$b = \alpha_1 v_1 + \cdots + \alpha_n v_n = \beta_1 v_1 + \cdots + \beta_n v_n,$$

em que $\alpha_1,\ldots,\alpha_n \in \mathbb{R}$ e $\beta_1,\ldots,\beta_n \in \mathbb{R}$. Então

$$\alpha_1 v_1 + \cdots + \alpha_n v_n - (\beta_1 v_1 + \cdots + \beta_n v_n) = b - b = \vec{0},$$

e portanto

$$\underbrace{(\alpha_1-\beta_1)}_{\in\mathbb{R}} v_1 + \cdots + \underbrace{(\alpha_n-\beta_n)}_{\in\mathbb{R}} v_n = \vec{0}.$$

[3]Note-se que $V = \mathcal{N}([1 \quad 1 \quad 1])$

Como $\{v_1 \ldots, v_n\}$ é linearmente independente (porque é base de V), a única combinação linear nula dos vetores $v_1 \ldots, v_n$ é a combinação **trivial**, isto é, com todos os coeficientes iguais a zero. Logo $\alpha_i - \beta_i = 0\ \forall i$, o que mostra que a combinação linear é única. □

Exemplo 39. Consideremos novamente $V = \{(x, y, z)\colon x + y + z = 0\}$ e a base $\{v_1, v_2\}$, com $v_1 = (-1, 1, 0)$ e $v_2 = (-1, 0, 1)$, do Exemplo 38.

Seja $b = (1, 2, -3)$. Então $b \in V$ pois satisfaz a relação $b_1 + b_2 + b_3 = 0$. Vejamos que b se pode exprimir de forma única como combinação linear de v_1, v_2. Ora,

$$[A|b] = \left[\begin{array}{cc|c} -1 & -1 & 1 \\ 1 & 0 & 2 \\ 0 & 1 & -3 \end{array}\right] \xrightarrow[L_2+L_1\to L_2]{} \left[\begin{array}{cc|c} -1 & -1 & 1 \\ 0 & -1 & 3 \\ 0 & 1 & -3 \end{array}\right] \xrightarrow[L_3+L_2\to L_3]{} \left[\begin{array}{cc|c} -1 & -1 & 1 \\ 0 & -1 & 3 \\ 0 & 0 & 0 \end{array}\right].$$

Logo, o sistema é possível determinado, e portanto b é combinação linear de v_1 e v_2 de uma única forma (determine os coeficientes desta combinação linear).

Terminamos esta secção com duas características importantes sobre os conjuntos de geradores e os conjuntos linearmente independentes de vetores, que permitem compreender melhor o conceito de base.

Seja $V = \langle v_1, \ldots, v_n\rangle$ um subespaço vetorial de $\mathbb{R}^m$, com $n \geq 2$. Se $\{v_1, \ldots, v_n\}$ for linearmente dependente então um dos vetores do conjunto $\{v_1, \ldots, v_n\}$ é combinação linear dos restantes. Para simplificar a notação podemos supôr que esse vetor é v_n (o raciocínio seria análogo se fosse qualquer outro vetor do conjunto). Nessa altura, resulta do Corolário 17 que

$$V = \langle v_1, \ldots, v_n\rangle = \langle v_1, \ldots, v_{n-1}\rangle,$$

e, portanto, que o vetor v_n não é necessário para gerar V, podendo ser removido do conjunto inicial de geradores.

Se $\{v_1, \ldots, v_{n-1}\}$ ainda for linearmente dependente podemos proceder de igual modo e remover um dos seus vetores de modo a continuar a gerar V e assim sucessivamente até se obter um conjunto linearmente independente de vetores que ainda gere V, isto é, até se obter uma base de V. Tem-se, portanto, o seguinte resultado.

Teorema 25. *Qualquer conjunto de geradores de um subespaço vetorial V pode ser reduzido, eliminando vetores desse conjunto, até formar uma base de V.*

O seguinte resultado relaciona os conjuntos linearmente independentes e as bases de um subespaço vetorial.

Teorema 26. *Qualquer conjunto linearmente independente de vetores de um subespaço vetorial V pode ser ampliado, acrescentado vetores de V, até formar uma base de V.*

Na próxima secção iremos ver que todas as bases de um subespaço vetorial possuem o mesmo número de vetores.

Dimensão de um subespaço vetorial

Segue-se um resultado que é apresentado sem demonstração.

Teorema 27. *Se $\{v_1, \ldots, v_n\}$ é uma base de um subespaço vetorial V, então qualquer subconjunto de V com mais que n vetores é linearmente dependente e qualquer subconjunto de V com menos que n vetores não gera V.*

Do resultado anterior conclui-se que qualquer outra base de V tem que possuir também n vetores. Tem-se portanto o seguinte resultado.

Teorema 28. *Todas as bases de um subespaço vetorial V têm a mesma cardinalidade, isto é, possuem o mesmo número de vetores.*

O resultado anterior valida a seguinte definição.

Definição 23. Chama-se *dimensão* de V, e denota-se por $\dim V$, ao número de vetores de qualquer base de V.

Exemplo 40. Vimos no Exemplo 34 que

$$V = \langle (1,-1,-2),(2,3,1) \rangle = \langle (1,1,0),(-1,0,1) \rangle.$$

Como os conjuntos de vetores $\{(1,-1,-2),(2,3,1)\}$ e $\{(1,1,0),(-1,0,1)\}$ geram V e são linearmente independentes (porque os vetores desses conjuntos não são múltiplos entre si), ambos os conjuntos são bases (distintas) de V. Em particular, $\dim V = 2$ (número de vetores de qualquer uma dessas bases).

Observação 18. Qualquer subespaço vetorial V não trivial ($V \neq \{\vec{0}\}$) admite uma *infinidade de bases distintas* (todas com a mesma cardinalidade).

Caracterização das bases de $\mathbb{R}^n$

Vimos na Observação 17 que $\mathbb{R}^n$ possui a base canónica, $\{e_1, \ldots, e_n\}$, que é constituída por n vetores e, portanto,

$$\boxed{\dim \mathbb{R}^n = n.}$$

Em particular, qualquer outra base de $\mathbb{R}^n$ possui necessariamente n vetores. Pretende-se caracterizar os conjuntos de n vetores que definem bases de $\mathbb{R}^n$. Sejam $v_1, \ldots, v_n \in \mathbb{R}^n$ e $A = [v_1 \ \cdots \ v_n]$. Têm-se as seguintes equivalências:

- Pela Proposição 21, $\{v_1, \ldots, v_n\}$ é linearmente independente se e só se $\text{car}(A)$ for igual ao número de colunas de A.
- Pela Proposição 18, $\langle v_1, \ldots, v_n \rangle$ gera $\mathbb{R}^n$ se e só se $\text{car}(A)$ for igual ao número de linhas de A.

Uma vez que A é uma matriz quadrada de ordem n e que uma base de $\mathbb{R}^n$ é um conjunto de n vetores de $\mathbb{R}^n$ linearmente independente e que gera $\mathbb{R}^n$, deduz-se imediatamente o seguinte resultado.

Teorema 29. *Sejam $v_1, \dots, v_n \in \mathbb{R}^n$. As seguintes afirmações são equivalentes:*

(i) $\{v_1, \dots, v_n\}$ *é uma base de* $\mathbb{R}^n$

(ii) $\{v_1, \dots, v_n\}$ *é linearmente independente.*

(iii) $\langle v_1, \dots, v_n \rangle = \mathbb{R}^n$.

Exemplo 41. Consideremos $v_1 = (1,2,1)$, $v_2 = (0,1,-1)$ e $v_3 = (2,3,4)$. Vimos no Exemplo 36.1 que $\{v_1, v_2, v_3\}$ é um conjunto linearmente independente de vetores de $\mathbb{R}^3$. Logo por (ii) do teorema anterior $\{v_1, v_2, v_3\}$ é uma base de $\mathbb{R}^3$.

2.5 Construção de bases para subespaços vetoriais

Comecemos por notar que um subespaço vetorial V de $\mathbb{R}^m$ pode ser definido de duas formas distintas:

- *Por um sistema de equações lineares homogéneas*, ou seja, como o *espaço nulo de uma matriz.* Por exemplo, o plano de $\mathbb{R}^3$ que passa na origem e é definido por uma equação cartesiana do tipo $ax + by + cz = 0$ (com a, b, c não todos nulos).

- *Gerado por um conjunto de vetores*, ou seja, definido como o *espaço das colunas de uma matriz.* Por exemplo, a reta de $\mathbb{R}^2$ que passa na origem e contém um dado vetor diretor $v \neq \vec{0}$.

Base para o espaço nulo de uma matriz

Pretende-se determinar uma base para um subespaço vetorial definido como *conjunto de soluções de sistema linear homogéneo.* Vejamos um exemplo.

Exemplo 42. Consideremos o hiperplano de $\mathbb{R}^5$,

$$V = \{(x_1, x_2, x_3, x_4, x_5) : x_1 - 2x_2 + x_4 + 3x_5 = 0\},$$

ou seja, $V = \mathcal{N}(A)$ em que $A = [1 \;\; -2 \;\; 0 \;\; 1 \;\; 3]$. Note-se que a matriz A já está reduzida com variável pivot x_1 e variáveis livres $x_2, \dots, x_5$. Então,

$$\begin{aligned}
V &= \{(x_1, x_2, x_3, x_4, x_5) \in \mathbb{R}^5 : x_1 - 2x_2 + x_4 + 3x_5 = 0\} \\
&= \{(x_1, x_2, x_3, x_4, x_5) \in \mathbb{R}^5 : x_1 = 2x_2 - x_4 - 3x_5,\ x_2 \in \mathbb{R}, x_3 \in \mathbb{R}, x_4 \in \mathbb{R}, x_5 \in \mathbb{R}\} \\
&= \{(2x_2 - x_4 - 3x_5, x_2, x_3, x_4, x_5) : x_2, \dots, x_5 \in \mathbb{R}\} \\
&= \{(2x_2, x_2, 0, 0, 0) + (0, 0, x_3, 0, 0) + (-x_4, 0, 0, x_4, 0) + (-3x_5, 0, 0, 0, x_5) : x_2, \dots, x_5 \in \mathbb{R}\} \\
&= \{x_2(2, 1, 0, 0, 0) + x_3(0, 0, 1, 0, 0) + x_4(-1, 0, 0, 1, 0) + x_5(-3, 0, 0, 0, 1) : x_2, \dots, x_5 \in \mathbb{R}\}.
\end{aligned}$$

A última expressão significa que $V = \langle v_1, v_2, v_3, v_4 \rangle$. Por outro lado, é imediato verificar que qualquer combinação linear nula dos vetores v_1, v_2, v_3 e v_4,

$$\alpha_1 \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \alpha_2 \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \alpha_3 \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + \alpha_4 \begin{bmatrix} -3 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

implica $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ e, portanto, que $\{v_1, v_2, v_3, v_4\}$ é um conjunto linearmente independente de vetores (alternativamente, pode-se mostrar a independência linear de $\{v_1, v_2, v_3, v_4\}$ aplicando o método de Gauss à matriz $[v_1 \ \ v_2 \ \ v_3 \ \ v_4]$, o que é deixado como exercício).

Logo $\{v_1, v_2, v_3, v_4\}$ é base para V, sendo $\dim V = 4$.

Repare-se que cada vetor desta base pode também ser obtido atribuíndo a uma das variáveis livres o valor 1, às restantes variáveis livres o valor zero e considerando a variável pivot $x_1 = 2x_2 - x_4 - 3x_5$.

Por exemplo, o vetor $v_1 = (2, 1, 0, 0, 0)$ pode ser obtido considerando $x_2 = 1$, $x_3 = x_4 = x_5 = 0$ e $x_1 = 2x_2 - x_4 - 3x_5 = 2$. Relativamente a $v_2 = (0, 0, 1, 0, 0)$, pode ser obtido considerando, $x_2 = 0$, $x_3 = 1$, $x_4 = x_5 = 0$ e $x_1 = 2x_2 - x_4 - 3x_5 = 0$. Analogamente para os restantes vetores da base (verifique!).

Tem-se que $\dim V = \dim \mathcal{N}(A) = 4$, que corresponde ao número de variáveis livres da equação homogénea $x_1 - 2x_2 + 0x_3 + x_4 + 3x_5 = 0$.

O procedimento descrito acima pode ser generalizado para construir bases do subespaço nulo de uma matriz.

Algoritmo 3. BASE PARA O ESPAÇO NULO DE UMA MATRIZ

`Input:` *Matriz A do tipo $m \times n$.*
`Objectivo:` *Base para $\mathcal{N}(A)$.*

1. *Seja A' matriz em escada obtida a partir de A por aplicação da fase descendente do método de eliminação de Gauss.*

2. *Se todas as colunas de A' têm pivot, $Ax = \vec{0}$ é determinado e $\mathcal{N}(A) = \{\vec{0}\}$. Nessa caso, $\dim \mathcal{N}(A) = 0$ e a base de $\mathcal{N}(A)$ é, por convenção, $\{\}$.*

3. *Se existem colunas sem pivot em A' aplicar a fase ascendente do método de Gauss para determinar $\mathcal{N}(A)$ e associar a cada variável livre do sistema (correspondente a cada coluna sem pivot de A') a solução de $Ax = \vec{0}$ em que essa variável livre toma o valor 1 (ou qualquer valor não nulo) e as restantes variáveis livres o valor zero. O conjunto de todas as soluções de $Ax = \vec{0}$ obtidas deste modo constitui uma base para $\mathcal{N}(A)$.*

Exemplo 43. Consideremos $A = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & 2 & -1 & -2 \end{bmatrix}$. Para calcular uma base de $\mathcal{N}(A)$ reduzimos o sistema $Ax = 0$ (note que este sistema nunca pode ser determinado):

$$A = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & 2 & -1 & -2 \end{bmatrix} \xrightarrow[L_2 - L_1 \to L_2]{} \begin{bmatrix} 1 & 1 & -1 & -1 \\ 0 & 1 & 0 & -1 \end{bmatrix} \xrightarrow[L_1 - L_2 \to L_1]{} \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix} = A'.$$

A última matriz corresponde ao sistema reduzido,

$$\begin{cases} x_1 \quad - x_3 \qquad = 0 \\ \qquad x_2 \quad - x_4 = 0 \end{cases} \Leftrightarrow \begin{cases} x_1 = x_3 \\ x_2 = x_4 \end{cases},$$

em que x_3, x_4 são as variáveis livres associadas às colunas sem pivot de A'.

Considerando $x_3 = 1$ e $x_4 = 0$ obtém-se a solução de $Ax = 0$, $v_1 = (1,0,1,0)$, e tomando $x_3 = 0$ e $x_4 = 1$ a solução $v_2 = (0,1,0,1)$. Logo, uma base para $\mathcal{N}(A)$ é $\{v_1, v_2\}$, tendo-se $\dim \mathcal{N}(A) = 2$.

Exercício. *Reobtenha a base $\{v_1, v_2\}$ do exemplo anterior usando um procedimento análogo ao que foi utilizado no Exemplo 42.*

Base para o espaço de colunas de uma matriz

Pretende-se obter uma base para um subespaço definido por um conjunto de geradores, isto é, como o espaço das colunas de uma matriz.

Algoritmo 4. BASE PARA O ESPAÇO GERADO POR VETORES

`Input:` *$v_1, \ldots, v_n \in \mathbb{R}^m$ ou $A = [v_1 \cdots v_n]$.*

`Objectivo:` *Base para $\langle v_1, \ldots, v_n \rangle = \mathscr{C}(A)$.*

1. *Seja A' matriz em escada obtida a partir de A por aplicação da fase descendente do método de eliminação de Gauss.*

2. *O subconjunto dos vetores de $\{v_1, \ldots, v_n\}$ constituído pelas colunas de A que correspondem às colunas com pivot em A' é uma base do subespaço vetorial $\langle v_1, \ldots, v_n \rangle = \mathscr{C}(A)$ contida no conjunto inicial de geradores.*

Exemplo 44. Vamos calcular uma base do subespaço vetorial $V \subset \mathbb{R}^3$ gerado pelos vetores $v_1 = (1,-1,-2)$, $v_2 = (2,3,1)$ e $v_3 = (-1,-2,-1)$. Ora tem-se,

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

Pelo algortimo 4, $\{v_1, v_2\}$ constitui uma base de $V = \langle v_1, v_2, v_3 \rangle$, contida no conjunto inicial de geradores de V. Em particular, $V = \langle v_1, v_2, v_3 \rangle = \langle v_1, v_2 \rangle$.

Observação 19. Também se pode obter uma base para um subespaço vetorial $V = \langle v_1, \dots, v_n \rangle$, começando por definir V como o conjunto de soluções de um sistema linear homogéneo (que pode ser obtido aplicando o método de eliminação de Gauss à matriz $\left[\begin{array}{ccc|c} v_1 & \cdots & v_n & b \end{array}\right]$, com b genérico) e depois aplicando o Algoritmo 3 para obter uma base de V. Note-se no entanto que esta base não está, em geral, contida no conjunto inicial de geradores de V.

Deixa-se como exercício obter uma base para o subespaço vetorial V do exemplo anterior determinando um sistema de equações lineares homogéneas para V e aplicando o Algoritmo 3.

Seja A matriz de tipo $m \times n$ e A' uma matriz em escada obtida a partir de A por aplicação do método de Gauss. Nos dois algorítmos anteriores viu-se que:

- $\dim \mathcal{N}(A) =$ número de variáveis livres do sistema homogéneo $Ax = \vec{0}$
  $=$ número de colunas sem *pivot* da matriz A'
  $= n - \text{car}(A)$
- $\dim \mathcal{C}(A) =$ número de variáveis *pivot* do sistema homogéneo $Ax = \vec{0}$
  $=$ número de colunas com *pivot* da matriz A'
  $= \text{car}(A)$

Daqui resulta seguinte relação fundamental entre as duas dimensões referidas anteriormente.

Proposição 30. *Se A é uma matriz de tipo $m \times n$, tem-se*

$$\boxed{\dim \mathcal{C}(A) + \dim \mathcal{N}(A) = \textit{número de colunas de } A = n}$$

Em particular, o conhecimento da dimensão de um dos subespaços $\mathcal{C}(A)$ ou da $\mathcal{N}(A)$ implica o conhecimento da dimensão do outro subespaço.

A informação sobre a dimensão de um subespaço vetorial tem diversas implicações importantes como mostram os seguintes resultados.

Teorema 31. *Sejam U e V subespaços vetoriais de $\mathbb{R}^m$ com $U \subset V$. Tem-se então que $U = V$ se e só se* $\dim U = \dim V$.

Como consequência do teorema anterior obtém-se o seguinte resultado que estabelece condições mais fracas para que um conjunto de vetores defina uma base de um subespaço vetorial.

Corolário 32. *Sejam V um subespaço vetorial de $\mathbb{R}^m$ e $v_1, \dots, v_n$ vetores de V. Para que $\{v_1, \dots, v_n\}$ constitua uma base de V basta que duas das 3 condições seguintes estejam verificadas:*

(i) *$\{v_1, \dots, v_n\}$ é linearmente independente.*

(i) $\langle v_1, \dots, v_n \rangle = V$.

(iii) $\dim V = n$.

Demonstração. (i) e (ii) implicam, por definição, que $\{v_1, \dots, v_n\}$ é base de V e portanto que $\dim V = n$.

Vejamos que (i) e (iii) também implicam que $\{v_1, \dots, v_n\}$ é base de V. Seja $U = \langle v_1, \dots, v_n \rangle \subset V$. Como $\{v_1, \dots, v_n\}$ é linearmente independente e gera U, $\{v_1, \dots, v_n\}$ é base de U. Logo, $\dim U = n = \dim V$. Como $U \subset V$, resulta do teorema anterior que $U = V$ e portanto que $\{v_1, \dots, v_n\}$ é base de V. Finalmente suponhamos que (ii) e (iii) são verificadas. Se $\{v_1, \dots, v_n\}$ fosse linearmente dependente, um dos vetores seria combinação linear dos restantes. Podemos supor, sem perda de generalidade, que esse vetor era v_n. Nessa altura ter-se-ia $V = \langle v_1, \dots, v_n \rangle = \langle v_1, \dots, v_{n-1} \rangle$, o que estaria em contradição com os resultados do Teorema 27, uma vez que as bases de V contêm n vetores. □

Note-se que o corolário anterior diz-nos, em particular, que quaisquer duas das 3 condições implica a condição que falta.

Terminamos esta secção com algumas observações sobre a caracterização do tipo de conjunto de um subespaço vetorial a partir da respetiva dimensão.

Em complemento do quadro da Observação 10 tem-se o seguinte.

	subespaços vetoriais	dimensão
$\mathbb{R}^2$	$\{\vec{0}\}$	0
	retas que passam na origem	1
	$\mathbb{R}^2$	2
$\mathbb{R}^3$	$\{\vec{0}\}$	0
	retas que passam na origem	1
	planos que passam na origem	2
	$\mathbb{R}^3$	3

Observação 20. Seja V é um subespaço vetorial de $\mathbb{R}^m$. Tem-se:

- $V = \{\vec{0}\}$ (subespaço minimal) $\iff$ $\dim V = 0$
- $V = \mathbb{R}^m$ (subespaço maximal) $\iff$ $\dim V = m$

Capítulo 3

Ortogonalidade e Projeção Ortogonal

No primeiro capítulo estendemos o conceito de ortogonalidade, já conhecido para vetores do plano e do espaço, a vetores com um número arbitrário de componentes. Neste capítulo vamos alargar o âmbito do estudo das relações de ortogonalidade aos subespaços vetoriais.

3.1 Complemento ortogonal

Definição 24. Seja $u \in \mathbb{R}^m$ e V subespaço vetorial de $\mathbb{R}^m$. Diz-se que u é *ortogonal* a V, e denota-se $u \perp V$, se u for ortogonal a todos os vetores de V, isto é, se $u \cdot v = 0\ \forall v \in V$.

Exemplos 45.

1. O vetor $(1,2,3)$ é ortogonal ao plano $V = \{(x_1, x_2, x_3) \colon x_1{+}2x_2{+}3x_3 = 0\}$, razão pela qual este vetor é também designado por *vetor normal* ao plano. De facto, tem-se
$$\begin{aligned} V &= \{(x_1,x_2,x_3) \in \mathbb{R}^3 \colon x_1 + 2x_2 + 3x_3 = 0\} \\ &= \{(x_1,x_2,x_3) \in \mathbb{R}^3 \colon (x_1,x_2,x_3) \cdot (1,2,3) = 0\} \\ &= \{(x_1,x_2,x_3) \in \mathbb{R}^3 \colon (x_1,x_2,x_3) \perp (1,2,3)\}, \end{aligned}$$
o que mostra que V consiste no conjunto dos vetores de $\mathbb{R}^3$ que são ortogonais ao vetor $(1,2,3)$. Portanto $(1,2,3) \perp V$.

2. O vetor nulo é o único vetor ortogonal ao subespaço $\mathbb{R}^n$. De facto, se u for ortogonal a todos os vetores de $\mathbb{R}^n$ tem-se, em particular, que $u \perp u$. Logo $u \cdot u = \|u\|^2 = 0$ e portanto $u = \vec{0}$.

O seguinte resultado mostra que para um vetor ser ortogonal a um subespaço vetorial é suficiente que seja ortogonal a um conjunto de geradores desse subespaço.

Teorema 33. *Sejam $u, v_1, \dots, v_n \in \mathbb{R}^m$ e $V = \langle v_1, \dots, v_n \rangle$. Então $u \perp V$ se e só se $u \perp v_1, \dots, u \perp v_n$.*

Demonstração. Se $u \perp V$ então $u \perp v$ para todo o $v \in V$. Em particular, $u \perp v_i$ para $i = 1, \dots, n$, pois $v_i \in V$.

Reciprocamente, suponhamos que $u \perp v_i$ para $i = 1, \dots, n$, isto é, $u \cdot v_i = 0$, $i = 1, \dots, n$. Seja $v \in V$ um vetor arbitário. Como $V = \langle v_1, \dots, v_n \rangle$, existem escalares $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ tais que $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$ e obtém-se

$$v \cdot u = (\alpha_1 v_1 + \cdots + \alpha_n v_n) \cdot u = \alpha_1 \underbrace{(v_1 \cdot u)}_{0} + \cdots + \alpha_n \underbrace{(v_n \cdot u)}_{0} = 0.$$

Logo $u \perp v$ para todo o $v \in V$ e portanto $u \perp V$. □

Exemplo 46. Seja $V = \langle v_1, v_2, v_3 \rangle$ com $v_1 = (1, 1, 2)$, $v_2 = (2, 3, 5)$ e $v_3 = (3, 4, 7)$ e consideremos o vetor $u = (-1, -1, 1)$. Tem-se, $u \cdot v_1 = 0$, $u \cdot v_2 = 0$ e $u \cdot v_3 = 0$ (verifique). Logo, pelo Teorema 33, $u \perp V$.

Como as bases constituem conjuntos geradores de um subespaço vetorial deduz-se imediatamente o seguinte resultado do Teorema 33.

Corolário 34. *Um vetor é ortogonal a um subespaço vetorial de $\mathbb{R}^m$ se e só se for ortogonal a todos os vetores de uma base desse subespaço.*

Definição 25. Seja V um subespaço vetorial de $\mathbb{R}^m$. Chama-se *complemento ortogonal* de V, e denota-se por $V^\perp$, ao conjunto de todos os vetores de $\mathbb{R}^m$ que são ortogonais a V, isto é,

$$V^\perp = \{u \in \mathbb{R}^m : u \perp V\}.$$

Exemplos 47.

1. $(\mathbb{R}^m)^\perp = \{\vec{0}\}$, pois pelo Exemplo 45.2 o vetor nulo é o único vetor ortogonal a $\mathbb{R}^m$.

2. $\{\vec{0}\}^\perp = \mathbb{R}^m$, pois qualquer vetor é ortogonal ao vetor nulo. De facto, $v \cdot \vec{0} = 0$, $\forall v \in \mathbb{R}^m$.

Recordemos que o produto escalar $x \cdot y$ entre dois vetores $x = (x_1, x_2, \dots, x_m)$ e $y = (y_1, y_2, \dots, y_m)$ de $\mathbb{R}^m$ pode ser escrito na forma matricial como $x^T y$ (ver a Observação 2).

Exemplo 48. Seja $V = \langle v_1, v_2, v_3 \rangle$ com $v_1 = (1, 1, 2)$, $v_2 = (2, 3, 5)$ e $v_3 = (3, 4, 7)$. Pretende-se determinar $V^\perp$. Denotando $A = [\, v_1 \;\; v_2 \;\; v_3 \,]$, tem-se $V = \mathscr{C}(A)$. Para todo o vetor $b=(b_1,b_2,b_3)\in\mathbb{R}^3$, obtém-se pelo Teorema 33,

$$
\begin{aligned}
b\perp V=\mathscr{C}(A) \quad &\Leftrightarrow \quad \begin{cases} v_1\perp b\\ v_2\perp b\\ v_3\perp b\end{cases} \Leftrightarrow \begin{cases} v_1\cdot b=0\\ v_2\cdot b=0\\ v_3\cdot b=0\end{cases} \Leftrightarrow \begin{cases} v_1^T b=0\\ v_2^T b=0\\ v_3^T b=0\end{cases}\\
&\Leftrightarrow \quad \begin{cases} [\,1\;\;1\;\;2\,]\begin{bmatrix} b_1\\ b_2\\ b_3\end{bmatrix}=0\\[2ex] [\,2\;\;3\;\;5\,]\begin{bmatrix} b_1\\ b_2\\ b_3\end{bmatrix}=0\\[2ex] [\,3\;\;4\;\;7\,]\begin{bmatrix} b_1\\ b_2\\ b_3\end{bmatrix}=0\end{cases}\\
&\Leftrightarrow \quad \begin{bmatrix} 1 & 1 & 2\\ 2 & 3 & 5\\ 3 & 4 & 7\end{bmatrix}\begin{bmatrix} b_1\\ b_2\\ b_3\end{bmatrix}=\begin{bmatrix} 0\\ 0\\ 0\end{bmatrix}.
\end{aligned}
$$

A matriz quadrada de ordem 3 anterior corresponde à matriz cujas linhas são os vetores v_1, v_2 e v_3, isto é, à matriz A^T. Conclui-se portanto que $b\perp V=\mathscr{C}(A)$ se e só se $b\in\mathscr{N}(A^T)$. Logo, $V^\perp=\mathscr{C}(A)^\perp=\mathscr{N}(A^T)$. Reduzindo a matriz A^T para determinar $V^\perp=\mathscr{N}(A^T)$ obtém-se,

$$
\begin{aligned}
&\begin{bmatrix} 1 & 1 & 2\\ 2 & 3 & 5\\ 3 & 4 & 7\end{bmatrix} \xrightarrow[L_3-3L_1\to L_3]{L_2-2L_1\to L_2} \begin{bmatrix} 1 & 1 & 2\\ 0 & 1 & 1\\ 0 & 1 & 1\end{bmatrix} \xrightarrow[L_3-L_2\to L_3]{}\\
&\begin{bmatrix} 1 & 1 & 2\\ 0 & 1 & 1\\ 0 & 0 & 0\end{bmatrix} \xrightarrow[L_1-L_2\to L_1]{} \begin{bmatrix} 1 & 0 & 1\\ 0 & 1 & 1\\ 0 & 0 & 0\end{bmatrix}.
\end{aligned}
$$

A matriz reduzida anterior corresponde ao sistema reduzido

$$
\begin{cases} x_1 \quad\quad + x_3 = 0\\ \quad\quad x_2 + x_3 = 0\end{cases}
$$

e portanto

$$
\begin{aligned}
V^\perp &= \{(x_1,x_2,x_3)\colon x_1=-x_3, x_2=-x_3,\ x_3\in\mathbb{R}\}=\{(-x_3,-x_3,x_3)\colon x_3\in\mathbb{R}\}\\
&= \{x_3(-1,-1,1)\colon x_3\in\mathbb{R}\}=\langle(-1,-1,1)\rangle.
\end{aligned}
$$

Logo $V^\perp$ define a reta em $\mathbb{R}^3$ que passa na origem com vetor diretor $(-1,-1,1)$.

O tipo de cálculos que foi aplicado no Exemplo 48 pode ser generalizado para subespaços vetoriais de $\mathbb{R}^m$, gerados por um número arbitrário de vetores. Mais precisamente, tem-se a seguinte observação.

Observação 21. Sejam $v_1,\dots,v_n \in \mathbb{R}^m$ e $A = [\, v_1 \ \cdots \ v_n \,]_{m\times n}$. Tem-se a relação fundamental,

$$\boxed{\langle v_1,\dots,v_n\rangle^\perp = \mathscr{C}(A)^\perp = \mathscr{N}(A^T).}$$

Uma vez que qualquer subespaço vetorial V de $\mathbb{R}^m$ pode ser obtido como espaço das colunas de uma matriz do tipo $m\times n$, cujas colunas são n geradores desse subespaço, o complemento ortogonal $V^\perp$ pode ser obtido como o espaço nulo de uma matriz do tipo $n\times m$. Em particular, deduz-se o seguinte resultado.

Teorema 35. *O complemento ortogonal de um subespaço vetorial de* $\mathbb{R}^m$ *define um subespaço vetorial de* $\mathbb{R}^m$.

Enunciam-se a seguir várias propriedades importantes do complemento ortogonal.

Teorema 36. *Seja* V *é subespaço vetorial de* $\mathbb{R}^m$. *Tem-se:*

1. $V\cap V^\perp = \{\vec{0}\}$.
2. $\dim V + \dim V^\perp = \dim \mathbb{R}^m = m$.
3. *Se* $\{v_1,\dots,v_k\}$ *é uma base de* V *e* $\{u_1,\dots,u_{m-k}\}$ *é uma base de* $V^\perp$, *então* $\{v_1,\dots,v_k,\ u_1,\dots,u_{m-k}\}$ *é uma base de* $\mathbb{R}^m$.
4. $(V^\perp)^\perp = V$, *ou seja, se* $U = V^\perp \Rightarrow U^\perp = V$.

Demonstração.

1. Se $v\in V\cap V^\perp$, então $v\in V$ e $v\in V^\perp$. Logo $v\perp v$, ou seja, $v\cdot v = \|v\|^2 = 0$. Portanto $v=\vec{0}$, o que mostra que $V\cap V^\perp = \{\vec{0}\}$.
2. Podemos supor $V\neq\{\vec{0}\}$, uma vez que no caso $V=\{\vec{0}\}$ tem-se $V^\perp = \mathbb{R}^m$ (ver o Exemplo 47), pelo que o resultado é verificado.

   Podemos escrever $V=\mathscr{C}(A)$, com A uma matriz cujas colunas de A são os vetores que constituem uma base de V. Nessa altura,

   $$\dim V = \dim \mathscr{C}(A) = \operatorname{car}(A). \tag{3.1}$$

   Por outro lado, pelo Teorema 21, tem-se

   $$\dim V^\perp = \dim \mathscr{C}(A)^\perp = \dim \mathscr{N}(A^T) = m - \dim \mathscr{C}(A^T) = m - \operatorname{car}(A^T). \tag{3.2}$$

   Como $\operatorname{car}(A) = \operatorname{car}(A^T)$, resulta das relações (3.1) e (3.2) que

   $$\dim V + \dim V^\perp = \operatorname{car}(A) + (m - \operatorname{car}(A)) = m.$$

3. Para mostrar que $\{v_1,\ldots,v_k,\, u_1,\ldots,u_{m-k}\}$ é uma base $\mathbb{R}^m$, basta mostrar que é um conjunto linearmente independente de vetores (ver a Proposição 29). Consideremos uma combinação linear nula dos vetores desse conjunto,

   $$\alpha_1 v_1 + \cdots + \alpha_k v_k + \beta_1 u_1 + \cdots + \beta_{m-k} u_{m-k} = \vec{0},$$

   com $\alpha_1,\ldots,\alpha_k,\beta_1,\ldots,\beta_{m-k} \in \mathbb{R}$. Então

   $$\alpha_1 v_1 + \cdots + \alpha_k v_k = -(\beta_1 u_1 + \cdots + \beta_{m-k} u_{m-k}).$$

   Como $\alpha_1 v_1 + \cdots + \alpha_k v_k \in V$ (por ser uma combinação de vetores de V) e $\beta_1 u_1 + \cdots + \beta_{m-k} u_{m-k} \in V^\perp$ (por ser uma combinação de vetores de $V^\perp$), resulta, da igualdade anterior, que

   $$\begin{gathered} \alpha_1 v_1 + \cdots + \alpha_k v_k \in V \cap V^\perp = \{\vec{0}\}, \\ \beta_1 u_1 + \cdots + \beta_{m-k} u_{m-k} \in V \cap V^\perp = \{\vec{0}\}. \end{gathered}$$

   Portanto

   $$\alpha_1 v_1 + \cdots + \alpha_k v_k = \vec{0} \quad \text{e} \quad \beta_1 u_1 + \cdots + \beta_{m-k} u_{m-k} = \vec{0}.$$

   Como $\{v_1,\ldots,v_k\}$ e $\{u_1,\ldots,u_{m-k}\}$ são linearmente independentes (pois são bases), tem-se $\alpha_1 = \cdots = \alpha_k = 0$ e $\beta_1 = \cdots = \beta_{m-k} = 0$, o que mostra que $\{v_1,\ldots,v_k,\, u_1,\ldots,u_{m-k}\}$ é um conjunto linearmente independente de vetores, e portanto que é uma base de $\mathbb{R}^m$.

4. Todo o elemento de V pertence a $(V^\perp)^\perp$, uma vez que por definição de $V^\perp$, todo o vetor de $V^\perp$ é ortogonal a todo o vetor de V. Logo $V \subset (V^\perp)^\perp$. Como $\dim V = k$ e $\dim V^\perp = m-k$, resulta que

   $$\dim(V^\perp)^\perp = \dim \mathbb{R}^m - \dim V^\perp = m-(m-k) = k.$$

   Logo $\dim V = \dim(V^\perp)^\perp$. Como $V \subset (V^\perp)^\perp$, conclui-se que $V = (V^\perp)^\perp$. □

Para qualquer matriz A tem-se pela Observação 21,

$$\mathscr{C}(A)^\perp = \mathscr{N}(A^T) \Rightarrow (\mathscr{C}(A)^\perp)^\perp = \mathscr{N}(A^T)^\perp$$

Como $(\mathscr{C}(A)^\perp)^\perp = \mathscr{C}(A)$, obtém-se $\mathscr{C}(A) = \mathscr{N}(A^T)^\perp$. Substituindo A por A^T na expressão anterior, obtém-se $\mathscr{C}(A^T) = \mathscr{N}((A^T)^T)^\perp = \mathscr{N}(A)^\perp$. Portanto $\mathscr{N}(A)^\perp = \mathscr{C}(A^T)$.

Observação 22. As relações de ortogonalidade estabelecidas anteriormente,

$$\boxed{\begin{gathered} \mathscr{C}(A)^\perp = \mathscr{N}(A^T) \\ \mathscr{N}(A)^\perp = \mathscr{C}(A^T) \end{gathered}}$$

permitem determinar o complemento ortogonal de um subespaço vetorial V definido por geradores, $V = \langle v_1, \ldots, v_n \rangle$, considerando $V = \mathscr{C}(A)$ em que $A = [\, v_1 \ \cdots \ v_n \,]$, e de um subespaço vetorial V definido por um sistema de equações lineares homogéneas, escrevendo $V = \mathscr{N}(A)$ em que A é a matriz dos coeficientes desse sistema.

Exemplo 49. Consideremos $V = \{(x, y, z) \in \mathbb{R}^3 : x + y + z = 0, x + 2y + 3z = 0\}$ que representa a intersecção de dois planos concorrentes numa reta que passa na origem. Podemos escrever $V = \mathscr{N}(A)$ com $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix}$ e portanto

$$V^\perp = \mathscr{N}(A)^\perp = \mathscr{C}(A^T) = \mathscr{C}\left(\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}\right) = \langle (1,1,1), (1,2,3) \rangle,$$

que representa o plano perpendicular à reta V, que passa na origem e definido pelas direções $(1,1,1)$ e $(1,2,3)$. Note-se que V pode também ser escrito como,

$$\begin{aligned} V &= \{(x, y, z) \in \mathbb{R}^3 : x + y + z = 0, x + 2y + 3z = 0\} \\ &= \{(x, y, z) \in \mathbb{R}^3 : (x, y, z) \cdot (1,1,1) = 0, (x, y, z) \cdot (1,2,3) = 0\} \\ &= \{(x, y, z) \in \mathbb{R}^3 : (x, y, z) \perp (1,1,1), (x, y, z) \perp (1,2,3)\} = \langle (1,1,1), (1,2,3) \rangle^\perp, \end{aligned}$$

reobtendo-se, $V^\perp = (\langle (1,1,1), (1,2,3) \rangle^\perp)^\perp = \langle (1,1,1), (1,2,3) \rangle$.

Terminamos esta secção com um resumo dos complementos ortogonais dos subespaços vetoriais do plano e do espaço.

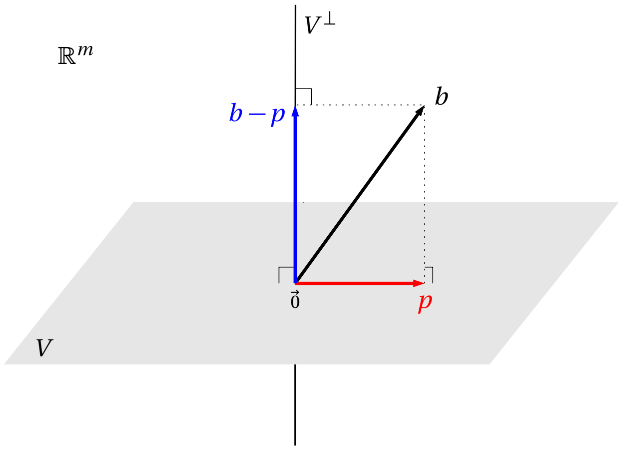
Observação 23. Quadro-resumo dos subespaços vetoriais de $\mathbb{R}^2$ (plano) e $\mathbb{R}^3$ (espaço) e respetivos complementos ortogonais.

	subespaço vetorial V	complemento ortogonal $V^\perp$
$\mathbb{R}^2$	$\{\vec{0}\}$	$\mathbb{R}^2$
	reta que passa na origem	reta perpendicular que passa na origem
	$\mathbb{R}^2$	$\{\vec{0}\}$
$\mathbb{R}^3$	$\{\vec{0}\}$	$\mathbb{R}^3$
	reta que passa na origem	plano perpendicular que passa na origem
	plano que passa na origem	reta perpendicular que passa na origem
	$\mathbb{R}^3$	$\{\vec{0}\}$

3.2 Projeção ortogonal

O seguinte resultado, que é consequência do Teorema 36, é fundamental para a definição de projeção ortogonal.

Teorema 37. *Se V é um subespaço vetorial de $\mathbb{R}^m$ e $b \in \mathbb{R}^m$ então existe um e um só $p \in V$ tal que $b - p \in V^\perp$.*



Demonstração. Vamos começar por provar a existência do vetor $p \in V$ tal que $b - p \in V^\perp$.

Se $V = \mathbb{R}^m$ tem-se $V^\perp = \{\vec{0}\}$ e basta considerar $p = b \in V$ e $b - p = \vec{0}$. De modo análogo, se $V = \{\vec{0}\}$ então $V^\perp = \mathbb{R}^m$, bastando considerar $p = \vec{0} \in V$ e $b - p = b \in \mathbb{R}^m$. Podemos então supor que $V \neq \mathbb{R}^m, \{\vec{0}\}$. Em particular, tem-se $1 \leq \dim V < m$ e $1 \leq \dim V^\perp < m$. Sejam $\{v_1, \ldots, v_k\}$ e $\{u_1, \ldots, u_{m-k}\}$ bases de V e $V^\perp$, respectivamente. Pelo Teorema 36 $\{v_1, \ldots, v_k, u_1, \ldots, u_{m-k}\}$ é uma base de $\mathbb{R}^m$ e portanto existem escalares (únicos) $\alpha_1, \ldots, \alpha_k, \beta_1, \ldots, \beta_{m-k} \in \mathbb{R}$ tais que

$$b = \alpha_1 v_1 + \cdots + \alpha_k v_k + \beta_1 u_1 + \cdots + \beta_{m-k} u_{m-k}.$$

Logo podemos considerar $p = \alpha_1 v_1 + \cdots + \alpha_k v_k \in V$ uma vez que $\{v_1, \ldots, v_k\}$ é base de V, tendo-se $b - p = \beta_1 u_1 + \cdots + \beta_{m-k} u_{m-k} \in V^\perp$, pois $\{u_1, \ldots, u_{m-k}\}$ é base de $V^\perp$.

Para mostrar a unicidade consideremos $p_1, p_2 \in V$ tais que $b - p_1 \in V^\perp$ e $b - p_2 \in V^\perp$. Nessa altura,

$$\vec{0} = b - b = (p_1 + b - p_1) - (p_2 + b - p_2) = (p_1 - p_2) + (b - p_1) - (b - p_2).$$

Como $p_1, p_2 \in V$, $p_1 - p_2 \in V$. Analogamente, como $b - p_1, b - p_2 \in V^\perp$ tem-se $(b - p_1) - (b - p_2) = p_2 - p_1 \in V^\perp$. Logo $p_1 - p_2 \in V \cap V^\perp = \{\vec{0}\}$ e portanto $p_1 - p_2 = \vec{0}$. Logo $p_1 = p_2$ o que mostra a unicidade da decomposição. □

O teorema anterior valida a seguinte definição.

Definição 26. Sejam $b \in \mathbb{R}^m$ e V subespaço vetorial de $\mathbb{R}^m$. Chama-se *projeção ortogonal* de b sobre V, e denota-se por $\text{proj}_V(b)$, ao único vetor $p \in V$ tal que $b - p \in V^\perp$.

Observação 24. Se $b \in \mathbb{R}^m$ e V é subespaço vetorial de $\mathbb{R}^m$, tem-se a decomposição,

$$\boxed{b = \text{proj}_V(b) + \text{proj}_{V^\perp}(b).}$$

Com efeito, sejam $p = \text{proj}_V(b) \in V$ e $q = b - p \in V^\perp$. Então,

$$b - q = b - (b - p) = p \in V = (V^\perp)^\perp.$$

Portanto $q \in V^\perp$ e $b - q \in (V^\perp)^\perp$ o que significa, por definição, que $q = \text{proj}_{V^\perp}(b)$. Logo, $b = p + (b - p) = p + q = \text{proj}_V(b) + \text{proj}_{V^\perp}(b)$.

Exemplo 50. Consideremos o subespaço vetorial $V = \langle v_1, v_2 \rangle$ com $v_1 = (1,1,2)$ e $v_2 = (-1,1,0)$, e seja $b = (-1,1,3)$. Pretende-se mostrar que $\text{proj}_V(b) = (0,2,2)$. Por definição temos que mostrar que $p = (0,2,2)$ verifica as duas condições:

(i) $p \in V$;

(ii) $b - p \perp V$, isto é, $b - p \in V^\perp$.

Ora,

(i) $p \in V = \langle v_1, v_2 \rangle = \mathscr{C}(A)$, com $A = \begin{bmatrix} v_1 & v_2 \end{bmatrix}$, se e só se o sistema linear $[\, A \,|\, p \,]$ for possível. Tem-se,

$$[\, A \,|\, p \,] = \left[\begin{array}{rr|r} 1 & -1 & 0 \\ 1 & 1 & 2 \\ 2 & 0 & 2 \end{array}\right] \xrightarrow[\substack{L_2 - L_1 \to L_2 \\ L_3 - 2L_1 \to L_3}]{} \left[\begin{array}{rr|r} 1 & -1 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 2 \end{array}\right] \xrightarrow[L_3 - L_2 \to L_3]{} \left[\begin{array}{rr|r} 1 & -1 & 0 \\ 0 & 2 & 2 \\ 0 & 0 & 0 \end{array}\right],$$

o que mostra que o sistema $[\, A \,|\, p \,]$ é possível, isto é, que $p \in V$.

(ii) $b - p = (-1,1,3) - (0,2,2) = (-1,-1,1) \in V^\perp = \mathscr{C}(A)^\perp = \mathscr{N}(A^T)$ se e só se $A^T(b - p) = \vec{0}$. De facto,

$$A^T(b - p) = \begin{bmatrix} 1 & 1 & 2 \\ -1 & 1 & 0 \end{bmatrix} \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Portanto $b - p \in V^\perp$.

Alternativamente, pode-se mostrar que $b - p \in V^\perp$ mostrando que $b - p$ é ortogonal aos geradores de V. De facto, tem-se

$$\begin{cases} (b - p) \cdot v_1 = (-1,-1,1) \cdot (1,1,2) = 0 \\ (b - p) \cdot v_2 = (-1,-1,1) \cdot (-1,1,0) = 0 \end{cases},$$

o que mostra que $(b - p) \perp v_1$ e $(b - p) \perp v_2$.

Como (i) e (ii) são ambas satisfeitas, $\text{proj}_V(b) = (0,2,2)$.

Nos casos em que $V = \mathbb{R}^m$ ou $V = \{\vec{0}\}$, a projeção ortogonal de qualquer vetor $b \in \mathbb{R}^m$ sobre V é imediata (cf. com a demonstração do Teorema 37):

- Se $V = \mathbb{R}^m$ então $\text{proj}_V(b) = \text{proj}_{\mathbb{R}^m}(b) = b$.
- Se $V = \{\vec{0}\}$ então $\text{proj}_V(b) = \text{proj}_{\{\vec{0}\}}(b) = \vec{0}$.

No caso em que as situações particulares anteriores não ocorrem, podemos usar o mesmo tipo de argumento que foi usado na demonstração do Teorema 37 para derivar um algoritmo, que vamos designar, informalmente, por *método das bases*, para calcular a projeção ortogonal de um vetor b sobre V.

Algoritmo 5. MÉTODO DAS BASES

`Input:` *Subespaço vetorial* V *de* $\mathbb{R}^m$, *com* $V \neq \{\vec{0}\}, \mathbb{R}^m$, *e* $b \in \mathbb{R}^m$.

`Objectivo:` *Calcular* $\text{proj}_V(b)$ *e* $\text{proj}_{V^\perp}(b)$

1. *Determinar uma base* $\{v_1, \ldots, v_k\}$ *de* V*;*
2. *Determinar uma base* $\{u_1, \ldots, u_{m-k}\}$ *de* $V^\perp$*;*
3. *Escrever* b *como combinação linear dos vetores da base de* $\mathbb{R}^m$ *obtida como* <u>*união das bases de* V *e* $V^\perp$</u>, $\{v_1, \ldots, v_k, u_1, \ldots, u_{m-k}\}$,
$$b = \alpha_1 v_1 + \cdots + \alpha_k v_k + \beta_1 u_1 + \cdots + \beta_{m-k} u_{m-k},$$
em que $(\alpha_1, \ldots, \alpha_k, \beta_1, \ldots, \beta_{m-k})$ *é a solução (única) do sistema linear com matriz ampliada*
$$\left[\begin{array}{cccccc|c} v_1 & \cdots & v_k & u_1 & \cdots & u_{m-k} & b \end{array}\right]. \tag{3.3}$$
4. *Tem-se então que* $\begin{cases} \text{proj}_V(b) = \alpha_1 v_1 + \cdots + \alpha_k v_k \\ \text{proj}_{V^\perp}(b) = \beta_1 u_1 + \cdots + \beta_{m-k} u_{m-k}. \end{cases}$

Exemplo 51. Consideremos $b = (3,2,0,3)$, $v_1 = (1,1,-1,0)$, $v_2 = (0,1,0,1)$, e o subespaço vetorial $V = \langle v_1, v_2 \rangle$. Pretende-se calcular $\text{proj}_V(b)$. Tem-se:

1. $\{v_1, v_2\}$ é base de V, uma vez que gera V e é linearmente independente (v_1 e v_2 são não colineares).
2. Considerando $A = [v_1\ v_2]$ tem-se $V = \langle v_1, v_2 \rangle = \mathscr{C}(A)$ e portanto $V^\perp = \mathscr{C}(A)^\perp = \mathscr{N}(A^T)$. Para calcular $\mathscr{N}(A^T)$ vamos reduzir a matriz A^T:
$$\begin{bmatrix} 1 & 1 & -1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \xrightarrow[L_1 - L_2 \to L_1]{} \begin{bmatrix} 1 & 0 & -1 & -1 \\ 0 & 1 & 0 & 1 \end{bmatrix}.$$

Logo,

$$\begin{aligned} V^{\perp} &= \{(x_1,x_2,x_3,x_4)\colon x_1 = x_3 + x_4, x_2 = -x_4, x_3, x_4 \in \mathbb{R}\} \\ &= \{(x_3 + x_4, -x_4, x_3, x_4)\colon x_3, x_4 \in \mathbb{R}\} \\ &= \{(x_3, 0, x_3, 0) + (x_4, -x_4, 0, x_4)\colon x_3, x_4 \in \mathbb{R}\} \\ &= \{x_3(1,0,1,0) + x_4(1,-1,0,1)\colon x_3, x_4 \in \mathbb{R}\} \\ &= \langle (1,0,1,0),(1,-1,0,1)\rangle. \end{aligned}$$

Como $u_1 = (1,0,1,0)$ e $u_2 = (1,-1,0,1)$ são não colineares, $\{u_1, u_2\}$ é base de $V^{\perp}$.

Alternativamente, pode-se utilizar o Algoritmo 3 para determinar uma base para o espaço nulo: considerando $x_3 = 1$ e $x_4 = 0$ obtém-se o vetor $u_1 = (1,0,1,0)$ e considerando $x_3 = 0$ e $x_4 = 1$ o vetor $u_2 = (1,-1,0,1)$.

3. Obtém-se então a base de $\mathbb{R}^4$, $\{v_1, v_2 u_1, u_2\}$, reunindo as bases de V e $V^{\perp}$. Para escrever b como combinação linear dos vetores desta base resolvemos o sistema linear (possível e determinado) $[\ v_1 \ \ v_2 \ \ u_1 \ \ u_2 \,|\, b\ ]$:

$$\left[\begin{array}{cccc|c} 1 & 0 & 1 & 1 & 3 \\ 1 & 1 & 0 & -1 & 2 \\ -1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 3 \end{array}\right] \xrightarrow[\substack{L_2-L_1\to L_2 \\ L_3+L_1\to L_3}]{} \left[\begin{array}{cccc|c} 1 & 0 & 1 & 1 & 3 \\ 0 & 1 & -1 & -2 & -1 \\ 0 & 0 & 2 & 1 & 3 \\ 0 & 1 & 0 & 1 & 3 \end{array}\right] \xrightarrow[L_4-L_2\to L_4]{} \left[\begin{array}{cccc|c} 1 & 0 & 1 & 1 & 3 \\ 0 & 1 & -1 & -2 & -1 \\ 0 & 0 & 2 & 1 & 3 \\ 0 & 0 & 1 & 3 & 4 \end{array}\right]$$

$$\xrightarrow[L_3\leftrightarrow L_4]{} \left[\begin{array}{cccc|c} 1 & 0 & 1 & 1 & 3 \\ 0 & 1 & -1 & -2 & -1 \\ 0 & 0 & 1 & 3 & 4 \\ 0 & 0 & 2 & 1 & 3 \end{array}\right] \xrightarrow[L_4-2L_3\to L_4]{} \left[\begin{array}{cccc|c} 1 & 0 & 1 & 1 & 3 \\ 0 & 1 & -1 & -2 & -1 \\ 0 & 0 & 1 & 3 & 4 \\ 0 & 0 & 0 & -5 & -5 \end{array}\right]$$

$$\xrightarrow[-\frac{1}{5}L_4\to L_4]{} \left[\begin{array}{cccc|c} 1 & 0 & 1 & 1 & 3 \\ 0 & 1 & -1 & -2 & -1 \\ 0 & 0 & 1 & 3 & 4 \\ 0 & 0 & 0 & 1 & 1 \end{array}\right] \xrightarrow[\substack{L_1-L_4\to L_1 \\ L_2+2L_4\to L_2 \\ L_3-3L_4\to L_3}]{} \left[\begin{array}{cccc|c} 1 & 0 & 1 & 0 & 2 \\ 0 & 1 & -1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{array}\right]$$

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

Logo

$$b = 1v_1 + 2v_2 + 1u_1 + 1u_2 = \underbrace{(1,1,-1,0) + 2(0,1,0,1)}_{\in V} + \underbrace{(1,0,1,0) + (1,-1,0,1)}_{\in V^{\perp}},$$

e portanto

$$\begin{aligned} \mathrm{proj}_V(b) &= (1,1,-1,0) + 2(0,1,0,1) = (1,3,-1,2), \\ \mathrm{proj}_{V^{\perp}}(b) &= (1,0,1,0) + (1,-1,0,1) = (2,-1,1,1). \end{aligned}$$

Exercício: confirmar que $\mathrm{proj}_V(b) = (1,3,-1,2)$*, mostrando que* $p = (1,3,-1,2) \in V$ *e* $(b-p) \in V^{\perp}$.

Projeção sobre um subespaço de dimensão 1 (reta)

O cálculo da projeção ortogonal sobre um subespaço vetorial gerado por um vetor não nulo (reta) é imediato. De facto, tem-se o seguinte resultado.

Teorema 38. *Se $V = \langle v \rangle$ com $v \in \mathbb{R}^m$, $v \neq \vec{0}$, então para qualquer $b \in \mathbb{R}^m$ tem-se*

$$\boxed{\operatorname{proj}_{\langle v \rangle}(b) = \frac{v \cdot b}{v \cdot v}\, v.}$$

Demonstração. Seja $p = \operatorname{proj}_{\langle v \rangle}(b)$. Por definição, $p \in V = \langle v \rangle$ e $(b-p) \in V^{\perp}$. Pela primeira relação conclui-se que existe um escalar $\alpha \in \mathbb{R}$ tal que $p = \alpha v$ e pela segunda relação obtém-se $(b - \alpha v) \in V^{\perp}$, isto é, $(b - \alpha v) \cdot v = 0$, ou seja, $b \cdot v - \alpha(v \cdot v) = 0$. Como $v \neq \vec{0}$, $v \cdot v = \|v\|^2 \neq 0$ e conclui-se pela relação anterior que $\alpha = \dfrac{v \cdot b}{v \cdot v}$. Logo,

$$\operatorname{proj}_{\langle v \rangle}(b) = p = \alpha v = \frac{v \cdot b}{v \cdot v}\, v. \qquad \square$$

Uma vez que a projeção ortogonal de um vetor sobre um subespaço vetorial é única, a expressão $\frac{b \cdot v}{v \cdot v} v$ não depende da escolha do gerador v de V.

Definição 27. Sejam $b, v \in \mathbb{R}^m$ com $v \neq \vec{0}$. Chama-se *projeção ortogonal* de b sobre o vetor v a

$$\operatorname{proj}_v(b) = \operatorname{proj}_{\langle v \rangle}(b) = \frac{v \cdot b}{v \cdot v}\, v.$$

Exemplo 52. Dados $b = (6,2)$ e $v = (1,2)$ tem-se

$$\operatorname{proj}_v(b) = \operatorname{proj}_{(1,2)}(6,2) = \operatorname{proj}_{\langle(1,2)\rangle}(6,2) = \frac{(1,2)\cdot(6,2)}{(1,2)\cdot(1,2)}(1,2) = 2(1,2) = (2,4).$$

Observação 25. O resultado do Teorema 38 pode também ser aplicado para calcular a $\operatorname{proj}_V(b)$ no caso em que $V^{\perp} = \langle u \rangle$ tem dimensão 1, isto é, no caso em que V é um subespaço vetorial de $\mathbb{R}^m$ de dimensão $m-1$([1]):

$$\operatorname{proj}_V(b) = b - \operatorname{proj}_{V^{\perp}}(b) = b - \operatorname{proj}_{\langle u \rangle}(b) = b - \frac{u \cdot b}{u \cdot u}\, u.$$

Exemplo 53. Pretende-se calcular a projeção ortogonal de $b = (3,1,-1)$ sobre $V = \langle (1,0,1), (0,1,2) \rangle$.

Como $(1,0,1)$ e $(0,1,2)$ são não colineares e geram V, $\{(1,0,1),(0,1,2)\}$ é base de V. Em particular, $\dim V = 2$. Logo

$$\dim V^{\perp} = \dim \mathbb{R}^3 - \dim V = 3 - 2 = 1.$$

Como $\dim V^{\perp} = 1$ vamos determinar a $\operatorname{proj}_V(b)$, usando a relação

$$\operatorname{proj}_V(b) = b - \operatorname{proj}_{V^{\perp}}(b).$$

[1] Diz-se então que V tem *codimensão* um.

Sejam $v_1 = (1,0,1)$ e $v_2 = (0,1,2)$ e $A = [v_1\ v_2]$. Então $V = \langle v_1, v_2\rangle = \mathscr{C}(A)$ e portanto $V^\perp = \mathscr{C}(A)^\perp = \mathscr{N}(A^T) = \mathscr{N}\left(\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix}\right)$. Como a matriz A^T já está reduzida, $\mathscr{N}(A^T)$ corresponde ao conjunto de soluções do sistema reduzido, com variável livre x_3,

$$\begin{cases} x_1 \quad\quad + \ x_3 = 0 \\ \quad\ x_2 + 2x_3 = 0 \end{cases},$$

e portanto,

$$\begin{aligned} V^\perp = \mathscr{N}(A^T) &= \{(x_1, x_2, x_3) \colon x_1 = -x_3, x_2 = -2x_3, x_3 \in \mathbb{R}\} \\ &= \{(-x_3, -2x_3, x_3) \colon x_3 \in \mathbb{R}\} \\ &= \{x_3(-1,-2,1) \colon x_3 \in \mathbb{R}\} \\ &= \langle(-1,-2,1)\rangle. \end{aligned}$$

Logo,

$$\begin{aligned} \mathrm{proj}_V(b) &= b - \mathrm{proj}_{V^\perp}(b) = (3,1,-1) - \mathrm{proj}_{\langle(-1,-2,1)\rangle}(3,1,-1) \\ &= (3,1,-1) - \frac{(-1,-2,1)\cdot(3,1,-1)}{(-1,-2,1)\cdot(-1,-2,1)}(-1,-2,1) \\ &= (3,1,-1) + (-1,-2,1) = (2,-1,0). \end{aligned}$$

3.3 Distância a um subespaço vetorial

Definição 28. Sejam $b \in \mathbb{R}^m$ e V subespaço vetorial de $\mathbb{R}^m$. A *distância de b a V*, denotada $d(b,V)$ ou $\mathrm{dist}(b,V)$, é a menor distância entre b e os vetores de V, isto é, é a distância entre b e o vetor p de V mais próximo de b:

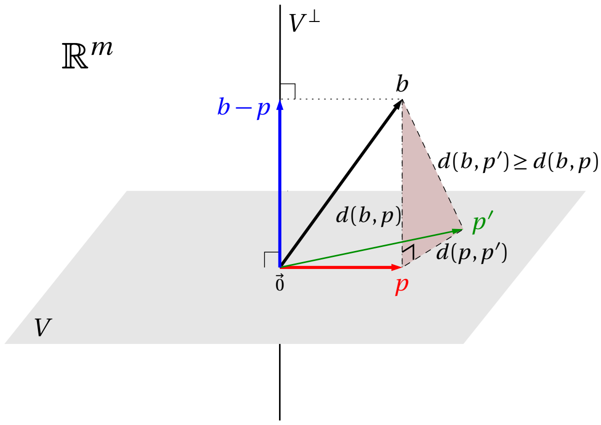
$$d(b,V) = d(b,p) = \min\{d(b,v) \colon v \in V\}.$$

O conceito de projeção ortogonal está intimamente relacionado com o conceito de distância de um vetor a um subespaço vetorial, tendo-se o seguinte resultado.

Teorema 39. *Seja V um subespaço vetorial de $\mathbb{R}^m$ e $b \in \mathbb{R}^m$. Tem-se que a* $\mathrm{proj}_V(b)$ *é o vetor de V à menor distância de b.*

Demonstração. Seja $p = \mathrm{proj}_V(b)$. Temos que mostrar que $\mathrm{d}(b,p') \geq \mathrm{d}(b,p)$ para todo o $p' \in V$.

Temos o triângulo assinalado a tracejado na seguinte figura, cujos lados medem $d(b,p) = \|b-p\|$, $d(p,p') = \|p'-p\|$ e $d(b,p') = \|b-p'\|$. Este triângulo é retângulo, pois $(b-p)\perp(p'-p)$, visto que $b-p\in V^{\perp}$ e $p'-p\in V$.



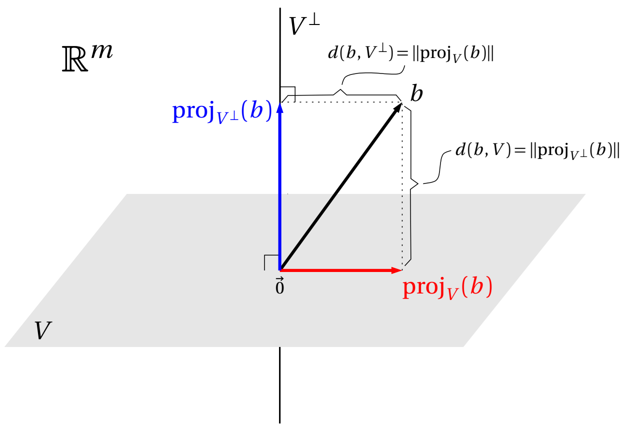
Pelo teorema de Pitágoras tem-se,

$$\|b-p'\|^2=\|b-p\|^2+\|p'-p\|^2.$$

Como $\|p'-p\|^2\geq 0$, $\|b-p'\|^2\geq\|b-p\|^2$. Logo $\|b-p'\|\geq\|b-p\|$, ou seja, $\mathrm{d}(b,p')\geq\mathrm{d}(b,p)$. □

Observação 26. Como consequência do Teorema 39 obtêm-se imediatamente as seguintes relações, que estendem as relações estabelecidas na página 83.

- $d(b,V)=\|b-\mathrm{proj}_V(b)\|=\|\mathrm{proj}_{V^{\perp}}(b)\|$.
- $d(b,V^{\perp})=\|b-\mathrm{proj}_{V^{\perp}}(b)\|=\|\mathrm{proj}_V(b)\|$.
- $\|b\|^2=\mathrm{d}^2(b,V)+\mathrm{d}^2(b,V^{\perp})$.



Das relações anteriores deduzem-se ainda as relações:

- $b\in V\Leftrightarrow\mathrm{d}(b,V)=0\Leftrightarrow\mathrm{proj}_{V^{\perp}}(b)=\vec{0}\Leftrightarrow\mathrm{proj}_V(b)=b$;
- $b\in V^{\perp}\Leftrightarrow\mathrm{d}(b,V^{\perp})=0\Leftrightarrow\mathrm{proj}_V(b)=\vec{0}$.

3.4 Equações normais

Embora o método das bases seja conceptualmente simples, o cálculo da projeção ortogonal por este método pode revelar-se moroso, sobretudo no caso de subespaços vetoriais de $\mathbb{R}^m$, com $m \geq 4$ (como sucede no exemplo 51), uma vez que involve a resolução de um sistema linear a m equações e m variáveis. Nesta secção vamos introduzir o método das equações normais que permite obter essa projeção através da resolução de um sistema linear 'mais simples'.

Sejam $b \in \mathbb{R}^m$ e A uma matriz de tipo $m \times n$. Pela definição 26, a projeção de b sobre $\mathscr{C}(A)$ é definida como sendo o único vetor $p \in \mathbb{R}^m$ que satisfaz as seguintes condições:

(i) $p \in \mathscr{C}(A)$, isto é, o sistema $Ax = p$ é possível;

(ii) $b - p \in \mathscr{C}(A)^{\perp} = \mathscr{N}(A^T)$, isto é, $A^T(b-p) = \vec{0}$.

Por (i) conclui-se que existe um vetor $\bar{x} \in \mathbb{R}^n$ tal que $A\bar{x} = p$ e por (ii) tem-se que $A^T(b-p) = A^T(b - A\bar{x}) = \vec{0}$, isto é, que $A^T A\bar{x} = A^T b$.

Daqui resulta que se $p = \text{proj}_{\mathscr{C}(A)}(b)$ então $p = A\bar{x}$ com $\bar{x}$ solução do sistema de equações lineares, ditas *equações normais*,

$$\boxed{A^T A x = A^T b}$$

Reciprocamente, se $\bar{x}$ é solução das equações normais, tem-se $A^T A\bar{x} = A^T b$, e portanto $p = A\bar{x} \in \mathscr{C}(A)$ e $A^T p = A^T b$, i.e., $A^T(b-p) = \vec{0}$. Logo $p = A\bar{x}$ verifica as condições (i) e (ii), i.e., $p = \text{proj}_{\mathscr{C}(A)}(b)$. Destas considerações resulta o seguinte método para calcular a projecção ortogonal sobre o espaço das colunas de uma matriz, que envolve apenas a determinação de uma solução das equações normais.

Algoritmo 6. MÉTODO DAS EQUAÇÕES NORMAIS

`Input:` *Matriz A do tipo $m \times n$ e $b \in \mathbb{R}^m$.*

`Objectivo:` *Calcular a proj$_{\mathscr{C}(A)}(b)$.*

1. *Determinar uma solução $\bar{x}$ do sistema das equações normais $A^T Ax = A^T b$;*
2. *Tem-se então* $\text{proj}_{\mathscr{C}(A)}(b) = A\bar{x}$

Exemplo 54. Consideremos $b = (1, -1, 1)$ e $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}$. Pretende-se calcular a $\text{proj}_{\mathscr{C}(A)}(b)$ usando o método das equações normais.

Ora,

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

Resolvendo o sistema $A^T Ax = A^T b$ vem:

$$\left[\begin{array}{cc|c} 3 & 4 & 1 \\ 4 & 6 & 2 \end{array}\right] \xrightarrow[L_2-L_1\to L_2]{} \left[\begin{array}{cc|c} 3 & 4 & 1 \\ 1 & 2 & 1 \end{array}\right] \xrightarrow[L_1\leftrightarrow L_2]{} \left[\begin{array}{cc|c} 1 & 2 & 1 \\ 3 & 4 & 1 \end{array}\right] \xrightarrow[L_2-3L_1\to L_2]{}$$

$$\left[\begin{array}{cc|c} 1 & 2 & 1 \\ 0 & -2 & -2 \end{array}\right] \xrightarrow[-\frac{1}{2}L_2\to L_2]{} \left[\begin{array}{cc|c} 1 & 2 & 1 \\ 0 & 1 & 1 \end{array}\right] \xrightarrow[L_1-2L_2\to L_1]{} \left[\begin{array}{cc|c} 1 & 0 & -1 \\ 0 & 1 & 1 \end{array}\right]$$

Logo a (única) solução das equações normais $A^T Ax = A^T b$ é $\bar{x} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$, e tem-se

$$\text{proj}_{\mathscr{C}(A)}(b) = A\bar{x} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

No exemplo anterior a solução $\bar{x}$ das equações normais $A^T Ax = A^T b$ era única. Vamos agora ver uma situação em que isso não ocorre.

Exemplo 55. Sejam $b = (1,0,3)$ e $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$. Pretende-se calcular a $\text{proj}_{\mathscr{C}(A)}(b)$ usando o método das equações normais. Ora, tem-se

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

$$A^T b = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \\ 4 \end{bmatrix}.$$

Resolvendo o sistema $A^T Ax = A^T b$ obtém-se (verifique),

$$\left[\begin{array}{ccc|c} 2 & 0 & 2 & 4 \\ 0 & 1 & 1 & 0 \\ 2 & 1 & 3 & 4 \end{array}\right] \to \cdots \to \left[\begin{array}{ccc|c} 1 & 0 & 1 & 2 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

Trata-se de um sistema indeterminado cujas soluções são dadas por

$$\begin{cases} x_1 = 2 - x_3 \\ x_2 = -x_3 \end{cases}$$

com $x_3 \in \mathbb{R}$ variável livre. Tomando, por exemplo, $x_3 = 0$, obtém-se a solução das equações normais $\bar{x} = (2,0,0)$ e portanto

$$\operatorname{proj}_{\mathscr{C}(A)}(b) = A\bar{x} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 2 \end{bmatrix}.$$

Em geral, considerando x_3 arbitrário, obtém-se a solução $\bar{x} = (2 - x_3, -x_3, x_3)$ das equações normais, tendo-se novamente,

$$\operatorname{proj}_{\mathscr{C}(A)}(b) = A\bar{x} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 - x_3 \\ -x_3 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 2 \end{bmatrix}.$$

Note-se que apesar do sistema das equações normais $A^T Ax = A^T b$ admitir uma infinidade de soluções, o cálculo da projeção ortogonal $\operatorname{proj}_{\mathscr{C}(A)}(b) = A\bar{x}$ não depende da escolha da solução $\bar{x}$ das equações normais $A^T Ax = A^T b$.

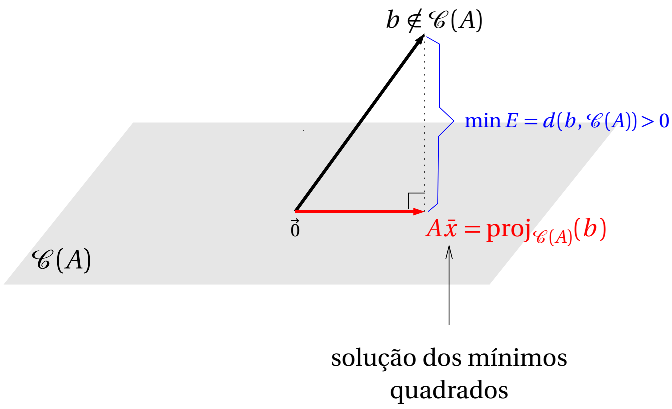
Observação 27. Se o subespaço vetorial V não estiver definido como espaço das colunas de uma matriz, ou seja, como espaço gerado por um conjunto de vetores, pode-se começar por determinar uma base $\{v_1, \ldots, v_k\}$ para V e aplicar então o método das equações normais para calcular a projeção ortogonal de b sobre $\mathscr{C}(A)$, em que $A = [\, v_1 \;\; v_2 \;\; \cdots \;\; v_k \,]$ é a matriz da base de V.

Solução dos mínimos quadrados de um sistema linear

No caso de um sistema linear $Ax = b$ ser impossível, o método das equações normais permite obter vetores $\bar{x}$ que melhor 'se aproximam' de serem solução desse sistema.

Seja A uma matriz do tipo $m \times n$ e $b \in \mathbb{R}^m$. Quando $b \notin \mathscr{C}(A)$, o sistema $Ax = b$ é impossível e portanto $\|Ax - b\| > 0$ para todo o $x \in \mathbb{R}^n$. Nessa altura, vamos estar interessados em determinar um vetor $\bar{x}$ que melhor se 'aproxime' de ser uma solução do sistema $Ax = b$, isto é, que minimize o 'erro' $E = \|Ax - b\|$. Um vetor $\bar{x}$ nessas condições designa-se por *solução dos mínimos quadrados* do sistema $Ax = b$ (a razão para esta designação ficará clara no Exemplo 56).

Ora, por um lado sabemos que os vetores da forma $A\bar{x}$ são combinações lineares das colunas de A com coeficientes dados pelas componentes do vetor $\bar{x}$ e portanto que são vetores de $\mathscr{C}(A)$. Por outro lado, sabemos que a $\operatorname{proj}_{\mathscr{C}(A)}(b)$ é o vetor de $\mathscr{C}(A)$ mais próximo de b. Portanto o valor mínimo do erro é atingido quando $A\bar{x} = \text{proj}_{\mathscr{C}(A)}(b)$, tendo-se nessa altura, $\min E = \|A\bar{x} - b\| = d(b, \mathscr{C}(A))$.



Tem-se portanto a seguinte observação.

Observação 28. As soluções dos mínimos quadrados de um sistema linear $Ax = b$ correspondem aos vetores $\bar{x}$ para os quais o 'vetor erro' $A\bar{x} - b$ é ortogonal a $\mathscr{C}(A)$, isto é, verifica $A^T(A\bar{x} - b) = \vec{0}$. Isto significa que as soluções $\bar{x}$ são dadas pelas soluções das equações normais $A^T Ax = A^T b$ e correspondem a soluções do sistema linear $Ax = \text{proj}_{\mathscr{C}(A)}(b)$.

O método das equações normais permite portanto obter, simultaneamente, a projeção ortogonal de um vetor b sobre $\mathscr{C}(A)$ e uma solução $\bar{x}$ do sistema linear $Ax = b$ no *sentido dos mínimos quadrados*.

Note-se que quando $b \in \mathscr{C}(A)$, o sistema linear $Ax = b$ é possível e qualquer solução $\bar{x}$ de $Ax = b$ no sentido dos mínimos quadrados corresponde a uma solução de $Ax = b$ no sentido usual, isto é, tem-se $A\bar{x} = b$, pois $b = \text{proj}_{\mathscr{C}(A)}(b)$. Neste caso obtém-se $\min E = \|A\bar{x} - b\| = d(b, \mathscr{C}(A)) = 0$.

Exemplo 56. Sejam $A = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}$ e $b = \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix}$ e consideremos o sistema linear $Ax = b$ ([2]). Tem-se então que,

$$Ax = b \Leftrightarrow \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} x = \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix} \Leftrightarrow \begin{cases} 1x = 2 \\ 2x = 3 \\ 4x = 5 \end{cases},$$

e portanto uma solução dos mínimos quadrados do sistema a uma variável $Ax = b$, corresponde a uma solução $\bar{x} \in \mathbb{R}$ que minimize o erro

$$\begin{aligned} E &= \|Ax - b\| = \|(x, 2x, 4x) - (2, 3, 5)\| \\ &= \|(x-2,\ 2x-3,\ 4x-5)\| = \sqrt{(x-2)^2 + (2x-3)^2 + (4x-5)^2} \end{aligned}$$

[2]Os sistemas com mais equações que variáveis designam-se por *sobredeterminados* e provêem muitas vezes de exemplos reais, correspondendo a sistemas lineares que incluem equações inconsistentes, i.e., sistemas impossíveis.

ou equivalentemente, que **minimize a soma de quadrados**,

$$E^2 = \|Ax - b\|^2 = \|(x-2,\, 2x-3,\, 4x-5)\|^2 = (x-2)^2 + (2x-3)^2 + (4x-5)^2.$$

Ora as equações normais vêm dadas, neste caso, por

$$\begin{aligned} A^T Ax = A^T b \quad &\Longleftrightarrow \quad [\,1 \quad 2 \quad 4\,] \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} x = [\,1 \quad 2 \quad 4\,] \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix} \\ &\Longleftrightarrow \quad (1,2,4)\cdot(1,2,4)x = (1,2,4)\cdot(2,3,5). \end{aligned}$$

Logo as equações normais anteriores possuem como solução (única),

$$\bar{x} = \frac{(1,2,4)\cdot(2,3,5)}{(1,2,4)\cdot(1,2,4)} = \frac{4}{3},$$

à qual corresponde o mínimo do quadrado do erro, isto é, o mínimo da soma de quadrados,

$$E^2 = \left\| (1,2,4)\frac{4}{3} - (2,3,5) \right\|^2 = \left(\frac{4}{3} - 2\right)^2 + \left(2\frac{4}{3} - 3\right)^2 + \left(4\frac{4}{3} - 5\right)^2 = \frac{2}{3}.$$

Em particular, o sistema $Ax = b$ é impossível e a solução $\bar{x}$ no sentido dos mínimos quadrados desse sistema não corresponde a uma solução no sentido usual.

Observação 29. Voltando aos exemplos 54 e 55, constatamos que, no primeiro destes exemplos, o sistema linear $Ax = b$ admite como única solução dos mínimos quadrados, $\bar{x} = (-1, 1)$ e no segundo, a infinidade de soluções no sentido dos mínimos quadrados, $\bar{x} = (2-a, -a, a)$, $a \in \mathbb{R}$. Nenhuma das soluções referidas corresponde a uma solução no sentido usual, uma vez que em ambos os exemplos $\mathrm{proj}_{\mathscr{C}(A)}(b) \neq b$ e portanto em ambos os casos o sistema linear $Ax = b$ é impossível.

3.5 Matriz de projeção

Nesta secção vamos considerar o caso em que as colunas da matriz A são linearmente independentes. Nessa altura o método das equações normais permite construir uma matriz, dita de *projeção*, sobre o subespaço vetorial $\mathscr{C}(A)$. Para isso necessitamos de mostrar alguns resultados preliminares.

Lema 40. *Para qualquer matriz A tem-se $\mathscr{N}(A) = \mathscr{N}(A^T A)$.*

Demonstração. Se $u \in \mathscr{N}(A)$ tem-se $Au = \vec{0}$ e portanto $A^T(Au) = (A^T A)u = \vec{0}$, ou seja, $u \in \mathscr{N}(A^T A)$. Portanto $\mathscr{N}(A) \subset \mathscr{N}(A^T A)$.

Reciprocamente, se $u \in \mathscr{N}(A^T A)$, então $(A^T A)u = \vec{0}$. Logo $u^T(A^T Au) = \vec{0}$, e portanto $(Au)^T(Au) = 0$, isto é, $\|Au\|^2 = 0$. Logo $Au = \vec{0}$, ou seja, $u \in \mathscr{N}(A)$. Portanto $\mathscr{N}(A^T A) \subset \mathscr{N}(A)$. □

Corolário 41. *Seja A uma matriz de tipo $m \times n$. Então $A^T A$ é invertível se e só se as colunas de A são linearmente independentes.*

Demonstração. $A^T A$ é invertível se e só se $\text{car}(A^T A) = n$ se e só se $\dim \mathcal{N}(A^T A) = \dim \mathcal{N}(A) = 0$ se e só se $\text{car}(A) = n$ se e só se as colunas de A são linearmente independentes. □

Se $V \neq \{\vec{0}\}$ é um subespaço vetorial de $\mathbb{R}^m$, $\{v_1, \ldots, v_n\}$ é uma base de V e $A = [\, v_1 \;\; v_2 \;\; \cdots \;\; v_n \,]$ é a matriz da base de V, deduz-se do Corolário 41 que a matriz $A^T A$ é invertível. Nessa altura, o método das equações normais permite obter o seguinte resultado.

Teorema 42. *Sejam $V \neq \{\vec{0}\}$ subespaço vetorial de $\mathbb{R}^m$, $\{v_1, \ldots, v_n\}$ uma base de V e $A = [\, v_1 \;\; v_2 \;\; \cdots \;\; v_n \,]$ matriz da base de V. Então para todo o $b \in \mathbb{R}^m$ tem-se*

$$\text{proj}_V(b) = \text{proj}_{\mathscr{C}(A)}(b) = A(A^T A)^{-1} A^T b.$$

Demonstração. Pelo método das equações normais tem-se $\text{proj}_{\mathscr{C}(A)}(b) = A\bar{x}$ em que $\bar{x}$ é uma solução arbitrária das equações normais $A^T A x = A^T b$. Como as colunas de A são linearmente independentes, $\text{car}(A) = n$ e conclui-se pelo Corolário 41 que $A^T A$ é invertível. Portanto

$$A^T A x = A^T b \Leftrightarrow \underbrace{(A^T A)^{-1} A^T A}_{I_n} x = (A^T A)^{-1} A^T b \Leftrightarrow x = (A^T A)^{-1} A^T b.$$

Logo a solução (única) das equações normais vem dada por $\bar{x} = (A^T A)^{-1} A^T b$, e tem-se

$$\text{proj}_{\mathscr{C}(A)}(b) = A\bar{x} = A(A^T A)^{-1} A^T b.$$

□

Definição 29. Sejam $V \neq \{\vec{0}\}$ subespaço vetorial de $\mathbb{R}^m$ e $\{v_1, \ldots, v_n\}$ uma <u>base</u> de V. Seja $A = [\, v_1 \;\; v_2 \;\; \cdots \;\; v_n \,]$ a matriz da base de V. A matriz

$$\boxed{P = A(A^T A)^{-1} A^T}$$

designa-se por *matriz de projeção* sobre o subespaço vetorial V.

Pode-se mostrar que a matriz de projeção não depende da escolha da base de V. Pelo Teorema 42 tem-se portanto,

$$\text{proj}_V(b) = Pb, \qquad \forall b \in \mathbb{R}^m.$$

Proposição 43. *Se P é uma matriz de projeção, então*

1. *P é simétrica, isto é, $P^T = P$;*
2. *P é idempotente, isto é, $P^2 = P$.*

Demonstração. Exercício. □

Corolário 44. *A projeção ortogonal sobre um subespaço vetorial V de $\mathbb{R}^m$ verifica as seguintes propriedades para todo $u, v \in \mathbb{R}^m$ e $\lambda \in \mathbb{R}$ ([3]):*

$$\boxed{\begin{array}{c}\operatorname{proj}_V(u+v) = \operatorname{proj}_V(u) + \operatorname{proj}_V(v)\\ \operatorname{proj}_V(\lambda u) = \lambda \operatorname{proj}_V(u)\end{array}}$$

Demonstração. Exercício (sugestão: utilize a matriz de projeção sobre V). □

Pelo Teorema 42 tem-se $\operatorname{proj}_V(b) = Pb$ para todo o $b \in \mathbb{R}^m$ e portanto

$$\operatorname{proj}_{V^\perp}(b) = b - \operatorname{proj}_V(b) = I_m b - Pb = (I_m - P)b.$$

Tem-se portanto a seguinte observação.

Observação 30. Se P é a matriz de projeção sobre um subespaço vetorial V de $\mathbb{R}^m$ então $I - P$ é a matriz de projeção sobre $V^\perp$.

Exemplo 57. Consideremos $A = \begin{bmatrix} 1 & 1\\ 0 & 1\\ 1 & 1\end{bmatrix}$. Pretende-se calcular as matrizes de projeção sobre $\mathscr{C}(A)$ e $\mathscr{C}(A)^\perp$. Tem-se

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

Aplicando o algoritmo da inversa obtém-se,

$$\left[\begin{array}{c|c} A^T A & I_2\end{array}\right] = \left[\begin{array}{cc|cc} 2 & 2 & 1 & 0\\ 2 & 3 & 0 & 1\end{array}\right] \xrightarrow[L_2 - L_1 \to L_2]{} \left[\begin{array}{cc|cc} 2 & 2 & 1 & 0\\ 0 & 1 & -1 & 1\end{array}\right] \xrightarrow[L_1 - 2L_2 \to L_1]{} \left[\begin{array}{cc|cc} 2 & 0 & 3 & -2\\ 0 & 1 & -1 & 1\end{array}\right]$$

$$\xrightarrow[\frac{1}{2}L_1 \to L_1]{} \left[\begin{array}{cc|cc} 1 & 0 & 3/2 & -1\\ 0 & 1 & -1 & 1\end{array}\right] = \left[\begin{array}{c|c} I_2 & (A^T A)^{-1}\end{array}\right]$$

e por conseguinte,

$$(A^T A)^{-1} = \begin{bmatrix} 3/2 & -1\\ -1 & 1\end{bmatrix} = \frac{1}{2}\begin{bmatrix} 3 & -2\\ -2 & 2\end{bmatrix}.$$

[3] Estas propriedades traduzem-se dizendo que projeção ortogonal define uma *transformação linear* em $\mathbb{R}^m$.

Portanto a matriz de projeção ortogonal sobre $\mathscr{C}(A)$ vem dada por

$$\begin{aligned}P = A(A^TA)^{-1}A^T &= \begin{bmatrix}1 & 1\\ 0 & 1\\ 1 & 1\end{bmatrix}\frac{1}{2}\begin{bmatrix}3 & -2\\ -2 & 2\end{bmatrix}\begin{bmatrix}1 & 0 & 1\\ 1 & 1 & 1\end{bmatrix}\\ &= \frac{1}{2}\begin{bmatrix}1 & 1\\ 0 & 1\\ 1 & 1\end{bmatrix}\left(\begin{bmatrix}3 & -2\\ -2 & 2\end{bmatrix}\begin{bmatrix}1 & 0 & 1\\ 1 & 1 & 1\end{bmatrix}\right)\\ &= \frac{1}{2}\begin{bmatrix}1 & 1\\ 0 & 1\\ 1 & 1\end{bmatrix}\begin{bmatrix}1 & -2 & 1\\ 0 & 2 & 0\end{bmatrix} = \frac{1}{2}\begin{bmatrix}1 & 0 & 1\\ 0 & 2 & 0\\ 1 & 0 & 1\end{bmatrix}.\end{aligned}$$

Pela Observação 30, a matriz de projeção ortogonal sobre $\mathscr{C}(A)^\perp$ vem dada por

$$I_3 - P = \begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix} - \frac{1}{2}\begin{bmatrix}1 & 0 & 1\\ 0 & 2 & 0\\ 1 & 0 & 1\end{bmatrix} = \frac{1}{2}\begin{bmatrix}1 & 0 & -1\\ 0 & 0 & 0\\ -1 & 0 & 1\end{bmatrix}.$$

Se, por exemplo, $b = (1,2,3) \in \mathbb{R}^3$, tem-se então,

$$\mathrm{proj}_{\mathscr{C}(A)}(b) = Pb = \frac{1}{2}\begin{bmatrix}1 & 0 & 1\\ 0 & 2 & 0\\ 1 & 0 & 1\end{bmatrix}\begin{bmatrix}1\\ 2\\ 3\end{bmatrix} = \begin{bmatrix}2\\ 2\\ 2\end{bmatrix},$$

e,

$$\mathrm{proj}_{\mathscr{C}(A)^\perp}(b) = (I_3 - P)b = \frac{1}{2}\begin{bmatrix}1 & 0 & -1\\ 0 & 0 & 0\\ -1 & 0 & 1\end{bmatrix}\begin{bmatrix}1\\ 2\\ 3\end{bmatrix} = \begin{bmatrix}-1\\ 0\\ 1\end{bmatrix}.$$

Note-se que,

$$b = (1,2,3) = (2,2,2) + (-1,0,1) = \mathrm{proj}_{\mathscr{C}(A)}(b) + \mathrm{proj}_{\mathscr{C}(A)^\perp}(b).$$

Terminamos esta secção com uma expressão muito elegante para a matriz de projeção sobre um subespaço vetorial de dimensão um (reta).

Proposição 45. *Consideremos o subespaço vetorial $V = \langle v \rangle$, com $v \in \mathbb{R}^m$, $v \neq \vec{0}$. Tem-se que a matriz de projeção P sobre V pode ser escrita como*

$$P = \frac{v\,v^T}{v^T v}.$$

Demonstração. Por definição tem-se, identificando $A = [\,v\,]$ com v,

$$P = A(A^TA)^{-1}A^T = v(v^Tv)^{-1}v^T = \frac{v\,v^T}{v^Tv},$$

uma vez que $v^Tv = \|v\|^2 \neq 0$. □

3.6 Método de ortogonalização de Gram-Schmidt

Conjunto ortogonal e ortonormado de vetores

Definição 30. Sejam $v_1 \ldots, v_n$ vetores não nulos de $\mathbb{R}^m$. Diz-se que o conjunto $\{v_1, \ldots, v_n\}$ é *ortogonal* se $v_i \perp v_j = 0$ para todo o $i \neq j$, isto é, se os vetores $v_1, \ldots, v_n$ forem ortogonais entre si 2 a 2. Se adicionalmente, $\|v_i\| = 1$ para todo o i, isto é, se os vetores $v_1, \ldots, v_n$ forem também unitários, $\{v_1, \ldots, v_n\}$ diz-se um conjunto *ortonormado* de vetores de $\mathbb{R}^m$.

Exemplos 58.

1. O conjunto $\{(0,1,1),(1,2,-2),(4,-1,1)\}$ é ortogonal. Com efeito,

$$(0,1,1)\cdot(1,2,-2)=0, \quad (0,1,1)\cdot(4,-1,1)=0, \quad (1,2,-2)\cdot(4,-1,1)=0.$$

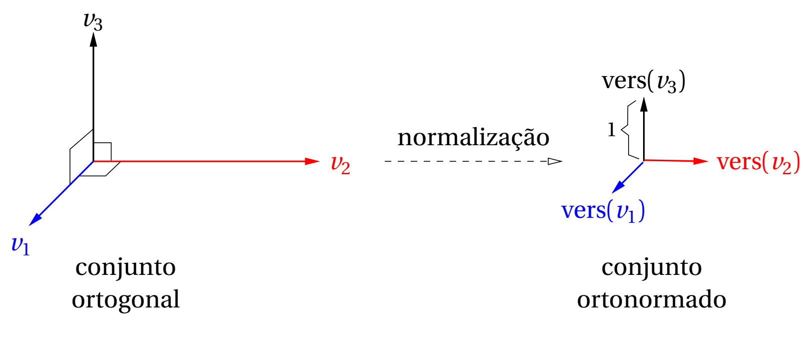
2. A base canónica de $\mathbb{R}^3$, $\{e_1, e_2, e_3\}$ em que $e_1 = (1,0,0)$, $e_2 = (0,1,0)$ e $e_3 = (0,0,1)$, define um conjunto ortonormado de vetores, pois $\|e_i\| = 1$ para $i = 1,2,3$ e $e_i \cdot e_j = 0$ para todo o $i \neq j$.

   Mais geralmente, a base a canónica $\{e_1, \ldots, e_m\}$ de $\mathbb{R}^m$ (formada pelos vetores que constituem as colunas da matriz identidade I_m) é um conjunto ortonormado de vetores de $\mathbb{R}^m$.

Observação 31. Se $\{v_1, \ldots, v_n\}$ é um conjunto ortogonal de vetores não nulos de $\mathbb{R}^m$, então

$$\Big\{\mathrm{vers}(v_1), \ldots, \mathrm{vers}(v_n)\Big\} = \left\{\frac{v_1}{\|v_1\|}, \ldots, \frac{v_n}{\|v_n\|}\right\},$$

é um conjunto ortonormado de vetores de $\mathbb{R}^m$.



Exemplo 59. Uma vez que o conjunto $\{(0,1,1),(1,2,-2),(4,-1,1)\}$ é ortogonal (ver o exemplo 58.1), o conjunto de vetores

$$\left\{\frac{(0,1,1)}{\|(0,1,1)\|}, \frac{(1,2,-2)}{\|(1,2,-2)\|}, \frac{(4,-1,1)}{\|(4,-1,1)\|}\right\} = \left\{\frac{1}{\sqrt{2}}(0,1,1), \frac{1}{3}(1,2,-2), \frac{1}{3\sqrt{2}}(4,-1,1)\right\}$$

é um conjunto ortonormado de vetores de $\mathbb{R}^3$.

Teorema 46. *Todo o conjunto ortogonal constituído por n vetores* <u>*não nulos*</u> *de $\mathbb{R}^m$ é linearmente independente. Em particular, $n \leq m$.*

Demonstração. Consideremos um conjunto ortogonal $\{v_1,\ldots,v_n\}$ formado por vetores não nulos. Dados escalares $\alpha_1,\ldots,\alpha_n \in \mathbb{R}$ tais que $\alpha_1 v_1 + \ldots + \alpha_n v_n = \vec{0}$, temos que mostrar $\alpha_1 = \cdots = \alpha_n = 0$. Ora, se $\alpha_1 v_1 + \cdots + \alpha_n v_n = \vec{0}$ tem-se,

$$\underbrace{(\alpha_1 v_1 + \cdots + \alpha_n v_n)}_{\vec{0}} \cdot v_1 = \alpha_1 \underbrace{(v_1 \cdot v_1)}_{\|v_1\|^2} + \alpha_2 \underbrace{(v_2 \cdot v_1)}_{0} + \cdots + \alpha_n \underbrace{(v_n \cdot v_1)}_{0} = 0,$$

uma vez que $v_i \perp v_1$, $i = 2,\ldots,n$. Logo $\alpha_1 \|v_1\|^2 = 0$. Como $v_1 \neq \vec{0}$, $\|v_1\|^2 \neq 0$ e portanto $\alpha_1 = 0$. De modo análogo se mostra que $\alpha_i = 0$, para $i = 2,\ldots,n$, considerando o produto escalar de $\alpha_1 v_1 + \cdots + \alpha_n v_n$ com v_i. Logo $\{v_1,\ldots,v_n\}$ é linearmente independente. Como em $\mathbb{R}^m$ os conjuntos linearmente independentes têm cardinalidade máxima m (Teorema 22), conclui-se que $n \leq m$. □

Definição 31. Um conjunto de vetores $\{v_1,\ldots,v_n\}$ diz-se uma *base ortogonal* de V, se for uma base de V e simultaneamente um conjunto ortogonal de vetores. Se adicionalmente os vetores $v_1,\ldots,v_n$ forem unitários, $\{v_1,\ldots,v_n\}$ diz-se uma *base ortonormada* de V.

Note-se que as bases de subespaços vetoriais de <u>dimensão um</u> são sempre ortogonais, uma vez que são constituídas por um único vetor.

Observação 32. Pelo Teorema 46, se $V = \langle v_1,\ldots,v_k\rangle$ com $\{v_1,\ldots,v_k\}$ conjunto ortogonal de vetores não nulos de $\mathbb{R}^m$, $\{v_1,\ldots,v_k\}$ é linearmente independente e portanto define uma base (ortogonal) de V.

Exemplos 60.

1. O conjunto ortonormado de vetores do Exemplo 59,

$$\left\{\frac{1}{\sqrt{2}}(0,1,1), \frac{1}{3}(1,2,-2), \frac{1}{3\sqrt{2}}(4,-1,1)\right\}$$

   define uma base ortonormada de $\mathbb{R}^3$.

2. A base canónica de $\mathbb{R}^m$ é ortonormada.

As base ortogonais permitem calcular a projeção ortogonal de um vetor de forma imediata. Mais precisamente, tem-se o seguinte resultado.

Teorema 47. *Seja V subespaço vetorial de $\mathbb{R}^m$ e seja $\{v_1,\ldots,v_k\}$* <u>*base ortogonal*</u> *de V. Então para todo o $b \in \mathbb{R}^m$ tem-se,*

$$\boxed{\operatorname{proj}_V(b) = \operatorname{proj}_{v_1}(b) + \cdots + \operatorname{proj}_{v_k}(b) = \frac{v_1 \cdot b}{v_1 \cdot v_1} v_1 + \cdots + \frac{v_k \cdot b}{v_k \cdot v_k} v_k.}$$

Demonstração. Seja $p = \text{proj}_V(b)$. Como $p \in V$ e $\{v_1, \ldots, v_k\}$ é base de V, existem escalares $\alpha_1, \ldots, \alpha_k \in \mathbb{R}$ tais que $p = \alpha_1 v_1 + \cdots + \alpha_k v_k$. Por outro lado, por definição de projeção, $b = p + (b - p)$ com $b - p \in V^\perp$. Logo,

$$\begin{aligned} v_1 \cdot b &= v_1 \cdot \big(p + (b-p)\big) = v_1 \cdot \big(\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_k v_k + (b-p)\big) \\ &= \alpha_1(v_1 \cdot v_1) + \alpha_2(v_1 \cdot v_2) + \cdots + \alpha_k(v_1 \cdot v_k) + v_1 \cdot (b-p). \end{aligned}$$

Como $v_1 \perp v_j$ para $j \geq 2$, $v_1 \cdot v_j = 0$ para $j \geq 2$. Como $v_1 \in V$ e $b - p \in V^\perp$, $v_1 \cdot (b-p) = 0$. Portanto $v_1 \cdot b = \alpha_1(v_1 \cdot v_1)$, donde se conclui que

$$\alpha_1 = \frac{v_1 \cdot b}{v_1 \cdot v_1},$$

uma vez que $v_1 \cdot v_1 = \|v_1\| \neq 0$. Considerando os produtos escalares de b com v_i, para $i = 2, \ldots, k$, deduz-se de modo análogo que

$$\alpha_i = \frac{v_i \cdot b}{v_i \cdot v_i},$$

e portanto tem-se

$$\begin{aligned} \text{proj}_V(b) &= p = \alpha_1 v_1 + \cdots + \alpha_k v_k \\ &= \frac{v_1 \cdot b}{v_1 \cdot v_1} v_1 + \cdots + \frac{v_k \cdot b}{v_k \cdot v_k} v_k = \text{proj}_{v_1}(b) + \cdots + \text{proj}_{v_k}(b). \end{aligned}$$ □

Observação 33. O resultado do teorema anterior apenas é válido se a base do subespaço vetorial V for ortogonal. Quando a base de V não é ortogonal, a projeção de um vetor b sobre V não é dada, em geral, pela soma das projeções de b sobre cada vetor da base!

Exemplo 61. Sejam $V = \langle (1,0,1), (1,1,-1) \rangle$ e $b = (1,2,3)$.

Como $(1,0,1) \cdot (1,1,-1) = 0$, $(1,0,1) \perp (1,1,-1)$ e portanto $\{(1,0,1), (1,1,-1)\}$ é uma base ortogonal de V. Logo

$$\begin{aligned} \text{proj}_V(b) &= \text{proj}_{\langle (1,0,1),(1,1,-1)\rangle}(1,2,3) \\ &= \text{proj}_{(1,0,1)}(1,2,3) + \text{proj}_{(1,1,-1)}(1,2,3) \\ &= \tfrac{(1,0,1)\cdot(1,2,3)}{(1,0,1)\cdot(1,0,1)}(1,0,1) + \tfrac{(1,1,-1)\cdot(1,2,3)}{(1,1,-1)\cdot(1,1,-1)}(1,1,-1) \\ &= \tfrac{4}{2}(1,0,1) + \tfrac{0}{3}(1,1,-1) = (2,0,2). \end{aligned}$$

Corolário 48. *Se $\{v_1, \ldots, v_k\}$ for uma base ortonormada de V, então para todo o $b \in \mathbb{R}^m$ tem-se*

$$\boxed{\text{proj}_V(b) = (v_1 \cdot b)v_1 + \cdots + (v_k \cdot b)v_k.}$$

Demonstração. Exercício. □

Método de ortogonalização de Gram-Schmidt

Nesta secção vamos estabelecer um método que permite obter, a partir de uma base $\{u_1, \ldots, u_n\}$ de um subespaço vetorial V, uma base ortogonal $\{v_1, \ldots, v_n\}$ para V.

Algoritmo 7. MÉTODO DE ORTOGONALIZAÇÃO DE GRAM-SCHMIDT

`Input:` $\{u_1, \ldots, u_n\}$ *base de um subespaço vetorial* V.

`Objectivo:` *Determinar uma base ortogonal de* V, $\{v_1, \ldots, v_n\}$.

Base de V

$\{u_1, u_2, \ldots, u_n\}$

$$v_1 = u_1$$

$$v_2 = u_2 - \operatorname{proj}_{v_1}(u_2) = u_2 - \frac{v_1 \cdot u_2}{v_1 \cdot v_1} v_1$$

$$\vdots$$

$$v_n = u_n - \operatorname{proj}_{v_1}(u_n) - \cdots - \operatorname{proj}_{v_{n-1}}(u_n) = u_n - \frac{v_1 \cdot u_n}{v_1 \cdot v_1} v_1 - \cdots - \frac{v_{n-1} \cdot u_n}{v_{n-1} \cdot v_{n-1}} v_{n-1}$$

$\{v_1, v_2, \ldots, v_n\}$

Base ortogonal de V

Notas:

- $v_2 = \operatorname{proj}_{\langle v_1 \rangle^\perp}(u_2) = u_2 - \operatorname{proj}_{v_1}(u_2)$ e portanto $v_2 \perp v_1$. Além disso, como $u_2 = v_2 + \dfrac{v_1 \cdot u_2}{v_1 \cdot v_1} v_1$, $\langle v_1, v_2 \rangle = \langle v_1, u_2 \rangle = \langle u_1, u_2 \rangle$.

- $v_3 = \operatorname{proj}_{\langle v_1, v_2 \rangle^\perp}(u_3) = u_3 - \operatorname{proj}_{\langle v_1, v_2 \rangle}(u_3) \overset{v_1 \perp v_2}{=} u_3 - (\operatorname{proj}_{v_1}(u_3) + \operatorname{proj}_{v_2}(u_3)) = u_3 - \operatorname{proj}_{v_1}(u_3) - \operatorname{proj}_{v_2}(u_3)$, e portanto $v_3 \perp v_1, v_2$. Além disso, como $u_1 = v_1$, $u_2 = v_2 + \dfrac{v_1 \cdot u_2}{v_1 \cdot v_1} v_1$ e $u_3 = v_3 + \dfrac{v_1 \cdot u_3}{v_1 \cdot v_1} + \dfrac{v_2 \cdot u_3}{v_2 \cdot v_2}$, $\langle v_1, v_2, v_3 \rangle = \langle v_1, v_2, u_3 \rangle = \langle u_1, u_2, u_3 \rangle$.

- De modo análogo, se mostra que, para cada $k = 2, \ldots, n$, se tem

$$\begin{aligned} v_k &= \operatorname{proj}_{\langle v_1, \ldots, v_{k-1} \rangle^\perp}(u_k) = u_k - \operatorname{proj}_{\langle v_1, \ldots, v_{k-1} \rangle}(u_k) \\ &= u_k - \operatorname{proj}_{v_1}(u_k) - \cdots - \operatorname{proj}_{v_{k-1}}(u_k), \end{aligned}$$

e portanto, que $v_k \perp v_1, \ldots, v_{k-1}$, $k = 2, \ldots, n$.

- Pode-se verificar ainda que $\langle v_1, \ldots, v_k \rangle = \langle u_1, \ldots, u_k \rangle$, para $k = 1, \ldots, n$.

  Em particular, $\langle v_1, \ldots, v_n \rangle = \langle u_1, \ldots, u_n \rangle = V$.

Observação 34. Se multiplicarmos os vetores de uma base ortogonal de V por escalares não nulos ainda obtemos uma base ortogonal de V (justifique).

Exemplo 62. Sabendo que $\{(1,-1,1),(1,0,1),(1,1,2)\}$ é base de $\mathbb{R}^3$, construir uma base ortogonal de $\mathbb{R}^3$.

Por aplicação do método de ortogonalização de Gram–Schmidt tem-se

Base de $\mathbb{R}^3$

$\{(1,-1,1),(1,0,1),(1,1,2)\}$

$$\to (1,-1,1)$$

$$\to (1,0,1) - \text{proj}_{(1,-1,1)}(1,0,1) =$$
$$= (1,0,1) - \frac{(1,-1,1)\cdot(1,0,1)}{(1,-1,1)\cdot(1,-1,1)}(1,-1,1)$$
$$= (1,0,1) - \frac{2}{3}(1,-1,1)$$
$$= \left(\frac{1}{3}, \frac{2}{3}, \frac{1}{3}\right)$$
$$= \frac{1}{3}(1,2,1) \rightsquigarrow (1,2,1)$$

$$\to (1,1,2) - \text{proj}_{(1,-1,1)}(1,1,2) - \text{proj}_{(1,2,1)}(1,1,2) =$$
$$= (1,1,2) - \frac{(1,-1,1)\cdot(1,1,2)}{(1,-1,1)\cdot(1,-1,1)}(1,-1,1) - \frac{(1,2,1)\cdot(1,1,2)}{(1,2,1)\cdot(1,2,1)}(1,2,1)$$
$$= (1,1,2) - \frac{1-1+2}{3}(1,-1,1) - \frac{1+2+2}{6}(1,2,1)$$
$$= (1,1,2) - \frac{2}{3}(1,-1,1) - \frac{5}{6}(1,2,1)$$
$$= \left(1 - \frac{2}{3} - \frac{5}{6}, 1 + \frac{2}{3} - \frac{10}{6}, 2 - \frac{2}{3} - \frac{5}{6}\right)$$
$$= \frac{1}{2}(-1,0,1) \rightsquigarrow (-1,0,1)$$

$\{(1,-1,1),(1,2,1),(-1,0,1)\}$

Base ortogonal de $\mathbb{R}^3$

Observação 35. Sejam V subespaço vetorial de $\mathbb{R}^m$, $\{v_1, \ldots, v_k\}$ base ortogonal de V e $\{u_1, \ldots, u_{m-k}\}$ base ortogonal de $V^\perp$. Então $\{v_1, \ldots, v_k, u_1, \ldots, u_{m-k}\}$ é uma base ortogonal de $\mathbb{R}^m$, uma vez que a ortogonalidade entre vetores de V e $V^\perp$ está garantida.

Exemplo 63. Dado um subespaço vetorial $V = \langle (1,0,1),(1,1,0) \rangle$, pretende-se determinar uma base ortogonal para $\mathbb{R}^3$ que contenha uma base de V.

Comecemos por notar que a base de V tem que ser ortogonal. Ora, tem-se que $\{(1,0,1),(1,1,0)\}$ é base de V, pois $(1,0,1)$ e $(1,1,0)$ são não colineares, logo linearmente independentes. Aplicando o método de ortogonalização de Gram-Schmidt a esta base, obtém-se a base ortogonal de V, $\{(1,0,1),(1,2,-1)\}$. De facto,

Base de V

$\{(1,0,1),(1,1,0)\}$

$\longrightarrow (1,0,1)$

$$\longrightarrow (1,1,0) - \operatorname{proj}_{(1,0,1)}(1,1,0) = (1,1,0) - \tfrac{(1,0,1)\cdot(1,1,0)}{(1,0,1)\cdot(1,0,1)}(1,0,1) =$$

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

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

$\{(1,0,1),(1,2,-1)\}$

Base ortogonal de V

Por outro lado, $V^\perp = \langle (1,0,1),(1,1,0) \rangle^\perp = \mathscr{C}\left(\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}\right)^\perp = \mathscr{N}\left(\begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}\right)$.

Reduzindo a matriz anterior obtém-se,

$$\begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \xrightarrow[L_2 - L_1 \to L_2]{} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix}$$

A matriz reduzida corresponde ao sistema homogéneo reduzido, com variável livre z,

$$\begin{cases} x \quad\;\; + z = 0 \\ \quad\; y - z = 0 \end{cases}$$

e portanto,

$$\begin{aligned} V^\perp &= \{(x,y,z) \colon x = -z, y = z, z \in \mathbb{R}\} = \{(-z,z,z) \colon z \in \mathbb{R}\} \\ &= \{z(-1,1,1) \colon z \in \mathbb{R}\} = \langle (-1,1,1) \rangle. \end{aligned}$$

Logo $\{(-1,1,1)\}$ é uma base de $V^{\perp}$, que é ortogonal, uma vez que é constituída por um único vetor (qualquer base formada por um único vetor, isto é, de um subespaço vetorial de dimensão um, é sempre ortogonal). Pela Observação 35 conclui-se que $\{(1,0,1),(1,2,-1),(-1,1,1)\}$ é uma base ortogonal de $\mathbb{R}^3$ que contém uma base de V, como se pretendia.

Capítulo 4

Determinantes

4.1 Determinante de matrizes de ordem $n \leq 3$

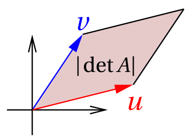
Consideremos uma matriz quadrada A de ordem n. Vamos associar à matriz A um número real, designado por *determinante* de A e denotado por $\det A$ (ou por $|A|$). Em particular, o determinante vai permitir decidir se A é invertível.

Se $n = 1$, tem-se $A = [a]$, com $a \in \mathbb{R}$ e define-se $\det[a] = a$.

Se $n = 2$ pode-se escrever $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ com $a, b, c, d \in \mathbb{R}$ e calcula-se o determinante de A subtraindo-se ao produto dos elementos da diagonal principal o produto dos elementos da anti-diagonal (a tracejado), ou seja,

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

Observação 36. Pode-se mostrar que o valor absoluto do determinante da matriz $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, corresponde à área do paralelogramo definido pelos vetores que constituem as colunas de A, $u = (a, c)$ e $v = (b, d)$:



Note-se que $\det A \neq 0$ se e só se a área do paralelogramo não for nula, isto é, se e só se as colunas de A forem não colineares formando um conjunto de vetores linearmente independente, concluindo-se portanto que $\det A \neq 0$ se e só se A for invertível.

Exemplo 64. Tem-se,

$$\det\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = 1\times 4 - 2\times 3 = -2 \neq 0.$$

Em particular, a matriz $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ é invertível e o paralelogramo definido pelos vetores $u=(1,3)$ e $v=(2,4)$ tem área 2.

Regra de Sarrus

Consideremos uma matriz quadrada de ordem 3,

$$A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}. \tag{4.1}$$

Para obter o determinante da matriz A pode-se aplicar o seguinte esquema, conhecido como *regra de Sarrus*: amplia-se a matriz A repetindo à direita de A as duas primeiras colunas de A. Em seguida, para calcular o determinante de A somam-se as parcelas (a vermelho) correspondentes aos produtos dos elementos que estão nas diagonais paralelas à diagonal principal de A e subtraem-se as parcelas (a azul) correspondentes aos produtos dos elementos que estão nas diagonais paralelas à anti-diagonal de A, como descrito abaixo:

$$\det(A) = \det\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}\begin{matrix} a & b \\ d & e \\ g & h \end{matrix}$$

$$= aei + bfg + cdh - (ceg + afh + bdi).$$

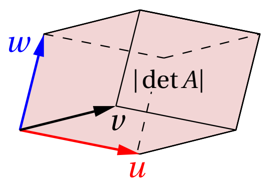
Exemplo 65. Consideremos a matriz $A = \begin{bmatrix} 1 & 0 & -2 \\ 2 & 1 & 1 \\ -1 & 3 & 1 \end{bmatrix}$. Aplicando a regra de Sarrus obtém-se,

$$\det(A) = \det \left[\begin{array}{ccc} 1 & 0 & -2 \\ 2 & 1 & 1 \\ -1 & 3 & 1 \end{array}\right] \begin{array}{cc} 1 & 0 \\ 2 & 1 \\ -1 & 3 \end{array}$$

$$= 1 \times 1 \times 1 + 0 \times 1 \times (-1) + (-2) \times 2 \times 3 - ((-2) \times 1 \times (-1) + 1 \times 1 \times 3 + 0 \times 2 \times 1)$$

$$= 1 + 0 - 12 - (2 + 3 + 0) = -16 \neq 0.$$

Observação 37. O valor absoluto do determinante de uma matriz quadrada de ordem 3, $A = [\, u \;\; v \;\; w \,]$, corresponde ao volume do paralelipípedo definido pelos vetores $u, v, w \in \mathbb{R}^3$ que constituem as colunas de A:



A demonstração deste resultado baseia-se nos conceitos de *produto externo* e *produto misto* de vetores que não serão dados neste ano letivo.

Note-se que o $\det A \neq 0$ se e só se o volume do paralelipípedo anterior for não nulo, isto é, se e só se os vetores u, v, w forem vetores não complanares. Daqui resulta que o $\det A \neq 0$ se e só se $\{u, v, w\}$ é um conjunto linearmente independente de vetores (cf. Observação 16), isto é, se e só se A é invertível.

Voltando ao Exemplo 65, conclui-se que o paralelipípedo definido pelos 3 vetores que constituem as colunas de A, $u = (1, 2, -1)$, $v = (0, 1, 3)$ e $w = (-2, 1, 1)$, tem volume 16. Daqui resulta, em particular, que $\{u, v, w\}$ é um conjunto de vetores linearmente independente e portanto que A é invertível.

4.2 Uma fórmula por recorrência

A regra de Sarrus não se generaliza para matrizes quadradas de ordem $n > 3$. Para definirmos o determinante de matrizes de ordem arbitrária $n \geq 2$ vamos dar uma fórmula por recorrência que permite obter esse determinante a partir de n determinantes de matrizes de ordem $n-1$([1]). Para isso necessitamos de introduzir alguns conceitos.

Definição 32. Sejam A uma matriz quadrada de ordem $n \geq 2$ e $1 \leq i, j \leq n$. Chama-se *menor complementar* do elemento (i, j) de A e denota-se por A_{ij}, ao determinante da submatriz de A que se obtém eliminando a linha i e a coluna j de A. Chama-se *complemento algébrico* ou *co-factor* do elemento (i, j) de A a

$$\Delta_{ij} = (-1)^{i+j} A_{ij}.$$

Exemplo 66. O menor complementar do elemento $(1,2)$ da matriz

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

é o determinante da submatriz de A que se obtém eliminando a linha 1 e a coluna 2 de A (a cinzento), isto é, é o determinante

$$A_{12} = \det \begin{bmatrix} 0 & 3 \\ -1 & 1 \end{bmatrix} = 0 \times 1 - 3 \times (-1) = 3,$$

e o respectivo co-factor é $\Delta_{12} = (-1)^{1+2} A_{12} = (-1) \times 3 = -3$.

Estamos agora em condições de enunciar a *regra de Laplace*.

Teorema 49. (Laplace) *Seja $A = [a_{ij}]$ uma matriz quadrada de ordem $n \geq 2$. Para qualquer linha $i = 1, \ldots, n$ e coluna $j = 1, \ldots, n$ tem-se:*

1. *Expansão do determinante ao longo da linha i:*

$$\det A = a_{i1}\Delta_{i1} + a_{i2}\Delta_{i2} + \cdots + a_{in}\Delta_{in}.$$

2. *Expansão do determinante ao longo da coluna j:*

$$\det A = a_{1j}\Delta_{1j} + a_{2j}\Delta_{2j} + \cdots + a_{nj}\Delta_{nj}.$$

Observação 38.

- O resultado **não depende** da escolha da linha i ou da coluna j da matriz.
- Devem-se escolher linhas ou colunas com o maior número possível de zeros uma vez que na expansão do determinante as parcelas associadas aos co-fatores de elementos nulos da matriz são também nulas. Em particular, matrizes com linhas ou colunas de zeros têm determinante nulo.

[1] O conceito de determinante de uma matriz de ordem n é definido usualmente recorrendo ao conjunto das permutações dos números $\{1, \ldots, n\}$, mas essa abordagem está fora do âmbito deste curso.

Vejamos nalguns exemplos como calcular o determinante usando a regra de Laplace.

Exemplos 67.

1. Consideremos a matriz $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} 1 & -2 & 3 \\ 0 & 2 & 3 \\ -1 & 2 & 1 \end{bmatrix}$.

   Aplicando a regra de Laplace ao longo da 2ª linha de A,

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} 1 & -2 & 3 \\ 0 & 2 & 3 \\ -1 & 2 & 1 \end{bmatrix},$$

   obtém-se

$$\begin{aligned} \det A &= a_{21}\Delta_{21} + a_{22}\Delta_{22} + a_{23}\Delta_{23} \\ &= 0 \times (-1)^{2+1} \det \begin{bmatrix} -2 & 3 \\ 2 & 1 \end{bmatrix} + 2 \times (-1)^{2+2} \det \begin{bmatrix} 1 & 3 \\ -1 & 1 \end{bmatrix} + \\ &\quad 3 \times (-1)^{2+3} \det \begin{bmatrix} 1 & -2 \\ -1 & 2 \end{bmatrix} \\ &= 0 + 2 \times (1+3) - 3 \times 0 = 8. \end{aligned}$$

   Expandindo agora o determinante ao longo da 1ª coluna de A,

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} 1 & -2 & 3 \\ 0 & 2 & 3 \\ -1 & 2 & 1 \end{bmatrix},$$

   obtém-se o mesmo valor. De facto,

$$\begin{aligned} \det A &= a_{11}\Delta_{11} + a_{21}\Delta_{21} + a_{31}\Delta_{31} = 1 \times (-1)^{1+1} \det \begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix} + \\ &\quad 0 \times (-1)^{2+1} \det \begin{bmatrix} -2 & 3 \\ 2 & 1 \end{bmatrix} + (-1) \times (-1)^{3+1} \det \begin{bmatrix} -2 & 3 \\ 2 & 3 \end{bmatrix} \\ &= 1 \times (2-6) + 0 + (-1) \times (-6-6) = 8. \end{aligned}$$

2. Consideremos a matriz $A = \begin{bmatrix} -1 & 2 & -1 & 0 \\ 0 & 1 & -1 & 2 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 2 & -3 \end{bmatrix}$.

Aplicando a regra de Laplace ao longo da 4ª coluna, obtém-se,

$$\begin{aligned}
\det A &= 0\times\Delta_{14}+2\times\Delta_{24}+0\times\Delta_{34}+(-3)\times\Delta_{44}\\
&= 2\times(-1)^{2+4}\begin{vmatrix} -1 & 2 & -1\\ 1 & 0 & 1\\ 0 & 1 & 2\end{vmatrix}+(-3)\times(-1)^{4+4}\begin{vmatrix} -1 & 2 & -1\\ 0 & 1 & -1\\ 1 & 0 & 1\end{vmatrix}\\
&= 2\times(-1)^{2+4}(-4)+(-3)\times(-1)^{4+4}(-2)=-2\\
&\quad \text{(o cálculo dos 2 determinantes } 3\times 3 \text{ fica como exercício).}
\end{aligned}$$

3. A regra de Laplace permite deduzir a regra de Sarrus dada anteriormente para matrizes de ordem 3. De facto, expandindo o determinante da matriz $A=\begin{bmatrix} a & b & c\\ d & e & f\\ g & h & i\end{bmatrix}$ ao longo da 1ª linha obtém-se,

$$\begin{aligned}
\det A &= a\Delta_{11}+b\Delta_{12}+c\Delta_{13}\\
&= a\times(-1)^{1+1}\det\begin{bmatrix} e & f\\ h & i\end{bmatrix}+b\times(-1)^{1+2}\det\begin{bmatrix} d & f\\ g & i\end{bmatrix}+\\
&\quad c\times(-1)^{1+3}\det\begin{bmatrix} d & e\\ g & h\end{bmatrix}\\
&= a(ei-fh)-b(di-fg)+c(dh-eg)\\
&= aei-afh-bdi+bfg+cdh-ceh\\
&= aei+bfg+cdh-(afh+bdi+ceh),
\end{aligned}$$

que é precisamente a expressão para o determinante de A obtida anteriormente pela regra de Sarrus. Deixa-se como exercício re-obter este determinante aplicando a regra de Laplace ao longo da 1ª coluna de A.

A regra de Laplace permite obter o determinante de matrizes triangulares, e em particular de matrizes diagonais e escalares, de forma imediata.

Proposição 50. *O determinante de uma matriz triangular superior ou inferior é o produto dos elementos da diagonal principal da matriz, isto é,*

$$\det\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n}\\ 0 & a_{22} & \cdots & a_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & a_{nn}\end{bmatrix}=\det\begin{bmatrix} a_{11} & 0 & \cdots & 0\\ a_{21} & a_{22} & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ a_{n1} & a_{n2} & \cdots & a_{nn}\end{bmatrix}=a_{11}a_{22}\cdots a_{nn}.$$

Em particular, obtém-se para matrizes diagonais e escalares:

1. $\det diag(a_1,\ldots,a_n) = \det \begin{bmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_n \end{bmatrix} = a_1 a_2 \cdots a_n.$

2. $\det(\lambda I_n) = \det diag(\lambda,\ldots,\lambda) = \lambda^n.$

3. $\det I_n = 1.$

Demonstração. Vamos considerar o caso das matrizes triangulares superiores (a demonstração para o caso das matrizes triangulares inferiores é feita de modo análogo). Aplicando a regra de Laplace ao longo da 1ª coluna obtém-se,

$$\begin{aligned} \det \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & 0 & a_{33} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{bmatrix} &= a_{11}(-1)^{1+1} \det \begin{bmatrix} a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix} + 0 + \cdots + 0 \\ &= a_{11} \times \det \begin{bmatrix} a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix}_{(n-1)\times(n-1)} \end{aligned}$$

Aplicando agora a regra de Laplace ao longo da 1ª coluna da submatriz de ordem $(n-1)$ que resultou de eliminar a 1ª linha e a 1ª coluna de A, obtém-se

$$\begin{aligned} \det \begin{bmatrix} a_{22} & a_{23} & \cdots & a_{1n} \\ 0 & a_{33} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix} &= a_{22}(-1)^{1+1} \det \begin{bmatrix} a_{33} & \cdots & a_{3n} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_{nn} \end{bmatrix} + 0 + \cdots + 0 \\ &= a_{22} \times \det \begin{bmatrix} a_{33} & \cdots & a_{3n} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_{nn} \end{bmatrix}_{(n-2)\times(n-2)} . \end{aligned}$$

e portanto

$$\det \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix} = a_{11} a_{22} \times \det \begin{bmatrix} a_{33} & \cdots & a_{3n} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_{nn} \end{bmatrix}_{(n-2)\times(n-2)} .$$

Procedendo de modo análogo relativamente à submatriz resultante de ordem $(n-2)$ e assim sucessivamente até se chegar à matriz de ordem 2, conclui-se que

$$\det\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix} = a_{11}a_{22}\cdots a_{n-2,n-2}\times\det\begin{bmatrix} a_{n-1,n-1} & a_{n-1,n} \\ 0 & a_{nn} \end{bmatrix}$$
$$= a_{11}a_{22}\cdots a_{n-2,n-2}a_{n-1,n-1}a_{nn},$$

o que prova o resultado. □

4.3 Propriedades do determinante

O seguinte resultado enuncia algumas propriedades importantes do determinante.

Proposição 51. *Sejam A e B matrizes quadradas de ordem n. Tem-se:*

1. $\det(AB) = \det A \det B$
2. $\det(A^T) = \det A$
3. $\det(\lambda A) = \lambda^n \det A$.
4. *A é invertível se e só se* $\det A \neq 0$, *tendo-se* $\det A^{-1} = \dfrac{1}{\det A}$.

A demonstração das duas primeiras propriedades é não trivial e será omitida. A terceira propriedade é consequência do determinante da matriz escalar $\det(\lambda I_n) = \lambda^n$ (ver a Proposição 50) e da primeira propriedade. De facto,

$$\det(\lambda A) = \det[\lambda(I_n A)] = \det[(\lambda I_n)A] = \det(\lambda I_n)\det A = \lambda^n \det A.$$

O determinante da matriz inversa decorre também da primeira propriedade, uma vez que

$$1 = \det I_n = \det(AA^{-1}) = \det A \det(A^{-1}).$$

Exemplo 68. Consideremos a matriz $A = \begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix}$ de ordem $n = 2$ e $\lambda = 5$.

Tem-se:

- $\det A^T = \det\begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix} = 1\times 1 - (-1)\times 2 = 3 = \det A.$
- $\det(\lambda A) = \det(5A) = \det\begin{bmatrix} 5 & 10 \\ -5 & 5 \end{bmatrix} = 75 = 5^2 \times 3 = \lambda^n \det A.$

- $A^{-1} = \begin{bmatrix} \frac{1}{3} & -\frac{2}{3} \\ \frac{1}{3} & \frac{1}{3} \end{bmatrix}$ (verifique) e $\det A^{-1} = \frac{1}{3} \times \frac{1}{3} - (-\frac{2}{3}) \times \frac{1}{3} = \frac{1}{3} = \dfrac{1}{\det A}$.

Observação 39.

- Ao contrário do que sucede para o produto de matrizes, a soma "não se comporta bem" com o determinante, isto é, em geral,

$$\det(A+B) \neq \det A + \det B.$$

- Se A possui linhas ou colunas **múltiplas entre si**, $\det A = 0$ uma vez que A é não invertível.

Cálculo do determinante via método de Gauss

O cálculo do determinante de uma matriz pode ser obtido usando o método de eliminação de Gauss. Para isso é necessário começar por analisar como o determinante é afetado quando se aplicam as 3 operações elementares do método de Gauss:

1. Adicionar um múltiplo de uma linha a outra linha **não afeta o valor do determinante**:

$$\det \begin{bmatrix} \vdots \\ \hline L_i \\ \hline \vdots \\ \hline L_j \\ \hline \vdots \end{bmatrix} = \det \begin{bmatrix} \vdots \\ \hline L_i \\ \hline \vdots \\ \hline L_j + \alpha L_i \\ \hline \vdots \end{bmatrix} \qquad (L_j \to L_j + \alpha L_i)$$

2. Trocar duas linhas entre si **troca o sinal do determinante**:

$$\det \begin{bmatrix} \vdots \\ \hline L_i \\ \hline \vdots \\ \hline L_j \\ \hline \vdots \end{bmatrix} = -\det \begin{bmatrix} \vdots \\ \hline L_j \\ \hline \vdots \\ \hline L_i \\ \hline \vdots \end{bmatrix} \qquad (L_i \longleftrightarrow L_j)$$

3. Multiplicar uma linha por um escalar não nulo, **multiplica o determinante por esse escalar**:

$$\det\begin{bmatrix} L_1 \\ \hline \vdots \\ \hline L_i \\ \hline \vdots \\ \hline L_n \end{bmatrix} = \frac{1}{\alpha}\det\begin{bmatrix} L_1 \\ \hline \vdots \\ \hline \alpha L_i \\ \hline \vdots \\ \hline L_n \end{bmatrix} \qquad (L_i \longrightarrow \alpha L_i)$$

Estas propriedades decorrem da chamada *linearidade do determinante nas linhas da matriz*, que será considerada mais adiante e que é uma consequência da regra de Laplace. Antes porém vamos ilustrar nalguns exemplos como se pode usar o método de eliminação de Gauss para calcular o determinante de uma matriz.

Exemplos 69.

1. Consideremos a matriz $\begin{bmatrix} 0 & 1 & 3 \\ 1 & 1 & 2 \\ 2 & 1 & -1 \end{bmatrix}$. Aplicando o método de Gauss obtém-se,

$$\begin{aligned}
\det\begin{bmatrix} 0 & 1 & 3 \\ 1 & 1 & 2 \\ 2 & 1 & -1 \end{bmatrix} &= -\det\begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & 3 \\ 2 & 1 & -1 \end{bmatrix} \quad (L_1 \longleftrightarrow L_2) \\
&= -\det\begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & 3 \\ 0 & -1 & -5 \end{bmatrix} \quad (L_3 - 2L_1 \longrightarrow L_3) \\
&= -\det\begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & 3 \\ 0 & 0 & -2 \end{bmatrix} \quad (L_3 + L_2 \longrightarrow L_3) \\
&= -(1 \times 1 \times (-2)) = 2 \quad (\text{det. da matriz triangular}).
\end{aligned}$$

2. Consideremos a matriz $\begin{bmatrix} 1 & 1 & 2 & 0 \\ -1 & 1 & 1 & 2 \\ 2 & -1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$. Vamos ver como podemos combinar diferentes métodos para facilitar o cálculo do seu determinante. Começamos por aplicar a primeira operação elementar do método de

Gauss para eliminar todos os elementos da 1ª coluna de A com exceção do pivot. De seguida aplicamos a regra de Laplace ao longo da 1ª coluna para reduzir o cálculo do determinante da matriz 4×4 ao cálculo do determinante de uma matriz de ordem 3 e por fim obtemos o valor do determinante dessa matriz de ordem 3 aplicando a regra de Sarrus:

$$
\begin{aligned}
\det \begin{bmatrix} 1 & 1 & 2 & 0 \\ -1 & 1 & 1 & 2 \\ 2 & -1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} &= \det \begin{bmatrix} 1 & 1 & 2 & 0 \\ 0 & 2 & 3 & 2 \\ 0 & -3 & -3 & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix} \\
&= 1 \times (-1)^{1+1} \det \begin{bmatrix} 2 & 3 & 2 \\ -3 & -3 & 1 \\ 0 & -1 & 1 \end{bmatrix} + 0 + 0 + 0 \\
&= \det \begin{bmatrix} 2 & 3 & 2 \\ -3 & -3 & 1 \\ 0 & -1 & 1 \end{bmatrix} \begin{matrix} 2 & 3 \\ -3 & -3 \\ 0 & -1 \end{matrix} \\
&= 2 \times (-3) \times 1 + 0 + 2 \times (-3) \times (-1) \\
&\quad -(0 + 2 \times 1 \times (-1) + 3 \times (-3) \times 1) = 11.
\end{aligned}
$$

Linearidade do determinante nas linhas de uma matriz

Denotando as matrizes pelas suas linhas e aplicando a regra de Laplace pode-se mostrar que o determinante verifica as seguintes propriedades:

1. $\det \begin{bmatrix} L_1 \\ \hline \vdots \\ \hline \color{red}{L_i'} + \color{blue}{L_i''} \\ \hline \vdots \\ \hline L_n \end{bmatrix} = \det \begin{bmatrix} L_1 \\ \hline \vdots \\ \hline \color{red}{L_i'} \\ \hline \vdots \\ \hline L_n \end{bmatrix} + \det \begin{bmatrix} L_1 \\ \hline \vdots \\ \hline \color{blue}{L_i''} \\ \hline \vdots \\ \hline L_n \end{bmatrix};$

2. $\det \begin{bmatrix} L_1 \\ \hline \vdots \\ \hline \color{red}{\alpha} L_i \\ \hline \vdots \\ \hline L_n \end{bmatrix} = \color{red}{\alpha} \det \begin{bmatrix} L_1 \\ \hline \vdots \\ \hline L_i \\ \hline \vdots \\ \hline L_n \end{bmatrix}, \forall \color{red}{\alpha} \in \mathbb{R}.$

Têm-se propriedades análogas relativamente às colunas da matriz, isto é, o determinante é também *linear nas colunas*, uma vez que o determinante de uma matriz é igual ao determinante da sua transposta. Deixa-se como exercício enunciar essas propriedades.

Para terminar este capítulo vamos utilizar a primeira propriedade da linearidade do determinante nas linhas para verificar num exemplo que o determinante fica invariante quando se aplica a 1ª operação elementar do método de Gauss.

Exemplo 70. Consideremos a matriz $A = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$. Multiplicando a 1ª linha de A por -2 e adicionando o resultado à 2ª linha, obtém-se o mesmo determinante,

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

De facto, atendendo a que o determinante de uma matriz com linhas múltiplas entre si é nulo (*) e à 1ª propriedade da linearidade do determinante nas linhas da matriz (**), pode-se escrever

$$\begin{aligned}
\det\begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix} &= \det\begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix} + \underbrace{\det\begin{bmatrix} 1 & 2 \\ (-2)\times 1 & (-2)\times 2 \end{bmatrix}}_{=0(*)} \\
&\overset{(**)}{=} \det\begin{bmatrix} 1 & 2 \\ 2+(-2)\times 1 & 3+(-2)\times 2 \end{bmatrix} \\
&= \det\begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix} = -1.
\end{aligned}$$

Capítulo 5

Valores e vetores próprios

5.1 Valores e vetores próprios

Definição 33. Sejam A matriz quadrada de ordem n e $v \in \mathbb{R}^n$, com $v \neq \vec{0}$. Dizemos que v é *vetor próprio* de A se existir $\lambda \in \mathbb{R}$ (ou $\lambda \in \mathbb{C}$) tal que $Av = \lambda v$. O escalar λ designa-se por *valor próprio* de A associado ao vetor próprio v.

Exemplo 71. O vetor $v = (10, 1, 5)$ é vetor próprio de $A = \begin{bmatrix} 3 & 0 & 2 \\ 0 & -1 & 1 \\ 2 & 0 & 0 \end{bmatrix}$ associado ao valor próprio $\lambda = 4$, uma vez que

$$Av = \begin{bmatrix} 3 & 0 & 2 \\ 0 & -1 & 1 \\ 2 & 0 & 0 \end{bmatrix} \begin{bmatrix} 10 \\ 1 \\ 5 \end{bmatrix} = \begin{bmatrix} 40 \\ 4 \\ 20 \end{bmatrix} = 4 \begin{bmatrix} 10 \\ 1 \\ 5 \end{bmatrix} = 4v.$$

Observação 40. Qualquer múltiplo não nulo de um vetor próprio v de A, $u = \alpha v$, $\alpha \neq 0$, é ainda vetor próprio de A associado ao mesmo valor próprio λ. De facto, se $v \in \mathbb{R}^n$, $v \neq \vec{0}$, é vetor próprio de A associado ao valor próprio λ e $\alpha \neq 0$, então

$$A(\alpha v) = \alpha A v = \alpha(\lambda v) = \lambda(\alpha v).$$

Isso significa que podemos falar em *direções próprias* associadas à matriz A, uma vez que os vetores não nulos da reta que passa na origem com direção de um vetor próprio de A são ainda vetores próprios de A associados ao mesmo valor próprio.

Subespaço próprio

Se A é matriz quadrada de ordem n, $v \in \mathbb{R}^n$ e $\lambda \in \mathbb{R}$, têm-se as equivalências

$$\begin{aligned} Av = \lambda v &\iff Av - \lambda v = \vec{0} \\ &\iff (A - \lambda I)v = \vec{0} \\ &\iff v \in \mathcal{N}(A - \lambda I). \end{aligned}$$

Logo $v \in \mathbb{R}^n$ é vetor próprio de A associado ao valor próprio λ se e só se

$$v \in \mathcal{N}(A - \lambda I) \setminus \{\vec{0}\}.$$

A observação anterior motiva a seguinte definição.

Definição 34. Sejam A matriz quadrada de ordem n e $\lambda \in \mathbb{R}$ valor próprio de A. Chama-se *subespaço próprio* de A associado ao valor próprio λ e denota-se por $E(\lambda)$, ao subespaço vetorial $\mathcal{N}(A - \lambda I)$. A dimensão de $E(\lambda)$ designa-se por *multiplicidade geométrica* de λ e denota-se por $m.g.(\lambda)$.

Pelas considerações anteriores $v \in E(\lambda)$ se e só se $Av = \lambda v$. Logo, $v \in \mathbb{R}^n$ é vetor próprio de A associado ao valor próprio λ se e só se for um vetor não nulo do subespaço próprio $E(\lambda)$.

Exemplo 72. Consideremos novamente a matriz A do Exemplo 71 que possui o valor próprio $\lambda = 4$. Vamos calcular $E(4) = N(A - 4I)$. Aplicando o método de Gauss tem-se,

$$A - 4I = \begin{bmatrix} -1 & 0 & 2 \\ 0 & -5 & 1 \\ 2 & 0 & -4 \end{bmatrix} \to \cdots \to \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & -\frac{1}{5} \\ 0 & 0 & 0 \end{bmatrix}.$$

Portanto $E(4) = N(A - 4I) = \left\{\left(2x_3, \frac{1}{5}x_3, x_3\right) : x_3 \in \mathbb{R}\right\} = \left\langle\left(2, \frac{1}{5}, 1\right)\right\rangle = \langle(10, 1, 5)\rangle$ e por conseguinte $E(4)$ define uma reta que passa na origem com vetor diretor $(10, 1, 5)$. Os vetores próprios de A associados $\lambda = 4$ são os vetores não nulos dessa reta, isto é, os múltiplos não nulos de $(10, 1, 5)$.

Polinómio característico

Seja A uma matriz quadrada de ordem n. Tem-se que $\lambda \in \mathbb{R}$ é valor próprio de A se e só existir $v \in \mathbb{R}^n$, com $v \neq \vec{0}$, tal que $Av = \lambda v$, ou seja, tal que $(A - \lambda I)v = \vec{0}$. Tal condição significa que o sistema homogéneo $(A - \lambda I)x = \vec{0}$ admite a solução <u>não trivial</u> $v \neq \vec{0}$ e portanto é indeterminado. Logo $\lambda \in \mathbb{R}$ é valor próprio de A se e só se a matriz $A - \lambda I$ for não invertível, isto é, se e só se, $\det(A - \lambda I) = 0$.

Observação 41. Pode-se mostrar que $p_A(x) = \det(A - xI)$ define um polinómio em x de grau n. Os valores próprios de A são portanto as soluções da equação $p_A(x) = 0$, isto é, as raízes do polinómio $p_A(x)$.

Definição 35. O polinómio $p_A(x) = \det(A - xI)$ é designado por *polinómio característico* de A. O número de vezes que um valor próprio λ aparece repetido como raíz de $p_A(x)$ designa-se por *multiplicidade algébrica* de λ e denota-se por m.a.(λ).

Observação 42. (Factorização do polinómio do 2º grau)

Recordemos que se o polinómio $p(x) = ax^2 + bx + c$ com $a \neq 0$, admite raízes α e β reais ou complexas e não necessariamente distintas, tem-se

$$p(x) = a(x - \alpha)(x - \beta).$$

Exemplo 73. Consideremos novamente a matriz $A = \begin{bmatrix} 3 & 0 & 2 \\ 0 & -1 & 1 \\ 2 & 0 & 0 \end{bmatrix}$. Aplicando a regra de Laplace ao longo da 2ª coluna, tem-se,

$$\begin{aligned}
p_A(x) &= \det(A - xI) = \begin{vmatrix} 3-x & 0 & 2 \\ 0 & -1-x & 1 \\ 2 & 0 & -x \end{vmatrix} \\
&= 0 + (-1-x)(-1)^{2+2} \begin{vmatrix} 3-x & 2 \\ 2 & -x \end{vmatrix} + 0 \\
&= (-1-x)(-1)^4[(3-x)(-x) - 4] = -(x+1)(x^2 - 3x - 4) \\
&= -(x+1)(x+1)(x-4) \quad \text{(pela fórmula resolvente e a Observação 42)} \\
&= -(x+1)^2(x-4).
\end{aligned}$$

Logo, $p_A(x) = 0$ se só se $x = -1$ (raíz dupla) ou $x = 4$ (raíz simples). Daqui resulta que A admite os valores próprios -1 e 4 tendo-se:

- m.a.$(-1) = 2$ e m.a.$(4) = 1$.
- $E(4) = \langle (10, 1, 5) \rangle$ (como vimos) e portanto m.g.$(4) = \dim E(4) = 1$.
- Relativamente ao valor próprio $\lambda = -1$ tem-se (verifique),

  $$A - (-I) = A + I = \begin{bmatrix} 4 & 0 & 2 \\ 0 & 0 & 1 \\ 2 & 0 & 1 \end{bmatrix} \longrightarrow \cdots \longrightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix},$$

  e portanto,

  $$\begin{aligned}
  E(-1) &= \mathcal{N}(A + I) \\
  &= \{(x_1, x_2, x_3) : x_1 = 0,\ x_3 = 0, x_2 \in \mathbb{R}\} \\
  &= \{(0, x_2, 0) : x_2 \in \mathbb{R}\} = \langle (0, 1, 0) \rangle.
  \end{aligned}$$

  Logo m.g.$(-1) = \dim E(-1) = 1$.

A informação sobre os valores e vetores próprios da matriz A pode ser organizada numa tabela da seguinte forma:

λ	m.a.(λ)	m.g.(λ)	base de $E(\lambda)$
-1	2	1	$\{(0,1,0)\}$
4	1	1	$\{(10,1,5)\}$

Observação 43. Para decidir se λ (real ou complexo) é valor próprio de uma matriz A, basta verificar se λ é raíz do polinómio característico $p_A(x)$, isto é, se

$$p_A(\lambda) = \det(A - \lambda I) = 0.$$

Não é necessário determinar as raízes de $p_A(x)$, o que pode nem ser possível!([1])

Exemplo 74. Para testar se -1 é valor próprio de $A = \begin{bmatrix} -1 & 1 & 0 & 2 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 2 & 0 & 0 & 1 \end{bmatrix}$, basta ver se é raíz de $p_A(x)$. Ora,

$$p_A(-1) = \det(A-(-1)I) = \det(A+I) = \det \begin{bmatrix} 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1 \\ 2 & 0 & 0 & 2 \end{bmatrix} = 0,$$

pois a 2ª linha nula da matriz $A+I$ é nula. Logo -1 é valor próprio de A.

Exemplo 75. Vejamos se i é valor próprio da matriz $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. Ora,

$$p_A(i) = \det(A - iI) = \det \begin{bmatrix} -i & 1 \\ -1 & -i \end{bmatrix} = (-i)^2 + 1 = i^2 + 1 = -1 + 1 = 0,$$

e portanto i é valor próprio de A.

[1] Para $n \geq 5$ não existe o análogo de uma "fórmula resolvente"que permita determinar as raízes de um polinómio arbitrário de grau n.

Propriedades dos valores próprios

Enunciam-se a seguir várias propriedades importantes dos valores próprios de uma matriz.

Proposição 52. *Seja A uma matriz quadrada de ordem n. Tem-se:*

1. *Para qualquer valor próprio λ de A, $1 \leq m.g.(\lambda) \leq m.a.(\lambda)$, isto é,*

$$\dim E(\lambda) \leq m.a.(\lambda).$$

2. *A soma das multiplicidades algébricas dos valores próprios distintos de A é igual à ordem da matriz A.*

3. *A soma dos valores próprios de A, contando com repetições, é igual à soma dos elementos da diagonal principal de A, que se designa por* traço *de A e se denota por $tr(A)$.*

4. *O produto dos valores próprios de A, contando com repetições, é igual ao determinante de A. Em particular, A é invertível se e só se 0 não é valor próprio de A.*

Vejamos as propriedades anteriores num exemplo.

Exemplo 76. Voltando ao Exemplo 73 em que $\Lambda(A) = \{-1, 4\}$, tem-se o seguinte:

1. $1 = \text{m.g.}(-1) \leq \text{m.a.}(-1) = 2$ e $1 = \text{m.g.}(4) \leq \text{m.a.}(4) = 1$.

2. $\text{m.a.}(-1) + \text{m.a.}(4) = 2 + 1 = 3 = n$.

3. $\text{tr}(A) = 3 + (-1) + 0 = 2$ que coincide com a soma dos valores próprios de A contando com repetições, $-1 + (-1) + 4$.

4. $\det(A) = 4$ (verifique), que coincide com o produto dos valores próprios de A contando com repetições, $(-1)\times(-1)\times 4$. Em particular, A é invertível pois 0 não é valor próprio de A.

5.2 Base própria e diagonalização

Vetores próprios e independência linear

Lema 53. *Se λ e μ são valores próprios distintos de A, então $E(\lambda) \cap E(\mu) = \{\vec{0}\}$.*

Demonstração. Seja $v \in E(\lambda) \cap E(\mu)$. Por um lado, tem-se $Av = \lambda v$, uma vez que $v \in E(\lambda)$. Por outro lado, como $v \in E(\mu)$ tem-se $Av = \mu v$. Logo $\lambda v = \mu v$, ou seja, $(\lambda - \mu)v = \vec{0}$. Por conseguinte $v = \vec{0}$ pois $\lambda \neq \mu$. □

Proposição 54. *Consideremos uma matriz quadrada A de ordem n e sejam $\lambda_1, \ldots, \lambda_k$, os valores próprios* <u>*distintos*</u> *de A. Tem-se:*

1. *Um conjunto formado por k vetores próprios de A associados aos valores próprios $\lambda_1, \ldots, \lambda_k$ é linearmente independente.*

2. *Um conjunto de vetores obtido reunindo bases dos subespaços próprios $E(\lambda_1), \ldots, E(\lambda_k)$ é linearmente independente.*

Demonstração. Vamos demonstrar os resultados para $k = 2$ denotando por λ e μ os valores próprios distintos de A.

1. Sejam $u \in E(\lambda)$ e $v \in E(\mu)$ vetores próprios de A associados a λ e μ respectivamente e consideremos uma combinação linear nula

$$\alpha u + \beta v = \vec{0}.$$

Então $\alpha u = -\beta v$ e portanto $\alpha u, \beta v \in E(\lambda) \cap E(\mu) = \{\vec{0}\}$ pelo Lema 53. Logo, $\alpha u = \beta v = \vec{0}$. Como $u, v \neq \vec{0}$, $\alpha = \beta = 0$. Logo o conjunto $\{u, v\}$ é linearmente independente.

2. Sejam $\{u_1, \ldots, u_\ell\}$ base de $E(\lambda)$ e $\{v_1, \ldots, u_p\}$ base de $E(\mu)$ e consideremos uma combinação linear nula

$$\alpha_1 u_1 + \cdots + \alpha_\ell u_\ell + \beta_1 v_1 + \cdots + \beta_p v_p = \vec{0}.$$

Então, $\alpha_1 u_1 + \cdots + \alpha_\ell u_\ell = -(\beta_1 v_1 + \cdots + \beta_p v_p) \in E(\lambda) \cap E(\mu) = \{\vec{0}\}$, ou seja, $\alpha_1 u_1 + \cdots + \alpha_\ell u_\ell = \beta_1 v_1 + \cdots + \beta_p v_p = \vec{0}$. Como $\{u_1, \ldots, u_\ell\}$ é base de $E(\lambda)$, é linearmente independente e portanto $\beta_1 = \cdots = \beta_p = 0$. Como $\{v_1, \ldots, v_p\}$ é base de $E(\mu)$ é linearmente independente e portanto $\beta_1 = \cdots = \beta_p = 0$. Logo $\alpha_1 = \cdots = \alpha_\ell = \beta_1 = \cdots = \beta_p = 0$ e portanto $\{u_1, \ldots, u_\ell, v_1, \ldots, u_p\}$ é um conjunto linearmente independente de vetores.

Para $k > 2$ os resultados 1. e 2. são demonstrados utilizando ideias semelhantes. $\square$

Observação 44. Seja A uma matriz quadrada de ordem n.

- Se A admitir n valores próprios distintos $\lambda_1, \ldots \lambda_n$, tem-se pela Proposição 54 que existe um conjunto linearmente independente constituído por n vetores próprios de A associados aos valores próprios $\lambda_1, \ldots, \lambda_n$, isto é, uma base de $\mathbb{R}^n$ formada por vetores próprios de A, que se designa por *base própria* de $\mathbb{R}^n$ associada à matriz A.

- Se A admitir k valores próprios distintos $\lambda_1, \ldots, \lambda_k$, com $k < n$, existe uma base própria de $\mathbb{R}^n$ associada à matriz A se e só se

$$\dim E(\lambda_1) + \cdots + \dim E(\lambda_k) = \text{m.g.}(\lambda_1) + \cdots + \text{m.g.}(\lambda_k) = n. \tag{5.1}$$

Nessa altura cada subespaço próprio $E(\lambda_i)$ contribui para a base própria com m.g.(λ_i) vetores próprios.

- A condição (5.1) é equivalente a pedir que m.g.(λ) = m.a.(λ) para todo o valor próprio λ de A, uma vez que $1 \leq$ m.g.(λ) $\leq$ m.a.(λ) e a soma das multiplicidades algébricas dos valores próprios distintos de A é igual a n.

Exemplo 77. Consideremos $A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 1 & 1 \end{bmatrix}$. Tem-se $p_A(x) = (1-x)^2(2-x)$ (verifique) e portanto A admite os valores próprios 1 e 2, com m.a.(1) = 2 e m.a.(2) = 1. Calculando os subespaços próprios obtém-se (verifique),

$$\begin{aligned} E(1) &= \mathcal{N}(A-I) = \langle (1,0,0),(0,0,1) \rangle, \\ E(2) &= \mathcal{N}(A-2I) = \langle (1,1,1) \rangle, \end{aligned}$$

tendo-se m.g.(1) = $\dim E(1) = 2$, m.g.(2) = $\dim E(2) = 1$ e m.g.(1) + m.g.(2) = 3 que é a ordem da matriz A. Portanto reunindo bases dos subespaços próprios $E(1)$ e $E(2)$ obtém-se uma base própria de $\mathbb{R}^3$ associada à matriz A:

$$\{u_1, u_2, u_3\} = \Big\{ \underbrace{(1,0,0),(0,0,1)}_{\text{base de } E(1)}, \underbrace{(1,1,1)}_{\text{base de } E(2)} \Big\}.$$

Exemplo 78. Consideremos novamente a matriz $A = \begin{bmatrix} 3 & 0 & 2 \\ 0 & -1 & 1 \\ 2 & 0 & 0 \end{bmatrix}$ do Exemplo 73, cujos valores próprios são -1 e 4, com m.a.(-1) = 2 e m.a.(4) = 1. Vimos nesse exemplo que m.g.(-1) = m.g.(4) = 1. Uma vez que m.g.(-1) < m.a.(-1), não existe uma base de $\mathbb{R}^3$ formada por vetores próprios de A.

Definição 36. Uma matriz A de ordem n diz-se *diagonalizável* se existir uma matriz invertível P e uma matriz diagonal D tal que $P^{-1}AP = D$. A matriz P designa-se por *matriz de diagonalização* para A.

O seguinte resultado estabelece a relação entre o conceito de diagonalização e a existência de bases próprias.

Teorema 55. *Uma matriz quadrada A de ordem n é diagonalizável se e só se existe uma base de* $\mathbb{R}^n$ *formada por vetores próprios de A. Nessa altura, qualquer matriz* $P = [\, v_1 \ \cdots \ v_n \,]$ *cujas colunas constituam uma base de* $\mathbb{R}^n$ *formada por vetores próprios de A, é uma matriz de diagonalização para A, tendo-se,*

$$P^{-1}AP = diag(\lambda_1, \ldots, \lambda_n),$$

em que $\lambda_1, \ldots, \lambda_n$ *são os valores próprios (possivelmente repetidos) associados aos vetores próprios* $v_1, \ldots, v_n$*, isto é,* $Av_i = \lambda_i v_i$*,* $i = 1, \ldots, n$.

Demonstração. Sejam $P = [\, v_1 \;\cdots\; v_n \,]$ invertível e $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ diagonal. Têm-se as equivalências,

$$P^{-1}AP = D \Leftrightarrow P\,P^{-1}AP = PD \Leftrightarrow AP = PD,$$

com

$$AP = A\begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} = \begin{bmatrix} Av_1 & Av_2 & \cdots & Av_n \end{bmatrix}$$

e

$$PD = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} = \begin{bmatrix} \lambda_1 v_1 & \lambda_2 v_2 & \cdots & \lambda_n v_n \end{bmatrix},$$

como decorre da Observação 3 e do facto do produto de uma matriz por um vetor corresponder à combinação linear das colunas da matriz com os coeficientes dados pelas componentes do vetor. Logo, $AP = PD$ se e só se $Av_i = \lambda_i v_i$ para $i = 1, \ldots, n$ se e só se v_i é vetor próprio de A associado ao valor próprio λ_i para todo o i. Daqui resulta que A é diagonalizável com matriz de diagonalização $P = [\, v_1 \;\cdots\; v_n \,]$ tal que $P^{-1}AP = D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ se e só se $\{v_1, \ldots, v_n\}$ é uma base de $\mathbb{R}^n$ formada por vetores próprios de A associados aos valores próprios $\lambda_1, \ldots, \lambda_n$, respectivamente. □

Exemplo 79. Consideremos novamente a matriz $A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 1 & 1 \end{bmatrix}$ do Exemplo 77, cuja informação espectral se encontra resumida na tabela seguinte,

λ	m.a.(λ)	m.g.(λ)	base de $E(\lambda)$
1	2	2	$\{(1,0,0),(0,0,1)\}$
2	1	1	$\{(1,1,1)\}$

Reunindo as bases dos subespaços próprios $E(1)$ e $E(2)$, obteve-se a base própria de $\mathbb{R}^3$ associada à matriz A,

$$\{u_1, u_2, u_3\} = \{(1,0,0,),(0,0,1),(1,1,1)\}.$$

Daqui resulta que A é diagonalizável e que a matriz da base própria anterior,

$$P = [\, u_1 \;\; u_2 \;\; u_3 \,] = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix},$$

constitui uma matriz de diagonalização para A, tendo-se (confirme),

$$P^{-1}AP = \begin{bmatrix} 1 & -1 & 0 \\ 0 & -1 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

Da Observação 44 e do Teorema 55 anterior deduz-se imediatamente o seguinte resultado.

Corolário 56. *Seja A uma matriz quadrada de ordem n.*

- *Se A admitir n valores próprios distintos então é diagonalizável.*
- *Se A admitir k valores próprios distintos, com $k < n$, então é diagonalizável se e só se m.g.(λ) = m.a.(λ) para todo o valor próprio λ de A.*

A matriz A do Exemplo 73, é não diagonalizável, uma vez que o valor próprio -1 de A verifica m.a.$(-1) = 2 \neq$ m.g.$(-1) = 1$.

Observação 45. A partir da informação constituída pelos *valores próprios* de uma matriz diagonalizável e pelas *bases dos subespaços próprios* associados aos valores próprios dessa matriz, podemos reconstruir a matriz.

De facto, se A é uma matriz diagonalizável de ordem n, P é a matriz de $\mathbb{R}^n$ obtida reunindo bases dos subespaços de próprios de A e D a matriz diagonal que contém os correspondentes valores próprios, tem-se

$$\boxed{A = PDP^{-1}.}$$

Com efeito,

$$\begin{aligned} P^{-1}AP = D &\Leftrightarrow P(P^{-1}AP) = PD \\ &\Leftrightarrow AP = PD \\ &\Leftrightarrow (AP)P^{-1} = (PD)P^{-1} \\ &\Leftrightarrow A(PP^{-1}) = PDP^{-1} \\ &\Leftrightarrow A = PDP^{-1}. \end{aligned}$$

Uma aplicação: cálculo de potências de matrizes diagonalizáveis

Seja A uma matriz diagonalizável de ordem n, $\lambda_1, \ldots, \lambda_n$ os valores próprios de A (possivelmente repetidos) e P uma matriz de diagonalização para A tal que

$$P^{-1}AP = D = \text{diag}(\lambda_1, \ldots, \lambda_n).$$

A partir da Observação 45 tem-se $A = PDP^{-1}$ e portanto

$$\begin{aligned} A^\ell &= (PDP^{-1})^\ell = \underbrace{(PDP^{-1})(PDP^{-1})(PDP^{-1})\cdots(PDP^{-1})}_{\ell \text{ vezes}} \\ &= PD\underbrace{(P^{-1}P)}_{I_n}D\underbrace{(P^{-1}P)}_{I_n}D\cdots D\underbrace{(P^{-1}P)}_{I_n}DP^{-1} \\ &= PD^\ell P^{-1}. \end{aligned}$$

Atendendo a que a potência de uma matriz diagonal é dada pela potência dos elementos que estão na diagonal dessa matriz, tem-se

$$D^\ell = (\operatorname{diag}(\lambda_1, \dots, \lambda_n))^\ell = \operatorname{diag}(\lambda_1^\ell, \dots, \lambda_n^\ell),$$

obtendo-se finalmente

$$\boxed{A^\ell = P \operatorname{diag}(\lambda_1^\ell, \dots, \lambda_n^\ell) P^{-1}.}$$

Exemplo 80. Pretende-se determinar A^{10} com $A = \begin{bmatrix} 11 & 6 \\ -18 & -10 \end{bmatrix}$.

Tem-se $p_A(x) = x^2 - x - 2$ cujas raízes são -1 e 2, concluindo-se que A é diagonalizável. Calculando os subespaços próprios associados aos valores próprios -1 e 2 obtém-se a tabela

λ	m.a.(λ)	m.g.(λ)	base de $E(\lambda)$
-1	1	1	$\{(-1,2)\}$
2	1	1	$\{(-2,3)\}$

Uma matriz de diagonalização P para A é portanto $P = \begin{bmatrix} -1 & -2 \\ 2 & 3 \end{bmatrix}$, cuja inversa é $P^{-1} = \begin{bmatrix} 3 & 2 \\ -2 & -1 \end{bmatrix}$. Designando por $D = \operatorname{diag}(-1, 2)$ a matriz dos valores próprios de A associados às colunas de P obtém-se,

$$\begin{aligned} A^{10} = PD^{10}P^{-1} &= \begin{bmatrix} -1 & -2 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} (-1)^{10} & 0 \\ 0 & 2^{10} \end{bmatrix}\begin{bmatrix} 3 & 2 \\ -2 & -1 \end{bmatrix} \\ &= \begin{bmatrix} -1 & -2 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1024 \end{bmatrix}\begin{bmatrix} 3 & 2 \\ -2 & -1 \end{bmatrix} \\ &= \begin{bmatrix} 4093 & 2046 \\ -6138 & -3068 \end{bmatrix}. \end{aligned}$$

5.3 Decomposição espectral de matrizes simétricas

Recordemos que uma matriz quadrada $A = [a_{ij}]$ diz-se *simétrica* se $A = A^T$, isto é, se $a_{ij} = a_{ji}$ para todo o $i \neq j$. Por exemplo, a matriz

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 5 & 6 \end{bmatrix},$$

é simétrica pois

$$a_{12} = a_{21} = 2, \quad a_{13} = a_{31} = 3, \quad a_{23} = a_{32} = 5.$$

As matrizes simétricas constituem uma classe muito importante de matrizes diagonalizáveis. Mais precisamente temos o seguinte resultado que é apresentado sem demonstração.

Teorema 57. *Toda a matriz simétrica é diagonalizável.*

Os valores e vetores próprios de matrizes simétricas gozam de propriedades importantes.

Teorema 58. *Seja A uma matriz simétrica. Tem-se o seguinte:*

1. *Os valores próprios de A são reais.*
2. *Vetores próprios de A associados a valores próprios distintos são ortogonais entre si.*

Demonstração. Comecemos por recordar que se $x \in \mathbb{R}^n$, $\|x\|^2 = x^T x$. Se z for um vetor com n componentes complexas, tem-se $\|z\|^2 = z^T \bar{z} = \bar{z}^T z$, onde $\bar{z}$ denota o vetor cujas componentes são os conjugados das componentes de z.

Provemos então a primeira afirmação. Seja $\lambda \in \mathbb{C}$ um valor próprio de A e v um vetor próprio (possivelmente com componentes complexas) associado a λ, isto é, $Av = \lambda v$. Como A é simétrica e tem entradas reais $A^T = A = \bar{A}$ e tem-se,

$$\begin{aligned} \|Av\|^2 &= \overline{Av}^T Av = (A\bar{v})^T Av = \bar{v}^T A^T Av = \bar{v}^T A(\lambda v) \\ &= \bar{v}^T \lambda Av = \bar{v}^T \lambda^2 v = \lambda^2 \bar{v}^T v = \lambda^2 \|v\|^2. \end{aligned}$$

Como $\|v\|$ é um número real positivo e $\|Av\|^2$ é um número real não negativo, λ^2 é também um número real não negativo, o que implica que $\lambda \in \mathbb{R}$([2])

Provemos a segunda afirmação. Sejam u e v vetores próprios de A associados aos valores próprios λ e μ, respectivamente. Se $\lambda \neq \mu$ tem-se,

$$\lambda u \cdot v = (Au) \cdot v = (Au)^T v = u^T A^T v = u^T Av = u^T (Av) = u^T \mu v = \mu u^T v = \mu u \cdot v$$

Logo $(\lambda - \mu) u \cdot v = 0$. Como $\lambda \neq \mu$, $u \cdot v = 0$, ou seja, $u \perp v$. □

[2]Escrevendo $\lambda = x + iy$, obtém-se $\lambda^2 = (x^2 - y^2) + 2ixy$. Como $\lambda^2 \in \mathbb{R}$ tem-se $x = 0$ ou $y = 0$. Se $y \neq 0$ ter-se-ia $\lambda^2 = -y^2 < 0$. Como $\lambda^2 \geq 0$, $y = 0$. Logo $\lambda \in \mathbb{R}$.

Matrizes ortogonalmente diagonalizáveis

Nesta secção vamos dar uma decomposição para matrizes simétricas que constitui um dos resultados mais importantes da Álgebra Linear.

Definição 37. Uma matriz P de ordem n diz-se *ortogonal* se $P^T P = I_n$.

Note-se que em particular P é invertível tendo-se $P^{-1} = P^T$.

Proposição 59. *Seja* $\{v_1, \ldots, v_n\}$ *uma base de* $\mathbb{R}^n$ *e* $P = [\, v_1 \; v_2 \; \ldots \; v_n \,]$ *a matriz da base. Tem-se:*

(i) $\{v_1, \ldots, v_n\}$ *é uma base ortogonal se s só se a matriz* $P^T P$ *é diagonal.*

(ii) $\{v_1, \ldots, v_n\}$ *é uma base ortonormada se s só se a matriz* $P^T P = I$*, isto é,* P *é uma matriz ortogonal.*([3])

Demonstração. Por definição do produto de matrizes o elemento na posição (i, j) de $P^T P$ é o produto escalar da linha i de P^T, v_i^T, com a coluna j de P, v_j, ou seja, $P^T P = \left[\, v_i^T v_j \,\right]$. No caso (i) tem-se que $\{v_1, \ldots, v_n\}$ é ortogonal se e só se $v_i^T v_j = v_i \cdot v_j = 0$ para $i \neq j$, se e só se, $P^T P$ é diagonal. No caso (ii) tem-se que $\{v_1, \ldots, v_n\}$ é ortonormada se e só se $v_i \cdot v_j = 0$ para $i \neq j$ e $\|v_i\| = 1$, isto é, se e só se $v_i^T v_j = 0$ para $i \neq j$ e $v_i^T v_i = \|v_i\|^2 = 1$, se e só se $P^T P = I_n$. □

Definição 38. Uma matriz A diz-se *ortogonalmente diagonalizável* se existir uma matriz ortogonal P e uma matriz diagonal D tal que $P^T AP = D$.

Uma consequência importante do Teorema 58 é o seguinte resultado.

Teorema 60. *Uma matriz quadrada* A *é ortogonalmente diagonalizável se e só se for simétrica.*

Demonstração. Seja A for uma matriz simétrica de ordem n. Vejamos que A é ortogonalmente diagonalizável. Uma vez que A é simétrica, A é diagonalizável com valores próprios reais. Sejam $\lambda_1, \ldots, \lambda_n$ os valores próprios de A (possivelmente repetidos).

Ortogonalizando bases dos subespaços próprios associados aos valores próprios distintos de A (aplicando, por exemplo, o método de Gram-Schmidt) e reunindo essas bases ortogonais obtém-se uma base $\{v_1, \ldots, v_n\}$ de $\mathbb{R}^n$ constituída por vetores próprios de A. Esta base é ortogonal uma vez que a ortogonalidade entre os vetores próprios associados valores distintos está garantida pelo Teorema 58. Normalizando esta base (dividindo cada vetor pela sua norma), obtém-se uma base ortonormada de $\mathbb{R}^n$ constituída por vetores próprios de A. Logo a matriz da base ortonormada P, verifica $P^T P = I_n$, isto é, $P^{-1} = P^T$ e tem-se

$$P^T AP = P^{-1} AP = D = \text{diag}(\lambda_1, \ldots, \lambda_n),$$

[3] Note-se que uma matriz diz-se *ortogonal* se as suas colunas formarem uma base ortonormada, e não apenas uma base ortogonal!

o que mostra que A é ortogonalmente diagonalizável.

Reciprocamente consideremos uma matriz ortogonalmente diagonalizável A. Vejamos que A é simétrica. Por definição existe uma matriz ortogonal P, tal que $P^TAP = D$ com D uma matriz diagonal. Multiplicando a relação anterior à esquerda por P e à direita por P^T, e atendendo a que $P^T = P^{-1}$ conclui-se que $A = PDP^T$. Logo

$$A^T = (PDP^T)^T = (P^T)^T D^T P^T = PD^T P^T = PDP^T = A,$$

uma vez que as matrizes diagonais são simétricas. Logo A é simétrica. □

Exemplo 81. Consideremos a matriz simétrica $A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$. Pretende-se uma construir uma matriz de diagonalização ortogonal P para A.

Tem-se $p_A(\lambda) = -\lambda(\lambda-2)^2$ (verifique) e portanto A possui valores próprios 0 e 2, com $m.a.(0) = 1$ e $m.a.(2) = 2$. Calculando os respetivos subespaços próprios obtém-se $E(0) = \langle(-1,1,0)\rangle$ e $E(2) = \langle(1,1,0),(0,0,1)\rangle$ (verifique). Reunindo as bases dos subespaços próprios anteriores obtém-se a base ortogonal de $\mathbb{R}^3$,

$$\{(-1,1,0),(1,1,0),(0,0,1)\}.$$

Dividindo cada vetor da base pela sua norma, obtém-se uma base ortonormada de $\mathbb{R}^3$, constituída por vetores próprios de A,

$$\{v_1, v_2, v_3\} = \left\{\left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right), \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right), (0,0,1)\right\}.$$

Considerando finalmente $P = [\, v_1 \;\; v_2 \;\; v_3 \,]$ tem-se $P^{-1} = P^T$ (verifique) e

$$P^TAP = \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} = \mathrm{diag}(0,2,2),$$

o que significa que A é ortogonalmente diagonalizável com matriz de diagonalização ortogonal P.

O próximo resultado (sem demonstração) decompõe o produto de duas matrizes A e B, numa soma de matrizes de característica máxima um, obtidas como produtos de colunas de A com linhas de B e generaliza o resultado que diz que o produto de uma matriz A por um vetor v corresponde à combinação linear dos vetores que constituem as colunas de A com coeficientes dados pelas componentes de v.

Proposição 61. *Dadas matrizes* $A_{m\times n} = [\, u_1\ u_2\ \cdots\ u_n \,]$ *e* $B_{p\times n} = [\, v_1\ v_2\ \cdots\ v_n \,]$, *tem-se*

$$AB^T = \begin{bmatrix} u_1 & u_2 & \cdots & u_n^T \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_n^T \end{bmatrix} = u_1 v_1^T + \cdots + u_n v_n^T.$$

Vejamos resultado anterior num caso particular.

Exemplo 82. Consideremos as matrizes

$$A = [\, u_1\ u_2\ u_3 \,] = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}_{3\times 3}, \qquad B = [\, v_1\ v_2\ v_3 \,] = \begin{bmatrix} x & y & z \\ u & v & w \end{bmatrix}_{2\times 3}.$$

Tem-se

$$\begin{aligned}
AB^T &= \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{bmatrix} x & u \\ y & v \\ z & w \end{bmatrix} \\
&= \begin{bmatrix} ax+by+cz & au+bv+cw \\ dx+ey+fz & du+ev+fw \\ gx+hy+iz & gu+hv+iw \end{bmatrix} \\
&= \begin{bmatrix} ax & au \\ dx & du \\ gx & gu \end{bmatrix} + \begin{bmatrix} by & bv \\ ey & ev \\ hy & hv \end{bmatrix} + \begin{bmatrix} cz & cw \\ fz & fw \\ iz & iw \end{bmatrix} \\
&= \begin{bmatrix} a \\ d \\ g \end{bmatrix} \begin{bmatrix} x & u \end{bmatrix} + \begin{bmatrix} b \\ e \\ h \end{bmatrix} \begin{bmatrix} y & v \end{bmatrix} + \begin{bmatrix} c \\ f \\ i \end{bmatrix} \begin{bmatrix} z & w \end{bmatrix} \\
&= u_1 v_1^T + u_2 v_2^T + u_3 v_3^T.
\end{aligned}$$

Como consequência do resultado anterior e do Teorema 60 obtém-se o seguinte resultado fundamental de Álgebra Linear. A generalização desse resultado para matrizes arbitrárias, é conhecida como *decomposição em valores singulares* e constitui um dos resultados mais importantes e com mais aplicações da Álgebra Linear.

Teorema 62. (Decomposição espectral de matrizes simétricas) *Seja A uma matriz simétrica de ordem n e consideremos uma base ortonormada $\{v_1, \ldots, v_n\}$ de $\mathbb{R}^n$, constituída por vetores próprios de A associados aos valores próprios* *(possivelmente repetidos) $\lambda_1,\ldots,\lambda_n$, isto é, $Av_i=\lambda_i v_i$, $i=1,\ldots,n$. Tem-se a decomposição,*

$$A=\lambda_1 v_1 v_1^T+\lambda_2 v_2 v_2^T+\cdots+\lambda_n v_n v_n^T.$$

Note-se que as matrizes $v_i v_i^T$, $i=1,\ldots,n$, são simétricas e de característica um.

Demonstração. Seja P a matriz ortogonal cujas colunas são os vetores da base ortonormada $\{v_1,\ldots,v_n\}$ e $D=\mathrm{diag}(\lambda_1,\ldots,\lambda_n)$. Atendendo ao facto que o produto da matriz P pela matriz diagonal D corresponde a multiplicar as colunas de P pelos elementos da diagonal de D, à Proposição 61 e à relação $A=PDP^T$ (ver a demonstração do Teorema 60) obtém-se o resultado pretendido:

$$\begin{aligned}A=PDP^T &= \begin{bmatrix} v_1 & \cdots & v_n^T\end{bmatrix}\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{bmatrix}\begin{bmatrix} v_1^T \\ \vdots \\ v_n^T\end{bmatrix}\\ &= \begin{bmatrix}\lambda_1 v_1 & \cdots & \lambda_n v_n^T\end{bmatrix}\begin{bmatrix} v_1^T \\ \vdots \\ v_n^T\end{bmatrix}\\ &= \lambda_1 v_1 v_1^T+\cdots+\lambda_n v_n v_n^T.\end{aligned}$$

□

Exemplo 83. Aplicando o teorema da decomposição espectral à matriz simétrica do Exemplo 81, obtém-se a decomposição (verifique),

$$\begin{bmatrix}1&1&0\\1&1&0\\0&0&2\end{bmatrix}=0\times\begin{bmatrix}-\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\\ 0\end{bmatrix}\begin{bmatrix}-\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0\end{bmatrix}+2\times\begin{bmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\\ 0\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0\end{bmatrix}+2\times\begin{bmatrix}0\\0\\1\end{bmatrix}\begin{bmatrix}0&0&1\end{bmatrix}.$$

Capítulo 6

Introdução à programação linear

Nesta secção vamos abordar problemas em que se pretende determinar a afectação ótima de um conjunto de recursos limitados a um conjunto de actividades que competem entre si por esses recursos. Muitos destes problemas de otimização podem ser traduzidos por modelos matemáticos de programação linear (PL). Vamos começar por apresentar três pequenos problemas, o primeiro dos quais foi retirado dos antigos apontamentos da UC de Álgebra Linear [1]. A seguir, apresentaremos a expressão geral de um problema de PL.

6.1 Problema 1

Uma exploração agrícola dispõe de 80 ha de terreno para produzir tomate e trigo. Para além do terreno, os recursos susceptíveis de limitar a produção das duas culturas são a água e a mão de obra: sabe-se que cada hectare de tomate necessita de 8000 m^3 de água e de 40 h de mão de obra e que cada hectare de trigo apenas requer 20 h de mão de obra. A exploração agrícola dispõe de 320000 m^3 de água e 2000 horas de mão de obra. As receitas, por cada hectare de tomate e trigo cultivados são, respetivamente, 300€ e 200€. Pretende-se determinar a área a destinar a cada cultura por forma a maximizar a receita total.

Dados

Para facilitar a construção do modelo matemático resumimos na Tabela 6.1 a informação sobre os recursos disponíveis e a receita obtida por hectare de área cultivada de tomate e trigo.

Construção do modelo matemático

Variáveis de decisão

Para determinar a áreas de terreno a destinar a cada uma das culturas vamos considerar duas variáveis, x e y, que representam os hectares a atribuir ao to-

	Recursos		
	Água	Mão de obra	Receita
Tomate	8000 m^3/ha	40 h/ha	300 €/ha
Trigo		20 h/ha	200 €/ha
	$\leq$ 320000 m^3	$\leq$ 2000 h	max

Tabela 6.1

mate e ao trigo, respetivamente.

Função objetivo

A função objetivo que vamos representar pela variável z, traduz a relação entre o valor da receita total (em €) e as receitas obtidas pelo cultivo de x hectares de tomate e y hectares de trigo:

$$z = 300x + 200y.$$

Restrições funcionais

As restrições funcionais traduzem as limitações dos recursos disponíveis tendo em conta as necessidades de cada cultura:

- A área total de **terreno** cultivado não pode exceder 80 ha:

$$x + y \leq 80.$$

- O consumo de **água** com o cultivo de x hectares de tomate não pode exceder 320000 m^3:

$$8000x \leq 320000.$$

- A **mão de obra** utilizada no cultivo de x ha de tomate e y ha de trigo não pode exceder 2000 h:

$$40x + 20y \leq 2000.$$

Restrições de sinal

Para completar o modelo falta acrescentar as **restrições de sinal** que vão impedir que as variáveis x e y, que representam áreas, tomem valores negativos:

$$x, y \geq 0.$$

Formulação do problema

O problema pode então ser formulado matematicamente como,

$$\begin{array}{lrcr} \text{max} & z = 300x + 200y & & \\ \text{sujeito a} & x + y & \leq & 80 \\ & 8000x & \leq & 320000 \\ & 40x + 20y & \leq & 2000 \\ & x, \quad y & \geq & 0 \end{array} \qquad \text{(PPL 1)}$$

em que

$x =$ área (em ha) destinada à cultura de tomate,

$y =$ área (em ha) destinada à cultura de trigo.

A formulação (PPL 1) pode ser simplificada, dividindo a segunda restrição por 8000 e a terceira por 20:

$$\begin{array}{lrcr} \text{max} & z = 300x + 200y & & \\ \text{sujeito a} & x + y & \leq & 80 \\ & x & \leq & 40 \\ & 2x + y & \leq & 100 \\ & x, \quad y & \geq & 0 \end{array} \qquad \text{(PPL 1')}$$

Repare-se que apesar da cultura de trigo não necessitar de água e requerer menos horas de mão de obra que a cultura de tomate, também gera menos receita, pelo que não é óbvia qual a área a destinar a cada uma das culturas de modo a maximizar a receita.

6.2 Problema 2

Uma empresa produz três tipos de fertilizantes, A, B e C. Cada tonelada de fertilizante A, B e C gera 50, 40 e 60 unidades de resíduos tóxicos e origina um lucro de 10, 5 e 10 euros, respetivamente. A empresa tem capacidade para produzir 15 mil toneladas de fertilizantes por mês. Compromissos já assumidos obrigam a empresa a entregar mensalmente 5 mil toneladas de fertilizante A a um cliente. Pretende-se determinar o plano de produção mensal que gera a menor quantidade possível de resíduos tóxicos de modo a obter-se um lucro mensal de pelo menos 100 mil euros e uma produção mensal nunca inferior a 80% da capacidade de produção da empresa.

Dados

A informação anterior encontra-se resumida na Tabela 6.2.

	Resíduos	Lucro	
Fertlizante A	50 unid./t	10 €/t	$\geq$ 5000 t/mês
Fertlizante B	40 unid./t	5 €/t	
Fertlizante C	60 unid./t	10 €/t	
	min	$\geq$ 100000 €	

Tabela 6.2

Sabemos ainda que a capacidade de produção mensal da empresa é 15 mil toneladas por mês e que a produção mensal não deve ser inferior a 80% da capacidade máxima de produção.

Construção do modelo matemático

Variáveis de decisão

Uma vez que se pretende determinar a produção mensal de cada tipo de fertilizante, vamos considerar variáveis x_A, x_B e x_C que representam, respetivamente, a quantidade em toneladas, de fertilizante dos tipos A, B e C a produzir mensalmente.

Função objetivo

A função objetivo, que se pretende minimizar, traduz a quantidade de resíduos tóxicos que são gerados mensalmente com a produção dos fertilizantes. A produção de x_A toneladas de fertilizante A gera $50x_A$ unidades de resíduos tóxicos. Por seu turno, a produção de x_B toneladas de fertilizante B e de x_C toneladas de fertilizante C originam, respetivamente, $40x_B$ e $60x_C$ unidades de resíduos. Logo, a função objetivo vem dada por

$$\min z = 50x_A + 40X_B + 60x_C.$$

Restrições funcionais

As restrições funcionais são as seguintes:

- A **capacidade mensal de produção** de fertilizantes da empresa é de 15000 t que se traduz, em termos das variáveis de decisão, por

$$x_A + x_B + x_C \leq 15000.$$

- Devido a um compromisso com um cliente, a **produção mensal de fertilizante A** deve ser pelo menos de 5000 t, o que se traduz pela restrição

$$x_A \geq 5000.$$

- O **lucro mensal** que deve ser pelo menos de 100000 €, traduz-se em termos das variáveis de decisão por,

$$10x_A + 5x_B + 10x_C \geq 100000.$$

- Finalmente a **produção mensal de fertilizantes**, que deve representar pelo menos 80 % da capacidade mensal de produção, pode ser formulada como

$$x_A + x_B + x_C \geq 15000 \times 0.80 = 12000.$$

Restrições de sinal

Falta acrescentar as **restrições de sinal** que vão impedir que as variáveis x_A, x_B e x_C, que representam quantidades, tomem valores negativos:

$$x_A, x_B, x_C \geq 0.$$

Formulação do problema

O problema pode então ser formulado em termos de programação linear como,

$$\begin{array}{ll} \min & z = 50x_A + 40x_B + 60x_C \\ \text{sujeito a} & x_A + x_B + x_C \leq 15000 \\ & x_A \geq 5000 \\ & 10x_A + 5x_B + 10x_C \geq 100000 \\ & x_A + x_B + x_C \geq 12000 \\ & x_A, \; x_B, \; x_C \geq 0 \end{array} \qquad \text{(PPL 2)}$$

em que x_i, $i = A, B, C$, é a quantidade, em toneladas, de fertilizante do tipo i a produzir mensalmente.

6.3 Problema 3

Duas fábricas de produtos alimentares A e B, fabricam um dado produto que é enviado para dois armazéns, E e F. As quantidades de produto fabricado diariamente em cada fábrica e as distâncias entre as fábricas e armazéns encontram-se

Fábrica	Quantidade de produto (ℓ/dia)	Distância aos armazéns (Km) E	F
A	500	30	20
B	400	36	42

Tabela 6.3: Produção diária e distâncias entre fábricas e armazéns.

Armazém	Capacidade (ℓ/dia)	Custo de armazenamento (€/kℓ)
E	500	40
F	600	30

Tabela 6.4: Capacidades e custos de armazenamento.

na Tabela 6.3. As capacidades diárias e respectivos custo de armazenamento na Tabela 6.4.

Pretende-se determinar o plano diário de transporte do produto alimentar, das fábricas A e B para os armazéns E e F, de modo a minimizar o custo total do transporte e armazenamento, sabendo que transportar 1 kℓ desse produto tem um custo de 10€/km.

Construção do modelo matemático

Variáveis de decisão

Uma vez que se pretende o plano diário de transporte das fábricas A e B para os armazéns E e F, consideram-se 4 variáveis de decisão x_{ij}, $i = A, B$ e $j = E, F$, em que x_{ij} representa a quantidade de produto (em ℓ) a transportar da fábrica i para o armazém j.

Função objetivo

A função objetivo deve traduzir o custo total de transporte e armazenamento, que se pretende minimizar. Ora, o custo diário de transportar x_{AE} litros de produto da fábrica A para o armazém E é dado por

$$\frac{10}{1000} 30 x_{AE} = 0.3 x_{AE} \text{€}.$$

Efectuando cálculos semelhantes para obter os custos diários de transporte de A para F e de B para E e F e adicionando esses custos ao custo de transporte da fábrica A para o armazém E, obtém-se o custo total de transporte

$$0.3x_{AE} + 0.2x_{AF} + 0.36x_{BE} + 0.42x_{BF}.$$

O custo de armazenar $x_{AE}+x_{BE}$ litros de produto no armazém E é dado por

$$\frac{40}{1000}(x_{AE}+x_{BE})=0.04(x_{AE}+x_{BE})€.$$

Efetuando cálculos análogos para determinar o custo de armazenar $x_{AF}+x_{BF}$ litros de produto no armazém F e adicionando esse valor ao valor obtido para o armazém E, obtém-se custo total de armazenamento

$$0.04(x_{AE}+x_{BE})+0.03(x_{AF}+x_{BF}).$$

A função objetivo, que corresponde à soma dos custos totais de transporte e armazenamento, vem dada, após simplificação, por

$$z=0.34x_{AE}+0.23x_{AF}+0.4x_{BE}+0.45x_{BF}.$$

Restrições funcionais

As restrições funcionais, que refletem as produções e as capacidades de armazenamento diárias, vêm dadas por:

- A **produção diária** de A é enviada para os dois armazéns, tal como a produção diária de B, o que se traduz pelas equações lineares,

$$\begin{aligned} x_{AE}+x_{AF}&=500,\\ x_{BE}+x_{BF}&=400. \end{aligned}$$

- As **capacidades diárias** de armazenamento das fábricas E e F que são, respectivamente 500 ℓ e 600 ℓ, traduzem-se pelas inequações lineares,

$$\begin{aligned} x_{AE}+x_{BE}&\leq 500,\\ x_{AF}+x_{BF}&\leq 600. \end{aligned}$$

Restrições de sinal

Finalmente acrescentam-se as **restrições de sinal** que forçam as variáveis de decisão, que representam quantidades, a tomarem valores não negativos:

$$x_{ij}\geq 0,\quad i=A,B \text{ e } j=E,F.$$

Formulação do problema

O problema pode então ser formulado em termos de programação linear como,

$$\begin{array}{lllllll} \min & z=0.34x_{AE} & +0.23x_{AF} & +0.4x_{BE} & +0.45x_{BF} & & \\ \text{sujeito a} & x_{AE} & +x_{AF} & & & =500 & \\ & & & x_{BE} & +x_{BF} & =400 & \\ & x_{AE} & & +x_{BE} & & \leq 500 & \text{(PPL 3)}\\ & & x_{AF} & & +x_{BF} & \leq 600 & \\ & x_{AE}, & x_{AF}, & x_{BE}, & x_{BF} & \geq 0 & \end{array}$$

em que x_{ij}, com $i = A, B$ e $j = E, F$, representa a quantidade de produto (em ℓ) a transportar diariamente da fábrica i para o armazém j.

6.4 Caso geral

Num problema de PL, pretende-se determinar o(s) valor(es) de um conjunto de variáveis x_j $(j = 1, \ldots, k)$ que otimizam (maximizam ou minimizam), uma dada função linear z designada por *função objetivo* (f.o.),

$$\text{max ou min}\, z = c_1 x_1 + c_2 x_2 + ... + c_k x_k \qquad (f.o.)$$

satisfazendo um conjunto de restrições lineares,

$$\begin{aligned} &a_{11} x_1 + a_{12} x_2 + \cdots + a_{1n} x_k \geq, \leq \text{ ou } = b_1, && (1)\\ &a_{21} x_1 + a_{22} x_2 + \cdots + a_{2n} x_k \geq, \leq \text{ ou } = b_2, && (2)\\ &\vdots \\ &a_{m1} x_1 + a_{m2} x_2 + \cdots + a_{mk} x_k \geq, \leq \text{ ou } = b_m, && (m) \end{aligned}$$

e de restrições de sinal,

$$x_j \geq 0, \leq 0 \text{ ou livre, com } j = 1, ..., n. \qquad (m+1)$$

As constantes c_j $(j = 1, \ldots, k)$, a_{ij} $(i = 1, ..., m,\ j = 1, \ldots, k)$ e b_i $(i = 1, ..., m)$ constituem os ***parâmetros*** do problema. As restrições (1), …, (m) designam-se por ***restrições funcionais*** e as restrições (m+1) por ***restrições de sinal***. O conjunto de pontos que satisfazem as restrições (1), …, (m+1) designa-se por ***região admissível*** do problema e denota-se por $\mathscr{R}$. Um ponto da região admissível $\mathscr{R}$ designa-se por ***solução admissível***. Uma solução admissível que otimize a função objetivo designa-se por ***solução ótima***.

Cada equação linear $a_{i1} x_1 + a_{i2} x_2 + ... + a_{ik} x_k = b_i$ associada a uma restrição funcional do tipo $a_{i1} x_1 + a_{i2} x_2 + ... + a_{ik} x_k \leq (\geq)\, b_i$ define um hiperplano em $\mathbb{R}^k$. Este hiperplano designa-se por ***hiperplano de suporte*** da região admissível $\mathscr{R}$ se intersectar a fronteira de $\mathscr{R}$. Nos problemas de PL com 2 e 3 variáveis, os hiperplanos de suporte são, respetivamente, retas e planos.

6.5 Resolução gráfica

Um problema de programação linear com apenas 2 variáveis pode ser resolvido graficamente. Para ilustrar isso vamos considerar novamente o Problema 1. As restrições de sinal

$$x, y \geq 0, \tag{6.1}$$

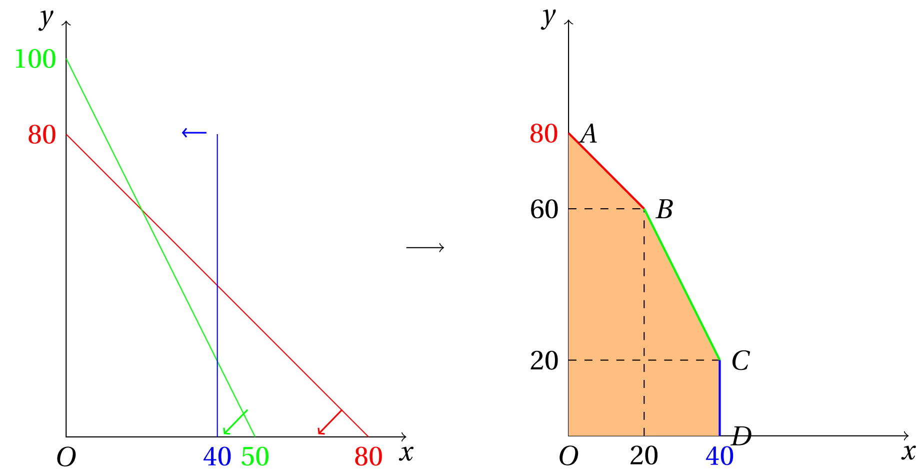
significam que a região admissível deste problema está contida no primeiro quadrante (incluindo eixos coordenados). A primeira restrição funcional da formulação (PPL 1),

$$x + y \leq 80, \tag{6.2}$$

define um *semi-espaço* de $\mathbb{R}^2$, isto é, um *semi-plano*. A fronteira deste semi-plano é a reta de equação

$$x + y = 80,$$

representada a vermelho na Figura 6.1 (a), que se designa por *reta de suporte* da restrição (6.2). Esta fronteira divide $\mathbb{R}^2$ nos dois semi-planos definidos pelas



(a) (b)

Figura 6.1: Região admissível do Problema 1. As retas de suporte da região admissível são dadas pelas equações $x + y = 80$, $8000x = 320000$, $40x + 20y = 2000$, $x = 0$ e $y = 0$.

inequações

$$x + y \leq 80 \qquad \text{e} \qquad x + y \geq 80.$$

O semi-plano que corresponde à restrição funcional (6.2) contém a origem uma vez que a origem satisfaz a desigualdade, sendo indicado por meio de uma seta na respectiva reta de suporte. Procedendo de igual modo identificam-se os semi-planos definidos pelas restantes restrições funcionais de (PPL 1),

$$8000x \leq 320000 \qquad \text{e} \qquad 40x + 20y \leq 2000, \tag{6.3}$$

cujas retas de suporte, representadas na Figura 6.1 (a) a azul e a verde, respetivamente, são definidas pelas equações,

$$8000x = 320000 \qquad \text{e} \qquad 40x + 20y = 2000,$$

ou equivalentemente, por

$$x = 40 \qquad \text{e} \qquad 2x + y = 100.$$

A região admissível do problema é o polígono OABCD assinalado a laranja na Figura 6.1 (b), que corresponde à interseção dos semi-planos definidos pelas restrições funcionais (6.2) e (6.3) e pelas restrições de sinal (6.1).

Vejamos como podemos determinar graficamente uma solução ótima do Problema 1, isto é, um ponto da região admissível que maximize a função objetivo. Para cada valor $c \in \mathbb{R}$ podemos considerar o conjunto de pontos para os quais a função objetivo toma o valor c,

$$\{(x, y) : 300x + 200y = c\},$$

que define uma reta em $\mathbb{R}^2$ que se designa por *conjunto de nível* c da função objectivo $z = 300x + 200y$. Por exemplo, o conjunto de nível $c = 5000$ é a reta de equação,

$$300x + 200y = 5000. \tag{6.4}$$

Os pontos do segmento de reta obtido como intersecção da reta (6.4) com a região admissível (representado pelo segmento de reta grosso a preto na Figura 6.2), correspondem às soluções admissíveis cuja receita é 5000€. Se incrementarmos o valor c da receita, obtemos uma reta de equação $300x + 200y = c$ paralela à reta (6.4), cuja intersecção com a região admissível está representada na Figura 6.2 pelo segmento de reta grosso a cinzento que se encontra mais à esquerda. Continuando a incrementar o valor de c, vamos obtendo novas retas paralelas que se vão aproximando do vértice $B = (20, 60)$, a que corresponde o valor da receita $c = 18000$€. Daqui se conclui que este vértice, é o ponto da região admissível para o qual a função objetivo atinge o valor máximo de 18000 €, uma vez que para valores $c > 18000$€, a reta $300x + 200y = c$ já não intersecta a região admissível do problema. O valor máximo da receita é portanto obtido atribuindo 20 ha à cultura de tomate e 60 ha à cultura de trigo, tendo-se que a única solução ótima do problema é o vertice B.

Note-se que a solução anterior utiliza a **totalidade** da área cultivável e a totalidade da mão de obra disponível, dizendo-se nessa altura que as restrições

$$x + y \leq 80, \qquad \text{e} \qquad 40x + 20y \leq 2000,$$

estão **saturadas**, isto é, satisfeitas com igualdade.
Se considerarmos valores de receita $0 < c < 5000$ vamos obter retas de equação $300x + 200y = c$, paralelas à reta (6.4) e localizadas à esquerda desta. À medida que estes valores se vão aproximando de zero essas retas vão-se aproximando da origem O, onde a função objectivo atinge o valor mínimo zero. Este vértice corresponde portanto à solução ótima do problema que consiste em **minimizar** a receita sujeita às mesmas restrições funcionais e de sinal.

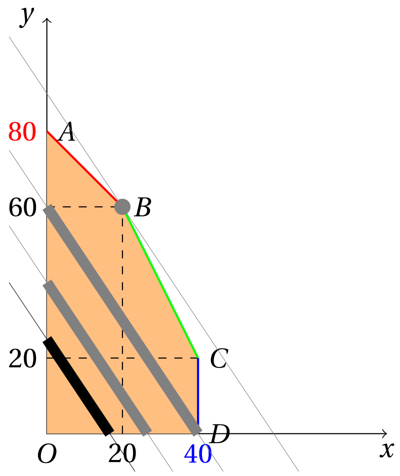


Figura 6.2: Resolução gráfica do Exemplo 1.

6.6 Problema 1 revisitado

Vamos modificar o Problema 1 para ilustrar algumas situações que podem ocorrer nos problemas de PL.

6.6.1 Problema com soluções ótimas alternativas

Consideremos o Problema 1 com o valor da receita do tomate por hectare alterado para 400€, ou seja, com a nova função objectivo

$$z = 400x + 200y.$$

Seguindo os mesmos passos do Problema 1 para determinar uma solução admissível que maximize a receita, constata-se que todos os pontos do segmento de recta $\overline{\text{BC}}$, (Figura 6.3), correspondem a soluções ótimas do problema gerando uma receita máxima comum de 20000€. Estas soluções ótimas correspondem no entanto a opções distintas: o vértice B, por exemplo, corresponde a cultivar 20 ha de tomate e 60 ha de trigo, enquanto que a opção C afecta 40 ha à cultura do tomate e 20 ha à cultura do trigo.

Cada uma das soluções ótimas do segmento de reta $\overline{\text{BC}}$, que se designa por **solução ótima alternativa**, pode escrever-se como *combinação convexa* dos vértices B e C, isto é, como uma combinação linear de B e C,

$$\lambda_1 B + \lambda_2 C, \qquad \text{em que} \qquad \lambda_1, \lambda_2 \geq 0 \quad \text{e} \quad \lambda_1 + \lambda_2 = 1.$$

O vértice B corresponde a tomar $\lambda_1 = 1$ e $\lambda_2 = 0$ e o vértice C a considerar $\lambda_1 = 0$ e $\lambda_2 = 1$. As combinações convexas de B e C, com $0 < \lambda_1, \lambda_2 < 1$, correspondem às restantes soluções ótimas que se encontram no interior do segmento de reta $\overline{\text{BC}}$. Por exemplo, tomando $\lambda_1 = \lambda_2 = 0.5$, obtém-se a solução ótima que consiste em atribuir 30 ha à cultura de tomate e 40 ha à cultura de trigo (verifique).

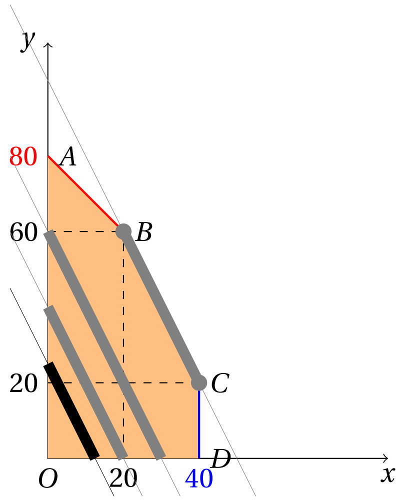


Figura 6.3: Resolução gráfica do Problema 1 com a nova função objectivo.

6.6.2 Problema sem soluções admissíveis

Consideremos agora o Problema 1 com a seguinte restrição adicional:

- A exploração agrícola tem clientes que pretendem comprar, pelo menos, 3250 t de tomate.

Considerando que a produção média do tomate por hectare é de 75 t, a restrição funcional adicional vem dada por

$$75x \geq 3250.$$

Ao adicionarmos à região admissível do Problema 1 os pontos que validam a restrição adicional anterior obtém-se o conjunto vazio (Figura 6.4), pelo que o problema modificado não possui soluções admissíveis.

6.7 Problema 4

Consideremos o problema de PL

$$\begin{array}{lrcl} \text{min} & z = x + 1.5y & & \\ \text{sujeito a} & x + y & \geq & 300 \\ & 100x + 200y & \geq & 40000 \\ & x & \leq & 300 \\ & x, \quad y & \geq & 0 \end{array}$$

cuja região admissível é não limitada (Figura 6.5 (a)). Comecemos por considerar o conjunto dos pontos de $\mathbb{R}^2$ para os quais a função objectivo $z = x+1.5y$ toma um dado valor c:

$$x + 1.5y = c. \tag{6.5}$$

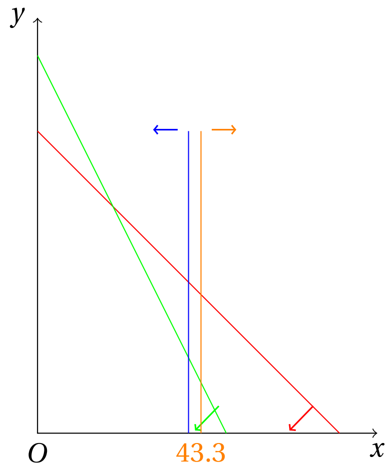


Figura 6.4: Região admissível do Problema 1 com a restrição adicional.

Se $c = 150$ obtém-se a reta $x + 1.5y = 150$ (reta mais próxima do vértice O na Figura 6.5 (b)), reta essa que não intersecta a região admissível. Isto significa que não há soluções admissíveis para as quais a função objectivo tome o valor 150. Se formos incrementando o valor de c vamos obtendo retas de equação $x+1.5y = c$, paralelas à reta $x+1.5y = 150$, que se vão aproximando do vértice $A = (200, 100)$ (Figura 6.5 (b)). Quando $c = 350$ a reta $x + 1.5y = c$ passa no vértice A, correspondendo esse valor de c ao menor valor para o qual a reta $x+1.5y = c$ intersecta a região admissível do problema. Uma vez que se pretende minimizar a função objectivo $z = x+1.5y$ a (única) solução ótima do problema é o vértice $A = (200, 100)$, a que corresponde o valor ótimo 350.

Notemos que se o problema consistisse em maximizar a função objetivo $z = x + 1.5y$, o problema não teria solução ótima, uma vez que para qualquer valor $c \geq 350$ a reta $x + 1.5y = c$ intersecta a região admissível. Neste caso o valor ótimo da função objetivo seria **infinito**.

Acabámos de apresentar vários exemplos que ilustram diversas situações que podem ocorrer na resolução de problemas de PL: uma única solução ótima (Problema 1), existência de soluções ótimas alternativas (Problema 1 modificado 6.6.1), não existência de soluções admissíveis (Problema 1 modificado 6.6.2) e região admissível do problema não limitada (Problema 6.7).

6.8 Análise de sensibilidade gráfica

Os modelos de PL apresentados assumem que os parâmetros tomam valores constantes. A análise de sensibilidade estuda o efeito que as alterações dos valores destes parâmetros produzem na solução ótima do problema.

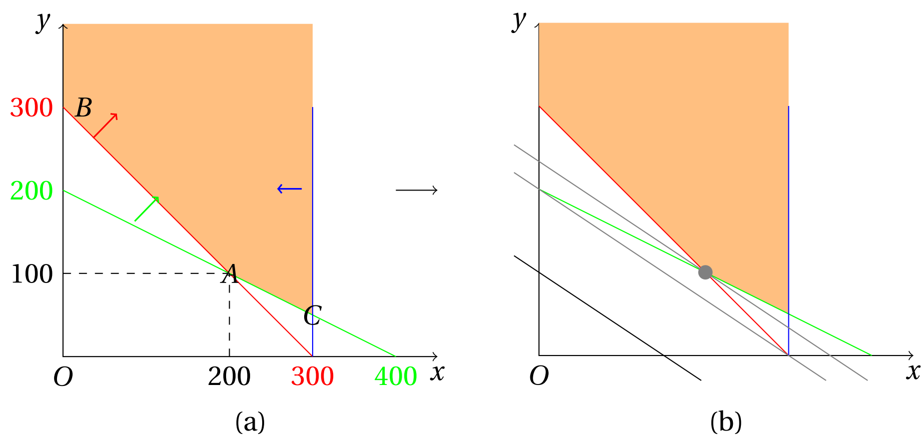


Figura 6.5: Resolução gráfica do Problema 4. As retas de suporte da região admissível são dadas pelas equações $x + y = 300$, $100x + 200y = 40000$, $x = 300$, $x = 0$ e $y = 0$.

6.8.1 Alteração de um coeficiente da função objetivo

Consideremos novamente o Problema 1. Vamos determinar os valores que a receita por cada hectare de tomate cultivado pode tomar de modo a manter como solução ótima a decisão anterior de cultivar 20 ha de tomate e 60 ha de trigo (assumindo que os restantes valores dos parâmetros se mantêm).

Designando por a, com $a \geq 0$, a receita obtida por cada hectare de tomate cultivado (em €/ha), tem-se que os pontos de x e y para os quais a receita total é, por exemplo, $c = 5000$€ (mantendo a receita obtida por hectare de trigo cultivado) definem a reta de

$$ax + 200y = 5000 \tag{6.6}$$

cujo declive é $-\dfrac{a}{200}$ e que se designa por **conjunto de nível** $c = 5000$ da f.o. $z = ax + 200y$. Fazendo variar o lucro $a \geq 0$ obtido por hectare de tomate cultivado, obtêm-se retas com distintos declives, como as que estão representadas a preto na Figura 6.6:

- Em (a), considerou-se um lucro de $a = 400$€ por hectare de tomate cultivado, de modo a que o conjunto de nível c da f.o. tenha o mesmo declive que a reta de suporte da restrição funcional $40x + 20y \leq 2000$, que está associada à mão de obra;

- Em (b), considerou-se um receita de $a = 200$€ por hectare, de modo a que conjunto de nível c da f.o. tenha o mesmo declive que a reta de suporte da restrição $x + y \leq 80$ que está associada ao terreno disponível;

- Em (c), considerou-se uma receita $a > 400$€, de modo a que o declive do conjunto nível c da f.o. seja maior que o da reta de suporte da restrição $40x + 20y \leq 2000$;
- Finalmente em (d), considerou-se uma receita $a < 200$€ de modo à curva de nível c ter um declive que seja inferior ao da reta de suporte da restrição $x + y \leq 80$.

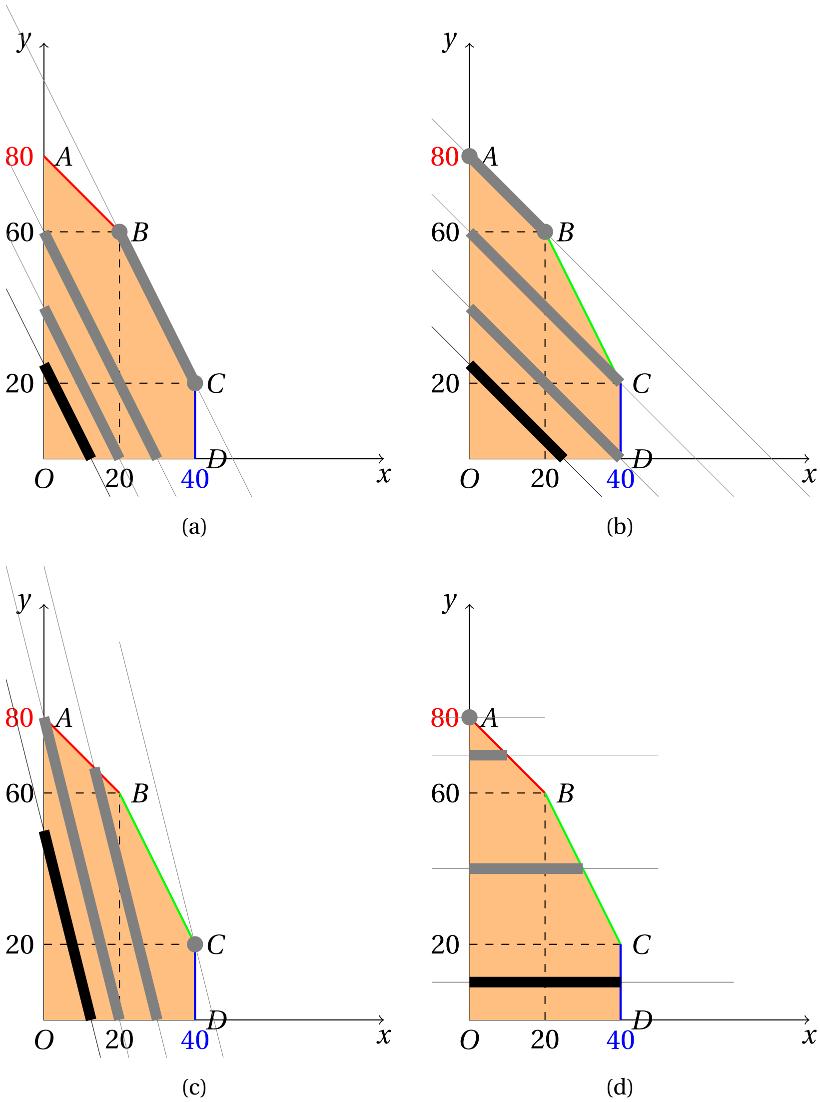


Figura 6.6: Resolução gráfica do Exemplo 1 para diferentes valores dos coeficientes da função objetivo.

Nos casos (a) e (b), o vértice $B = (20, 60)$ mantém-se como solução ótima. Mas em ambos os casos, o problema passa a ter uma infinidade de soluções ótimas: os pontos do segmento de reta BC em (a) e os pontos do segmento de reta AB em (b). Nos casos (c) e (d), a solução ótima não se mantém: em (c) o vértice $C = (40, 20)$ passa a ser a única solução ótima e em (d) é o vértice $A = (0, 80)$ que maximiza a f.o.

De facto, o vértice B mantém-se ótimo se e só se o declive dos conjuntos de nível c que representam a receita total,

$$ax + 200y = c,$$

está compreendido entre os declives das retas

$$40x + 20y = 2000 \qquad \text{e} \qquad x + y = 80,$$

respetivamente, ou seja, se só se

$$2 \leq \frac{a}{200} \leq -1,$$

isto é,

$$200 \leq a \leq 400.$$

Repare que o valor ótimo da função objetivo altera-se para $a \neq 300$. Por exemplo, quando $a = 400$ a valor ótimo da receita passa a ser 20000€.

Uma análise semelhante pode ser efetuada relativamente aos valores da receita por cada hectare de trigo cultivado, que se deixa como exercício.

6.8.2 Alteração do membro direito de uma restrição funcional

Vamos estudar o impacto que o aumento de uma unidade do membro direito de uma restrição funcional tem no valor ótimo da função objetivo, assumindo que os restantes valores dos parâmetros se mantêm.

Vamos continuar com o Problema 1 e efetuar a análise relativamente às restrições funcionais,

$$8000x \leq 320000, \qquad \text{e} \qquad x + y \leq 80.$$

Relativamente à restrição $8000x \leq 320000$, se aumentarmos o membro direito (a quantidade de água disponível) em uma unidade (1 m^3), a solução ótima mantém-se. Repare-se que no entanto região admissível é alterada, uma vez que a reta de suporte $x = 40$ é substituída pela reta $x = 40.000125$, obtendo-se novos vértices C' e D', de coordenadas $(40.000125, 19.99975)$ e $(40.000125, 0)$, respetivamente. Uma vez que a solução ótima se mantém, esta alteração não tem impacto na solução ótima.

Relativamente à restrição $x + y \leq 80$, se aumentarmos o membro direito (quantidade de terreno disponível) em uma unidade (1 ha), a solução ótima altera-se (Figura 6.7), passando a ser o vértice B', que consiste em atribuir 19 ha à cultura do tomate e 62 ha à cultura do trigo. A correspondente receita ótima é agora de 18100€. Logo, um incremento de 1 ha de terreno disponível produz um aumento de 100€ na receita total.

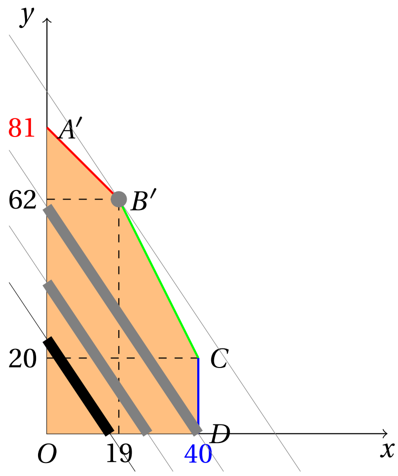


Figura 6.7: Resolução gráfica do Exemplo 1 em que a disponibilidade de terreno é de 81 ha.

6.9 Vértices da região admissível

Verificou-se nos exemplos anteriores que uma solução ótima, quando existia, era atingida num vértice da região admissível. Nesta secção vamos ver como se pode generalizar essa propriedade um problema genérico de PL.

Denotemos por (P) o problema de PL,

$$\begin{array}{rl} \max \ (\text{ou} \ \min) & z = c_1 x_1 + c_2 x_2 + \cdots + c_n x_n \\ \text{sujeito a} & (x_1, x_2, \ldots, x_n) \in \mathscr{R} \end{array}$$

em que c_j $(j = 1, ..., n)$ é constante e $\mathscr{R}$ é a região admissível. Tem-se o seguinte resultado fundamental.

Teorema 63. *Se $\mathscr{R}$ for limitada e não vazia tem-se:*

1. *Existe um vértice de $\mathscr{R}$ que é solução ótima de (P);*

2. *Se q vértices de $\mathscr{R}$, $v_1, ..., v_q$, são soluções ótimas de (P) então qualquer combinação convexa destes vértices,*

$$\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_q v_q, \quad \text{em que} \quad \lambda_1, \lambda_2, ..., \lambda_q \geq 0 \quad e \quad \lambda_1 + \lambda_2 + \cdots + \lambda_q = 1,$$

é também solução ótima de (P).

O teorema anterior reduz o problema de determinar uma solução ótima de um problema de PL com região admissível limitada e não vazia, ao problema de determinar os vértices dessa região admissível (que são em número finito) e identificar um dos vértices em que a função objectivo tome o maior ou menor valor, consoante o problema seja de maximização ou minimização.

Vamos exemplificar como podemos aplicar o teorema para determinar uma solução ótima considerando novamente o Problema 1.

Cada vértice da região admissível do Problema 1 (Figura 6.1) é determinado resolvendo um sistema de duas equações lineares que representem retas de suporte de duas restrições funcionais (Figura 6.8).

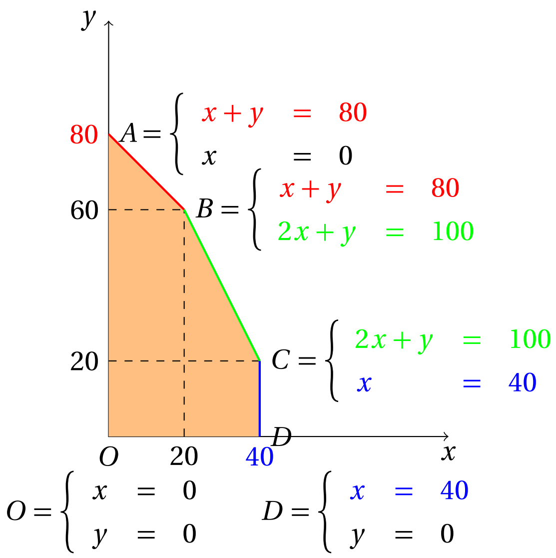


Figura 6.8: Determinação dos vértices da região admissível do Exemplo 1.

Por exemplo, o vértice B, que corresponde à interseção das retas $x+y=80$ e $2x+y=100$, determina-se resolvendo o sistema linear

$$\begin{cases} x+y &= 80 \\ 2x+y &= 100. \end{cases}$$

Como a região admissível é limitada e não vazia, uma solução ótima do problema ocorre num dos vértices bastando por isso calcular o valor da receita em cada um dos 5 vértices e selecionar o(s) vértice(s) com a receita mais elevada. A Tabela 6.5 mostra os valores obtidos da receita para cada vértice, concluindo-se que uma solução ótima é atingida no vértice $B=(20,60)$, resultado esse que já tinhamos obtido anteriormente através da resolução gráfica do problema.

Note-se que existem $\binom{5}{2}=10$ sistemas lineares formados por duas equações que representam retas de suporte de restrições lineares e de sinal do problema. No entanto, apenas 5 sistemas lineares têm como solução um vértice. Os restantes 5 sistemas lineares ou são impossíveis

$$\begin{cases} x &= 0 \\ x &= 40 \end{cases},$$

Vértice (x, y)	$z = 300x + 200y$
O=(0,0)	0
A=(0,80)	16000
B=(20,60)	18000
C=(40,20)	16000
D=(40,0)	12000

Tabela 6.5: Vértices e correspondentes valores da função objetivo (Exemplo 1).

ou têm como solução um ponto que não satisfaz todas as restrições do problema, isto é, que não é solução admissível (Figura 6.9).

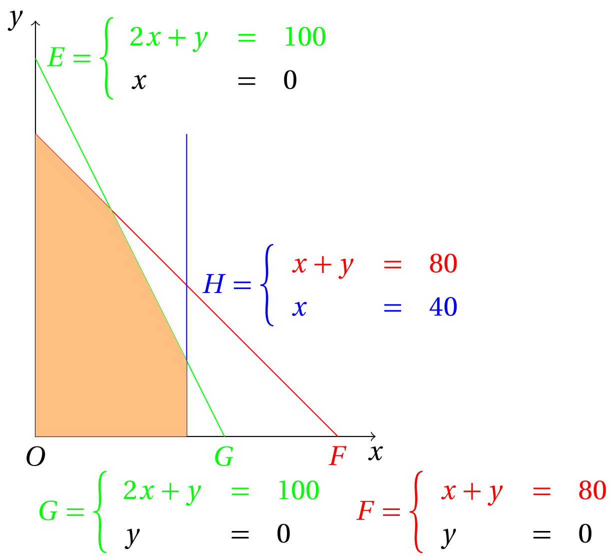


Figura 6.9: Sistemas lineares possíveis que não representam vértices da região admissível do Problema 1.

6.10 Forma *standard*

Nesta secção vamos mostrar como podemos converter o problema descrito em termos de variáveis de decisão num problema em que todas as variáveis tomam valores não negativos e todas restrições funcionais são definidas por equações lineares, dizendo-se nessa altura que o problema está na forma ***standard***. Uma vantagem importante desta nova formulação é que a transformação das variáveis do problema original nas variáveis do problema na forma *standard* estabelece uma correspondência entre vértices da região admissível $\mathscr{R}$ do problema original e um subconjunto de soluções do sistema linear que define a região admissível $\mathscr{F}$ do problema na forma *standard*, permitindo desse modo identificar os vértices de $\mathscr{R}$.

Para converter um problema de PL para a forma *standard* aplicam-se as seguintes regras:

1. Cada restrição de sinal do tipo $x_j \leq 0$ é substituída pela restrição de sinal $x'_j \geq 0$, sendo a variável x_j substituída pela expressão $-x'_j$ nas restrições funcionais e na função objetivo;

2. Para cada variável sem restrições de sinal x_j ([1]), acrescentamos duas variáveis $x'_j, x''_j \geq 0$, sendo a variável x_j substituída pela expressão $x'_j - x''_j$ nas restrições funcionais e na função objetivo;

3. Para cada restrição funcional do tipo,

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n \leq b_i,$$

acrescentamos uma nova variável f_i e substituímos essa restrição pelas restrições linear e de sinal,

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n + f_i = b_i, \qquad f_i \geq 0;$$

4. Para cada restrição funcional do tipo,

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n \geq b_i,$$

acrescentamos uma nova variável f_i e substituímos essa restrição pelas restrições linear e de sinal,

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n - f_i = b_i, \qquad f_i \geq 0;$$

5. Os coeficientes das variáveis f_i da função objetivo na forma *standard* são nulos.

Note-se que as restrições funcionais do tipo $a_{i1}x_1 + a_{i2}x_2 + ... + a_{in}x_n = b_i$, não são alteradas, uma vez que já são definidas por equações lineares.

As variáveis f_i associadas às restrições 3. e 4. designam-se por **variáveis de folga**.

A primeira regra resulta da equivalência:

$$x_j \leq 0 \quad \Leftrightarrow \quad x'_j = -x_j \geq 0.$$

A segunda regra é mais subtil. A partir de uma solução ótima do problema

[1]Estas variáveis são usualmente designadas por variáveis *livres*, mas preferimos não usar essa designação neste contexto, uma vez que tem sido utilizada ao longo do texto com outro significado.

transformado, isto é, do problema com variáveis x'_j e x''_j, obtém-se uma solução ótima do problema original (com o mesmo valor da função objetivo), considerando $x_j = x'_j - x''_j$. Reciprocamente, a partir de uma solução ótima do problema original podemos obter uma infinidade de soluções ótimas distintas do problema transformado. Uma dessas soluções pode ser obtida, por exemplo, considerando $x'_j = \dfrac{|x_j| + x_j}{2}$ e $x'_j = \dfrac{|x_j| - x_j}{2}$, tendo-se obviamente, $x'_j, x''_j \geq 0$ e $x_j = x'_j - x''_j$.

As regras 3 e 4 resultam, respectivamente, das seguintes equivalências:

$$
\begin{aligned}
a_{i1}x_1 + a_{i2}x_2 + ... + a_{in}x_n \leq b_i \quad &\Leftrightarrow \quad \overbrace{b_i - (a_{i1}x_1 + a_{i2}x_2 + ... + a_{in}x_n)}^{f_i} \geq 0 \\
&\Leftrightarrow \quad \begin{cases} f_i &= b_i - (a_{i1}x_1 + a_{i2}x_2 + ... + a_{in}x_n) \\ f_i &\geq 0 \end{cases} \\
&\Leftrightarrow \quad \begin{cases} a_{i1}x_1 + a_{i2}x_2 + ... + a_{in}x_n + f_i &= b_i \\ f_i &\geq 0 \end{cases}
\end{aligned}
$$

e

$$
\begin{aligned}
a_{i1}x_1 + a_{i2}x_2 + ... + a_{in}x_n \geq b_i \quad &\Leftrightarrow \quad \overbrace{a_{i1}x_1 + a_{i2}x_2 + ... + a_{in}x_n - b_i}^{f_i} \geq 0 \\
&\Leftrightarrow \quad \begin{cases} f_i &= a_{i1}x_1 + a_{i2}x_2 + ... + a_{in}x_n - b_i \\ f_i &\geq 0 \end{cases} \\
&\Leftrightarrow \quad \begin{cases} a_{i1}x_1 + a_{i2}x_2 + ... + a_{in}x_n - f_i &= b_i \\ f_i &\geq 0 \end{cases} .
\end{aligned}
$$

Para ilustrar como se aplicam as regras anteriores, vamos converter à forma *standard* os problemas de programação linear 1, 2 e 3.

Problema 1 na forma *standard*

Como as variáveis de decisão da formulação (PPL 1) são não negativas e as restrições funcionais são do tipo descrito na 3ª regra, o problema 1 na forma *standard* vem dado, aplicando a 3ª e 5ª regras por,

$$
\begin{array}{lllllll}
\max & z = 300x + 200y & & & & \\
\text{sujeito a} & x + y & + \ f_1 & & & = 80 \\
 & 8000x & & + f_2 & & = 320000 \\
 & 40x + 20y & & & + f_3 & = 2000 \\
 & x, \quad y, & f_1, & f_2, & f_3 & \geq 0
\end{array}
$$

em que as variáveis de folga f_1, f_2 e f_3 têm o seguinte significado:

$f_1 = 80 - (x + y)$ representa a área de terreno (ha) não cultivada;

$f_2 = 320000 - 8000x$ representa a quantidade de água (m^3) não utilizada;

$f_3 = 2000$–$(40x+2y)$ representa o número de horas de trabalho não usadas.

Problema 2 na forma *standard*

Na formulação (PPL 2) as variáveis de decisão são não negativas e as restrições funcionais são dos tipos descritos nas regras 3 e 4, pelo que problema na forma *standard* vem dado, aplicando essas regras juntamente com a regra 5 por,

$$
\begin{array}{rl}
\min & z = 50x_A + 40x_B + 60x_C \\
\text{sujeito a} & x_A + x_B + x_C + f_1 = 15000 \\
 & x_A - f_2 = 5000 \\
 & 10x_A + 5x_B + 10x_C - f_3 = 100000 \\
 & x_A + x_B + x_C - f_4 = 12000 \\
 & x_A,\ x_B,\ x_C,\ f_1,\ f_2,\ f_3,\ f_4, \geq 0
\end{array}
$$

em que,

$f_1 = 15000 - (x_A + x_B + x_c)$ representa a capacidade (t) mensal de produção não utilizada;

$f_2 = x_A - 5000$ representa quantidade (t) de fertilizante A produzido mensalmente para além das 5000 t;

$f_3 = 10x_A + 5x_B + 10x_c - 100000$ é o lucro (€) obtido mensalmente para além de 100000 €;

$f_4 = x_A + x_B + x_c - 12000$ representa a quantidade (t) de fertilizantes produzidos mensalmente para além das 12000 t.

Problema 3 na forma *standard*

O problema na forma *standard* vem formulado como,

$$
\begin{array}{rl}
\min & z = 0.34x_{AE} + 0.23x_{AF} + 0.4x_{BE} + 0.45x_{BF} \\
\text{sujeito a} & x_{AE} + x_{AF} = 500 \\
 & x_{BE} + x_{BF} = 400 \\
 & x_{AE} + x_{BE} + f_1 = 500 \\
 & x_{AF} + x_{BF} + f_2 = 600 \\
 & x_{AE},\ x_{AF},\ x_{BE},\ x_{BF},\ f_1,\ f_2, \geq 0
\end{array}
$$

em que,

$f_1 = 500 - (x_{AE} + x_{BE})$ representa a capacidade (ℓ) diária do armazém E não utilizada;

$f_2 = 600 - (x_{AF} + x_{BF})$ representa a capacidade (ℓ) diária do armazém F não utilizada.

6.11 Vértices e soluções básicas admissíveis

Consideremos o problema (P) de PL na forma *standard*, descrito em notação matricial por,

$$\begin{array}{rrcl} \max & z & = & c^T\tilde{x} \\ \text{sujeito a} & A\tilde{x} & = & b \\ & \tilde{x} & \geq & 0 \end{array}$$

em que A é uma matriz de tipo $m\times n$, $b\in\mathbb{R}^m$, $c\in\mathbb{R}^n$ e

$$\mathscr{F} = \{\tilde{x}\in\mathbb{R}^n : A\tilde{x}=b,\ \tilde{x}\geq 0\},$$

é a região admissível de (P) na forma *standard.* A cada solução da região admissível $\mathscr{R}$ do problema formulado em termos de variáveis de decisão associamos uma e uma só solução da região admissível $\mathscr{F}$ do problema na forma *standard*, que é obtida transformando as variáveis de decisão como descrito pelas 5 regras da secção anterior.

Exemplo 84. No Problema 1, tem-se uma correspondência

$$(x,y)\in\mathscr{R}\longleftrightarrow(x,y,f_1,f_2,f_3)\in\mathscr{F},$$

que associa a cada solução admissível w descrita em termos de variáveis de decisão do problema, a solução admissível $\tilde{w}$ que é obtida acrescentando as variáveis de folga de acordo com as regras 3 e 5,

$$\begin{array}{rcl} f_1 & = & 80-(x+y), \\ f_2 & = & 320000-8000x, \\ f_3 & = & 2000-(40x+20y). \end{array}$$

Por exemplo, se considerarmos a solução admissível $w=(20,20)\in\mathscr{R}$ obtemos a solução admissível $\tilde{w}=(20,20,40,160000,800)\in\mathscr{F}$, como se verifica calculando os valores das folgas quando $x=y=20$:

$$\begin{array}{rcl} f_1 & = & 80-(20+20)=40, \\ f_2 & = & 320000-8000\times 20=160000, \\ f_3 & = & 2000-(40\times 20+20\times 20)=800. \end{array}$$

No que se segue estaremos sempre a admitir que $\text{car}(A)=m$.

Dado um conjunto $\beta=\{A_{j_1},A_{j_2},\ldots,A_{j_m}\}$ de m colunas de A **linearmente independente**, em que $1\leq j_1<\cdots<j_m\leq n$, chamamos *variáveis básicas* às variáveis que estão associadas às colunas A_{j_i}, $i=1,\ldots,m$, e chamamos *variáveis não básicas* ou *livres* às restantes $n-m$ variáveis. Uma solução do sistema $A\tilde{x}=b$ em que todas as variáveis não básicas são nulas para alguma escolha

de m variáveis básicas designa-se por *solução básica* (sb). Vamos denotar por $\tilde{x}_\beta$ a solução básica que é obtida considerando como variáveis básicas as variáveis que estão associadas às m colunas β de A. Uma solução básica diz-se *admissível* (sba) se todas as suas componentes forem não negativas.

Por definição cada solução básica de um sistema linear $A\tilde{x} = b$ em que $\text{car}(A) = m$ contém, pelo menos, $n - m$ componentes nulas uma vez que algumas das variáveis básicas podem também ser nulas. Aqui m designa o número de linhas da matriz e n o número de colunas de A, isto é, o número de variáveis do sistema. Mais precisamente tem-se a seguinte observação.

Observação 46. Admitindo que a $\text{car}(A) = m$, em que m é o número de linhas de A, tem-se que uma solução $\widetilde{w}$ do sistema $A\tilde{x} = b$ que define a região admissível $\mathscr{F}$ de um problema na forma *standard*, é básica admissível se e só se verificar as seguintes condições:

(i) Todas as componentes de $\widetilde{w}$ são não negativas.

(ii) $\widetilde{w}$ tem pelo menos $n - m$ componentes nulas, com m o número de restrições funcionais e n o número de variáveis contando com variáveis de folga.

(iii) O conjunto das colunas de A que estão associadas às componentes não nulas de $\widetilde{w}$ é linearmente independente([2]).

Teorema 64. *A transformação definida pelas 5 regras da secção anterior, faz corresponder a cada vértice da região admissível $\mathscr{R}$ do problema original, uma solução básica admissível da região admissível $\mathscr{F}$ do problema na forma standard e vice-versa.*

Exemplo 85. Com base no Teorema 64 vamos enumerar todas as sba_s do Problema 1. Ora, a região admissível deste problema na forma *standard* pode ser escrita matricialmente da seguinte forma:

$$\mathscr{F} = \left\{ \begin{array}{c} (x, y, f_1, f_2, f_3) \in \mathbb{R}^5 : \overbrace{\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 8000 & 0 & 0 & 1 & 0 \\ 40 & 20 & 0 & 0 & 1 \end{bmatrix}}^{A} \overbrace{\begin{bmatrix} x \\ y \\ f_1 \\ f_2 \\ f_3 \end{bmatrix}}^{w} = \overbrace{\begin{bmatrix} 80 \\ 320000 \\ 2000 \end{bmatrix}}^{b}, \\ x, y, f_1, f_2.f_3 \geq 0 \end{array} \right\}.$$

[2]A sba $\widetilde{w}$ diz-se **degenerada** se o número de componentes nulas de $\widetilde{w}$ for superior a $n - m$, isto é, se uma ou mais variáveis básicas tomarem o valor zero. Neste caso, o número de colunas de A associadas às componentes não nulas de $\widetilde{w}$ é inferior a m. No entanto, como a $\text{car}(A) = m$, é sempre possível acrescentar outras colunas de A de modo a obter-se um conjunto linearmente independente com m vetores.

A Tabela 6.6 apresenta vários pontos da região admissível do problema $\mathscr{R}$ (incluindo os seus vértices) e as correspondentes soluções da região admissível na forma *standard* $\mathscr{F}$. As soluções básicas admissíveis são as soluções de $\mathscr{F}$ que correspondem a vértices de $\mathscr{R}$. Em particular, estas soluções têm (pelo menos) 2 componentes nulas (5 variáveis - 3 restrições funcionais) e o conjunto das colunas de A associadas às respectivas componentes não nulas é linearmente independente. As duas soluções de $\mathscr{F}$ associadas a I e J não são soluções básicas.

	$(x,y)\in\mathscr{R}$	$(x,y,f_1,f_2,f_3)\in\mathscr{F}$	
Vértices	$O=(0,0)$	$(0,0,80,320000,2000)$	sba$_s$
	$A=(0,80)$	$(0,80,0,320000,400)$	
	$B=(20,60)$	$(20,60,0,160000,0)$	
	$C=(40,20)$	$(40,20,20,0,0)$	
	$D=(40,0)$	$(40,0,40,0,400)$	
	$I=(30,40)$	$(30,40,10,80000,0)$	
	$J=(20,20)$	$(20,20,40,160000,800)$	

Tabela 6.6: Soluções admissíveis de $\mathscr{R}$ e correspondentes soluções de $\mathscr{F}$.

Fica como exercício verificar que o ponto E assinalado na Figura 6.9 está associado a uma solução básica não admissível.

Exemplo 86. Consideremos o Problema 2 com uma restrição adicional redundante, que resulta de somarmos à 3ª restrição funcional a 2ª restrição multiplicada por 10:

$$20x_A + 5x_B + 10x_C \geq 150000.$$

A solução $(x_A, x_B, x_C) = (5000, 5000, 2500)$ é admissível uma vez que verifica as restrições funcionais e de sinal do problema (justifique). Vejamos que no entanto não corresponde a um vértice da região admissível, ou seja, que a correspondente solução da região admissível $\mathscr{F}$ na forma *standard* não é básica admissível.

Convertendo o problema para a forma *standard*, obtém-se

$$\begin{array}{lrcl}
\min & z = 50x_A + 40x_B + 60x_C & & \\
\text{sujeito a} & x_A + x_B + x_C + f_1 & = & 15000 \\
 & x_A - f_2 & = & 5000 \\
 & 10x_A + 5x_B + 10x_C - f_3 & = & 100000 \\
 & x_A + x_B + x_C - f_4 & = & 12000 \\
 & 20x_A + 5x_B + 10x_C - f_5 & = & 150000 \\
 & x_A,\ x_B,\ x_C,\ f_1,\ f_2,\ f_3,\ f_4,\ f_5 & \geq & 0.
\end{array}$$

Denotemos por $A\tilde{x} = b$ o sistema linear dado pelas 5 equações anteriores. Comecemos por observar que o conjunto das 5 colunas de A associadas às variáveis de folga é linearmente independente o que implica que a car $A = 5$ (nº de linhas de A). Podemos então afirmar que uma solução de $A\tilde{x} = b$ é básica admissível se e só se verificar as condições da Observação 46, isto é, se e só se: (i) todas as componentes forem não negativas; (ii) possuir pelo menos três (8 variáveis-5 restrições funcionais) componentes nulas; e (iii) o conjunto das colunas de A associadas às componentes não nulas for linearmente independente.

A solução com $(x_A, x_B, x_C) = (5000, 5000, 2500) \in \mathscr{R}$ corresponde à solução do sistema $A\tilde{x} = b$,

$$(x_A, x_B, x_C, f_1, f_2, f_3, f_4, f_5) = (5000, 5000, 2500, 2500, 0, 0, 500, 0) \in \mathscr{F}, \qquad (6.7)$$

como se pode verificar calculando as folgas. Obviamente (i) e (ii) são verificadas. Relativamente à condição (iii), tem-se que o conjunto das colunas de A associadas às componentes não nulas do vetor definido em (6.7) é linearmente dependente, uma vez que

$$\det \begin{bmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 10 & 5 & 10 & 0 & 0 \\ 1 & 1 & 1 & 0 & -1 \\ 20 & 5 & 10 & 0 & 0 \end{bmatrix} = 0.$$

Logo, $(5000, 5000, 2500, 2500, 0, 0, 500, 0)$ não é sba de $\mathscr{F}$, pelo que $(5000, 5000, 2500)$ não é vértice da região admissível $\mathscr{R}$.

Apêndice A

Método do simplex

O método do simplex foi desenvolvido em 1949 pelo matemático americano George B. Dantzig para resolver problemas de PL. A ideia central do método consiste em mover-se de uma sba para sba adjacente enquanto o valor da f.o. for melhorando, até atingir uma solução ótima, em que as soluções adjacentes já não melhoram esse valor da f.o.

Considere-se novamente um problema de PL na forma *standard*

$$\begin{array}{rrcl} \max & z & = & c^T \tilde{x} \\ \text{sujeito a} & A\tilde{x} & = & b \\ & \tilde{x} & \geq & 0 \end{array}$$

em que A é uma matriz de tipo $m \times n$, $b \in R^m$ e $c \in R^n$. Duas sba$_s$, $\tilde{x}_\alpha$ e $\tilde{x}_\beta$, com variáveis básicas associadas aos conjuntos α e β de m colunas de A linearmente independentes, dizem-se **adjacentes** se α e β têm $m-1$ colunas em comum.

O método do simplex considera o problema na forma *standard*. Começa-se por determinar, caso exista, uma sba inicial. Caso não exista o método termina, concluindo-se que o problema é impossível. Caso exista, o método prossegue para uma sba adjacente que melhore o valor da função objetivo. O passo anterior é repetido até que a solução ótima seja atingida (i.e., que não existam soluções adjacentes que melhorem o valor da f.o.) ou se conclua que a região admissível do problema é não limitada e o problema não tem solução ótima.

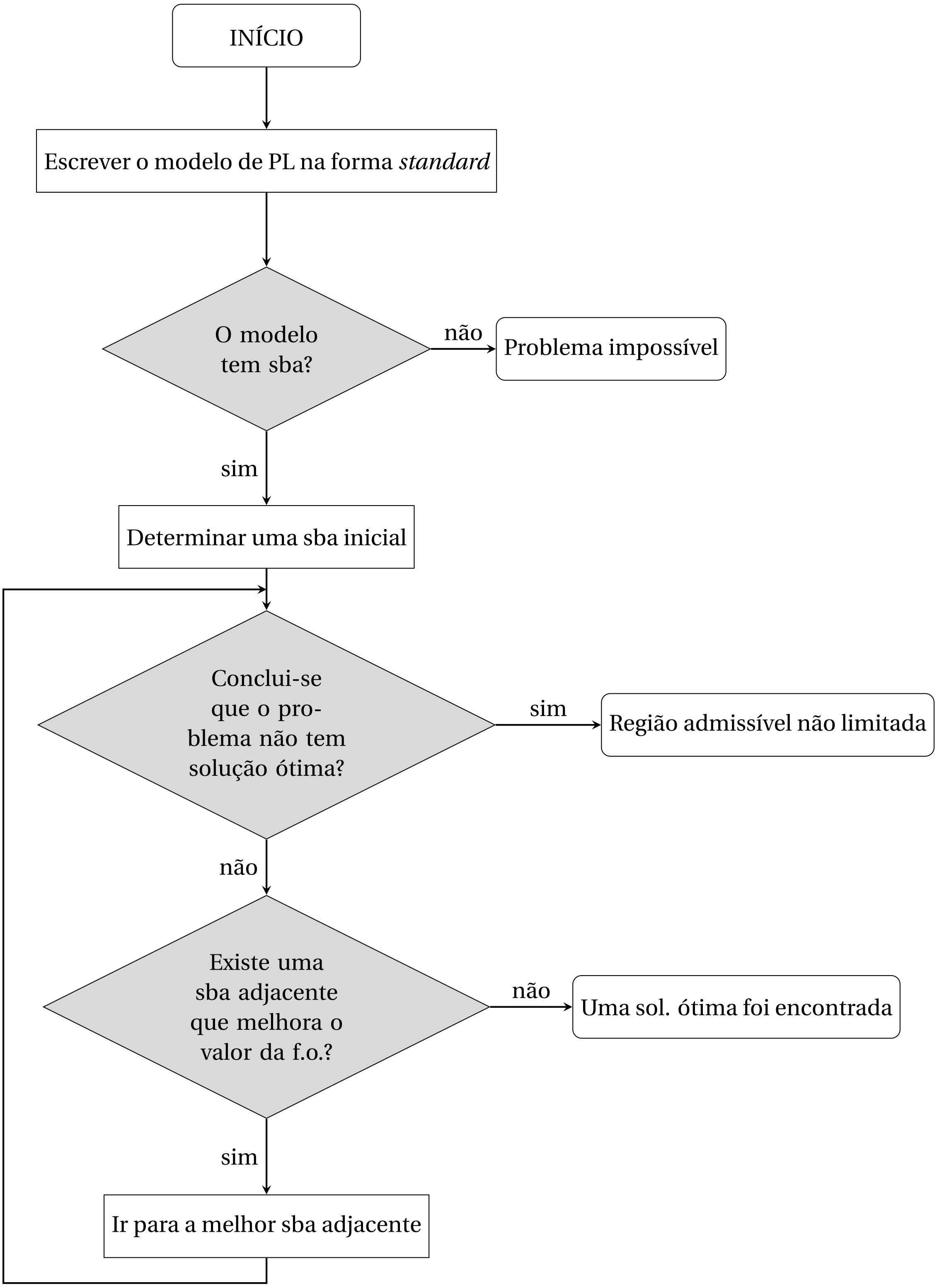


Figura A.1: Fluxograma do método do simplex.

Bibliografia

[1] Jorge Orestes, *Álgebra Linear*, Instituto Superior de Agronomia (2012).

[2] Gilbert Strang *Linear Algebra and its Applications* (3rd Edition), Harcourt Brace Jovanovich, Inc, ISBN: 0-15-551005-3 (1988).